Response to Referee comment on "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 Alek Aaron Petty et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2022-39-RC1, 2022

Original referee comment in black, our responses in blue

General comments

The paper assesses the impacts of a number of changes to ICESat-2 ATL10 processing and to the NESOSIM snow model on estimates of along-track and gridded sea ice freeboard and ice thickness. This assessment is important for users of high level sea ice products such as ATL20 gridded sea ice. Overall the paper is well conceived and written. However, there are a nuber of issues that need to be addressed before the paper is ready for publications. I list these below. I also have a number of specific comments.

We sincerely appreciate the time taken by the referee to provide the review of our paper and the thoughtful suggestions. Please see below for our responses.

Overall, the quality of the figures is good. However, some of them could be improved by adding descriptive titles/labels to each panel. For example, figurure 7 has titles but these appear to be file variable names. Rather than "ice_thickness_unc", it would be more helpful to readers to have "Ice thickness uncertainty" spelled out. Likewise with panel (i) "ice thickness int" would be better as "Interpolated ice thickness". The authors might also want to think about a better layout and if all panels are necessary.

Yes, good points about the panel labels. We have improved and/or added these in the revised manuscript. We have decided to keep the panels as is other than edits/changes made based on other review comments. We have also removed the ice type panels from the time-series plots based on a later suggestion.

The Jupyter notebook is an excellent addition as is making the code available.

Thank you, we are excited about this aspect of our work!

Different releases are used for different evaluations. The authors show that there is little difference between releases 003 through 005 but it would make for a cleaner, and more up to date, analysis to use release 005 throughout. The only exception being to show differences between releases 002 and subsequent releases.

We have now updated the CryoSat-2 comparisons to use rel005 instead of rel003 data (the only place other than the release comparison where rel003 was used). Differences were negligible and we have updated the figures in the text accordingly. Based on a later comment we have also updated the CryoSat-2 comparison figure to improve readability. We still use rel004 in the snow comparison analysis however as that involved various different configurations that were all varied out during rel004 processing. Re-doing this would be very time-consuming and not add much value considering the negligible freeboard differences between rel004 and rel005.

I would like to see a map in the main paper showing the "Inner Arctic Ocean" region as the study region introduced as part of the methods. This would focus readers attention on the analysis region up front.

Agreed, we have now added this to the main manuscript as Figure 5.

Figure 8 is another example of a figure that would benefit from having labels such as a) sea ice freeboard. Parameter names are on the y-axes but they are small. Panels a, b, etc should be referenced in the text.

We have now added these figure panel labels to Figure 8, and also to Figure 12 and 13. We did not feel it was necessary to provide further labels as these overly cluttered the figures when we tried this.

There are a number of places in the text where important statements are put in parentheses. I think it would improve readability to rewrite these statements as part of the main text. Some of these parenthetical statements are unnecessary.

Agreed, we have removed several of these parentheses from the manuscript. It is a bad habit of the primary author!

Specific comments

L60. "is *being* developed"

Added, added.

L63. Suggest "collected to estimate sea ice thickness"

Added, changed.

Section 2. I think it would be helpful to summarise upgrades to IS2 processing, NESOSIM and ATL20 gridding in a simple table.

We didn't feel like a table was the best place to provide all this information. We currently include the Table (Table 1) summarizing the NESOSIM configuration upgrades, but IS2 processing upgrades were more descriptive so we think it's best to keep this to the description already included in the manuscript.

L111 prefer "km" to be consistent.

We followed the convention of the NSIDC user guides here and feel this is easiest to interpret for the user.

L124 "0-3 cm freeboard changes at basin scales". Does "basin-scales" refer to the Inner Arctic region used in the current paper? Maybe say "an increase in basin average freeboard of up to 3 cm.

Added, changed.

L139 Suggest "New releases of ATL07 and ATL10 also reflect upgrades to the underlying ATL03 processing, such as improvements in geolocation.

Added, changed.

L141 and 110. ATBD for ATL07/10 use "surface reference" rather than "reference sea surface". To avoid confusion it might be better to use the same terminology as the ATBD.

Added. We have used “sea surface reference” to be more consistent with the ATBD which sometimes uses this and sometimes shortens it to just sea surface.

Figure S1. Would it be better to have this figure in the main text? Also, the point here is that the number of reference surfaces is reduced from rel002 to rel003 because dark leads are not used. However, the count difference is positive. It make more sense to me to have this reduction as a negative number.

Agreed, we have moved this to the main manuscript and changed this to show counts relative to rel002. Now Figure 2.

