Response to Reviewer #1

Note that the reviewer comments are italicized and our responses are in blue. Where we make changes to existing quoted passages from the text, additions are underlined and deletions are struck through.

Review of manuscript

"Investigating the impact of reanalysis snow input on an observationally calibrated snow-on-sea-ice reconstruction" by Cabaj et al.

The paper addresses the critical issue of relying on reanalysis data in undersampled regions like the Arctic, where limited observations—especially during harsh winters—lead to significant uncertainties in key surface variables such as snowfall and snow depth. This is problematic since reanalysis data are often used to benchmark models, drive satellite-derived algorithms, and reconstruct products like snow-on-sea-ice estimates. Studies show significant variability in Arctic snowfall across reanalyses due to observational gaps and differing model physics, which can lead to misleading evaluations and derived trends.

The authors focus on understanding discrepancies in snowfall estimates from three major reanalysis products—ERA5, JRA-55, and MERRA-2—when used as inputs for the NASA Eulerian Snow On Sea Ice Model (NESOSIM). They employ Markov chain Monte Carlo (MCMC) techniques to calibrate NESOSIM's snow depth and density parameters, addressing complexities not extensively covered in previous studies. While this work highlights the complexity of modeling snow on Arctic sea ice and the challenges associated with choosing and calibrating reanalysis products. It would benefit from a stronger emphasis on the need for improved observational data and robust calibration methodologies to enhance the reliability of model outputs.

The manuscript is well-structured and generally clear, but defining technical terms earlier would improve readability.

I thus recommend accepting this submission after revision that considers a few major and several minor comments.

We thank the reviewer for this helpful feedback, and we are including our responses below. Regarding the general point on the emphasis for the need of observation data, one major change we are planning to include to address this is the inclusion of a discussion on a comparison of MOSAiC snow depth and density observations to NESOSIM and SnowModel-LG, discussed further below.

Major comments

1. *Line 23-28. The introduction highlights the challenges posed by biases in reanalysis products. I would think the motivation was quite clear. However, a more focused explanation of how previous research improves on model calibration efforts, such as those by SnowModel-LG, would provide clearer context.*

Thank you for the suggestion. We do discuss model calibration in later paragraphs, and as such, will add some additional explanation at line 46 as follows:

"The implementation of this calibration approach was motivated by the fact that the free parameters in NESOSIM had been previously manually calibrated by comparing model output to observations (Petty et al., 2018). Between NESOSIM model versions, the ERA-Interim dataset, previously used for input to the model, was superseded by ERA5. Updating this dataset provided motivation for the development of a new automated parameter calibration process. The need for estimates of snow-on-sea-ice uncertainty for sea ice thickness retrieval applications further motivated the choice of the MCMC approach, since the MCMC process provides estimates of parameter uncertainty."

We also propose modifying Line 59 as follows:

“SnowModel-LG likewise includes observation-based calibration, namely an assimilation-based bias correction to precipitation to bring modelled snow depth into agreement with ground-based and remote sensing observations, including Operation IceBridge measurements (Liston et al., 2020, Stroeve et al., 2020).”

References:

Liston, G. E., Itkin, P., Stroeve, J., Tschudi, M., Stewart, J. S., Pedersen, S. H., Reinking, A. K., and Elder, K.: A Lagrangian Snow-Evolution System for Sea-Ice Applications (SnowModel-LG): Part I—Model Description, *Journal of Geophysical Research: Oceans*, 125, e2019JC015913, https://doi.org/10.1029/2019JC015913, 2020.

Petty, A. A., Webster, M., Boisvert, L., and Markus, T.: The NASA Eulerian Snow on Sea Ice Model (NESOSIM) v1.0: initial model development and analysis, *Geoscientific Model Development*, 11, 4577–4602, https://doi.org/10.5194/gmd-11-4577-2018, 2018.

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, *Journal of Geophysical Research: Oceans*, 125, e2019JC015900, https://doi.org/10.1029/2019JC015900, 2020.

2. *Table 1. The issue of coarse resolution does not seem to be explicitly addressed in the discussion. I think one limitation of this study is the coarse spatial resolution used for both the NASA Eulerian Snow On Sea Ice Model (NESOSIM) and the reanalysis data (ERA5, MERRA-2, JRA-55). NESOSIM operates on a 100 km × 100 km grid, and the reanalysis data ranges from 0.25° to 1.25° grids, which may obscure important sub-grid scale processes.*

   The authors should provide more specific examples of the sub-grid scale processes impacted by the coarse spatial resolution, such as snow redistribution due to ice ridge formation, wind-blown snow dynamics, or small-scale leads that can significantly alter local snow accumulation and density. These sub-grid scale processes could significantly influence snow depth, density, and heat fluxes, particularly in heterogeneous regions like the marginal seas or areas with dynamic ice cover. The coarse resolution also limits the model's ability to capture fine spatial variability, which may result in oversimplifications when calibrating model parameters, especially for regions with rapid snow accumulation or melt.

We agree with the reviewer that the coarse spatial resolution is a limitation of this work, and based on the reviewer’s comment, we suggest the following addition to the discussion at line 575:

“The relatively coarse resolution of NESOSIM may impact its representativeness, since some snow-on-sea-ice processes operate on very small spatial scales and short timescales. The sea ice advection and divergence processes in NESOSIM represent a spatially-averaged tendency of snow to be redistributed with sea ice motion, but may fail to capture small-scale effects from localized ridging and small-scale leads often seen in observational studies (Itkin et al., 2023; Macfarlane et al., 2023). The amount of blowing-snow loss due to leads has been observed to be influenced by strong winds and warm air temperatures from Arctic cyclone events, which may be challenging to capture in the current configuration of NESOSIM (Clemens-Sewall et al., 2023). The coarse time resolution also limits the model’s ability to capture rapid changes in snow depth due to short-term accumulation events. In a broader modelling context, high-resolution modelling may be necessary to adequately capture small-scale processes (Lecomte et al., 2015). NESOSIM could be run at a higher resolution to take advantage of the higher resolution of available drift products to better capture the influence of sea ice motion. However, sub-gridscale parameterization would still be necessary to better capture smaller-scale effects.

