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.