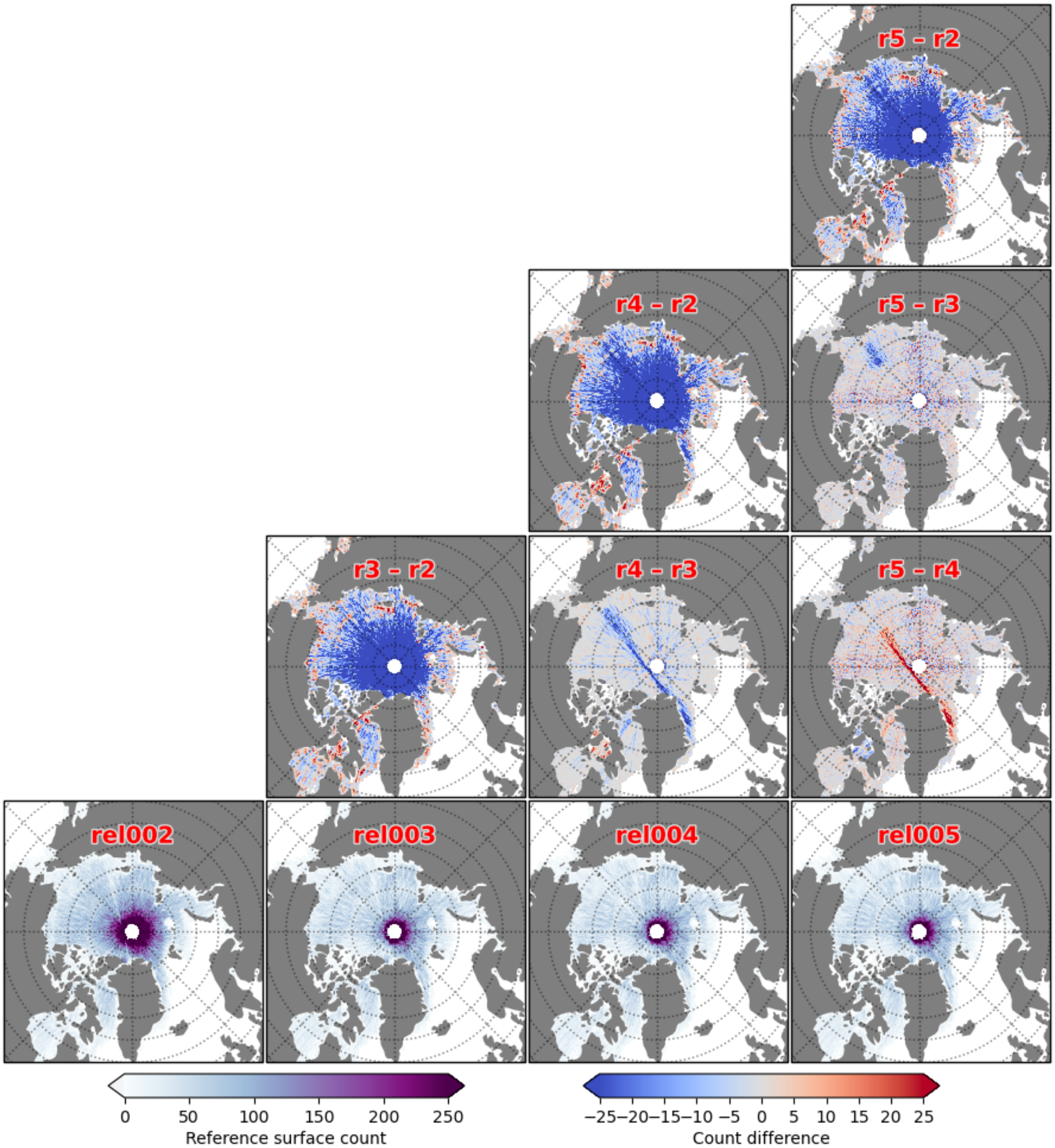


Figure 2: (bottom row) number of 10 km along-track reference surfaces from the three strong beams from November 2018 to April 2019 for Release 002/rel002 (left) to Release 005/rel005 (right). Panels above show the difference in reference surface counts from releases rel003, rel004 and rel005 relative to rel002.

L190 Effectively the /beta and /gamma terms in equation 1 are corrections to solid precipitation. It is not clear to me what the difference is between the two terms. They could be combined into a single loss coefficient.

We introduced the atmosphere snow loss term as an added blowing snow to atmosphere loss term as we expect some fraction of snow that is disturbed during high winds will be sublimated away to the atmosphere, which we developed to be an added function of the new unitless atmosphere loss parameter and not a function of ice concentration.

We agree that for simplicity it makes more sense to combine these coefficients in this equation which we have now done. We now refer to a “blowing snow atmosphere loss coefficient” which has units of s^{-1} as in the now renamed blowing snow open water loss coefficient. The value has been reassigned by combining the two terms into one, now with a value of 2×10^{-8}.

L217 Do you mean "For each OIB snow depth product, snow depths are binned into 100 km grid cells using a drop-in-the-bucket averaging procedure. For each grid cell, the median snow depth of the three products is then assigned as the grid cell snow depth". So in all cases, you are taking the middle value. If the number of products was larger, I can see this as an acceptable approach to avoid outliers but for just three values, you can't really identify an outlier. It would seem that the mean is a better estimator.

Yes, that is the correct interpretation. We decided on using the median and not the mean as we are aware of some odd behavior in some regions and some dates for some of the algorithms. For example, Kwok et al., (2017) shows that the SRLD-derived snow depths exhibit a strong positive trend including very thin snow depths at the start of the OIB time-series, which appears unrealistic (as also noted by Kwok et al., 2017). We admit that three products is not much of a distribution to truly assess outliers, but we were keen for our results to not be consistently skewed by one algorithm being constantly biased compared to the others, hence our approach here.

L 230. "within reason" This needs some clarification. Are there limits you can set on depth or start date?

This is a concept we are currently exploring in the more sophisticated NESOSIM calibration efforts cited in the paper (Cabaj et al., 2021). Due to the lack of modern and reliable late summer/early fall snow depths at basin scales our knowledge is mainly heuristic: Arctic sea ice refreeze generally begins sometime in September following the Arctic sea ice extent minimum.

In response to this comment and a comment from reviewer #2 we have included more discussion on the SnowModel-LG end-of-summer snow depths, which generally show a complete removal of snow in August in those simulations, with snow depth increasing in September onwards.

“as another tuning parameter, constrained mainly by limited evidence in the literature. For example, the Warren et al., (1999) climatology (W99) shows a mean snow depth of 3 cm in August including depths of up to 8 cm near the Greenland/Canadian Arctic coastline based on the quadratic fit to observations. However, output from SnowModel-LG presented in Stroeve et al., (2020) shows zero snow depths in August in the earlier (1985/1986) and later (2015/2016) time periods of that time-series. As NESOSIM includes no snow melt terms, we prefer instead to initialize later in the year (Sep 1st) and prescribe an expected end of August mean snow depth based on our original temperature scaled W99 August climatology.”