References:

Clemens-Sewall, D., Polashenski, C., Frey, M. M., Cox, C. J., Granskog, M. A., Macfarlane, A. R., Fons, S. W., Schmale, J., Hutchings, J. K., von Albedyll, L., Arndt, S., Schneebeli, M., and Perovich, D.: Snow Loss Into Leads in

Arctic Sea Ice: Minimal in Typical Wintertime Conditions, but High During a Warm and Windy Snowfall Event, *Geophysical Research Letters*, 50, e2023GL102816, https://doi.org/10.1029/2023GL102816, 2023.

Itkin, P., Hendricks, S., Webster, M., et al.: Sea ice and snow characteristics from year-long transects at the MOSAiC Central Observatory, *Elementa: Science of the Anthropocene*, 11, 00048, https://doi.org/10.1525/elementa.2022.00048, 2023.

Lecomte, O., Fichefet, T., Flocco, D., Schroeder, D., and Vancoppenolle, M.: Interactions between wind-blown snow redistribution and melt ponds in a coupled ocean–sea ice model, *Ocean Modelling*, 87, 67–80, https://doi.org/10.1016/j.ocemod.2014.12.003, 2015.

Macfarlane, A. R., Schneebeli, M., Dadic, R. et al: A Database of Snow on Sea Ice in the Central Arctic Collected during the MOSAiC expedition, *Sci Data*, 10, 398, https://doi.org/10.1038/s41597-023-02273-1, 2023.

Including references to recent high-resolution modeling studies that have explored these processes would help contextualize this limitation. Future work could benefit from incorporating downscaling techniques or nested models to better resolve local variability. High-resolution observational data from satellites or in-situ measurements could also improve model validation and enhance regional accuracy. Addressing these limitations would help refine the model's ability to capture critical snow-ice-atmosphere interactions at finer scales, improving both regional forecasts and large-scale trend assessments.

We agree that higher resolution would be beneficial, and that future work could benefit from incorporating downscaling approaches/nested models. We have included reference to modelling and observation studies describing small-scale processes in our suggested revision above. The relative simplicity of NESOSIM (including its comparatively coarse resolution) was a decision to enable to rapid production of snow-on-sea-ice estimates for operational purposes; particularly the production of sea ice thickness estimates from ICESat-2 altimetry measurements.

Regarding high-resolution observational data, taking into account this comment and comments from other reviewers, we propose to also include some discussion comparing NESOSIM and SnowModel-LG to snow depth and density observations from the MOSAiC campaign which we will briefly summarize below. We find that this comparison highlights some of the challenges the models have with coarse resolution not necessarily representing small-scale processes. Observed snow depth and density tend to be highly variable within a single model grid cell.

Comparison of NESOSIM and SnowModel-LG to MOSAiC observations

Below is a brief summary with key points for a comparison to observations we intend to include in our revised manuscript; more detail and discussion will be added when we prepare the revised manuscript. We compare output from NESOSIM and SnowModel-LG to snow depth and density measurements (Macfarlane et al., 2023) obtained during the 2019-2020 Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) campaign (Nicolaus et al., 2022). Snow depths were measured using magnaprobes (Itkin et al., 2021), and in previous studies have been noted to be relatively thin (Itkin et al., 2023). Bulk snow densities used in this comparison were calculated from density cutter measurements, which sample densities at varying depths within a snow pit (Macfarlane et al., 2022). Snow was sampled over a variety of conditions, including ridges and leads, and snow over first-year and multi-year ice (Macfarlane et al., 2023).

To compare with gridded snow model outputs, MOSAiC observations are collocated to the nearest model grid point, and then averaged by day for each grid point. These values are then compared to the corresponding model grid point value, excluding dates during which NESOSIM output is unavailable. Below, we present figures aggregated by month.

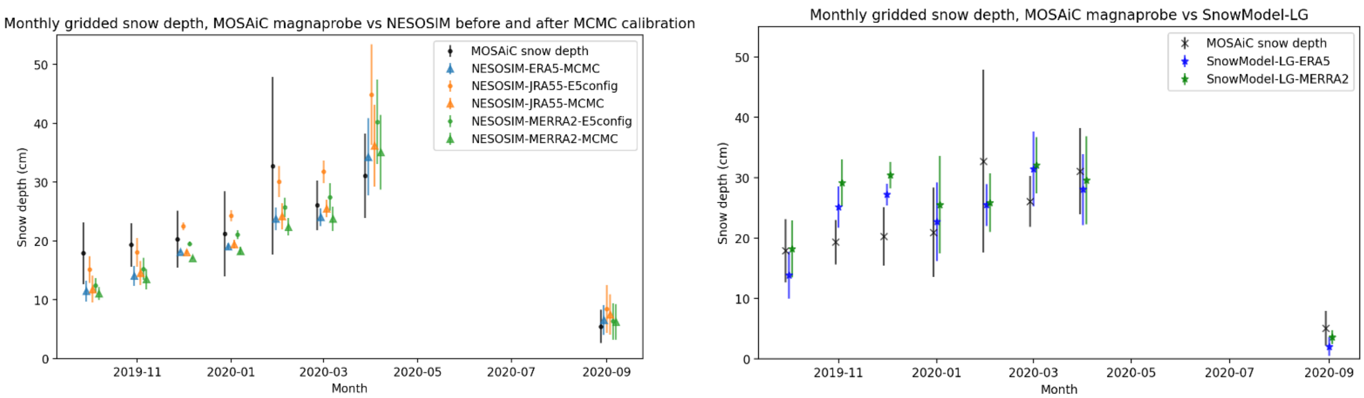


Figure A: (left): MOSAiC magnaprobe snow depth (Itkin et al., 2021) compared to NESOSIM snow depth, before ("E5config") and after ("MCMC") individual dataset calibration to observations; triangles indicate individually-calibrated datasets. (right): MOSAiC snow depth compared to SnowModel-LG snow depth with different snow forcings. May-August are excluded due to the absence of NESOSIM data. Error bars represent 1 standard deviation of the monthly mean (with MOSAiC data also including contributions from the daily standard deviation).

