Response to Reviewer #2

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.

This study aims to investigate the impact of different reanalysis snow input in snowon- sea-ice reconstructions (snow depth and density) provided by NASA's Eulerian Snow On Sea Ice Model (NESOSIM) (Petty et al., 2018). In an earlier study by Cabaj et al. (2023) wind packing and blowing snow parameters were calibrated in NESOSIM using a Markov chain Monte Carlo (MCMC) approach, with ERA5 snowfall. The same MCMC approach is used here, with snowfall inputs from additional reanalysis (MERRA-2 and JRA-55). All reanalysis snowfall was first calibrated to CloudSat following Cabaj et al. (2020). The impact of the MCMC calibration to NESOSIM and the snow depth and density uncertainties are assessed and discussed. NESOSIM outputs of snow depth and density are regionally compared to SnowModel-LG (Liston et al., 2020), a Langragian snow evolution modeling system for sea ice applications. Pan-Arctic and regional monthly trends of snow depth, density and volume derived from both NESOSIM and SnowModel-LG are examined and discussed.

The paper addresses scientific questions within the scope of TC. The title reflects the contents and the abstract provides a concise summary of the study. The overall presentation of the paper is well-structured and the language is clear and coherent. However, while the central findings and conclusions are supported by the data, there are areas where the manuscript would benefit from additional context and clarification. My recommendation is to reconsider the paper after major revisions, to improve its overall quality and precision.

We thank the reviewer for the comments, and appreciate the assistance on providing additional context and clarification. Our responses are below. One major change we are planning to address some of the reviewer comments is to compare MOSAiC snow depth and density observations to NESOSIM and SnowModel-LG, discussed further below.

Major comments:

1. *NESOSIM and MCMC calibration*

Attention should be payed in this section to ensure a coherent description of the methodology used for calibrating the parameters, so that the method can be repeated by anyone to reproduce the results. Some examples are given:

We appreciate the specific examples, which enable us to enhance the clarity of this work. We will go line by line and describe our implementations of the suggested changes.

Line 145: Rewrite into "MCMC is an algorithm applied to Bayesian problems where, given prior information of the parameters..."

Thank you for the suggestion, we will rewrite Line 145 to "MCMC is an algorithm applied to Bayesian problems where, given prior information of the parameters..." as suggested.

Line 147: "in this case, a log-likelihood function". Function of what?

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.

Line 149: Replace "prior" with "initial"? In Bayesian problems "prior" refers to a distribution, when in your case you provide a single value. Same in the caption of Figure 3.

Thank you for the suggestion. We would like to clarify that we do mean this as a Bayesian prior; the initial value is effectively the mean of the prior distribution, for which we also provide a prior standard deviation (prior parameter uncertainty estimate). We will add the following to clarify this point as follows:

"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: "with step size chosen from the distribution". Need to be more specific.

Line 152: We suggest rephrasing to better clarify: "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: You can add the uncertainty related to errors of representativeness of the observations, in each term of the equation.

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: "all distributions are assumed to be Gaussian". Specify which distributions.

Line 174: rephrasing to "all distributions (the prior parameter distribution, the likelihood function, and the posterior distribution) are assumed to be Gaussian"

Figure 3: Discuss why we see correlation between the properties, especially in JRA-55.

We note that this was also discussed in previous work, but we propose discussing this as follows: "The correlation between the wind packing and blowing snow parameters may be a consequence of the processes compensating for each other, as described in Cabaj et al., 2023. The wind packing process transfers snow to the lower layer, where the blowing snow process cannot remove snow, so if wind packing is strengthened, the blowing snow process may also be strengthened to compensate and enable additional snow depth reduction."

Reference:

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.

2. The scaling issue and the sub-grid variability of the snow properties is not discussed enough in the paper. More attention should be paid in this, especially as point measurements of density are used in the calibration.

We did consider these points when conducting the analysis but we agree that it would be beneficial to mention them in the paper. Following this comment and a comment from another reviewer, 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 it does not capture small-scale effects from localized ridging and small-scale leads often seen in observational studies (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). 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.

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.

Regarding the density measurements used in the calibration, we would like to clarify that although the source of the density values is point measurements, density observations are aggregated to a monthly climatology to address representativeness. This is stated in Line 162: "Basin-averaged monthly climatologies are used for the drifting station and buoy measurements, and OIB measurements are aggregated to daily averages over the NESOSIM model grid." That said, we will add the following to further clarify: "This aggregation helps to mitigate the impact of observational biases due to the relatively sparse and infrequent observations in these datasets."

Why do you use ERA5 wind in all runs? Consider using wind inputs from different reanalysis data to investigate their effect in the calibration.

