Response to Reviewer #3

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: Cabaj et al, Investigating the impact of reanalysis snow input on an observationally calibrated snow-on-sea-ice reconstruction

General

The paper investigates the impact of snowfall rates from three reanalysis products on two calibrated parameter values in the NASA Eulerian Snow on Sea Ice Model (NESOSIM). NESOSIM is a two-layer snow budget model used to calculate depth and density of snow on sea ice for polar (mostly Arctic) oceans. NESOSIM is forced by snowfall, wind speed, sea ice concentration, and ice motion. It accounts for horizontal transport of snow cover by ice drift, densification and metamorphosis of the snow pack, and loss of snow cover to leads through wind erosion and transport. The current paper focuses on calibrating paramaters that control these last two processes, and on correcting reanalysis snowfall. It builds on two earlier papers: one paper that presented a snowfall bias correction procedure using CloudSat as the target snowfall; and a second paper that presented a Markov Chain Monte Carlo automated calibration method.

It is not clear to me what new insights have been gained by the current study or if any new method or solution is presented. The automated calibration procedure using Monte Carlo Markov Chain is described in Cabaj et al (2023) and the snowfall bias correction is presented in Cabaj et al (2020). Although the current study evaluates using multiple reanalysis products, I do not think this evaluation in it's current form adds much that is new.

We thank the reviewer for the feedback. This study builds on previous work in important ways by quantifying how uncertainties from distinctive sources of forcing could be strongly linked to parameter uncertainty in snow on sea ice modeling. The particular application, here, combining NESOSIM and MCMC optimization, enables automated parameter calibration that can account for differences in forcing (which we have chosen to limit to snowfall in this case to keep the paper manageable). We are not aware in the current literature of this kind of systematic model development approach, including uncertainty quantification, at the level provided here. Without the relatively low cost of NESOSIM, such an approach would have been impossible (see next reply). This study provides a strong example for how to calibrate models and account for forcing uncertainty for an important and uncertain climate metric (snow on Arctic sea ice). Because of reviewer interest, and our interest, in the critical measurements from MOSAiC, we will now also include a limited comparison with field measurements from the campaign, to further place our results and the uncertainty analysis in context of recent observations independent of the calibration process.

NESOSIM is a relatively simple conceptual model that uses simple parameterizations of blowing snow losses and densification. At a fundamental level the model can be seen as a transfer function that corrects for biases in snowfall products. The paper demonstrates that parameter values for blowing snow and densification are sensitive to the snowfall product. It also demonstrates that the two parameters chosen for calibration are highly interdependent. The first point is well known. There are numerous examples of model sensitivity to input in the statistical and Earth science literature, including in the terrestrial snow modelling and snow hydrology literature. The second point should not be surprising because the blowing snow parameterization removes snow (reducing the bias in snowfall input) and the densification parameterization (the rate at which low density new snow is "transformed" into high density old snow) increases/retains snow (increasing the bias in snowfall input). I suspect a "brute-force" evaluation on the parameter space would have shown this interdependence. The question for the modeller is how to constrain this interdependence.

We respectfully disagree with the suggestion that *NESOSIM is a relatively simple conceptual model*. In fact, NESOSIM is regularly applied operationally (providing snow depth and density estimates on Arctic sea ice for estimation of sea ice thickness from ICESat-2 altimetry (Petty et al., 2023)), and has been shown to perform well quantitatively in previous publications, provided suitable sampling and interpretation are applied (Zhou et al., 2021). Even with its simplicity, NESOSIM is not dissimilar from other commonly-used reanalysis-based approaches used for satellite sea ice thickness estimation, and it improves on simpler approaches such as the widely-used fractional scaling of the Warren et al. (1999) climatology (as used in e.g. Kwok and Cunningham, 2015; though still used as a point of comparison in more recent studies such as Zhou et al., 2021). By contrast, *conceptual models* are used primarily for pedagogical studies or for idealized analytical studies. Our study capitalizes on NESOSIM's low cost and simplicity to carry out a thorough characterization of its uncertainty across the important dimensions of its free-parameter uncertainty and snow-input related forcing uncertainty. In this sense, NESOSIM serves as a benchmark model that can be improved upon for similar automated calibration approaches using more comprehensive models. But we disagree that the model should be dismissed in and of itself as is suggested here.

The motivation for using an MCMC approach was to avoid the very "brute-force" approach the reviewer suggests using. Prior to the use of this approach, we did not know that the parameter space would be smooth, and we are glad that the inter-dependence makes sense to the reviewer, as it does to us. The quantitative estimate of this dependency and its dependence on forcing dataset are new results, even if the sign of the dependence could have been anticipated physically. Using automated parameter calibration positions us well for future changes in forcing datasets and model updates. For example, when NESOSIM was previously updated to use ERA5 input instead of ERA-Interim, the free parameters needed to be adjusted, and the parameters were previously selected by hand based on best estimates to agreement with selected observations. An automated approach was sought out to streamline and better justify this process, and to quantify parameter uncertainty around the optimal estimate (which brute-force approaches are incapable of doing). As ERA5 and other updated reanalysis products come onstream, we will be able to re-calibrate NESOSIM and its updates relatively quickly, guided by the experience outlined in this study. We will add some mention of these points in the discussion.

References: Kwok, R., & Cunningham, G. F. (2015). Variability of Arctic sea ice thickness and volume from CryoSat-2. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, *373*(2045), 20140157. https://doi.org/10.1098/rsta.2014.0157

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.

Warren, S. G., Rigor, I. G., Untersteiner, N., Radionov, V. F., Bryazgin, N. N., Aleksandrov, Y. I., and Colony, R.: Snow Depth on Arctic Sea Ice, *Journal of Climate*, 12, 1814-1829, https://doi.org/10.1175/1520-0442(1999)012<1814:SDOASI>2.0.CO;2, 1999.