Figure 3. The left panel is busy. I suggest having a separate panel for October and April. The horizontal grid-lines should be lighter or removed.

Agreed, we have now split up the panels into October and April which has increased the readability of the figure.

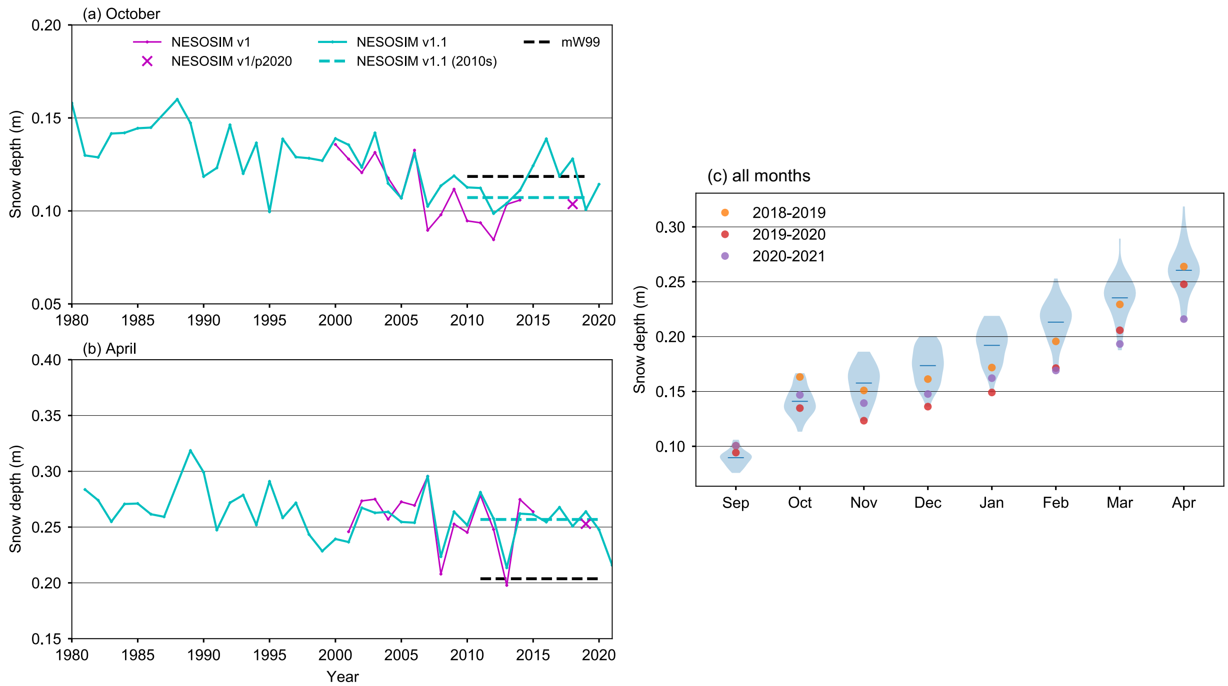


Figure 4: (Mean Arctic snow depths in October (a) and April (b) from NESOSIM v1.0 and v1.1 within an Inner Arctic Ocean domain (Figure 5). NESOSIM now depths are also masked where concentration (from passive microwave) is less than 25%. The cross markers show the extended ICESat-2 NESOSIM v1.0 results used in (Petty et al., 2020). The dashed cyan horizontal lines show the NESOSIMv1.1_2010-2020ave snow depths averaged across the respective month, while the dashed black lines show the modified Warren climatology (mW99) in October and April respectively for regions of coincident NESOSIM v1.1 coverage. (c) violin plots showing interannual distributions of monthly mean snow depths from NESOSIM v1.1 within an Inner Arctic Ocean domain from 1980-2021, colored markers indicate mean monthly snow depths for recent (ICESat-2) years.

Note that there was an error in the original figures in how the Inner Arctic Ocean region mask was applied which = has been fixed for these new plots. The main difference is a small shift in all snow depths and a small relative increase in the April mW99 snow depths, which were lower in the original manuscript, and some slight differences to the violin plot ICESat-2 snow depths.

We have also added new maps to the SI (now Figure S1) showing maps of the differences between the NESOSIM 2010-2020ave and mW99 snow depths for October and April too – highlighting the regional differences including strong differences in the Kara Sea region especially.

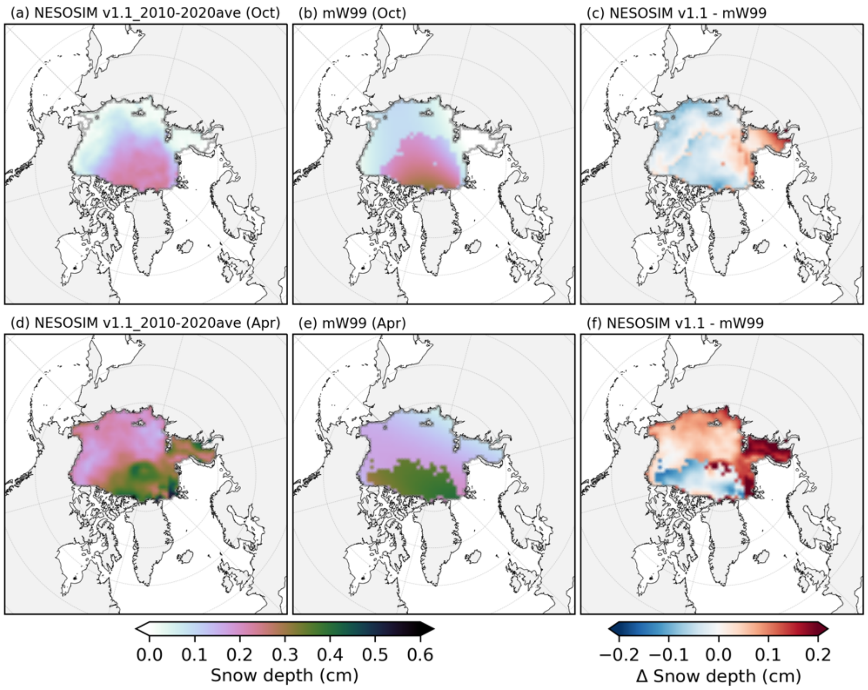


Figure S3: Comparisons between NESOSIM v1.1_2010-2020ave and the modified Warren snow depth climatology (mW99) within our Inner Arctic Ocean domain. The mW99 snow depths are limited to valid NESOSIM grid-cells. Top row shows October (Oct) comparisons while the bottom row shows April (Apr).

