**Review of Winter Arctic sea ice thickness from ICESat-2: upgrades to freeboard and snow loading estimates and an assessment of the first three winters of data collection by Petty, A. A. (review #3)**

Review responses to reviewer #3 (technically a second review round). Review in black our responses in blue.

Summary: With the ICESat-2 satellite laser altimeter launched in September 2018 the sea ice community has access to very high resolution observations of the sea ice topography and hence novel means to retrieve the sea ice thickness distribution. This manuscript details the impact of recent changes in the processing of the ICESat-2 data relevant for such retrieval. It details further the impact of updates of an important data set of ancillary information required for this retrieval: the snow loading. The manuscript convinces with comparably clear messages of these impacts mentioned and comes up with a presentation, interpretation and discussion of the seasonal development of the freeboard, snow and sea ice thickness values at basin scale.

The present manuscript falls into the category of work typically connected to new satellite missions when the (raw) processing is still ramping up and updates in the derived products are issued at a comparably high frequency. It is therefore of a quite technical nature - albeit it has a considerable fraction of geophysical interpretation. I was wondering therefore, whether the authors have considered to hand in this manuscript also into the journal "Earth System Science Data"? I'd say it is at the verge between both journals.

I should note that I read this manuscript in the second round of reviews - which explains why I refer to comments of the previous review at a few occasions.

Thank you for taking the considerable time to provide this detailed review, your effort is much appreciated, as well as the positive comments about our paper above.

Re: journal selection - we felt that presenting comparisons across the first three winters - including reporting on the significant thickness decline in the 2020-2021 winter - made it much more suitable to The Cryosphere than a more data description journal. In response to this review we have also added new comparisons with BGEP upward looking sonar draft measurements (only recently made available) and additional CryoSat-2 comparisons which have added considerably to the paper and we think adds even more to the suitability of this study for The Cryosphere. We describe these major upgrades here:

1. **BGEP ULS comparisons:**

These were not added in direct response to a review comment, but do help respond to some of the concerns about the value of the CryoSat-2 comparisons and their utility for validating our thickness estimates. As stated in response to the direct comments, we never intended to use CS-2 as a validation of our data, it was instead an intercomparison to other commonly used thickness data. The BGEP ULS data are more commonly used for validation purposes, however, as we do here also. Note that the data were only made available earlier this summer and we recently completed the analysis we believe is very suitable for inclusion here for providing crucial early validation of our ICESat-2 ice thickness data which is lacking from the literature to-date (due mainly to the lack of available data and the early lifetime of the mission). The comparisons are very strong and we think add considerable weight to our study, see Figure 10 below. New data description, results and discussion have been added to the revised manuscript to highlight these results.

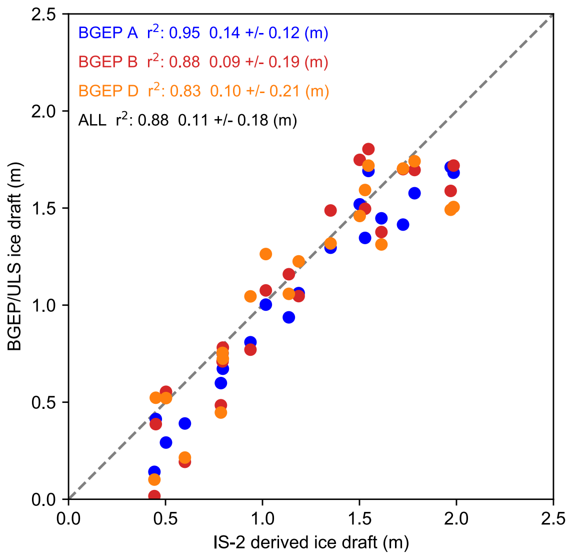
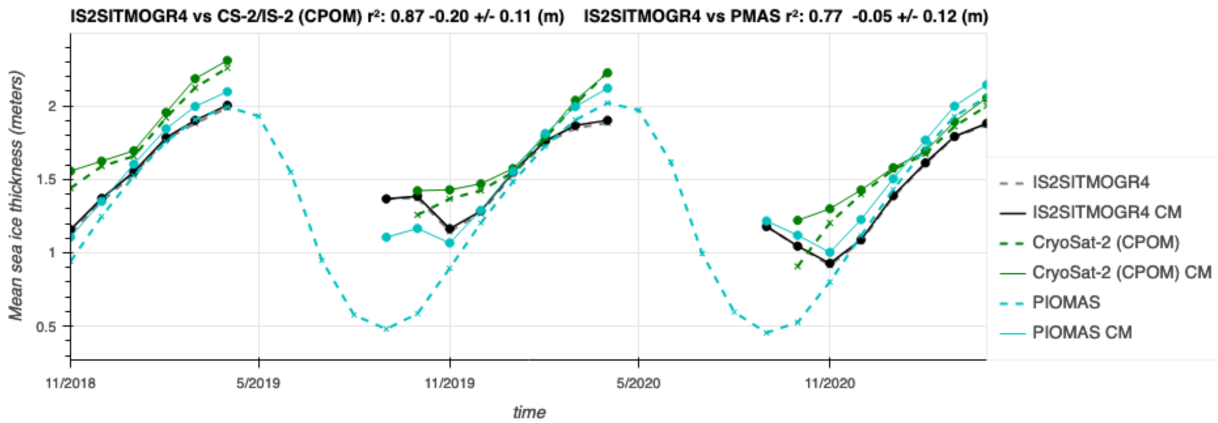



**Figure 10**: Comparisons of IS2SITMOGR4, v002 (Nov 2018-April 2019, September 2019-April 2020, September 2020-April 2021) converted to ice draft against ice draft measurements obtained by Beaufort Gyre Exploration Project (BGEP) upward looking sonar moorings. The mean of all IS2SITMOGR4 data within 100 km of the given mooring are used in this comparison.