Figure A shows monthly-averaged MOSAiC snow depth measurements (Itkin et al., 2021) compared to NESOSIM and SnowModel-LG. Note that aggregated MOSAiC outputs may differ slightly between figure panels because the model output is being aggregated to different model grids. Both models show general good agreement with the observations, with some products showing slight biases. The uncalibrated NESOSIM output driven by JRA55 has a general high bias relative to the other products (and a daily mean bias of 3.2 cm relative to MOSAiC). Differences in seasonal cycles are apparent between the models. Compared to MOSAiC, several NESOSIM products are biased low in October-November 2019, and some products are biased high in March-April. SnowModel-LG (particularly when driven by MERRA-2) is conversely biased slightly high in November and December. Nevertheless, overall agreement is close, with daily root-mean-square difference not exceeding 10 cm for all products relative to MOSAiC.

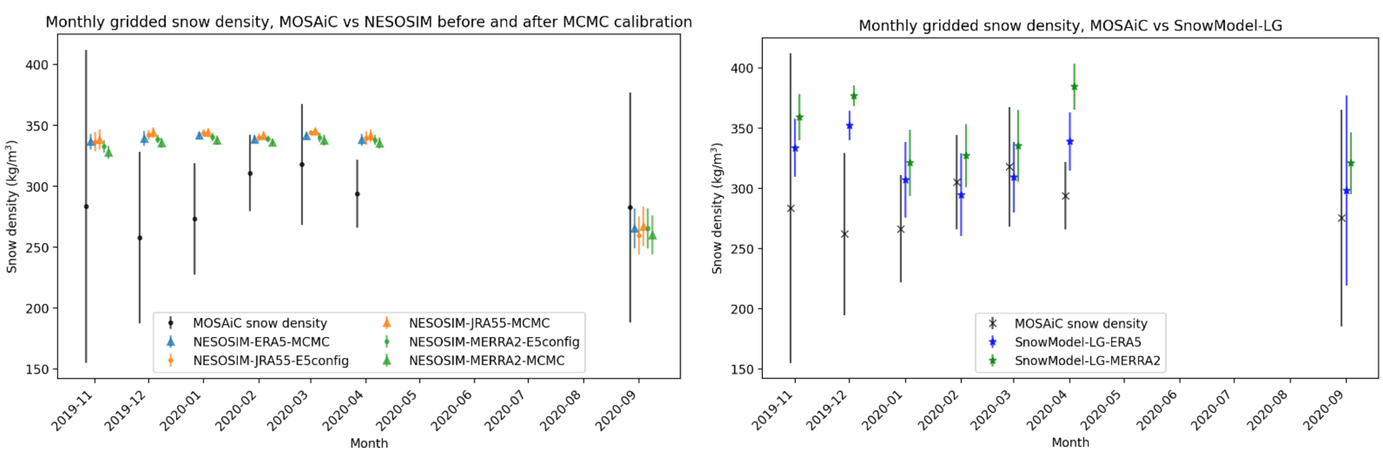


Figure B: Monthly averages of gridded MOSAiC snow density cutter data vs. monthly averages of coincident model output data. Error bars indicate 1 standard deviation. (left): MOSAiC compared to NESOSIM, before ("E5config") and after (MCMC) the MCMC calibration. (right): MOSAiC compared to SnowModel-LG. Only months with at least 8 measurements are shown.

Figure B shows monthly averages of MOSAiC snow density cutter measurements (Macfarlane et al., 2022) compared to NESOSIM and SnowModel-LG. Prior to gridded collocation with the models, bulk density for each measurement event was calculated from the average of sampled densities weighted by sample snow thickness. NESOSIM snow density from all products shows relatively little variation over the time period, whereas SnowModel-LG snow density shows more seasonality. Both models show a high mean bias relative to observed values, with SnowModel-LG driven by MERRA-2 having the largest daily mean bias (60 kg/m^3). The comparatively high variability of the observed values is also apparent.

Below are tables with daily (gridded) comparison statistics for NESOSIM and SnowModel-LG with respect to MOSAiC, including correlation, root-mean-square difference (RMSD) and mean bias error (MBE), for reference.

Table A: Daily comparison statistics for NESOSIM and SnowModel-LG comparisons to MOSAiC snow depth observations

	NESOSIM-ERA5-MCMC	NESOSIM-JRA55-E5config	NESOSIM-JRA55-MCMC	NESOSIM-MERRA2-E5config	NESOSIM-MERRA2-MCMC	SnowModel-LG-ERA5	SnowModel-LG-MERRA2
Pearson Correlation to MOSAiC	0.68	0.67	0.67	0.67	0.67	0.64	0.58
RMSD (cm)	8.5	9.5	8.6	8.7	8.9	9.3	10
MBE (cm)	-2.4	3.2	-1.6	-0.24	-2.9	-0.26	1.8

Table B: Daily comparison statistics for NESOSIM and SnowModel-LG comparisons to MOSAiC snow density observations

	NESOSIM-ERA5-MCMC	NESOSIM-JRA55-E5config	NESOSIM-JRA55-MCMC	NESOSIM-MERRA2-E5config	NESOSIM-MERRA2-MCMC	SnowModel-LG-ERA5	SnowModel-LG-MERRA2
Pearson Correlation to MOSAiC	0.22	0.20	0.20	0.15	0.15	0.24	0.16
RMSD (kgm^{-3})	79	80	80	80	79	80	93
MBE (kgm^{-3})	32	32	35	32	28	32	60

References (abbreviated for this response, full references will be included in the article):

Itkin, P., Hendricks, S., Webster, M., et al.: Sea ice and snow characteristics from year-long transects at the MOSAiC Central Observatory, *Elementa: Science of the Anthropocene*, 11, 00048, https://doi.org/10.1525/elementa.2022.00048, 2023.

Macfarlane, A. R., Schneebeli, M., Dadic, R. et al.: A Database of Snow on Sea Ice in the Central Arctic Collected during the MOSAiC expedition, *Sci Data*, 10, 398, https://doi.org/10.1038/s41597-023-02273-1, 2023.

Nicolaus, M., Perovich, D. K., Spreen, G. et al.: Overview of the MOSAiC expedition: Snow and sea ice, *Elementa: Science of the Anthropocene,* 10, 000046, https://doi.org/10.1525/elementa.2021.000046, 2022.