L254 One of the arguments for not using the Warren climatology for snow depth is that it is not representative of the present day conditions. The previous paragraph and Figure 3 have been used to argue that recent years snow depth are also lower than average and may be declining. So why would you use a climatology of NESOSIM.

We were wrong to refer to this as a climatology. We now refer to this as a 'modern era representation' and have labelled this dataset NESOSIM_v1.1_2010-2020ave to be clear we only use data from the recent 10-year period to better reflect current conditions.

Wouldn't using output from an operation product or low latency reanalysis be a better option?

It is our understanding that there is a wide-spread in snow depths across available operational products which are generally not calibrated to available observations (e.g. Operation IceBridge). We agree there is potential there but are unaware of a reliable product for this purpose with consistent output.

L266. The redistribution method needs a reference.

Added the Petty et al., (2020) and original Kurtz et al., (2009) citation.

L296 The smoothing/gridding procedure needs more explanation. It would be helpful to say why each of the steps are done. Why use Delaunay triangulation - generally this method is used to interpolate unstructured data? Presumably the KDTree algorithm is to speed up the search for neighboring cells.

We have reproduced all maps and the time-series analysis with the non-interpolated data. This was previously a mix, however our testing has shown the difference in results to be negligible when considering basin-scale aggregates. Our main motivation for introducing the interpolation was to fill in the pole hole and increase spatial coverage and mitigate spatial biases when comparing across months/years, especially considering some of the declines in coverage since the removal of dark leads from the freeboard derivation. However, we admit our method here is crude and do not wish this to distract from the main analysis which is now entirely based on non-interpolated gridded data and we have removed and reordered the discussion of the gridded dataset and the interpolated variables.

L467: "We also include in this Version 2 IS2SITMOGR4 dataset smoothed and interpolated variables of freeboard, snow depth and thickness in an initial attempt to fill in the pole hole and mitigate the spatial sampling biases. These preliminary variables are not used in the subsequent analysis presented here but are available to interested users and shown in the Jupyter Book discussed below. We expect that future work will explore more sophisticated interpolation procedures and blending with other thickness datasets, which we discuss more in the summary section."

Some notes on the method: Delaunay triangulation interpolation benefits from being a flexible interpolation procedure for our needs. The input data is already gridded but the interpolation is able to fill in gaps across variable sizes and directions, e.g. gaps across vertices and more than one grid-cell away. Although we admit simpler algorithms could have been used, this is already a built-in function with the core scientific Python ecosystem making it easy to implement for our needs. Similarly, the KDTree algorithm is perhaps overkill as the problem being solved (finding distance of each grid-cell to the nearest raw valid grid-cell) is a relatively simple one considering the fact the data is already gridded and not too large. We've decided to drop the mention of KDTree as that isn't needed here. Note that the scripts used to generate the entire gridded datasets will be made available so interested readers will be able to reproduce our exact methodology if needed. The interpolation code is already contained in the Jupyter Book which we have now highlighted in the revised paper.

Figure 4, L343. How do these look for other months and for other years? No need to show them but a comment in the text would be helpful.

It was actually challenging to maintain all the needed datasets across different versions and releases so we only have these months on-hand to discuss. We feel they represent the key changes through the winter accumulation season. We look at monthly changes in the gridded comparisons which show no obvious step-change or anomalies in the results.

L354. Significant or major?

Agreed that is better language, made the change.

L356. Prefer peak rather than mode. Mode could be confused with operating mode.

Agreed that is better language, made the change.

Figure 4. How many segments are used to generate these plots? Are dark leads more common in November?

The issue of lead counts and dark vs specular in ATL07 was discussed more in Kwok et al., (2021) and Petty et al., (2021). PDFs of January, June and October 2019 SSH separated by leads counts are shown in Figure 3 of Kwok et al., (2021) showing specular lead counts dominating over dark lead counts. Dark lead counts increased through to January as overall coverage increases also.

In our thickness processing we do not store information regarding the surface type (e.g. specular or dark lead) as we're only using freeboard and associated variables, e.g. segment length, to generate our thickness estimates. We could add the freeboard segment counts to the plots if desired - it is a lot, but drops from rel002 onwards as the ref surf count plots highlight - but we do not believe this would add much value.

L370. The name NESOSIMv1.1clim has not been introduced yet.

Agreed. In Section 2.2.1. we have now referred to this as the modern-era mean and included the relevant label 'NESOSIM v1.1_2010-202ave' here and elsewhere in the manuscript and figures.

L389. Suggest "In Figure 6, we show the correlation coefficients, mean bias and standard deviations of ICESat-2 monthly gridded ice thickness from rel002 and rel003 compared with ESAs CryoSat-2."

Agreed, we have used this and also added a bit more clarity to this and the following sentence.

What are the standard deviations of? Why mask data less than 0.25 m?

Standard deviation of the differences, so the standard deviation between the two after removing the mean bias. We have clarified this based on the previous comment.

Also added "to focus more on the representation of consolidated pack ice between the two sensors, rather than the added complexities of thin and marginal ice."