Zhou, L., Stroeve, J., Xu, S., Petty, A., Tilling, R., Winstrup, M., Rostosky, P., Lawrence, I. R., Liston, G. E., Ridout, A., Tsamados, M., and Nandan, V.: Inter-Comparison of Snow Depth over Arctic Sea Ice from Reanalysis Reconstructions and Satellite Retrieval, *The Cryosphere*, 15, 345-367, https://doi.org/10.5194/tc-15-345-2021, 2021.

I also do not think that the approach described here gets at the question of uncertainty. There are (at least) three sources of uncertainty in models: uncertainty related to input fields, parameter uncertainty, and model structural uncertainty. Although the scaling with CloudSat address some of the input bias and uncertainty, the different parameter distributions for the different reanalysis products suggest that input uncertainty is also included in the assessment of parameter uncertainty. I would suggest that a better understanding of input and parameter uncertainty would be gained by calibrating the model on measured snowfall from MOSAiC or the drifting stations, and then using these parameter estimates in a bias correction step. The model (parameterizations and bias correction) should then be tested against data that has not been used for calibration. This is an important step in any modelling study but has not been included in the present study.

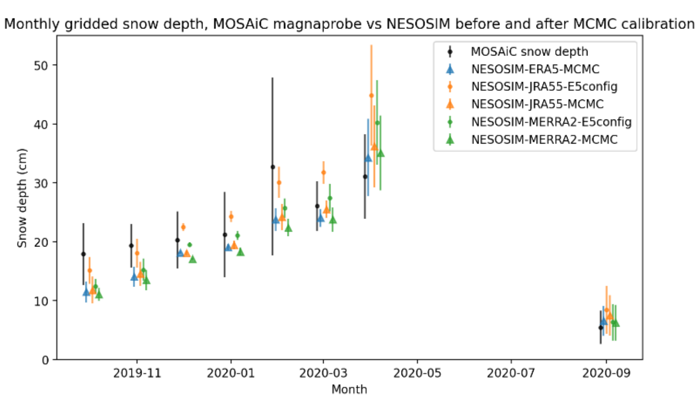
It is of course not possible within a single study to fully address all sources of uncertainty that the reviewer leads with here, nor did we claim to do so. Our approach was to look at an existing snow-on-sea-ice estimation workflow and identify key parts of it that could be improved through automated calibration, so we limited our analysis to that part of the reanalysis we thought would play the strongest controlling role – the snowfall input. CloudSat was chosen for snowfall calibration because it represents a multi-year dataset with comparatively widespread observational coverage. MOSAiC data, although of high quality, is highly localized and only available for a single year. The key issue with any one campaign like MOSAiC is that of representation of its measurements in our effort to produce a multi-decadal pan-Arctic dataset. Drifting station snowfall measurements are likewise relatively localized compared to CloudSat measurements, and we would prefer to use contemporary observations for this work.

We agree that out-of-sample testing can be improved and we are using the most recent seasons to generate tests of the current parameter settings that we will briefly summarize. The prompt to use MOSAiC data is a good idea; rather than use it for calibration, we will use it as an independent dataset for comparison and we will provide a summary analysis that will not add too much to the length of the paper. Interestingly, the comparison of NESOSIM and SnowModel-LG with MOSAiC seems to reveal as many characteristics of the MOSAiC sampling strategy as of the models themselves. This analysis provides a significant impetus for future work in the area. This will also be touched on in the discussion. A brief summary of the analysis is included 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.



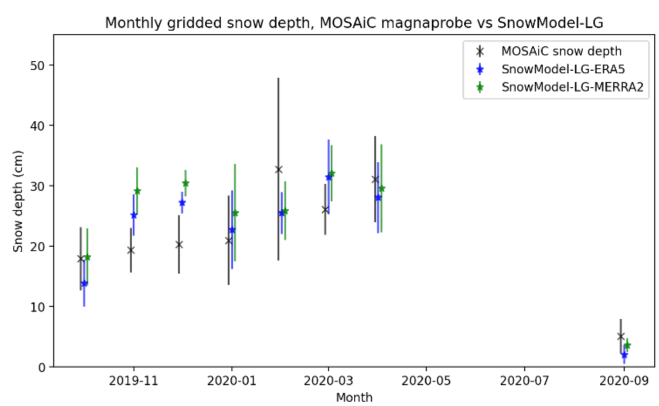


Figure A: (left): MOSAiC magnaprobe snow depth (Itkin et al., 2021) compared to NESOSIM snow depth, before ("E5config") and after ("MCMC") individual dataset calibration to observations; triangles indicate individually-calibrated datasets. (right): MOSAiC snow depth compared to SnowModel-LG snow depth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