Dataset references:

Itkin, Polona; Webster, Melinda; Hendricks, Stefan, et al. (2021): Magnaprobe snow and melt pond depth measurements from the 2019-2020 MOSAiC expedition [dataset]. *PANGAEA*, https://doi.org/10.1594/PANGAEA.937781

Macfarlane, Amy R; Schneebeli, Martin; Dadic, Ruzica, et al. (2022): Snowpit snow density cutter profiles measured during the MOSAiC expedition [dataset]. *PANGAEA*, https://doi.org/10.1594/PANGAEA.940214, In: Macfarlane, AR et al. (2021): Snowpit raw data collected during the MOSAiC expedition [dataset bundled publication]. *PANGAEA*, https://doi.org/10.1594/PANGAEA.935934

The use of ERA5 wind input for all NESOSIM runs can introduce biases. One concern is inconsistencies in sea ice representation between datasets—ERA5 uses SST/SIC from HadISST2/OSI SAF, while NESOSIM uses NSIDC SIC data for all runs. These discrepancies can alter surface roughness and wind stress, affecting snow redistribution and compaction. This mismatch can lead to errors in snow depth estimates, particularly in regions with dynamic ice conditions. Would you comment on the potential effect of using ERA5A wind inputs combined with different reanalysis data? Since wind patterns influence snow redistribution and compaction, the choice of a single wind product may not fully capture the sensitivity of snow depth to wind dynamics. Considering product-specific wind inputs or conducting sensitivity tests with different wind datasets could strengthen the reliability of the model results. Furthermore, in Line 84, ERA5 uses a threshold of SIC>20% to distinguish between open ocean and sea ice cover, whereas MERRA-2 uses a 50% threshold (Line 93). Could this difference in SIC thresholds introduce artifacts, particularly in regions with marginal sea ice cover, potentially influencing the weakest or strongest trends in snow depth as discussed in Lines 450-451?

Regarding the different SIC representation in the reanalysis products, we agree that these could potentially introduce some artefacts if the precipitation processes in reanalyses are impacted via e.g. surface moisture fluxes, although we would consider this as one of the contributing factors to inter-product differences in reanalysis snowfall rates, among other factors. We briefly discussed this at line 608. We note that NESOSIM does not account for surface roughness; snow distribution within model grid cells is considered to be uniform. For running NESOSIM, we do not use different SIC products for different snow products. We will add the following to the discussion at line 575 to address this:

"The representation of sea ice differs between reanalysis products, and may not be coincident with the observational sea ice concentration used as input to NESOSIM in this work. This, in conjunction with regridding, may introduce some artefacts in regions of marginal sea ice cover such as the Greenland Sea region."

We kept the wind inputs the same to isolate the contribution of snow input to NESOSIM model output. We note that ERA5 wind performs relatively well compared to other reanalysis products in Arctic observational studies over sea ice (cf. Graham et al. 2019a, 2019b). The focus on snow was motivated by the fact that snowfall is the primary input to the NESOSIM snow budget, and when we were updating the model reanalysis input, we found that changing snowfall had a stronger impact on the snow output. However, we will acknowledge the impact of wind as follows in the discussion after line 575:

"The ERA5 wind product was used in all configurations in this study to isolate the contribution of snowfall to NESOSIM, since snowfall is the primary input to the NESOSIM budget. In observational comparisons in the Arctic, ERA5 has been found to perform relatively well compared to other reanalysis products, including JRA-55 and MERRA-2, which motivates the choice of ERA5 over other products (Graham et al., 2019a, 2019b). However, the choice of reanalysis wind input may also have an impact on NESOSIM output. The wind packing and blowing snow processes take effect only when wind speed exceeds the 5 m/s wind action threshold. If wind speeds from different input products are on differing sides of the threshold, wind-related snow processes may take effect at a given location and time for one product and not another. The strength of the blowing snow process is also dependent on wind speed. Future work could investigate the impact of differing wind input products to NESOSIM."

References:

Graham, R. M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R. G., Rinke, A., Walden, V. P., Granskog, M. A., and Hudson, S. R.: Evaluation of Six Atmospheric Reanalyses over Arctic Sea Ice from Winter to Early Summer, *Journal of Climate*, 32, 4121–4143, https://doi.org/10.1175/JCLI-D-18-0643.1, 2019a.

Graham, R. M., Hudson, S. R., and Maturilli, M.: Improved Performance of ERA5 in Arctic Gateway Relative to Four Global Atmospheric Reanalyses, *Geophysical Research Letters*, 46, 6138–6147, https://doi.org/10.1029/2019GL082781, 2019b.

Additionally, the choice of initializing the model in September each year may overlook key early-season snow accumulation events, especially in regions where snow can start accumulating in late summer. Considering the importance of accurately capturing the initial snow state, it would be beneficial to assess the impact of starting the model earlier in the season or adjusting the initialization timing based on regional climatologies.

The choice was made to initialize NESOSIM in September based on several factors. Firstly, August is a complex month for snow-on-sea-ice representation, since melt may be occurring, which may not be represented in NESOSIM. Instead of attempting to represent August snow evolution, NESOSIM is initialized with initial snow depth values which attempt to capture climatological snow-on-sea-ice conditions. Furthermore, September was chosen for initialization since it is the month during which Arctic sea ice attains its minimum extent, so that NESOSIM can model snow-on-sea-ice from the start of seasonal sea ice growth to the melt season. The motivation for this choice is discussed in Petty et al., 2023. We also note that we have previously investigated calibrating initial conditions, (Cabaj et al., 2023) although our previous investigation found that the initial condition values were underconstrained.

References:

Cabaj, A., Kushner, P. J., and Petty, A. A.: Automated Calibration of a Snow-On-Sea-Ice Model, *Earth and Space Science*, 10, e2022EA002655, https://doi.org/10.1029/2022EA002655, 2023.

Petty, A. A., Keeney, N., Cabaj, A., Kushner, P., and Bagnardi, M.: 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, *The Cryosphere*, 17, 127–156, https://doi.org/10.5194/tc-17-127-2023, 2023.