## 2. Updated CryoSat-2 comparisons:

In response to comments here (and partly in response to the original reviewers) we have decided to change the CryoSat-2 comparison approach. This was also motivated by the fact we now have BGEP ULS data that provides a robust assessment of our IS-2 thickness data. We now use CS-2 in two ways:

1. We provide the same CS-2 'regional/monthly' comparison as before, but we substantially reduce the description of this analysis as recommended. Essentially, we show the figure and highlight the in all three stats the 'agreement' with CS-2 has improved since our version 1 thickness data – we think this is still a good sign and important to communicate. We also still think there is value in showing this comparison using the same input assumptions to remove the impact of different snow assumptions and to get a sense of the monthly/regional biases between the two retrieval approaches.
2. We now provide a new figure showing various comparisons (one example shown below) between CryoSat-2 and our monthly gridded ICESat-2 data but for the mean Inner Arctic Ocean thickness timeseries. In essence this is a comparison of the various products' best guess at the state of Arctic winter sea ice thickness and for this we use our final thickness product too (NESOSIM v1.1 snow loading). This comparison builds on (and integrates) the PIOMAS thickness comparison we originally included and is intended to provide a sense of seasonal agreement between these various thickness products. This also includes upgraded AWI/SMOS thickness data (Ricker et al., 2017) , a new all-season CryoSat-2 thickness dataset (Landy et al., 2022) and merged ICESat-2/CryoSat-2 thickness data (Kwok and Kacimi, 2022). We think this analysis, albeit basic, is useful for the community in providing a broad sense of how our thickness data compares to other well-used products. Again, more regional information is provided in the maps that are on the online Jupyter Book for more engaged users.




More discussion of the CS-2 additions are included below in response to the specific review comments.

General comments:
GC1: The authors need - throughout the manuscript - clarify and correct the usage of "total (sea ice plus snow) freeboard" and "sea ice freeboard". Currently, these terms appear to be mixed and they are in part misleading the discussion of the results. A laser altimeter observes the total freeboard. A radar altimeter is supposed to observe the sea ice freeboard. This must be corrected in the manuscript both in the text as well as in all figures and tables where this applies. You will find repeated notion of this in my specific comments. This is the main reason why I suggest that this manuscript requires another round of major revisions because it might require some time to correct this in an appropriate way.

'sea ice freeboard' is the language used in the ATL10/NSIDC data descriptions: https://nsidc.org/data/atl10/versions/5 so we had followed that convention. Our general view is that when talking to a broad audience it is helpful to include 'sea ice' to clearly indicate what the geophysical quantity represents, as total freeboard may not mean much to non-specialists. However, it is true that we include discussion of laser and radar profiling in this paper and agree it can be helpful in these more specific instances to use 'total freeboard' to refer to the freeboard including the snow then 'ice freeboard' to refer to the freeboard without the snow, similar to what you suggest.

In the introduction we now introduce the idea of total freeboard. We have also edited the rest of the manuscript to refer to just 'freeboard' when we are talking about ATL10 measured freeboards, as repeated use of total seemed superfluous.

We have also added the following clarification to the data section:
*"The laser returns are expected to track the snow-covered ice surface, so ATL10 is expected to provide a measure of 'total' (ice plus snow) freeboard. "*

GC2: Despite a high fraction of technical details there are a few technical and/or methodological issues that, to my opinion, are not laid out sufficiently well. To these belong i) the calibration of the parameter gamma, ii) the evaluation of the NESOSIMv1.1 product for other months than April, and iii) the description of how you derived the sea-ice thickness uncertainties.

We tried to strike a balance in this paper between providing a clear documentation of changes to NESOSIM and the sea ice thickness datasets, without getting too technical. We address these comments in detail where they come up in the specific comments below.

Specific comments:

L19-21: Is an enhanced agreement with CS-2 based sea-ice thickness data that desirable given the fact that penetration depth of the radar signal into snow plus ice-snow interface processes and properties play a much larger role therein?

We don't use CS-2 as a validation dataset, but we believe increased correspondence between the two satellite-derived products is generally encouraging and worth noting, especially as our results strongly suggest this occurred primarily due to improvements in the ATL10 freeboard determination. CS-2 has challenges with possible biases from ice-snow interface processes, but the data have also been well validated against available airborne/in-situ datasets to produce what we see as a reliable indicator of seasonal sea ice thickness change across the Arctic.

In response to a further comment below, we have now added additional CS-2 comparisons, including those from newer state-of-the-art CS-2 products to those originally presented in the manuscript as discussed above.

L35-46: In this paragraph you speak of "sea ice freeboard" in the context of laser altimetry. I strongly recommend to clearly differentiate between sea ice freeboard and total (sea ice plus snow) freeboard which is the quantity derived from the ICESat-2 elevation measurements first. Any conversion into sea ice freeboard (if need be) requires knowledge of snow thickness and density. It could be the that product you are using is named "sea ice freeboard" but physically it is not.

Yes, see main comment response above.

- What I am missing in front of this paragraph is that classical paragraph that tells us why satellite altimeters are such an important tool to observe the polar regions and why this is important. This would move your paper away from the impression one of the reviewers had that this is merely a technical report. In other words: So far the paper is not put into a larger context sufficiently well.