We chose to use ERA5 wind in all runs because we wanted to isolate the contribution of snow input to NESOSIM model output. 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. We note that ERA5 has been found to perform well relative to other reanalysis wind products in studies over Arctic sea ice (Graham et al., 2019a, 2019b), which motivates the choice of ERA5. 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 nevertheless may 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.

You compare post calibration results from NESOSIM to SnowModel-LG but not to independent measurements like passive microwave products or airborne campaigns (OIB/IceBird). Adding comparison to independent measurements will strengthen the study. Regarding the comparison of NESOSIM to SnowModel-LG you should consider the effect of Eulerian (NESOSIM) vs. Langragian (SnowModel-LG) approach when discussing the *differences between the model results. Is SnowModel-LG also forced with CloudSat scaled reanalysis forcing? If not, this is another aspect that needs to be emphasized and discussed.*

We agree that comparison to independent measurements would be beneficial to the study. Motivated by suggestions below (and from other reviewers), we have decided to compare to snow depth and density measurements from MOSAiC. We have also conducted some preliminary comparisons with IceBird, but due to the coarseness of the model grids and the proximity of IceBird observations to coastal areas, the coincident coverage between IceBird observations and model output is comparatively limited. We will present some results from a preliminary comparison of MOSAiC output with NESOSIM and SnowModel-LG below after we address the other points in this paragraph.

We agree that making note of the differences between the Eulerian and Lagrangian approaches would be helpful to discuss the differences between model results. We will include the following at line 560:

"inter-product differences. NESOSIM and SnowModel-LG also have different approaches for addressing sea ice drift. The Lagrangian approach of SnowModel-LG allows it to track individual ice parcels and thus can more robustly account for the contribution of sea ice dynamics to the snow budget. NESOSIM presents a more simplified approach, but is less computationally expensive to run. Limitations in NESOSIM's [...]"

SnowModel-LG is not forced with CloudSat-scaled reanalysis forcing; whereas we run NESOSIM ourselves, the SnowModel-LG output we use is provided by the NSIDC. We will clarify in our manuscript that SnowModel-LG does not use CloudSat-scaled inputs, although it does include other observation-based adjustments of its input data. We propose a modification to 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.

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.

Returning back to the topic of comparison to independent measurements, below we present a brief summary of a comparison to MOSAiC observations, which we intend to incorporate into the subsequent revision of our manuscript, below.

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.

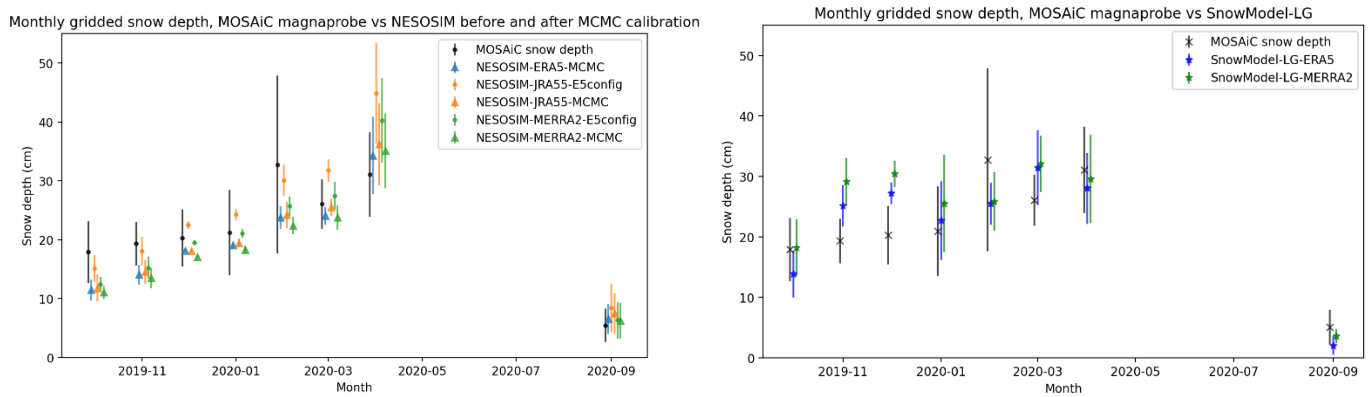

Figure A: (left): MOSAiC magnaprobe snow depth (Itkin et al., 2021) compared to NESOSIM snow depth, before ("E5config") and after ("MCMC") individual dataset calibration to observations; triangles indicate individually-calibrated datasets. (right): MOSAiC snow depth compared to SnowModel-LG snow depth 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.