3. *The results are presented clearly but raise questions about the handling of snow density, where the model calibration reconciles snow depths well but leaves density poorly constrained. The authors highlight that this might be due to a lack of density observations, but a more detailed exploration of potential biases introduced by the model's simplicity could be useful.*

We agree that representation of snow density is an ongoing challenge. We have discussed part of what we suspect is a contributing factor to biases in the model; namely, the imposed maximum snow density value, at line 560. Further to this, our proposed inclusion of comparisons to in situ MOSAiC snow density cutter observations (included above) can provide some insight to this point. We note that this comparison has associated caveats due to representational differences between comparatively coarse-resolution models and point observations. Nevertheless, both SnowModel-LG and NESOSIM have comparable difficulty reproducing snow density from these observations. However, NESOSIM snow density has low seasonal variability relative to both SnowModel-LG and MOSAiC. Since snow in NESOSIM cannot be removed from the lower layer (for a given grid cell, it can only decrease as a consequence of sea ice motion), end-of-season densities are expected to approach 350 kg/m^3 as an increasing proportion of the snow in each grid cell is old (lower-layer) snow. We will further expand our discussion to include these points.

Given the current limitations of the MCMC calibration for snow density, the authors could consider exploring multi-level Bayesian models that integrate different observational datasets (e.g., buoy measurements, satellite data). This approach could provide more reliable density estimates by better accounting for observational gaps and biases.

Thank you for the suggestion. We agree that such an exploration could be of interest in general, but we do not necessarily think this would be a beneficial modification to our current study, given the simplicity of the NESOSIM model parameter space and the limited ability of the NESOSIM model to represent a wide range of snow densities, since the model is limited to two layers and one densification process. However, this could be an option for calibrating a possible future version of NESOSIM with a more complex representation of snow density or additional snow processes, and thus would be an approach we would consider for a future study.

In addition to wind packing and blowing snow, other snow-atmosphere interactions, such as sublimation, melt, and refreeze processes, are also simplified in NESOSIM. These processes are crucial for understanding seasonal changes in snow density and depth, and their absence may contribute to biases. Future work could explore parameterizations for these interactions to improve density and depth estimates across diverse environmental conditions.

We agree, and hope that this study can be a motivation for additional work to address these points. Guided by the reviewer's comment, we will expand the discussion at line 571 to mention this: "Several processes, including snow redistribution by wind, sublimation, and melt and refreeze processes are simplified in NESOSIM." and at line 575: "Future work could explore parameterizations of additional processes, such as sublimation and snow redistribution by wind to improve snow depth and density estimates across a variety of environmental conditions."

4. *The model's representation of snow density relies heavily on parameterizations of processes like wind packing and blowing snow, which may not fully capture the complex physical processes occurring at different scales or in diverse environmental conditions. This can lead to uncertainties in the snow density estimates, which might not be fully representative of actual conditions across the Arctic. By integrating additional independent datasets, such as in-situ observations from buoys, satellite-derived snow density estimates, or regional field campaigns, the model calibration can be further refined, reducing the risk of biases and providing a more robust validation of the snow density outputs.*

We agree, however, given that NESOSIM in its current state has a very simplified representation of snow density, including additional measurements at this time may not yield major improvements. We would like to note that the calibration already incorporates in situ observations from buoys (as described in e.g. Line 159). Also, the challenge with incorporating observations from field campaigns is that such measurements tend to be highly localized, which may lead to "overfitting" if a large number of measurements are located in a small geographical region, or if only a single season is sampled. The primary dataset used for the calibration is OIB airborne measurements, which were repeated over several seasons and cover a relatively wide spatial range. The buoys likewise cover a wide spatial range, and historical density measurements are used to also provide relatively widespread density estimates which span several decades, to provide an estimate of the snow density seasonal cycle. Localized density measurements may be difficult to incorporate since it is likely that values of snow density larger than what NESOSIM in its current configuration can represent may be present.

Based on several reviewer suggestions, we propose including a comparison to MOSAiC in situ measurements as discussed above, which highlights the difficulty in comparing comparatively coarse-resolution snow-on-sea-ice models such as NESOSIM and SnowModel-LG against localized measurements.

It would also be useful to discuss how the observed variability in snow depth trends between reanalysis products could alter climate sensitivity estimates. For example, different snow depth trends could affect sea ice model sensitivity to atmospheric drivers such as warming temperatures or shifting storm patterns. A brief discussion on this topic would highlight the broader implications of inter-product variability.

Thank you for the suggestion, we propose expanding the discussion at line 608 to include the following: "The differing snow depth trends between model outputs due to different reanalysis snowfall inputs may have impacts on climate sensitivity estimates due to its influences on sea ice. Coupled climate model simulations have found contrasting climate impacts of snow on Arctic sea ice due to competing influences on congelation sea ice growth and surface melt (Holland et al., 2021), but snow-free summers may increase sea ice melt (Webster et al., 2021). Thus, by influencing sea ice thickness, a declining snow depth trend could influence trends in atmosphere-ice heat fluxes, which in turn could influence sea ice extent and other climate variables."

References:

Holland, M. M., Clemens-Sewall, D., Landrum, L., Light, B., Perovich, D., Polashenski, C., Smith, M., and Webster, M.: The influence of snow on sea ice as assessed from simulations of CESM2, *The Cryosphere*, 15, 4981–4998, https://doi.org/10.5194/tc-15-4981-2021, 2021.

Webster, M. A., DuVivier, A. K., Holland, M. M., and Bailey, D. A.: Snow on Arctic Sea Ice in a Warming Climate as Simulated in CESM, *Journal of Geophysical Research: Oceans*, 126, e2020JC016308, https://doi.org/10.1029/2020JC016308, 2021.

Since snow depth and density are crucial for calculating sea ice thickness from satellite altimetry, the variability observed between reanalysis products could have substantial impacts on the interpretation of sea ice thickness trends. Expanding the discussion on the implications for sea ice thickness estimates would emphasize the broader significance of these findings and strengthen the rationale for using multi-product approaches.

We appreciate the reviewer's acknowledgement of the importance of this context, and agree that this context is good to keep in mind, since NESOSIM was primarily developed for deriving ice thickness from sea ice altimetry. We will clarify in the NESOSIM model description that the model was developed for this purpose, and we propose also adding the following to the discussion at line 604:

"Inter-product differences in snow depth and density may have substantial impacts on estimates of sea ice thickness from sea ice altimetry measurements. For example, given representative values of lidar freeboard, and representative densities of snow, ice, and water, if snow depth estimates with a 5 cm difference are used to estimate sea ice thickness, the difference in derived sea ice thickness can be as large as 30 cm (Giles et al, 2007). Thus, if trends differ between snow products, trends in derived sea ice thickness will be impacted as well. For sea ice freeboard, a snow product with a decreasing trend would impose an increasing derived ice thickness trend on top of any trend in the freeboard itself. Interannual variability in snow was found to strongly influence sea ice volume derived from CryoSat-2 altimetry measurements (Bunzel et al., 2018). Hence, differing snow depth trends (or lack thereof) between products could lead to differing conclusions on trends in derived sea ice thickness."