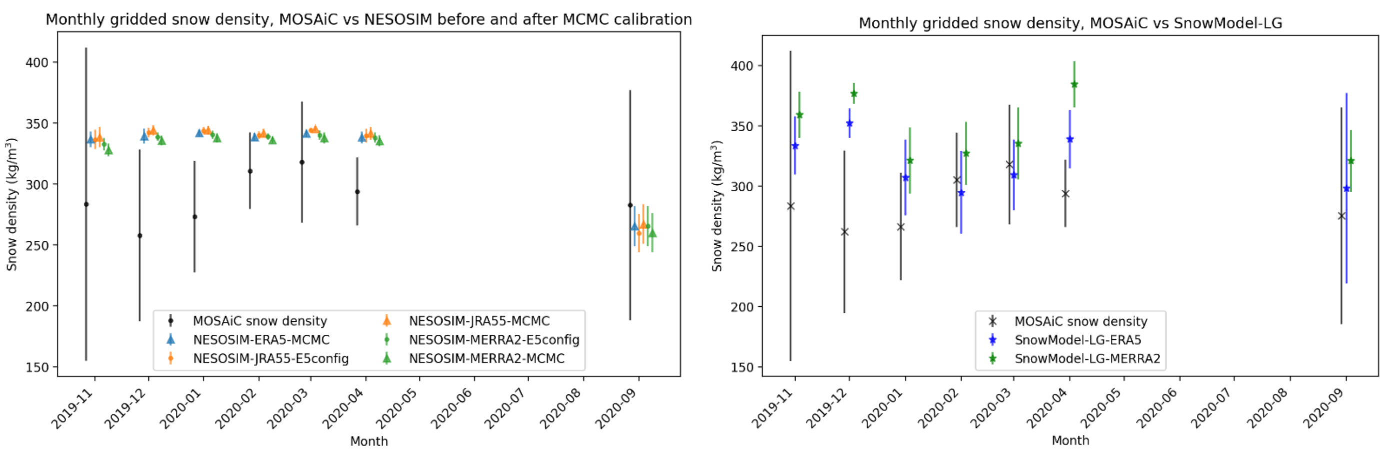


Figure B: Monthly averages of gridded MOSAiC snow density cutter data vs. monthly averages of coincident model output data. Error bars indicate 1 standard deviation. (left): MOSAiC compared to NESOSIM, before ("E5config") and after (MCMC) the MCMC calibration. (right): MOSAiC compared to SnowModel-LG.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

Another issue with the paper is the section on trend analysis. I don't think this section adds anything to the paper. Moreover, it is not always clear from the discussion when trends have passed tests for statistical significance. An example of this is:

> Although the trend magnitudes and seasonal cycles for snowfall vary by region, most of the trends for most products are not statistically significant at a 95% confidence interval, likely due to high interannual variability of snowfall.

In trend analysis, the most common purpose of a significance test is to decide whether or not the null hyposthesis, that the trend is zero (\alpha = 0), can be rejected. If it cannot be rejected (i.e. it is "insignificant"), the null hypothesis has to be accepted; no trend distinguishable from zero can be detected. Put another way, there is no trend, it is zero and it cannot be positive or negative value. So in the example above, trends for the regionds cannot be different if they are not statistically significant.

If the authors can demonstrate that this section is relevant to the paper, I would suggest only discussing statistically significant trends in the text and only showing statistically significant trends in the figures. Note, the grey hatching is not visible in the pdf version of the paper. I suggest not showing regions that do not pass the significance test. Also the overlapping shaded regions in the line plots obscure the information. As the plots show trends for each month, the statistically significant trends for each product should be shown as symbols with bars to show the 95% confidence intervals, grouped by month, because they are individual data points.

We respectfully disagree that the section does not add anything to the paper. The motivation for this section was to examine how trends interpreted from snow-on-sea-ice products can differ depending on which product and which reanalysis input is used, and to examine if trends differ more between products or between snow forcings. There is precedent for using reanalysis products to examine trends in Arctic precipitation (e.g. Boisvert et al, 2018) and given the lack of widespread long-term continuous observations of snow on Arctic sea ice, we find it justifiable to consider the use of reanalysis-based snow-on-sea-ice reconstructions such as NESOSIM (and SnowModel-LG, for that matter) to examine trends and variability. In fact, SnowModel-LG has recently been used to examine the influence of the Arctic Oscillation on summer snow on sea ice (Webster et al. 2024). Furthermore, previous sea ice studies have found that snow can influence sea ice volume (and therefore thickness) trends (Bunzel et al., 2018), and as such, for a product with sea ice thickness applications, we believe that investigating these trends is relevant.

References:

Boisvert, L. N., Webster, M. A., Petty, A. A., Markus, T., Bromwich, D. H., and Cullather, R. I.: Intercomparison of Precipitation Estimates over the Arctic Ocean and Its Peripheral Seas from Reanalyses, *J. Climate*, 31, 8441–8462, https://doi.org/10.1175/JCLI-D-18-0125.1, 2018.

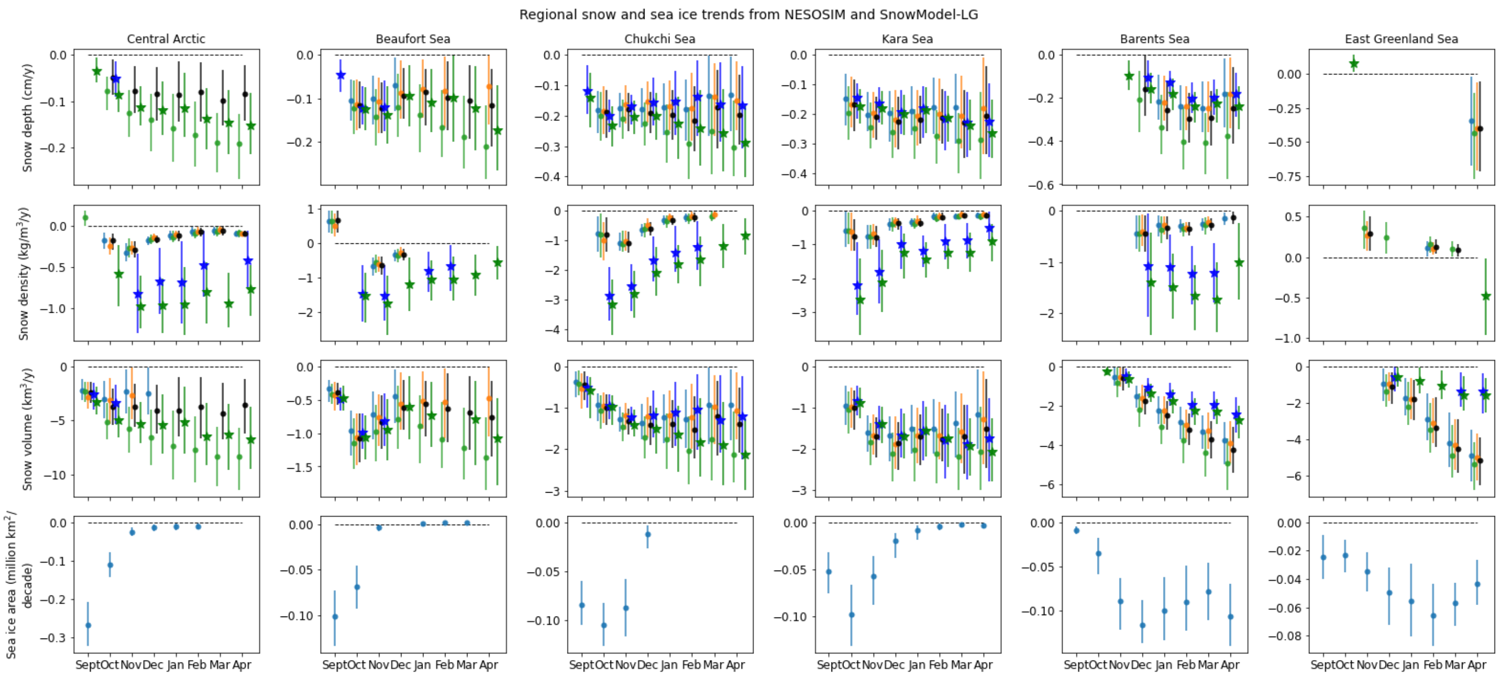
Bunzel, F., Notz, D., and Pedersen, L. T.: Retrievals of Arctic Sea-Ice Volume and Its Trend Significantly Affected by Interannual Snow Variability, *Geophysical Research Letters*, 45, 11,751-11,759, https://doi.org/10.1029/2018GL078867, 2018.