Agreed, we have now added an additional introductory paragraph about the importance of sea ice in the climate system and the need for new estimates of sea ice thickness.

L51-53: "Additional .... procedure" --> You could stress perhaps, that this is a high skill for daylight, clearsky conditions.

Yes, good point, we have added this now.

L108: "Sea ice freeboard" --> I assume this is in fact the total (sea ice plus snow) freeboard and should be referenced like this in the paper.

Yes, see general comment response above.

Section 2.1.1: I can understand that one of the previous reviewers got the impression to read a technical report rather than a scientific paper. This section appears to be quite long given the comparably little information that appears to be relevant for the content of the paper. To my opinion, this section could be condensed such that the main changes that determine the differences between rel002 and rel005 are highlighted while the reader is referred to the (regularly updated) technical documentation and change log associated with every release for the less relevant changes.

One of the main purposes of this paper was to highlight the impact of ATL10 changes on our estimates of sea ice thickness, hence the inclusion of what we see as very brief summaries of these pertinent changes. We made this as condensed as possible, and already include a link to the ATBD change log where the more minor issues or issues not relevant to freeboard are described, so we don't think there is much more we can really do here.

L195: How is the calibration of gamma carried out?

This is discussed further down in this section " We heuristically calibrated NESOSIM v1.1 using the daily OIB consensus gridded snow depths with the aim of removing the mean bias relative to OIB when using the default NESOSIM v1.0 parameter settings … tuning the new atmosphere snow loss coefficient, $\gamma$,"

L212/213: "Here we choose instead ..." --> This recalibration has the disadvantage that is is only valid for April snow conditions. Is that correct? How reliable can NESOSIM v1.1 then be for the rest of the freezing season. How is this period evaluated?

Unfortunately, this is a consistent challenge with simulating snow on sea ice – a real lack of regional-scale observations beyond what is offered by Operation IceBridge which are mainly at the end of the accumulation season.

We feel the following text included in the manuscript, which we have also added 'early-season' to make the point clearer, is about all we can say:
*"In the absence of contemporary early-season ground-truth data, we view the initial conditions (either their distribution or the representative start date) as another tuning parameter"*

L232-241: It is still not clear to me after these lines what the initial snow thickness on the sea ice on September 1 is.

Some example maps for 2012/2013 are given in Petty et al., (2018). We have edited the test her to point the reader to that:
*"based on the temperature scaled W99 August climatology shown in Petty et al., (2018, see Figure 2)."*

L232: It is still not clear how you tuned gamma.

See response to the comment above on heuristic/manual calibration to the median OIB snow depths.

L251: "declining trend" --> Is suggest to either write "decline" or "decrease" or "negative trend". A declining trend is a trend which is not stable over time but where the change of the parameter with time is decreasing over time.

Good point, changed.

- I am sure you are aware that this decline in snow thickness has several causes, beginning with the change in ice age, over changes in snow accumulation (period), actual ice drift vectors and retrieved ice drift vectors, to potential inconsistencies and spurious trends in the input data to NESOSOM v1.1. Because of these it might be a good idea to not overinterpret this decline. In addition: The tuning / evaluation of the NESOSIM v1.1 is limited to April ... and hence the time series shown for October could be less credible / reliable than the time series shown for April.

Yes, good points, agreed, we have further reduced this trend discussion in the revised manuscript.

L272-276 / Figure S3: Please check these lines; it seems they contain two times almost the same sentence.

Yes thanks, a similar line was erroneously added in the review responses. Removed now.

- How much ice is in the Kara Sea in October? I am not overly convinced that this particular region is well suited looking at October snow thickness on sea ice. I am wondering where the ice edge drawn in the Kara Sea for October comes from.

Yes, good point, ice concentration is very low in the Kara Sea in October and NESOSIM (and Warren) are highly uncertain this region. We have now masked this plot using an 80% threshold (needs to be high as NESOSIM only provides data where ice concentration > 15%) – now virtually all the grid-cells in October in the Kara Sea have gone.

However, following other comments and feedback we have now dropped the climatology analysis from the manuscript.

We have included a new figure showing the mean NESOSIM v1.1 vs mW99 for the years analysed here , which we include instead

- Figure S3 units are cms. Is this correct?

This has now been removed from the manuscript as we removed the NESOSIM clim discussion entirely as it distracted from our other messages. We have instead included a new figure showing mW99 vs NESOSIM for the years analysed here (2018-2021) and have fixed the units accordingly (new Figure S4)!

- The color table used in the snow thickness maps on the left does not convince me that there is "measurable" snow thickness on Kara Sea sea ice. It looks white.

See above comments.

- Apart from these comments, I guess I have a conceptual problem with on the one hand striving to produce a temporally reasonably fine resolved snow thickness data set for proper sea-ice thickness retrieval and on the other hand generating (again) kind of a climatology.
Shouldn't the strategy be to use the auxiliary data sets when these are fully available? I mean, you also switched to a different daily sea-ice motion product for the time period that NSIDC drift is not available. You are not computing a climatology. Also, since you don't show these NRT products in your paper and only deal with the winter season 2020-2021 as the most recent, I believe you could condense this part of the investigation considerably. Perhaps you could mention at the side that a climatology based on NESOSIMv1.1 based on the 2010s (which would exclude 2020 by the way) provides an October snow thickness which is about 1 cm smaller than mW99 and an April snow thickness which is about 5 cm larger than mW99 for the inner-Arctic region you selected. You could then argue that most likely - if one still wants to use a climatology - then the NESOSIMv1.1 climatology might be a better choice ... but your results in fact only show the difference but do not indicate which of the two data sets is the more accurate one. I therefore find this comparison between mW99 and the 2010s snow thickness climatology not overly useful and I cannot recommend to keep this in the manuscript. To my opinion this sets the wrong signal.