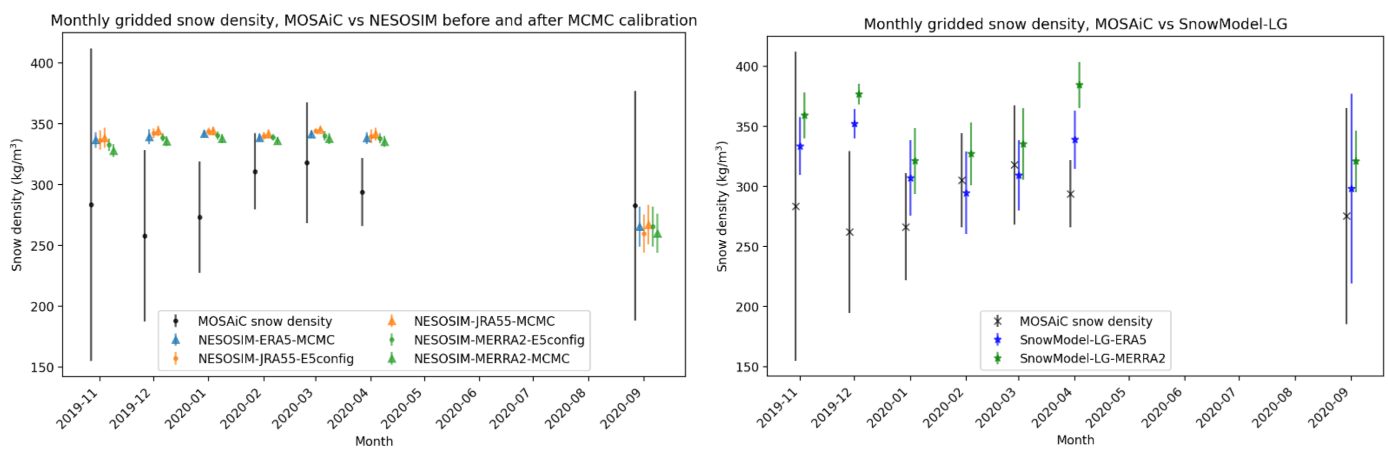

Figure B: Monthly averages of gridded MOSAiC snow density cutter data vs. monthly averages of coincident model output data. Error bars indicate 1 standard deviation. (left): MOSAiC compared to NESOSIM, before ("E5config") and after (MCMC) the MCMC calibration. (right): MOSAiC compared to SnowModel-LG.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

Why does the analysis stop in 2019? Consider extending to 2022, so MOSAiC observations can be included in the MCMC calibration.

We initially intended to include 4 decades in this work; hence, we performed the analysis from 1980-2019. Following the reviewer suggestion, we have produced NESOSIM output up to April 2023, the most recent period for which all required NESOSIM input forcings are available. We will update the dataset associated with this article to include the updated NESOSIM output. For the purposes of the SnowModel-LG comparisons, however, we need to limit the analysis to end in April 2021 for consistency (since SnowModel-LG output is not available after July 2021). *To the editor: We will adjust all NESOSIM and SnowModel-LG output/comparison figures to include these additional years in the subsequent manuscript revision and revise our manuscript accordingly.*

A challenge when selecting datasets to use for the MCMC calibration is to avoid overfitting to observational datasets. The observations we chose for the MCMC calibration were chosen because they were available over many years, and because they covered a wide spatial extent. For example, Operation IceBridge, although limited to the spring season, covers several different regions with measurements spanning multiple NESOSIM grid cells for each measurement day, and several seasons of data are available. Since MOSAiC measurements are available for only a single season, and are generally localized to a single grid point per day, incorporating them into the MCMC calibration may result in overfitting.

However, as discussed above, we are now proposing including a brief comparison of NESOSIM and SnowModel-LG snow depth and snow density to MOSAiC into our article, to provide a set of independent measurements to assess the impact of the MCMC calibration, and to compare the representation of snow in NESOSIM and SnowModel-LG. This comparison reveals some of the challenges with comparing gridded products with point measurements, since there is high variability in the measured snow depths and densities which is not represented by the models.

Minor comments:

1. Need to specify the blowing snow parameter better. I assume it refers to a snow loss term to the atmosphere (i.e., sublimation) and the open ocean. Make clear that snow is not blown from one 100 km x 100 km grid to another.

Yes, the blowing snow term refers to loss to the atmosphere and open ocean. We will add two sentences at Line 139 to clarify this. “The blowing snow term is exclusively a loss term and does not include redistribution. When snow is lost from a grid cell via this process, it is removed, not redistributed to another grid cell.”

Following suggestions from other reviewers, we also propose expanding the text at Line 134 as follows for additional clarity:

“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 4 and 5 include only one season. Consider an inter-annual average monthly evolution plot for all properties and their uncertainties.

We wanted to show a representative single season so that the uncertainties due specifically to the different products would be apparent, since in the inter-annual average, the monthly uncertainties are less than the standard deviation of the monthly climatology. *To the editor: In the revision, we propose to include some representative inter-annual average monthly plots of this in a supplement.*