Webster, M. A., Riihelä, A., Kacimi, S., Ballinger, T. J., Blanchard-Wrigglesworth, E., Parker, C. L., and Boisvert, L.: Summer snow on Arctic sea ice modulated by the Arctic Oscillation, *Nat. Geosci*., 17, 995–1002, https://doi.org/10.1038/s41561-024-01525-y, 2024.

As with any reanalysis-based product, caution is required when interpreting trends. Our work demonstrates that different conclusions may be drawn about basin-wide and regional trends depending on what product is used. As discussed above, differing trends in snow on Arctic sea ice products have implications for ice thickness retrieval applications. We also find it interesting to explore how snow depth trends do not necessarily reflect reanalysis snowfall trends (or lack thereof).

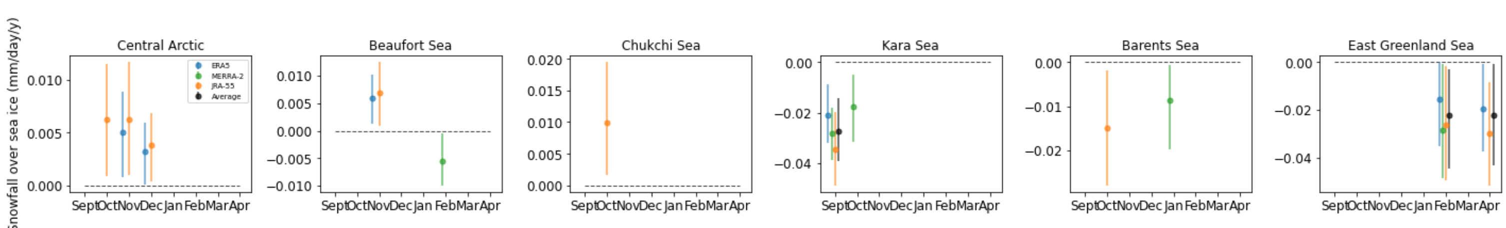
Regarding the phrasing around statistical significance, we agree that some of our phrasing was unclear, and we did not intend to overstate results that did not represent actual trends (i.e. were not statistically significant). We propose some substantial rephrasing of our section on trends further below to address this issue (refer to our comments further below addressing Line 414 and onwards), and we will likewise adjust phrasing in the discussion accordingly.

Regarding trend plots, as suggested, we will remove the values that are not statistically significant. As the reviewer suggested, we propose the following modification to Fig. 10; substituting the shading for bars. Note that this is an illustrative example; we will adjust the subplot horizontal spacing and include a legend for better legibility in the next revision of the manuscript. By suggestion of another reviewer, we will also include tables of values corresponding to these plots in a supplement. Also, we have removed the row with snowfall trends and will show that in a supplement, since snowfall trends are generally absent.