Due to the inclusion of the new analysis and the issues you raise about the use of the NESOSIM average here, we've decided to drop this analysis from the paper.

- Did you try to compute a similar time series as shown in Figure 4 using W99 / mW99 and available multiyear ice areal fractions for "your" Inner Arctic region to see whether the negative trend in NESOSIM snow thickness is also visible in W99 snow thickness (simply via the change of the multiyear ice versus first-year ice partition)?

This was a good idea, although we're limited by the fact OSI SAF ice types are only available back to 2009. We have now calculated this using the interannual/monthly ice type and we think it is nice to show this to highlight the ice type dependency on the mW99 results and how that compares to NESOSIM. The results are quite intuitive, the variability appears similar in October, when snow is thinner and largely ice type dependent, but NESOSIM appears to have higher variability later in the season. We have made some comments on this in the revised manuscript.

L279-298: Given the coarseness of the NESOSIM snow thickness data I am wondering whether from the view point of spatial scales involved it would make much more sense to discard the interpolation from 100 km x 100 km over more than 4 orders of magnitude to 30 meters and instead work with 10 km along track mean freeboard estimates. With that the interpolation would just be about one order of magnitude - at least into one direction.
But lets see what you will write about this in your discussion section.

We do not really interpolate the 100 km data to the segment scale, what we explored in Petty et al., (2020) is fitting a distribution of segment-scale data within 100 km aggregates. As explored in Glissenaar et al., (2020, see references below) there is no real skill in this redistribution when using CS-2 like footprints (scales of kilometers), we think the skill really only starts to be shown when you redistribute to much higher resolutions, e.g. IS-2, and the difference between low and high snow depths over different types of ice surface becomes clear in the distributions.

- I note that your description of the error estimation in Lines 284/285 is not overly specific. If there are no further details given elsewhere I recommend to specify better how you carried out this step.

This information is provided in the linked to Petty et al., (2020) reference, so we do not believe it is worth repeating this here.

L307: So in the monthly sea ice thickness data you use the CDR version 4 but in NESOSIMv1.1 you use version 3. This difference in versions is possibly not dramatic, isn't it?

Yes, in our conversations with the CDR sea ice data producers it did not seem like there was a big difference. We switched the thickness processing mainly just to ensure our system is up-to-date for future processing needs.

L326-332: I am sorry, I don't get why you prefer to compare ICESat-2 sea ice thickness data based on rel005 of ATL10 with the four CS-2 sea ice thickness products using THEIR snow loading. What is the motivation? What are you aiming to show here? It seems you are using the CS-2 sea-ice thickness products as a benchmark against which you would like to reference your product. Is this a viable approach given the difficulties / assumptions these radar altimeter products need to deal with? If you find a large bias / RMSD ... I'd say this is fine ...

I think your last point gets at the approach - we did observe large biases between CryoSat-2 and our ICESat-2 thickness results in Petty et al., (2020) which were concerning and part of our rationale for looking into this again with our new thickness data. We think it's important to document clearly that the changes in freeboard have significantly reduced the large thickness biases. We agree that we do not intend for CryoSat-2 to be used as a benchmark to assess the ICESat-2 data, so following another comment have reduced the description of the comparisons shown in Figure 8 and focus more on the fact the biases have been substantially reduced.

As discussed earlier, we have now also introduced new CryoSat-2 comparisons, and crucially comparisons against BGEP Upward Looking Sonar draft measurements that are much more suitable to validate our thickness estimates (see major comment at the start of this response).

In addition, you only use data of strong beam #1 here - to be consistent with your previous work. Are you not interested in how your updated sea-ice thickness product (see the many updates you wrote about so far in the manuscript!) compares to all these other products?

Yes, based on various comments in this review and the release of new CryoSat-2 sea ice products (e.g. Landy et al., 2022), we have now updated our approach here to 1. Briefly show the improvements from Petty et al., (2020), before comparisons with ULS draft data and also seasonal CryoSat-2/PIOMAS thickness comparisons. In the latter we focus more on comparisons of 'best guess Inner Arctic Ocean sea ice thickness' on basin-scales and use the updated (v2) ISTSITMOGR4 data which does use all three beams. Please see description of this change at the top of the review responses.

Section 2.6: This section appears to need some more work at it seems not to be complete. Also, it is not clear what the ERA5 data are for here. "To assess the winter Arctic atmospheric conditions" reads as if you made an investigation of the atmospheric conditions such as comparing ERA5 with rawinsonde and in-situ observations or the like. You need to work on your wording. Also, the last sentence does not fit quite well. I guess you merely perform a consistency check whether ERA5 2-m air temperatures support your assumptions about the physics and conditions but the nature of what you do is far away from an "assessment". It is an inter-comparison.

We agree the analysis presented here is brief and our aim was to really provide some extra insight into whether the thickness changes (seasonally and interannually) are consistent with the primary atmospheric forcing differences. This use of 'assessment' seems to align with the definition of '*the evaluation or estimation of the nature, quality, or ability of someone or something*.' Although again we acknowledge this effort is far from exhaustive. I am not sure inter-comparison is correct here (that would imply a comparison across different products perhaps?) so instead we have stated our aim is to "understand the possible relationships between seasonal/interannual differences in ice thickness and winter Arctic atmospheric conditions" and have worked on further clarifying our goals and the main results of this analysis.