References:

Bunzel, F., Notz, D., and Pedersen, L. T.: Retrievals of Arctic Sea-Ice Volume and Its Trend Significantly Affected by Interannual Snow Variability, *Geophysical Research Letters*, 45, 11,751-11,759, https://doi.org/10.1029/2018GL078867, 2018.

Giles, K. A., Laxon, S. W., Wingham, D. J., Wallis, D. W., Krabill, W. B., Leuschen, C. J., McAdoo, D., Manizade, S. S., and Raney, R. K.: Combined airborne laser and radar altimeter measurements over the Fram Strait in May 2002, *Remote Sensing of Environment*, 111, 182–194, https://doi.org/10.1016/j.rse.2007.02.037, 2007.

Minor comments:

1. From the analyses, both wind packing and blowing snow appear to be critical processes in this study. Therefore, their relevance, physical processes, and mechanisms affecting snow depth and density should be clearly explained in the introduction. Additionally, how these processes are represented and utilized within the NASA Eulerian Snow On Sea Ice Model (NESOSIM) should be outlined. Considering the broader audience of this study, beyond just NESOSIM users, a brief explanation in the introduction or methods section would significantly enhance clarity and understanding.

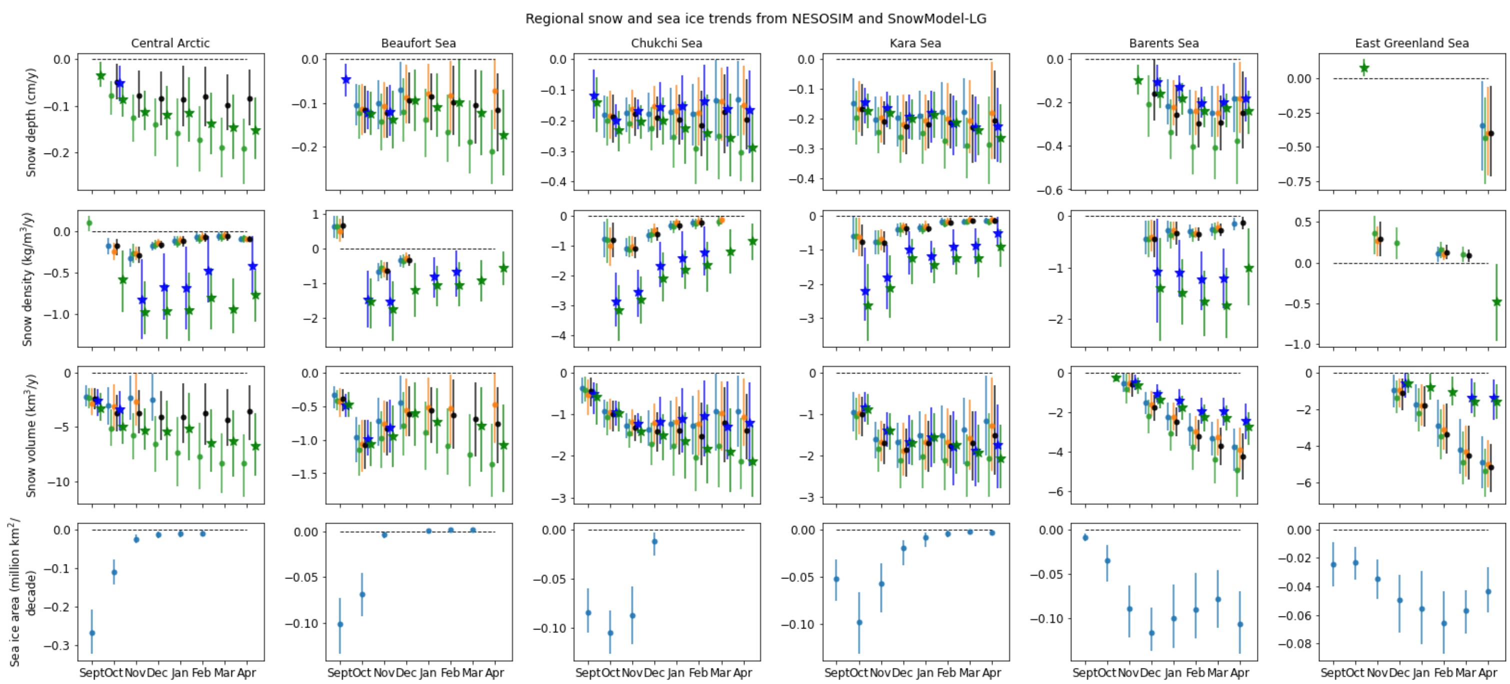
We attempted to explain these processes in Section 2.3 starting at line 134, and propose changing it to more clearly explain as follows:

"Wind packing controls the amount of snow transferred between layers, ~~impacting~~ decreasing the snow depth and increasing the bulk snow density as snow is transferred from the upper (less dense) layer to the lower (denser) layer. The blowing snow process acts only on the upper snow layer, and decreases the snow depth in the upper layer linearly with wind speed. The blowing snow term includes an atmosphere loss and an open-water loss term, which are prescribed separately in NESOSIM v1.1 (Petty et al., 2023). The open-water loss term accounts for sea ice concentration, with regions of lower sea ice concentration experiencing more open water loss. For the purpose of [...]"

2.Figures 8 and 10 are visually dense, and the small font size makes them challenging to interpret. Including a statistical summary for each panel (e.g., mean and standard deviation) either alongside the figures or in a supplementary table would greatly enhance clarity. In particular, the Figure 8 caption lacks sufficient detail: the overlapping colored lines and shaded areas are difficult to distinguish and not clearly defined in the caption. It should be explicitly stated what each represents, and the method used to quantify the interannual variability indicated by the shaded area should also be clarified.

Thank you for the feedback. We will adjust the font size in the subsequent revision of the article, and will also include as a supplement a summary table of the numerical values of climatological means and standard deviations (or trend values with confidence intervals, as applicable) for each quantity and each month, since we agree this would be helpful information to have for clarity.