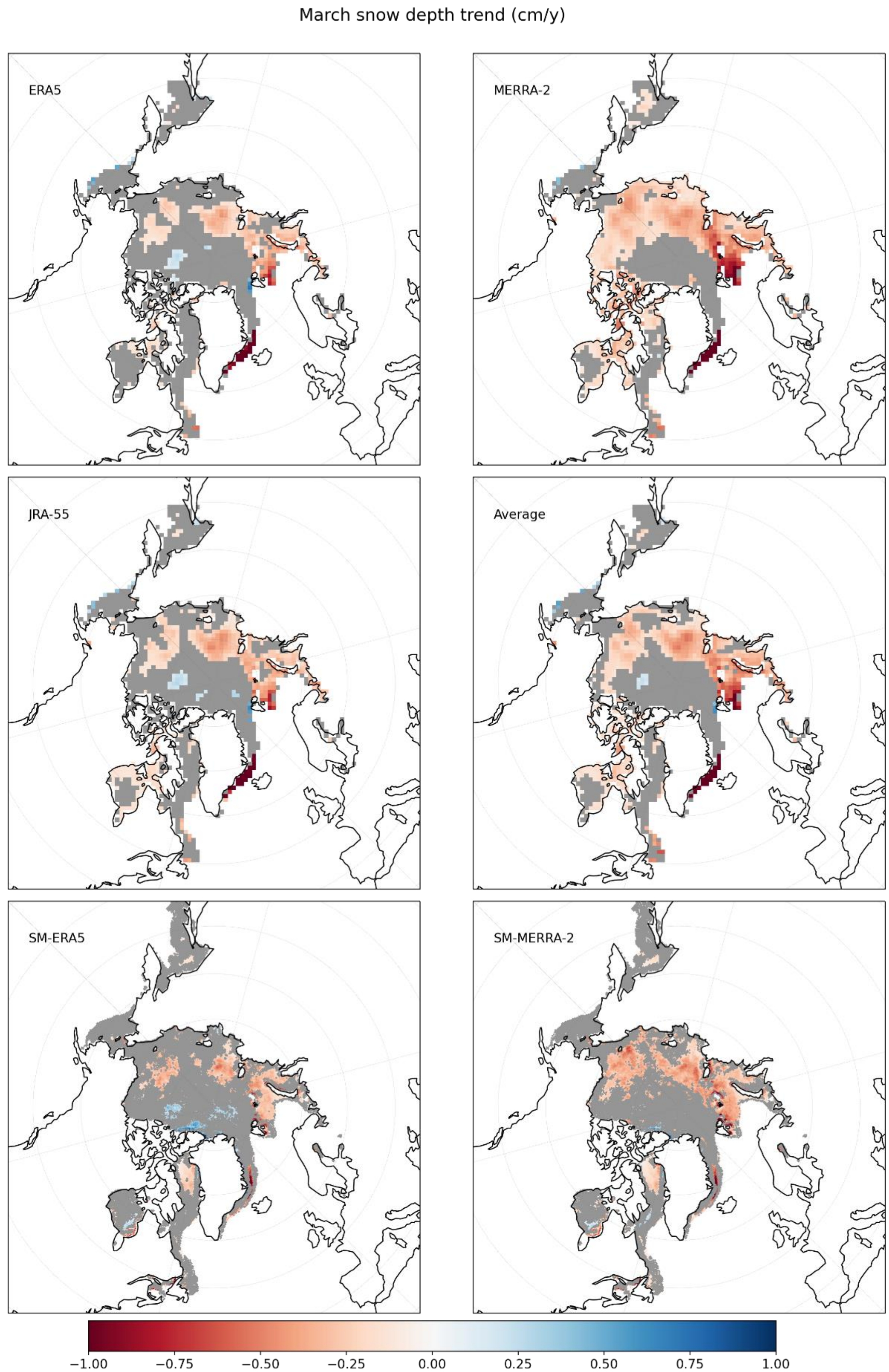


“Figure 10. Monthly trends for regionally-averaged quantities over the 1980-2019 time period: snow depth, density and volume (from NESOSIM and SnowModel-LG), and sea ice area (from the Climate Data Record product). Error bars indicate a 95% confidence interval as given by the trend estimator; points where there are no trends (the interval overlaps the zero line) are not shown.”

For reference, the regional snowfall trends over sea ice are shown below:



Likewise, we will fully grey-out regions with no significant trends in Fig 11 and 12, and will move Fig. 11 to a supporting figure. We show the updated Fig. 12 for reference; Fig. 11 will be updated similarly:



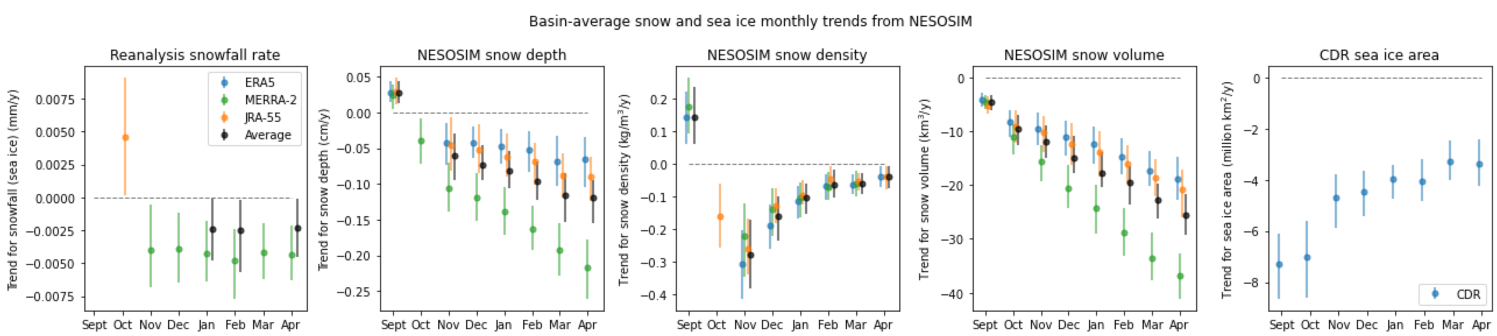
"Figure 12: Snow depth trend maps for March 1980-2019 from NESOSIM run with snowfall input from ERA5, MERRA-2, and JRA-55, and SnowModel-LG with snowfall input from ERA5 and MERRA-2. The snow depth is output from NESOSIM with parameters specific to each separate reanalysis product. The trend in the average of the output of the three NESOSIM runs is also plotted (Average). Regions with no trends (not significant to a 95% confidence interval) are shaded in grey. Note that SnowModel-LG is not provided within the Canadian Arctic Archipelago, so data from that region is absent in this map."

We propose modifying the text of the section on trends as follows (paragraphs labelled by line):

Line 414: "In the following discussion, trends are considered significant if the 95% confidence interval does not overlap with zero. If the confidence interval overlaps with zero, we consider there to be no trend."

Line 416: "Basin-average trends from NESOSIM for snowfall over sea ice, snow depth, snow density, snow volume, and sea ice area are shown in Fig. 9. The trends in snowfall over sea ice are not statistically significant for most products except for a ~~significant~~ decline for MERRA-2 from November onwards and an increasing snowfall trend in October for JRA-55. ~~The products also disagree on the sign of the trend for several months.~~ The basin-average trends in snow depth from MCMC-calibrated NESOSIM output vary in magnitude by product, but are all broadly similar in sign. MERRA-2 has the strongest trends in the basin-average overall. The trend is found to be negative (declining snow depth) in all months except September, where the trend is significantly positive for all products, and October, where the trend is not significant for the multi-product average; MERRA-2 shows a ~~significant~~ decline but the other products show no ~~significant~~ trend in that month. Snow density trends are generally ~~consistent~~ similar between the products, aside from October, where only JRA-55 shows a trend. ~~which is consistent with the fact that~~ Similarity between the snow density trends is expected, since snow density in NESOSIM is less sensitive to snow input, being primarily dependent on wind speed."