L362-365: The derivation of individual freeboard values from surface heights and approximated sea surface level implies that there are negative freeboard values as well. What is their fraction and what did you do with these?

In the ATL10 algorithm, negative freeboards are set to zero. We have clarified this in the data section: "*Negative freeboards are set to zero.*"

L366: The results shown in Figure 6 are for your inner Arctic region or the entire Arctic?

Thanks, results are shown in an Inner Arctic domain. We have clarified this in the figure caption.

L399: I see the "performance" of the v1.1_2010s snow thickness climatology in a slightly different way than you. I would state that in April it appears to be a good representation of what Nv1.0 and Nv1.1 provide. But in January and particularly in November the distributions differ considerably, casting doubts about the credibility of this climatology; these doubts are justified looking at the sea-ice thickness distributions where usage of the climatology provides a substantially smaller modal (1st mode) sea ice thickness. If used for model initialization, assimilation or intercomparison studies the sea ice thickness data based on the climatology are certainly more problematic - especially at this critical time of the freezing season. I cordially invite you to tone your statements about the climatology into this direction rather than saying, well, all data sets look just fine with some minor differences.

Based on the earlier comments about the NESOSIM climatology, and the inclusion of our new analysis, we have decided to drop this from the manuscript so this is now redundant.

L403/404: "In general ..." --> I suggest to refer to a table or figure in your manuscript which supports this very general comment.

Yes this was a bit loosely framed. We have now added some extra lines in the plots to show rel002 thickness and have improved the wording to make it clear the thickness distribution difference is greater in the rel002 to rel003 and onwards runs compared to the use of different NESOSIM versions. We have not dropped the climatology from this discussion in part to increase focus on this point.

L468-470: How do the W99 snow densities compare to your densities? Can you identify hotspots in space and/or time where the W99 snow density would be considerably off compared to your results?

A NESOSIM vs mW99 snow density analysis was included in the original Petty et al., (2020) paper (Figures S2-S4). We have now included a reference to this in the revised paper.
*" Note that differences between NESOSIM and mW99 snow density are given in P2020 (Supplemental Figures S2-S4). The difference between NESOSIM v1.0 and v1.1 snow density is minimal (as this was not the focus of the v1.1. upgrades) and is expected to have a negligible impact on our thickness results, so we opt against an additional density comparison here."*

L478-481: "For example: thinner ... mean thickness." --> consider revisiting and perhaps correcting these statements in light of so far failing to adequately discriminate between total (sea ice plus snow) freeboard and sea ice freeboard in your manuscript.

See above, we were following the ATL10/NSIDC convention here of 'ATL10 sea ice freeboard' but have made various edits based on your suggestions.

L486-510: I am sorry, but I do not see the added value of anomalies computed from three winters of data. To my opinion this Figure 12 and this paragraph can be deleted without substantially changing the relevance of the manuscript.
If you decide to keep both, then I recommend to condense the text considerably and focus on those highlights that seem most obvious to be related with each other. Please keep in mind that a positive snow thickness anomaly in April does not need to coincide with a negative sea ice thickness anomaly - simply because the snow thickness during earlier in the winter determine how much the ice had the chance to grow thermodynamically. There is a temporal dimension involved which cannot be interpreted from the maps shown. Also consider to mention the retrieval noise of the parameters presented to foster readers to disentangle what is noise from what is a real signal.

We would like to keep these maps as we believe they are quite informative. We agree that much of our discussion can be condensed as we have now done, focussing on the main features, and we have also tried to make it clear that when we talk about the impact of snow here we are really referring to the impact on our thickness retrievals, you are right that snow also provides a time-varying impact on the thickness evolution we do not describe here! By reducing this we have also added more weight to the potential uncertainties with NESOSIM in regions like the Barents Sea.

New text: "*The most notable feature of these maps are the positive freeboard and thickness anomalies in 2018-2019 and negative freeboard and thickness anomalies in 2020-2021 north of Greenland and the Canadian Arctic Archipelago (CAA), the region of the Arctic where we generally expect to observe the thickest freeboard, snow depth and thickness. On more regional-scales, there are noteworthy examples of the impact of the time-varying NESOSIM snow depths on our thickness retrievals, e.g., the positive snow depth anomalies in the Laptev Sea and the Central Arctic in 2018-2019 modulate the impact of the positive ATL10 freeboard anomalies on our retrieved thickness anomalies. In contrast, the strong negative NESOSIM snow depth anomalies in 2020-2021 modulate the negative ATL10 freeboard anomalies.*"

L514-516: "In general ... thicker" --> Does this intercomparison result fit to other results where, e.g. PIOMAS data were compared to in-situ, sub-marine and ICESat data? In my mind there is this result from somewhere in the published literature that PIOMAS over-estimates thinner ice and under-estimates thicker ice - a result that is not that well confirmed by your results - particularly not for the second winter period.

Similar results are shown in Schweiger et al., (2011) in comparisons with the original ICESat mission – e.g., PIOMAS thinner than ICESat in the thickest ice regime and thicker in the thinner ice regime. Similar results were also shown in Wang et al., (2016). A similar, albeit weaker – regional difference pattern has been noted in comparisons with CryoSat-2 (Wang et al., 2016, Petty et al., 2018). We now include a reference to this in the revised manuscript. This section has also undergone changes based on our introduction of new CryoSat-2 comparisons.

L545-551: "These ice type .. this limited record" --> What I am clearly missing here in the discussion is the role of different snow thickness conditions. And this discussion could be also linked better with the 2m-air temperature and longwave radiation data from ERA5.