Figure 6. I suggest removing the shading and, for each month, plot release 002 and release 003 as separate columns. That way you can see the ovelap. The shading suggests the data is continuous rather than discrete monthly data.

Agreed, we have now reproduced these plots as bar plots which we think has improved the readability. Based on the earlier comment these are now also base don rel005 ATL10s.

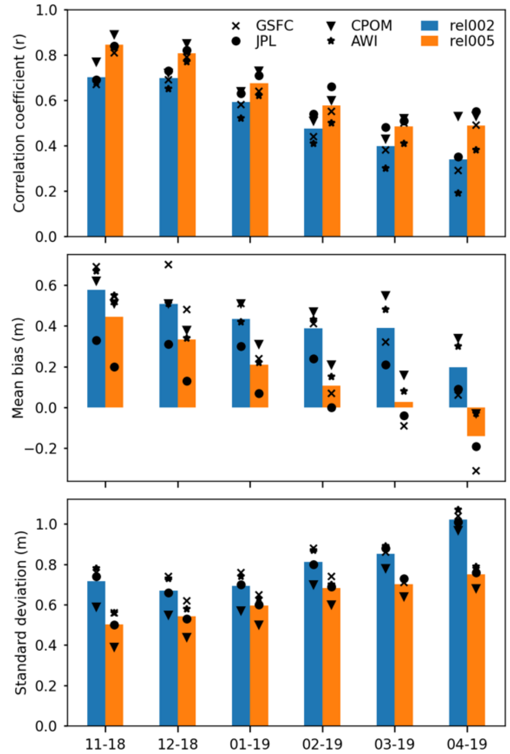


Figure 8: *Comparison statistics of monthly gridded CryoSat-2 thickness for four different CryoSat-2 products (GSFC, JPL, CPOM, AWI) with monthly gridded ICESat-2 sea ice thickness using rel005 ATL10 and the same snow loading and ice density input assumptions from November 2018 (11-18) to April 2019 (04-19). Data are compared within our Inner Arctic Ocean domain and for grid-cells in both datasets that contain thicknesses > 0.25 m.*

L445. NESOSIM presecribes snow density for new and old snow. The bulk density is a weighted average of these two values. How much can be read into variations in density?

This is a fair point, the parameterization is crude and we offered this up more to understand its impact on thickness. We have added the following:

"Due to the crude nature of the NESOSIM density parameterization, we do not view this analysis as a reliable interannual snow density assessment but highlight this more to understand the density variability impact on our ice thickness estimates."

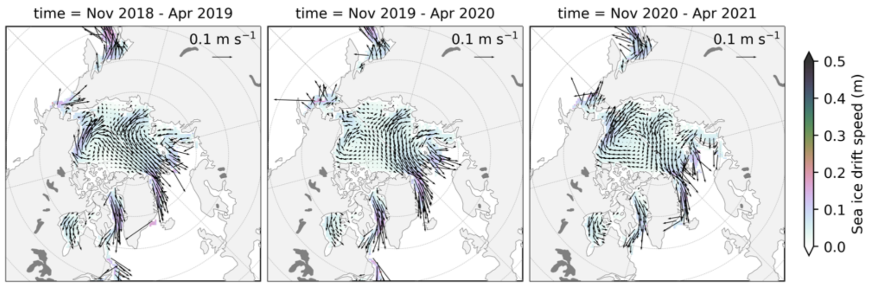
Figure 8. Why is sea ice concentration lowest in October? Is this an artifact of averaging.

This is largely an artifact of coverage changes due to sea ice refreeze and the fact we don't include grid-cells with a mean concentration <50% in ATL10. We have added the following:

"Note that the concentration decline from September to October is due to changes in data coverage as regions with ice concentrations < 50% are not included in ATL10. "

Figure 9. The flow vectors obscure the thickness data. They are not really discussed. Are they necessary? Could they be relegated to supplemental material?

We have now added these overlaid on drift magnitude as a new row in the figure (now Figure 11).



Line 535. Care needs to be taken with ERA5 (or any reanalysis) near-surface variables over snow. ERA5 snow parameterisation is still a single layer, which does not produce realistic surface fluxes (Arduini et al 2019).

Agreed, we did not include surface turbulent fluxes partly for this reason.

L540. Are three years of data enough to make a statement about strength of coupling?

Our point here was that near-surface conditions are generally expected to be strongly coupled with the sea ice state. We did not mean to imply we are discovering this coupling by our analysis here. We have re-worded this line and other elements of this paragraph to make clearer what our limited analysis has shown.

L591. This seems to contradict what is shown in Figure 4.

We state height biases here which are different to the relative measurement of freeboard. We have added ‘absolute’ to make this clearer.

Figure 12 and 13. The multi-year ice fraction panel is not needed.

Agreed, we have dropped this and the FYI panel in the new figures.

Arduini, G., Balsamo, G., Dutra, E., Day, J. J., Sandu, I., Boussetta, S., & Haiden, T. (2019). Impact of a Multi-Layer Snow Scheme on Near-Surface Weather Forecasts. Journal of Advances in Modeling Earth Systems, 11(12), 4687–4710. https://doi.org/10.1029/2019MS

References

Kwok, R., Kurtz, N. T., Brucker, L., Ivanoff, A., Newman, T., Farrell, S. L., King, J., Howell, S., Webster, M. A., Paden, J., Leuschen, C., MacGregor, J. A., Richter-Menge, J., Harbeck, J., and Tschudi, M.: Intercomparison of snow depth retrievals over Arctic sea ice from radar data acquired by Operation IceBridge, The Cryosphere, 11, 2571–2593, https://doi.org/10.5194/tc-11-2571-2017, 2017.