We modify Figure 9 as follows:



"Figure 9. Basin-average monthly trends from 1980-2019 for snowfall over sea ice from reanalysis products, MCMC-calibrated NESOSIM snow depth, density, and volume, and CDR sea ice concentration, calculated using a Theil-Sen trend estimator for all products. "Average" denotes the multi-product average. Error bars indicate a 95% confidence interval as given by the trend estimator; points where there are no trends (the interval overlaps the zero line) are not shown. The grey dashed lines indicate the zero line for reference. SnowModel-LG is excluded from this plot due to differences in model domains."

Line 431: "Regional trends in ~~reanalysis snowfall over sea ice are shown in Fig. 10 along with trends in~~ snow depth, snow density, snow volume, and sea ice area are shown in Fig. 10. Regional trends in snowfall are shown in Fig. A#; although there is regional variation in snowfall trends, ~~the trend magnitudes and seasonal cycles for snowfall vary by region, most of the trends for most products are not statistically significant at a 95% confidence interval,~~ most products show no trend for most months, likely due to high interannual variability of snowfall. ~~Different reanalysis products disagree on the sign of snowfall trends in the Central Arctic, with MERRA-2 indicating declining trends throughout the season, and ERA5 and JRA-55 suggesting early-season increases. Trends in the East Greenland Sea region are large in some months but generally not significant, except for some products in February~~

~~and April, where a declining trend is observed.~~ A large and significant early-season snowfall decline is apparent in the Kara Sea region, but only for the month of September for most products."

Line 441: "The East Greenland Sea region differs noticeably ~~in seasonality~~ from the other regions shown~~, with a slight but not significant increase until February-March, where a declining trend is found~~ with no trend for most months except for a slight increase in November for SnowModel-LG driven by MERRA-2, and declines in April only for NESOSIM products~~. This declining trend is weak in SnowModel-LG, but strong and significant in NESOSIM.~~ In the Central Arctic region, ~~all products show a declining trend after September, but this trend is not significant in SnowModel-LG driven by ERA5 and NESOSIM driven by JRA-55.~~ declines are generally seen only for products driven by MERRA-2, with no trend for products driven by JRAA-55, and only a slight October decline for SnowModel-LG driven by ERA5."

Line 450: change sentence to "In most regions, the strongest declining trends are found in MERRA-2, whereas trends often tend to be smaller or absent for ERA5, for both NESOSIM and SnowModel-LG."

Line 452: "Several regions demonstrate ~~near-zero~~ small or absent snow depth trends in September followed by an abrupt decline in the following months. These patterns are not reflected in the snowfall trends, and are likely to be related to sea ice decline. In the Central Arctic region, where more sea ice is present during the early months, the early-season change in the trend is less pronounced ~~steep~~. However, in the Kara Sea region, which experienced a significant declining trend in snowfall over sea ice in September for all snowfall products ~~and October~~, a corresponding decline in snow depth is not observed in September in either NESOSIM or SnowModel-LG. Nevertheless, significant declines are found in this region for ~~other~~ later months."

Line 458: "For snow density trends, inter-model differences tend to be larger than inter-product differences. Declining trends are largest ~~strongest~~ around October-November for most products and regions, except in the Barents and East Greenland Seas. Densities in SnowModel-LG tend to show large ~~and significant~~ declines relative to NESOSIM. As discussed previously, NESOSIM end-of-season density trends may be spuriously low due to NESOSIM snow densities approaching their maximum towards the end of the season, although end-of-season density trends as represented in SnowModel-LG also tend to be smaller. ~~weaker~~."

Line 463: "Trends in snow volume closely mirror snow depth trends in several regions, though differences are nevertheless apparent. In the Central Arctic, Beaufort Sea and Chukchi Sea regions, there is a notable significant early-season decline in snow volume for all products, whereas such a decline is not necessarily found in the corresponding snow depths for these regions. ~~the snow depth declines, if present, are not so significant in the early season for these regions~~. These volume declines are associated with strong early-season sea ice area declines. The inter-product spread in trends increases towards the end of the season, however. In the Kara Sea region, ~~September snow volume trends remain not significant, but become significant from~~ snow volume trends are declining from October onward. In the Barents and East Greenland Sea regions, ~~early-season trends are small and not significant, but~~ later-season declines are also apparent for most products. Snow volume trends in these two regions differ considerably in seasonality from the snow depth trends, with stronger late-season declines, likely influenced by sea ice area decline in these regions. There is a large inter-model difference in trend magnitude between NESOSIM and SnowModel-LG in the East Greenland Sea region, with NESOSIM showing much larger declines overall, and SnowModel-LG driven by ERA5 not showing any declining trend until March. The largest snow volume declines are found in the Central Arctic for NESOSIM driven by MERRA-2, although this region also has a very wide inter-product spread, with NESOSIM driven by ERA5 and JRA-55 and SnowModel-LG driven by ERA5 not showing ~~significant~~ late-season snow volume declines. Thus, the choice of reanalysis also has an impact on snow volume trends, though inter-model differences are more readily apparent in some regions."

Line 477: "Sea ice area trends vary by region, but strong declines are found for at least part of the season in all regions shown. In the Central Arctic and the Siberian sector, as well as the Beaufort Sea, the ~~strongest~~ largest declining trends are in the earlier months of the cold season. (~~Stronger~~ Larger trends may be present in months outside of the NESOSIM study period.) When sea ice in these regions attains its maximum extent, the trends largely vanish, suggesting a persistent cold-season cover. Towards the North Atlantic (Barents, East Greenland), ~~stronger~~ larger declines are seen in later months."