As suggested by another reviewer, we have adjusted Figure 10 and other trend figures to exclude non-significant values, and to represent the values with bars, as follows. Note that this figure is not finalized and we will make further adjustments with legend placement and font size prior to the next revision, but we include it for visual reference. Our proposed changes to Fig. 8 and its caption are shown further below.



Another concern is the interpretation of overlapping uncertainty/internal variability envelopes in the figures (Figs.2,4,6,7,8,9,10). When the envelopes for different reanalysis products overlap, it can be challenging to visually assess whether the observed differences are statistically significant.

Thank you for the feedback. As mentioned above, by suggestion from another reviewer, we have now left out values that are not statistically significant in the trend plots, and for the climatology plots, we have also opted to show error bars instead of shading as shown below, with Fig. 8 serving as a representative figure for the proposed changes. The modified caption for Figure 8 is also shown below. As previously stated, note that some changes to e.g. text size and legend placement are still outstanding, but will be finalized for the next revision of the article.

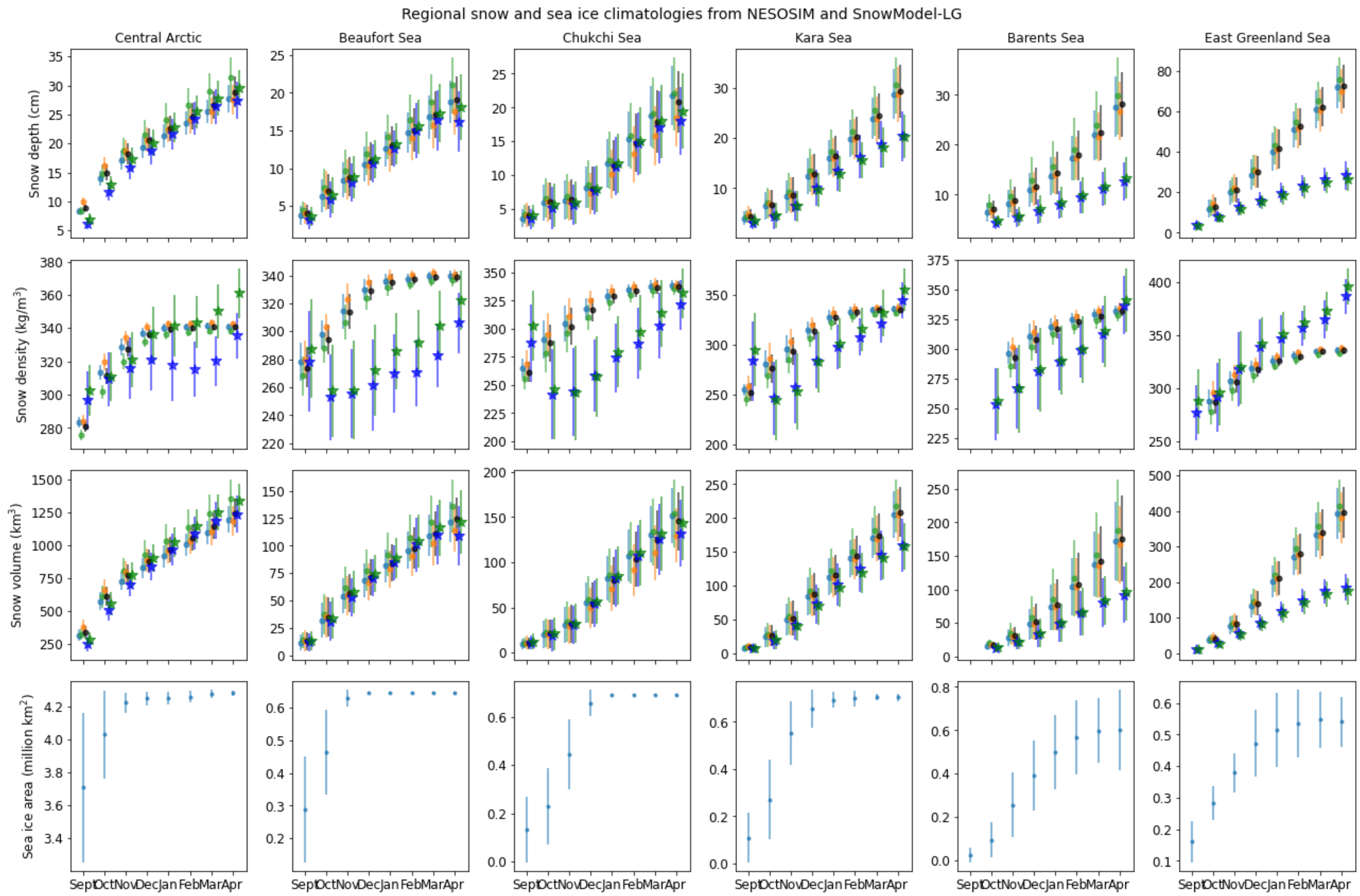


Figure 8: Climatologies of regionally-averaged snow depth, density, and volume from MCMC-calibrated NESOSIM output and SnowModel-LG output, for 1980-2019. Regional CDR sea ice area climatologies also shown. “Average” indicates the inter-product average for the three NESOSIM configurations. Climatologies from SnowModel-LG driven with ERA5 and MERRA-2 are also shown, with dashed lines. Regions are as described in Fig. A1. Bars indicate interannual variability of each respective climatology, which is quantified by the standard deviation of the climatology.

3. Table 2 could benefit from a clearer introduction in the main text discussing the importance of the acceptance rates and coefficients of variation in MCMC results, especially for readers less familiar with Bayesian technique

Thank you for the suggestion. We described acceptance rate further later in the article but agree that it would benefit from an earlier introduction. As such, we propose modifying at line 273 as follows:

“The acceptance rate, calculated from the ratio of accepted parameters to the total number of iterations, indicates the efficiency of the MCMC process, with an optimal efficiency for a 2-parameter MCMC process being approximately 23% (Gelman et al., 2013). Coefficients of variation are calculated from the standard deviation of the posterior distribution divided by the posterior parameter value, and quantify the relative spread of the posterior distribution. This provides a quantitative indication of how well-constrained the parameters are by the MCMC calibration. The posterior distribution of ERA5 [...]”