Kwok, R., A. A. Petty, M. Bagnardi, N. T. Kurtz, G. F. Cunningham, A. Ivanoff (2021), Refining the sea surface identification approach for determining freeboard in the ICESat-2 sea ice products, The Cryosphere, 15, 821–833, doi:10.5194/tc-15-821-2021

Petty, A. A., M. Bagnardi, N. T. Kurtz, R. Tilling, S. Fons, T. Armitage, C. Horvat, R. Kwok (2021), Assessment of ICESat-2 sea ice surface classification with Sentinel-2 imagery: implications for freeboard and new estimates of lead and floe geometry Earth and Space Science, 8, e2020EA001491. doi:10.1029/2020EA001491.

Response to Referee comment on "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 Alek Aaron Petty et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2022-39-RC2, 2022

Original referee comment in black, our responses in blue

This paper discusses improvements in the ICESat-2 (IS2) processing of sea ice thickness retrievals from different releases of the IS2 products. As such it feels more like a NASA technical report that discusses how the different versions change the thickness retrievals. The author previous published in 2020 on the processing chain to IS2 and I do not find that with the changes this now warrants an updated assessment of thickness changes and a new publication. The question is what do we really gain from this paper vs. having a NASA technical report on the changes in data processing?

This is in part because any sea ice thickness (SIT) assessment depends strongly on the choice of snow loading used. It also depends strongly on the choice of snow depth processing applied to OIB data for validation of your snow loading, and the seasonality of this validation period. It seems that with the changes presented to NESOSIM there are minimal changes anyway to the snow loading and thus it is the changes to the lad detection that seem to have the largest influence. To make this paper more impactful and not just a technical report on updates to IS2 data processing, one way forward could be to assess the choice of snow loading in the IS2 SIT retrievals. Since Zhou et al. (2020) already showed how different these data products can be, and other studies such as Mallett et al. (2021) and Glissenaar et al. (2021) detailed how using different snow loading can lead to different trends, one really cannot trust any assessment of thickness changes over the 3 years evaluated here without addressing the uncertainty in the snow loading. How would Figures 8-10 look using different snow data sets for example? You state that it's the freeboard processing that results in the largest changes (again indicative that this should be a technical report), but given the wide variety of snow depth data sets out there, the 3 years analysed here may be quite different depending on data set applied. And is analysing 3 years of data really useful for assessing drivers of SIT variability? At the moment I really do not see much value in having this as a publication in The Cryopshere for an incremental update to the IS2 processing chain. That doesn't mean it shouldn't be published someplace, but The Cryosphere should be for more impactful papers.

We thank the reviewer for taking the time to provide this review. There are a few general issues raised here that we address first:

Technical report: we feel that a peer-reviewed paper in The Cryosphere is a highly suitable place for this work. The thickness data we present here is not an official mission product so has no such Algorithm Theoretical Basis Document (ATBD) or detailed technical reporting infrastructure in-place. The associated NSIDC user guides for both the along-track and gridded datasets shown here provide only top-level information regarding data production and notable changes as in other NSIDC products. All official ICESat-2 products including ATL07 and ATL10 are described in Algorithm Theoretical Basis Documents (ATBDs) which include change logs and descriptions of updates to the underlying data processing. However even they do not include results of these changes on the data output or downstream impact assessments, hence the need for papers like Kwok et al., (2021, The Cryosphere) for highlighting the rationale and impacts of algorithm changes on basin-scale freeboard distribution.

A major aim of this paper is to highlight the changes to the ATL10 freeboard product across several releases (rel002 to rel005), together with updates to NESOSIM and, most importantly, the impacts of these changes on our estimates of winter Arctic sea ice thickness. As freeboard is largely measured by satellite altimeters like ICESat-2 towards the goal of inferring estimates of sea ice thickness, we believe a manuscript detailing these changes and their impacts is highly warranted and also scientifically insightful, especially considering the three years of data we now have and show from ICESat-2.

We do not plan to assess every new release of ATL10 in this manner, but considering this is the end of the ICESat-2 3-year prime mission period, assessments regarding data quality and impacts on higher-level products are urgently needed considering the importance of this mission and potential for improving our understanding of sea ice conditions.

Differences in snow loading: the primary author was involved in both Zhou (2020) and Glissenaar (2021) studies so is well aware of issues surrounding snow loading uncertainty and impacts on thickness. We use NESOSIM as it is configured to produce daily data across the entire Arctic Ocean including its peripheral seas and data is available for our entire study period. NESOSIM v1.1 was calibrated against a new consensus snow depth estimate from OIB giving us additional confidence regarding its reliability compared to other products available (some are calibrated, others are not). The Zhou (2020) study showed that NESOSIM output was largely consistent with other products but assessments of accuracy are still hindered by the lack of contemporary ground-truth data. We have added the following to the summary section:

"Recent studies leveraging newly generated Arctic snow reconstructions and satellite-derived data products, including the joint ICESat-2/CryoSat-2 derived snow depths, are helping collectively provide new insights into snow depth variability and its impacts on sea ice thickness and its contribution to total thickness uncertainty (Zhou et al., 2021; Mallett et al., 2021; Glissenaar et al., 2021). While these datasets, including NESOSIM, are still generally limited by a lack of contemporary ground-truth data for assessing data accuracy, the creation of new operational, i.e., continuously updated and disseminated, snow products should help enable more comprehensive assessments of systematic snow loading uncertainties."

Our derived thickness data includes an estimate of thickness uncertainty which includes a contribution from both random and systematic uncertainty from snow loading. We use NESOSIM together with the modified Warren climatology to deduce the latter which is represented in the shading in our thickness time series plots together with the other potential sources of systematic uncertainty. A similar paper using CryoSat-2 with ICESat-2 to infer snow and thickness concurrently for the three-year period of data we have was also published in GRL after this discussion period started (https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2021GL097448). They show very similar thickness results to our NESOSIM-derived estimates which we now highlight in our revised manuscript (in the results and in the summary) as an indirect validation of our results. It is also worth noting that this analysis did not include any thickness uncertainty estimates.