Line 482 (A# is a placeholder for the number of the figure in the final manuscript): "To provide a more regional perspective on snow trends, ~~Fig. 11 shows maps of March snowfall trends in the reanalysis products and their average. M~~ maps of snow depth trends in NESOSIM and SnowModel-LG output are shown in Fig. 12. Corresponding snowfall trends are shown in Supporting Fig. A#. For these plots, trends were also calculated using a Theil-Sen estimator, but only grid squares containing at least 20 years of values were included to exclude spurious trends. Consistently with results from the regional monthly trend plots, ~~snowfall trends are not significant for~~ there is a lack of snowfall trends over most of the Arctic basin, due to the high interannual variability of Arctic snowfall relative to the magnitude of the trends. The depth trends are more robust, highlighting a decline in the peripheral seas consistent with the results shown in the regional plots, as well as some slight declines around Hudson Bay and Labrador Sea. Some significant increasing depth trends north of the Beaufort Sea are found in both SnowModel-LG products, as well as in NESOSIM driven by ERA5 and JRA-55, though the products differ on the existence ~~significance~~ of the increasing trend near the North Pole."

Line 494: "There is broad consistency, however, in the ~~strong~~ declining trend found ~~observed~~ in the Barents Sea region. The overall ~~strong~~ large declining trends in depth derived from MERRA-2 are particularly apparent in Fig. 12~~; the increasing depth trends around the central Arctic are not significant and have a narrower spatial extent than those in ERA5 and JRA-55~~. ERA5 and JRA-55 agree better on the spatial pattern of the snow depth trends compared to MERRA-2."

Line 499: "There are small but significant increases in snow depth in Hudson Bay and some strong declines east of Greenland that are absent from the NESOSIM output.~~, and some strong significant increases east of Greenland that are not significant in the NESOSIM output."~~

The reader is left to do a lot of work to understand what was done and how the NESOSIM model works. The description of the scaling using CloudSat is very brief (Lines 211 to 215). Cabaj et al (2020) describes the scaling but I cannot find how it is applied to the four Arctic quadrants in Cabaj et al (2023). In my opinion, papers should contain sufficient information to allow a reader to understand what was done. I would recommend including a brief description of how CloudSat scaling is applied and how blowing snow and wind packing parameters are applied.

We thank the reviewer for indicating where clarification is needed, and will make changes accordingly. It appears that we have erroneously left in some text which was supposed to be relegated to a different section; we are moving the description of the quadrants to Section 3.2 and will expand the description to clarify. We will also modify Fig. A1 to better illustrate the quadrants (including the associated interpolation points). We propose the following changes to the text below:

From line 205:

Figure 1 shows regionally-aggregated monthly-mean snowfall rates from reanalysis products and CloudSat, from 1980-2016, without and with scaling to the CloudSat monthly climatology (Cabaj et al., 2020).~~ as described in Cabaj et al. (2020).~~ The scaling entails taking the monthly reanalysis snowfall rate for each month and multiplying it by a scaling factor, which consists of the CloudSat climatological monthly mean snowfall rate divided by the reanalysis climatological monthly mean snowfall rate for each respective month. The climatological means for this scaling are taken from 2006-2016, excluding months in 2011 where CloudSat observations are absent due to instrument malfunctions. Further details of this scaling are provided in Cabaj et al. (2020). Before the scaling is applied in Fig. 1, […]"

From line 214, text moved to line 205 above:

“As in Cabaj et al. (2020), we bias-adjust reanalysis snowfall input for NESOSIM to climatological CloudSat snowfall for 2006-2016 ~~(excluding months in 2011 where CloudSat observations are absent due to instrument malfunctions). The adjustment uses multiplicative scaling interpolated across four Arctic quadrants, a level of aggregation that was found to be a necessary to obtain robust results (Cabaj et al., 2023).~~ CloudSat scaling [...]”

From Line 222 revising as follows:

“To apply CloudSat scaling over the NESOSIM model domain, reanalysis snowfall rates are scaled to CloudSat measurements from 60-82°N over four quadrants, as described in Cabaj et al. (2020). The scaling coefficients for these quadrants were obtained by dividing the NESOSIM model domain north of 60N into four quadrants (illustrated in Fig. A1), calculating monthly reanalysis and CloudSat snowfall climatologies for each respective quadrant to obtain climatological scaling factors for each quadrant following the approach in Cabaj et al. (2020). Then, the scaling factors were linearly interpolated over the model domain from the outer corners of the model domain towards the centre, with one scaling factor at each corner. The CloudSat scaling was found to improve agreement in basin-averaged and regionally-averaged snow depths in NESOSIM v1.0, as was discussed in Cabaj et al. (2020). Some adjustments were made to the scaling for NESOSIM v1.1, which has a larger model domain (Petty et al., 2023), extending down to 50°N, compared to 60°N for NESOSIM v1.0 (Petty et al., 2018). The model domain of NESOSIM v1.1 is shown in the map in Fig. A1, with shading indicating the 60-82°N latitude band, and reference points for the scaling coefficient interpolation indicated. ~~To apply CloudSat scaling over the NESOSIM model domain, reanalysis snowfall rates are scaled to CloudSat measurements from 60-82°N over four quadrants, as discussed in Cabaj et al. (2020), and are linearly interpolated over the model domain from the corners.~~ The longitudinal boundaries of the quadrants are at longitudes 135°W, 45 °W, 45°E, and 135 °E, respectively, as illustrated in Fig. A1. For this work, the scaling is set up as follows. The same scaling coefficients as previously calculated for Cabaj et al. (2020) are used. Over a central rectangular subdomain with corners at 45°N (longitudes of 90°W, 0°E, 90°E, and 180°E), the scaling factors are linearly interpolated from the corners across the pole. For the rest of the domain, the factors ~~In NESOSIM v1.1, the scaling factors are unchanged, but~~ are extrapolated southward as constant values to cover the extended model domain.”

Regarding describing how wind packing and blowing snow factors are applied, we are adding the following to clarify based on this comment and suggestions from other reviewers, at Line 134:

“using an MCMC process (Cabaj et al., 2023). At each model step and grid point, if the input wind speed exceeds 5 m/s, the wind packing and blowing snow processes act on the snow in NESOSIM. Wind packing transfers snow from the upper (less dense) model layer to the lower (denser) model layer, ~~controls the amount of snow transferred between layers, impacting~~ decreasing the snow depth and increasing the bulk snow density. The wind packing factor scales the amount of snow transferred between layers at each timestep. 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. 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. For the purpose of this study, the blowing snow term parameters are treated as a single term, as was done in previous work (Cabaj et al., 2023), with the atmospheric loss factor being 0.15 times the blowing snow parameter. ~~Both the wind packing and blowing snow processes are subject to a wind action threshold of 5 m/s.~~ This current study will extend previous parameter calibration work by investigating the impact of using different reanalysis snowfall input products in NESOSIM.”