- What is interesting to see, for instance, is that the sea ice thickness increase between November and February is larger in 2018/19 than in 2019/20 despite a) the sea ice itself being thicker and b) the snow thickness being larger by about 5 cm. Wouldnt' one expect that thicker ice with a thicker snow load grows less thermodynamically than comparably thin ice with a thinner snow load - provided the oceanic and atmospheric forcings are roughly the same?

We think this comment is in relation to the MYI graph, as that is the only one that shows notable ice thickness differences. We don't see that thickness increase as being notably different to the other years considering the uncertainties involved so have refrained from discussion that idea in the paper. However there is perhaps some suggestion of faster ice thickness increase through the middle of winter which is perhaps counter-intuitive given the fact the ice is thicker. It is worth noting however that this is showing ice thickness change, not growth, so dynamics (convergence into the region) may be a likely cause. Our drift maps do not show anything clearly obvious in this regard but this is something an interested user could explore in the Jupyter book!

- Another observations is that from November through January the freeboard values are the same for 2019/20 and 2020/21. At the same time snow thickness is about 5 cm larger for 2020/21. This means the sea-ice freeboard which is your total (sea ice plus snow) freeboard minus snow thickness is actually smaller by 5 cm for 2020/21 than 2019/20 which would point to about 40 cm thinner sea ice during these months in 2020/21 than 2019/20 and is actually confirmed by the respective panel. So here your observations are consistent in themselves. The fact that over these 3 months there is 20 cm more ice growth in 2020/21 compared to 2019/20 despite a more or less constant snow load could be discussed more specifically in light of the balance between thicker snow isolating better while thinner sea ice isolates less and the ancillary data used (drift, ERA5).

We have added an extra note on this in the paper as it is a good highlight of the impact of dynamic snow loading:
*The MYI results also highlight the important role of dynamic snow loading in constraining regional/monthly thickness variability, e.g., the November-December 2019-2020 and 2020-2021 freeboards are near identical, but NESOSIM indicates 2020-2021 snow depths are ~ 5 cm thicker, resulting in a ~30 cm thicker ice estimate.*

- I guess my recommendation is to discuss specific observations in your data time series in a comprehensive way, taking ERA5 and drift information of Fig. 11 into account instead of kind of listing what is shown in the respective panels without connecting the information well to your observations.

We have reworded some of this analysis and also moved this figure to the Supplement, in-part to keep the focus on the thickness data and the extra comparisons we have introduced, and also because to make it clear this is not an exhaustive analysis, but a simple assessment of the link between thickness variability and core atmospheric conditions.

- Did you check, by the way, how ERA5 treats sea ice and its snow cover? How well is the snow thickness on top of sea ice resolved and is the sea ice allowed to grow and melt? Has it a variable thickness that would support discussions about feedback between 2m-air temperatures and ice and snow thickness?

ERA5, as in other reanalyses, has a simple fixed sea ice representation (1.5 m thickness, no snow but fixed albedos). This has impacts on possible warm biases which have been noted in previous studies. We've now made a reference to this in the data and analysis section of the revised manuscript.

L562-564: "For example ... " --> As always the interpretation of such maps offers potential for subjectivity and might depend stongly on the observer. I for my part rather see that the 2019/20 winter is different in terms of the strength and extent of the Beaufort Gyre - including the drift speed of its southern limb - as compared to the other two winters considered. Since the anomaly maps of these two other winters do not provide consistent information, i.e. exhibit different spatial patterns and signs of the anomalies, I am not convinced that the possible relationship between these anomalies and the ice drift pattern should be discussed the way you did. I note in this context, that the anomalies you are referring to here are mostly below 0.5 m in magnitude. How large an anomaly needs to be to be larger than the retrieval noise?

We have tried to not be overly firm in our conclusions we have drawn from analyses such as this. We are not sure there is much we can really change here as we think it best for the reader to further interpret these results beyond the analysis we have presented.

L615-617: This statement is not sufficiently well supported by the results given in the manuscript.

We have changed this from accurate to 'important role of dynamic snow loading' instead.

L619/620: "although ... scales." --> Also this part is not sufficiently backed up by the results presented in your manuscript.

We think our analysis did show the differences in seasonal cycles and we noted the regional differences (that are available in the Jupyter Book and have been highlighted in previous studies we now cite based on an earlier comment).

L638-644: I am wondering whether you perhaps could be a bit more specific here because there are snow depth observations from several sovjet drift stations and in addition the N-ICE2015 and the MOSAiC campaigns provided quite a lot of useful snow thickness data to be used for a better calibration and/or evaluation of NESOSIM.

Good idea, we have added the following: "*The comprehensive in-situ snow observations collected from recent campaigns, including the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) expedition (Wagner et al., 2022) can hopefully aid with the continued refinement of these new snow reconstructions and redistribution methods*."

L556-674: I am not sure this "sea ice thickness reconciliation" paragraph should be kept as is. My impression is that it could be potentially misleading. What I am missing is i) a more clearly formulated statement that ICESat-2 could perhaps be the benchmark sensor for sea-ice thickness retrieval rather than CryoSat-2 - simply because of the more well defined main reflection horizon and the smaller number of snow processes and properties influencing already the freeboard retrieval. ii) What could also be more in the focus of future developments is a better handling of the different spatio-temporal scales involved - both in the retrieval process but also in the evaluation of the products. All parameters, freeboard, snow and ice thickness have their specific distributions. Two issues that I do not find solved satisfactorily is the error propagation of the downscaling approach from 100 km to 30 m and the propagation of the influence of substantial difference in the acquisition time of ICESat-2 data within one grid cell. These (and others) would be enough material for further improvement and I currently do not see the need to write that much about the potential of radar altimetry. But this is clearly my personal view based on this manuscript and on the results presented in similar papers.