It is our belief that the increased focus on snow on sea ice in recent years will help provide a more complete estimate of its contribution to thickness uncertainty, but more work is needed to ensure timely and consistent data access as we now discuss more in the summary.

Three years of data: That is unfortunately all we have from ICESat-2, and we strongly feel that the results we show will be highly informative to a wide spectrum of readers interested in Arctic sea ice variability.

More specific major comments:

It is stated that NESOSIM is updated to use ERA5 calibrated against CloudSat and a new blowing snow term. However, there is no validation of this blowing snow loss term, or discussion on how the coefficients, i.e. wind action threshold, blowing snow loss coefficient and atmosphere snow loss coefficients are derived and validated. There is no in situ evidence that a significant amount of snow is lost to leads in the winter (any lead in winter quickly refreezes in a matter of a few hours), and there is no assessment here of the magnitude of this new snow loss term, and comparison to the old (and presumably still used) snow loss term to leads. Since SIT retrievals depend very strongly on the snow loading, at a minimum some quantitative analysis is needed on what these changes represent in terms of the overall snow mass, and some science justification is needed for doing this in the first place. It seems that some artificial tuning is based on trying to reduce the mean difference with OIB snow depths, but of course those are not perfect either. And they are done only in the springtime, and the question is how valid this bias- correction is for other months during the winter season?

We discuss the rationale behind NESOSIM development in the original manuscript (Petty et al., 2018) including the use of the snow loss terms as largely unconstrained free parameters. The maps included in Petty et al., (2018) show that the impact of the blowing snow lost to leads term, which we now refer to as "blowing snow open water loss" is isolated to regions of lower concentrations in the more peripheral Arctic seas where lead counts and widths are higher and large stretches of open water are prevalent, and where temperatures are warmer and winds can be stronger too.

We have added an additional comment about this in the revised manuscript:

L196: "As discussed in the original NESOSIM study (Petty et al., 2018), these snow loss terms are crude representations of complex physical processes that we introduce primarily to remove snow and improve correspondence with the limited observations we have for calibration purposes.

Based on review #1's comments we have also re-worded and simplified the discussion of the new blowing snow atmosphere loss term.

The author is wrong about what SM-LG does at the end of summer as it keeps the snow cover in places where it doesn't entirely melt out. Also, snow can start to accumulate before September in the Arctic, and thus it seems these changes are made purely to reduce your bias but there is no physical reason to justify these changes. I do not think that because NESOSIM matches mW99 in October that you can conclude you have "good" snow depths. In fact given delays in freeze-up, I would expect much thinner snow in October compared to mW99 based on the fact that ice is forming later than it used to.

We were incorrect when we stated that SnowModel explicitly removes all snow at the start of the simulation year (August 1st) as yes, theoretically, the model converts snow that is isothermal (0 ^{o}C) and saturated with meltwater at the end of a given simulation year to superimposed ice and enables the remaining snow to persist through to the following year. However, the related manuscript showing the output from SnowModel-LG (using ERA5 and MERRA-2 forcings) by Stroeve et al., (2020) shows zero snow depths across the entire Arctic in August and in some cases no snow in July either (Figure 2 of Stroeve et al., 2020). We have re-worded this line accordingly in the revised manuscript.

"For example, the Warren et al., (1999) climatology (W99) shows a mean snow depth of 3 cm in August including depths of up to 8 cm near the Greenland/Canadian Arctic coastline based on the quadratic fit to observations. However output from SnowModel-LG presented in Stroeve et al., (2020) shows zero snow depths in August in the earlier (1985/1986) and later (2015/2016) time periods of that model output."

We have provided an updated comparison of our modern-era NESOSIM mean output with modified Warren (Figure S3 in the SI) which shows thinner October snow in NESOSIM across much of the Central Inner Arctic, but thicker snow in the Kara Sea – a region where mW99 was largely produced through extrapolation of the observations collected in the more central Arctic through the quadratic fit. We have made a note of this in the revised manuscript.

Zhou et al. (2020) showed large differences between the various atmospheric reanalysis- based approaches to snow loading as well as the remote sensing-based retrievals, with the SM-LG (Liston et al. 2020) providing more spatial structure to the snow depth/density distributions, whereas products such as NESOSIM are artificially smoothed products. I see you get around this by taking your smoothed products and then adding some artifical spatial structure to match IS2 resolution, but why regrid to 100km in the first palce? Anyone who has spent time on sea ice knows the snow is very heterogeneous and thus the artificially smoothed 100km NESOSIM product seems unrealistic. Some justification for regridding the snow depth to 100km is needed and why you think this artificially smoothed data set is a good representation of snow over sea ice. Also, the impact of the redistribution then to 30m resolution is needed.

There is a significant spatial scale issue between the meter-scale information obtained from ICESat-2 freeboard altimetry measurements and basin-scale snow reconstructions, e.g. NESOSIM, which are largely based on satellite input data with resolutions of 10s of kms. This is a big challenge!

To reconcile this scale gap, our approach has been to utilize high-resolution snow depth and freeboard measurements from Operation IceBridge obtained across the Arctic which, despite uncertainties, we believe is really our best means of bridging this scale gap using a redistribution/downscaling approach. NESOSIM thus provides our estimate of the seasonal/regional snow depth and density distribution, the redistribution scheme then helps us attempt to bridge the scale gap. The motivation behind the snow redistribution was discussed more in the original ICESat-2 thickness study (Petty et al., 2020).

This is an imperfect state of affairs but it seems that significantly finer resolution snow modelling will require much lower resolution input data and/or more comprehensive statistical distributions of snow properties that are validated against field data to capture the small-scale dynamic sea ice/snow processes. Advances like this are only now being incorporated into state-of-the-art sea ice models, e.g. CICE, but we do hope to explore this more in future work.