Reference:

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: Bayesian Data Analysis, *CRC Texts in Statistical Science*, CRC Press, Boca Raton, third edn., 2013.

4. Introduction to MCMC: The explanation of the MCMC process in the methods section is detailed but slightly dense. Consider simplifying the language in the initial description to cater to a broader audience or moving more technical details to a supplementary section.

Adding a brief explanation of why MCMC was chosen over other calibration methods would further support the use of Bayesian techniques and enhance the methodological discussion.

Following the reviewer's suggestion, we have decided to move the detailed explanation of the MCMC process to a supplemental section. For an explanation of why MCMC was chosen of other methods, we propose adding the following at line 147:

"Using an MCMC approach for calibrating NESOSIM allows for the automated estimation of free parameters, which were previously manually estimated via comparison to observations (Petty et al., 2018). An added benefit of this approach is that it yields posterior distributions of the parameters, which provide an estimate of parameter uncertainty."

Per another reviewer's suggestions, note that we will make the following changes to the explanation of the MCMC process, for added clarity:

- Line 145 rewritten as "MCMC is an algorithm applied to Bayesian problems where, given prior information of the parameters..." as suggested
- Line 147: replacing with "a log-likelihood function of the difference between model output and selected aggregated observations used for the calibration, weighted by the uncertainty in the observations." to clarify
- Added at Line 149: "The prior parameter values are associated with prior parameter distributions $p(a_0)$ for which the mean is a_0 and the uncertainty is a prescribed prior parameter uncertainty value. These prior values are given in Table 1." We will also add the prescribed prior parameter uncertainties (1×10^{-8} s^{-1} for wind packing and 1×10^{-8} m^{-1} for blowing snow) to Table 1 for reference.
- Line 152: rephrase to: "with the subsequent step chosen from $p(a_0)$; a normal distribution centered at a_0 whose standard deviation is determined by the prior parameter uncertainty"
- Equation 1: we will add a clarification in the description that observation uncertainties also account for estimated errors of representativeness in each term.
- Line 174: rephrasing to "all distributions (the prior parameter distribution, the likelihood function, and the posterior distribution) are assumed to be Gaussian"

5. Trends in units of 'per decade.': This adjustment will help avoid the need for four decimal places in Figs.9-12 and improve readability.

Thank you for the helpful suggestion. We will adjust the units accordingly in the subsequent revision of the manuscript.

6. CloudSat Discussion: The section discussing the use of CloudSat data might benefit from clarification on the limitations of this dataset, particularly the reduced reliability for latitudes north of 82°N. This could be stated earlier in the methods subsection 2.2.

We agree that these are necessary details to mention, and propose adding the following:

Line 105: "CloudSat's ground track had latitudinal coverage between 82°N and 82°S, and hence, no measurements are available near the pole. To mitigate ground clutter contamination of near-surface returns, near-surface snowfall rate measurements are retrieved from the 3rd vertical bin above ocean surfaces, or the 5th vertical bin above sea ice (as determined by a climatological sea ice mask) (Wood & L'Ecuyer, 2018). Data quality flags are applied to exclude potentially contaminated observations as described in Cabaj et al., 2020."

References:

Cabaj, A., Kushner, P. J., Fletcher, C. G., Howell, S., and Petty, A. A.: Constraining Reanalysis Snowfall Over the Arctic Ocean Using CloudSat Observations, *Geophysical Research Letters*, 47, e2019GL086426, https://doi.org/10.1029/2019GL086426, 2020.

Wood, N. B., & L'Ecuyer, T. S.. Level 2C Snow Profile Process Description and Interface Control Document, Product Version P1R05. NASA JPL CloudSat project (Document revision 0), 2018. Retrieved from http://www.cloudsat.cira.colostate.edu/sites/default/files/products/files/2C-SNOW-PROFILE_PDICD.P1_R05.rev0_.pdf

7. In the results section discussing snow density, consider adding a few more sentences to highlight how the differences in reanalysis products might specifically affect the observed snow depths. Additionally, expanding on the implications of snow depth differences for sea ice thickness estimates would add clarity for practical application

Thank you for the suggestion. Regarding the snow density section, assuming that the reviewer meant to ask for us to highlight how differing reanalysis products might affect the snow densities, we note that as we discuss, we do not expect snowfall to have as much of a contribution to snow density. However, we can state the following at Line 358:

"The slight differences between the uncalibrated snow density outputs may result from the influence of snowfall. For example, depending on the timing of snowfall, reanalysis snowfall may impact snow density in different ways; a high snow accumulation event will reduce the overall bulk density in the short term, but if this accumulation occurs early in the season, more snow may subsequently be transferred to the lower layer, increasing the bulk snow density in the long term if subsequent accumulation is lower. For example, NESOSIM driven by JRA55 shows deep snow in the early season relative to other products, which may contribute to its high later-season snow density bias as seen Fig. 7a."

Regarding reanalysis impacts on snow depths, we can further expand as follows at Line 337, since we did intend to include this in our manuscript but may not have been explicit enough in our discussion:

"Some of the relative biases between the products persist; JRA-55 continues to have a relatively large early-season snow depth which is not seen in the other products, consistent with its early-season snowfall bias. Conversely, at the end of the season, JRA-55 and MERRA-2~~, which~~ previously both exceeded ERA5 at the end of the season, consistent with snowfall biases over sea ice in most regions, particularly over the central Arctic. ~~now~~ Following the MCMC calibration, JRA-55 and MERRA-2 bracket ERA5 snow depth ~~it~~ on either side, with the multi-product average closely matching the ERA5 values. "

The impact on sea ice thickness is connected to our response earlier to Major Comment 4 as discussed above; we will expand to include a discussion of impacts on sea ice thickness from differences in snow depth.

8. When referring to previous studies that employed ERA5 or other reanalysis products, try to explicitly state how the multi-product approach improves over previous single-product studies. This would strengthen the rationale for using multiple reanalysis inputs rather than relying solely on a single product.

Thank you for the suggestion, we have previously stated this briefly in our discussion at Line 610 but will also restate this earlier in the motivation at Line 57, and propose expanding further:

"Multi-dataset approaches help to reveal biases between datasets, and facilitate the characterization of dataset uncertainties."

We will expand similarly where other previous single-product studies are mentioned, as suggested by the reviewer.