We included more discussion of CS-2 issues in part due to previous reviewer comments and also the increased use of CS-2 comparisons in this revised manuscript. We think these are good points to highlight.

We appreciate the ideas you include here and have attempted to include some of these in a revised section, e.g., a comment on ATL10 total freeboard providing a limit on snow loading, and the need to better understand uncertainties and error propagation.

Figure 1: I don't find this figure particularly well developed. Neither is clear that sea ice comprises level and ridged parts nor is clear that a radar measures sea ice freeboard while a laser measures total (sea ice plus snow) freeboard. The fact that the laser signal might also penetrate into the snow a bit is neglected. The role of the "internal ice stresses" is not clear, neither is mentioning of keel depth for satellite altimetry of sea ice. Not represented well is that radar waves may in fact be reflected at the snow-ice interface or even from within the sea ice but that they may also not reach to that interface at all. While

the figure is for winter, which is good, the caption refers to the key challenges and one of these definitely is to measure freeboard during summer in the presence of melt ponds. So there is in fact a lot more one could and potentially should include into this schematic figure.

It is obviously challenging to include all those issues and processes in one simple diagram. We've tried to identify the main issues in our view – the laser (total) and radar (ice) freeboard profiling, clouds scattering laser, variable topography, but we appreciate the feedback in trying to improve this.

We have thus adapted the figure to include extra issues affecting sea ice altimetry. We have also added winter to the caption as it was always our intention to focus on winter issues here. Please see updated Figure 1.

Figure 2: Particularly in the bottommost row the white circular area is not a well defined circle. What is the reason for that? Is there a higher probability for data drop-outs near the pole?
The main aim of the paper is to show the differences between rel002 used Petty et al. 2020 and rel005 used in this paper. I am wondering what can learn in terms of science from the many other panels and whether it wouldn't be sufficient and more straightforward to concentrate on the differences between rel002 and rel005 only.

The uneven hole is just a function of the binning procedure. We have since simplified this figure to show just a single row of rel002 to rel005 coverage then just the rel002 to rel005 difference. We have included the original figure (which includes all rel002 to rel005 differences) in the SI.

Figure 3: What about the median biases for both cases? I note that (b) shows more cases than (a). Why?

Mean bias is a common diagnostic for comparisons like this so we prefer to stick with that for clarity. The SD captures deviations beyond that. The reason b showed more grid-cells in the comparison is because of slight differences in the number of valid grid-cells from the different configuration. As 2 is negligible in terms of its impact on these statistics, we have dropped N from the panels.

- Personally, I would make the lengths of x- and y-axes the same when the data range is the same - here 80 cm for both quantities shown.

We did not have time to re-do this figure but the aspect ratio seems close enough to 1:1.

Figure 4 c) What is the reason for the violins for Oct. through Jan. to look cut off at the side facing higher snow thickness values?

This is simply a product of the KDE fitting procedure (the violin plot cuts off outside of the raw data input), but is still able to highlight the skewness of the distributions.

Figure 5: This map contains many regions that are not used in the paper. It might make sense to color all regions not used in the same color as open water, only explain the acronyms / names of the regions actually used, and add that this map is "adopted" from a region mask provided by ...

Agreed, we have changed this figure accordingly and decided to keep the colors but only annotate the regions used in our study domain.

Figure 6: The x-axis title needs to be changed into "total freeboard". Please provide the meaning of the dashed lines in the caption.

Dashed lines are the means and we have now included that in the caption. Thanks!

- I strongly recommend to be consistent in the notation between the figure and the caption. You write r002, r003 and so forth in the figure but write "rel002" and "rel005" in the caption text; this is inconsistent. Similarly, in the figure you write "bnum1", "bnum3" and so on, in the caption text you write "beam #1, #3" and so on. This is inconsistent as well.

Good point, we have rectified this in the revised version.

Figure 9: I note that the sea ice thickess uncertainty seems not to be influenced by the substantial differences in the "mean day of the month" varying over a very short distance the ICESat-2 data. Hence in one 25 km grid cell ICESat-2 might come (on average) from the first few days of a month while in the adjacent grid cell this data might stem from the end of the month. I would expect that thermodynamic sea ice growth and dynamic processes can add substantially to the uncertainty - most likely especially during months October through January. Did you check this?

This is a good point, and is something we have considered, although it is hard to exactly quantify this. It is our belief that with more data we can start to understand expected monthly changes in thickness and quantify this sampling error, which we hope to include in a future data release.
*"We hope to incorporate a more comprehensive and accurate accounting of the various error contributions where possible, e.g., accounting for the clear sampling/representation error associated with grid-cell data differences, while also exploring more sophisticated uncertainty quantification methods (e.g. Monte Carlo approaches)".*

Figure 14: I guess it would make sense to not show values for September but rather begin with October as this is when the freezing season commences. There is room for one more panel in which you could put the fraction of the FYI relative to the total sea ice area of your Inner Arctic region. The same comment applies to Figure 15, but here for the MYI fraction of course.

This is a good point, we have dropped September from the FYI figure and made a note on this in the caption and manuscript. We don't feel the need for another panel on FYI vs total area as interested readers can make this out from the MYI fraction in Figure 10.