We are also not convinced that the 25 km-scale 'spatial structure' is related to improved accuracy and do not believe it should be used as a metric like this. The ice drift products are noisy at daily time-scales and this is a primary factor for us smoothing these data when used by NESOSIM, as was discussed in the original peer-reviewed NESOSIM v1.0 paper (Petty et al., 2018). Even lowering NESOSIM to 25 km would still require us to consider a redistribution/downscaling. Our aim with NESOSIM is thus to generate seasonal snow depth (and density) estimates constrained by the available basin-scale, but very limited, OIB observations.

Some assessment of the impact of using different ice motion products is also needed. It is not true that updated ice motion from NSIDC is not available, and the author could have contacted the data provider for updated ice motion fields. Since OSI SAF and NSIDC ice motion vectors to not agree, how does this influence your results? It is also unclear now how the Warren et al. climatology is used, are you assigning MYI snow depths on September 1 based on W99 and then accumulating snow? And finally, I'm not sure why so much smoothing is applied to both the snow and SIT retrievals, and some justification for this is needed. What does your SIT data product really give us if so much smoothing is applied? Snow and ice are highly spatially variable and thus is this a data product that is really useful to the community if it is artificially smoothed? Wanting "pretty" maps is not a reason to do this.

The original Petty et al., (2018) study undertook a comprehensive sensitivity study into the impact of differences in ice drift, using 4 different datasets (NSIDC, OSI SAF, CERSAT, KIMURA, Figure 11 and 12) concluding that at basin-scales this is of secondary impact to snow accumulation, but can have important regional impacts. It is challenging to discern a clearly optimal drift data product, so our choice in forcing is primarily driven by data availability. OSI SAF and NSIDC show reasonable agreement as shown in Petty et al., (2018) and our subsequent assessments of both products.

The reviewer's suggestion that we seek pre-release versions of the NSIDC drifts through independently contacting the providers seems problematic to us. The typical lag time for release of these products has, from our past experience, been about a year, so it is apparent that considerable time and effort needs to be taken in the processing and validation of these products prior to release. Furthermore, our philosophy of taking a more transparent and open-science approach (using whichever datasets are fully publicly available, carrying out reproducible and verifiable analysis through Jupyter notebooks, etc.) precludes this kind of exclusive approach to obtaining data.

I do not find much value in the CS2 to IS2 comparison. In particular, now suddenly the mW99 climatology is applied after spending much time discussing updates to NESOSIM. This seems to be only because you want to use existing products out there, which we already know are not realistic because they do not have a realistic snow loading representations. Instead, maybe comparison of the freeboards would be a better thing to do, as you can convert the IS2 snow freeboards to ice freeboards with your snow loading from NESOSIM. Then we can better understand differences on the ice freeboard level, and may be get some insights into where the dominant scattering surface from CS2 is located as well as the influence of surface roughness on the freeboard retrievals. The use of PIOMAS is also not useful in my opinion, it's a model and has known biases, so adding it here just distracts from the overall paper.
The abstract is too long and reads more like a technical report.

We do not believe this was a sudden jump, we explain the motivation and approach in detail in Section 2.4 as making sure we use consistent input data is crucial when carrying out these thickness comparisons, and the effort involved was not trivial. It was also the same approach as that taken in our original ICESat-2 thickness paper (Petty et al., 2020).

This was not a paper investigating CryoSat-2 scattering issues but products of basin-scale sea ice thickness, hence the focus on that higher-level data variable instead of freeboard. We also believe the thickness biases are also more intuitive to understand for the interested reader. Based on the comments of reviewer #1 we have now updated this to use rel005 data to be consistent with the rest of our analysis (differences are negligible) and have changed the comparison figure to bar plots to improve readability.

PIOMAS is well-used by the community so simply highlighting the seasonal and regional differences we believe to be a useful exercise. It is a model, but it is constrained by observations (mainly SST) and has been well-tuned over the years to provide useful thickness estimates used across various recent studies for assessing climate-scale variability and more regional sea ice changes. We have now adapted Section 2.5 to state this more clearly and provide further citations:

"PIOMAS data is commonly used in the sea ice community for assessments of Arctic sea ice thickness variability at regional and basin-scales (Tilling et al., 2015; Labe et al., 2018; Petty et al., 2018b; Schweiger et al., 2021; Moore et al., 2018)."

We have made some small edits to the abstract to improve readability.

References

Kwok, R., A. A. Petty, M. Bagnardi, N. T. Kurtz, G. F. Cunningham, A. Ivanoff (2021), Refining the sea surface identification approach for determining freeboard in the ICESat-2 sea ice products, The Cryosphere, 15, 821–833, doi:10.5194/tc-15-821-2021

Petty, A. A., M. Webster, L. N. Boisvert, T. Markus (2018), The NASA Eulerian Snow on Sea Ice Model (NESOSIM) v1.0: Initial model development and analysis, Geosci. Model Dev., doi: 10.5194/gmd-11-4577-2018.

Petty, A. A., N. T. Kurtz, R. Kwok, T. Markus, T. A. Neumann (2020), Winter Arctic sea ice thickness from ICESat-2 freeboards, Journal of Geophysical Research: Oceans, 125, e2019JC015764. doi:10.1029/2019JC015764

Stroeve, J., Liston, G. E., Buzzard, S., Zhou, L., Mallett, R., Barrett, A., Tschudi, M., Tsamados, M., Itkin, P., and Stewart, J. S.: A Lagrangian Snow Evolution System for Sea Ice Applications (SnowModel-LG): Part II—Analyses, 125, e2019JC015900, https://doi.org/10.1029/2019JC015900, 2020.