Pure typos / editoral comments:
A general editoral comment upfront: Since you advertize at the end that basically all your results and the processing is available as jupyter notebook you might want to scan you paper and reduce the number of repeated mentioning of this. I think you overdo it a little bit.

L28/29: "Mean first-year ... negligible" --> possibly you refer to changes or differences?

Added differences, thanks.

L139: "2we" --> "2, we"

Changed, thanks.

L218: "consensus" --> In the modeling world one would speak of an ensemble mean - or in your case apparently an ensemble median. I would find this a better expression because "consensus" is something that usually involves some discussion, some weighing perhaps, some additional constraints. What you do, however, is taking mean and median, i.e. use a statistical tool. Hence, I would recommend to change the wording accordingly - here and later in the text.

Agreed, changed this language to refer mainly to 'median' gridded OIB. data.

L231: "halving the blowing snow open water coefficient" --> would you mind to show the respective equation in this paper as well so that everybody immediately understand the effect of what you stated here?

Yes we have now added a new Eq. 1 to the revised manuscript.

L247-249: You are providing enough information about this region map in the context of Figure 5 I guess and can delete the sentence "Out Inner ... Figure 5" and instead refer to this figure at the end of the previous sentence.

Agreed, removed.

L304-306: "Our monthly ... ATL10 data" --> namely what ancillary data?

The ancillary data is what follows in this sentence, the mean day and number of segments.

L309: "on to" --> "onto"

We do not think this correct so have kept the language as is.

L311-317: This reads a lot like again being in a technical report. Have you thought about publishing this manuscript in the journal Earth System Science Data? Would that be a more adequate journal for your manuscript being comparably rich in processing and product details, bug fixes, novel masks, and so forth? I slowly begin to second that other reviewer who had difficulties to see the merit of this manuscript in The Cryosphere.

See initial discussion and new analysis included!

L374-376: Since you do not consider June I suggest to remove June and the respective value from this sentence.

Agreed, removed.

L407: Here you write rel003, in L413++ you write rel005. What is correct?

It should be rel005, this was changed in the revisions but not updated here. Corrected now, thanks for spotting that.

L413-424: I am wondering whether you need to list all ranges for all parameters shown in Figure 8. Do you see a chance to simply let speak Figure 8 for itself and perhaps only put the highlights here - perhaps along the lines that correlation coefficients in general increased by around 0.1, that biases change (not necessarily reduce) by about 30 cm and that standard deviations reduce by, on average, 0.2 m. This might make this part of the paper easier to read.

Agreed, we have condensed this dramatically as we have now included the new CryoSat-2 thickness comparisons. See discussion at the start.

L417: "58" --> "0.58"
Thanks, changed.

L434: "i-j" --> "i-k"

Thanks, changed.

L450/451: "The data within this domain ... Figure 10e)" --> You need to put labels a) to f) to the panels in Figure 10 if you want to refer to them.

Thanks, added.

- I cannot see any sea-ice concentrations within the range given here (40-60%) in your Figure 10 e; something needs to be corrected here.

Thanks, this was an error that has been corrected in the revised paper.

L533: "snow depth in November ... winters" --> you could denote in addition, that in Feb to April, however, differences in the snow thickness on the sea ice are between 2 cm and 4 cm.

Agreed, we've added this extra detail now.

L611: I suggest to use "significant" only in the context of statistical tests which I could not see in your manuscript. Hence "considerable", "notable", or "substantial" might be a better wording here.

Agreed, we've removed as much use of significance here and elsewhere and replaced with phrase slike notable, or large.

L654/655: "The extension ... estimates" --> While this is true it would require first and foremost that we have (more) accurate freeboard estimates during summer from ICESat-2?

Thanks, we've added this clarification and a comment on the recent summer cal/val campaign. "*together with improved understanding of the performance of ICESat-2 over the complex summer melt surface for summer freeboard determination (Tilling et al., 2020). A recently completed 2022 ICESat-2 summer airborne cal/val campaign should also provide important insights towards this goal*."

L920: "now depths" --> "snow depths"

Thanks, changed.

L958/959: Perhaps better (and more correct): "The background dark grey shading in panels a to k is the CDR sea ice concentration shown in panel l."

Thanks, changed.

**References**

Glissenaar, I. A., J. C. Landy, A. A. Petty, N. T. Kurtz, and J. C. Stroeve (2021), Impacts of snow data and processing methods on the interpretation of long-term changes in Baffin Bay sea ice thickness, The Cryosphere, 15, 4909–4927, doi:10.5194/tc-15-4909-2021.

Landy, J. C., Dawson, G. J., Tsamados, M., Bushuk, M., Stroeve, J. C., Howell, S. E. L., Krumpen, T., Babb, D. G., Komarov, A. S., Heorton, H. D. B. S., Belter, H. J., and Aksenov, Y.: A year-round satellite sea-ice thickness record from CryoSat-2, Nature, 609, 517–522, https://doi.org/10.1038/s41586-022-05058-5, 2022.

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., and Haas, C.: A weekly Arctic sea-ice thickness data record from merged CryoSat-2 and SMOS satellite data, The Cryosphere, 11, 1607–1623, https://doi.org/10.5194/tc-11-1607-2017, 2017.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R. (2011), Uncertainty in modeled Arctic sea ice volume, *J. Geophys. Res.*, 116, C00D06, doi:10.1029/2011JC007084.

Wang, X.; Key, J.; Kwok, R.; Zhang, J. Comparison of Arctic Sea Ice Thickness from Satellites, Aircraft, and PIOMAS Data. Remote Sens. 2016, 8, 713. https://doi.org/10.3390/rs8090713