The authors should use the correct citations for datasets and I would encourage them to cite DOIs. For example. The correct citation and title for the NOAA/NSIDC sea ice concentration is:

Meier, W. N., F. Fetterer, A. K. Windnagel, and S. Stewart. 2021. NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 4. [Indicate subset used]. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center https://doi.org/10.7265/efmz-2t65. [Date Accessed].

This allows the data to be found easily and quickly instead of digging through the paper describing the dataset. They should should check that the data citations are correct for the other datasets they use.

We appreciate the correction and will double-check and revise our dataset citations, and include DOIs where possible. We note that some of the datasets (e.g. the CRREL buoy data) do not have DOIs (we have cited those data in the format requested by the dataset website, though we have noticed a duplicate reference there which we will clean up).

It is not clear to me why the CRREL Buoy snow depths are treated as monthly climatologies but the OIB snow depths are daily. Why not use daily data from the buoys.

This is somewhat of a subtle point, and relates to the spatial extent of the observations. In our previous model calibration work (Cabaj et al., 2023), we found that using daily buoy data led to the calibration being overconstrained; each buoy is localized at a single point (rather than having a comparatively large spatial extent, as OIB did). Since capturing small-scale local variations is challenging for a model with a 100-km grid, we opted to aggregate the buoy measurements for the purpose of the calibration. The OIB measurements are also aggregated, but to a daily 100 km grid. We did also test using climatological values for the OIB measurements in our 2023 publication, but we ultimately found that the daily measurements provided a good balance between data coverage and constraint.

Specific Comments

Line 227: Suggest "described" instead of discussed.

Thank you for the suggestion; we have moved this line up as mentioned in the revisions above, (starting from line 222) but have rephrased this sentence accordingly.

Line 228: I have no idea what "interpolated from the corners" means here. Corners of the quadrants or of what?

We intended this to mean from the corners of the quadrants towards the centre of the pole, although offset at a higher latitude. We have rephrased as described earlier in our revision (from line 222) and will update Figure A1 with the relevant regions outlined and the corner points identified to clarify.

Line 242: Which regions and months?

Over sea ice, the overall inter-product spread increases for Sept-Nov in the Central Arctic, October in the Beaufort Sea, Oct-November in the Chukchi Sea, all months except September in the Kara and Barents Seas, and all months except April in the East Greenland Sea. We will specify this in the revised manuscript.

Line 243: I am trying to understand why the inter-reanalysis spread is not reduced by scaling with CloudSat. The discussion indicates that over the central Arctic, the inflation of JRA-55 may be because CloudSat does not extend above 82N. This suggests to me that an alternative scaling strategy for this region should be explored.

Overall, the CloudSat scaling performs best in aggregate over large regions, since smaller regions may have high variability which may not necessarily be accounted for by the scaling. When the analysis is relegated to smaller regions and restricted to include only sea-ice-covered regions (which also decreases the size of regions analyzed), we find that the scaling performs worse. We considered alternate scaling approaches, but there is a tradeoff with using smaller reference regions for calculating scaling factors, since then, spatial undersampling from CloudSat may be encountered. In particular, one approach we attempted was calculating CloudSat scaling factors using only reanalysis and CloudSat snowfall over ice-covered grid cells as determined by a sea ice area climatology. This did improve the agreement even in smaller regions, but as we discussed in the article, we encountered sampling issues as shown in Figure A3: Due to limited coincident CloudSat measurements coincident with the (limited-extent) sea ice in the Greenland/Norwegian Sea quadrant, CloudSat failed to reproduce the monthly climatology in that region, and this biased the precipitation excessively low. We do agree that exploration of additional scaling strategies would be of interest, but we leave this as a consideration for future work.

Line 250: How well does the CloudSat snowfall rate algorithm perform over sea ice?

The primary factor influencing CloudSat retrieval performance over sea ice is enhanced ground clutter due to the higher elevation of sea ice (relative to open ocean). The CloudSat retrieval algorithm estimates surface snowfall from vertical radar profiles by selecting a designated near-surface bin. To mitigate adverse impacts from sea ice, for locations over a climatological sea ice mask, the retrieval algorithm selects snowfall from the 5th radar bin above the (retrieval-designated) surface rather than the 3rd radar bin from the surface (as is usually the case over ocean). We would expect some biases to be introduced here if sea ice were present in regions that did not coincide with the climatological mask. In an effort to mitigate this, we filter data based on retrieval quality flags as was done in Cabaj et al. (2020) but we are aware that some biases may still be present. We will add mention of this to Section 2.2.

Line 267: It is not clear if the first iterations refer to the burn-in period or to the first iterations after the burn in period. Please clarify in the text.

We will rephrase as follows: “the optimal posterior parameter~~s~~ values did not differ significantly between the first (burn-in) and subsequent (after burn-in) sets of iterations,”

Figure 3: It is difficult to see the individual marginal distributions. I would suggest using lines rather than bars.

We agree, and will change the marginal distributions to be shown with lines for the next revision.

Line 277: I suggest avoiding the "a(b) for c(d)" pattern and write this out in full. It is much easier to read. E.g. "Coefficients of variation for the wind packing parameters are 15% for ERA5, 42% for JRA-55... Coefficients of variation for blowing snow parameters are 13% for ERA5, 38% for JRA-55..." Or better yet, use a table.

We note that we do already use a table—these values are shown in Table 2—but we will rephrase the lines in the text for clarity as suggested: “The respective coefficients of variation for the wind packing parameters, as indicated in Table 2, are 15% for ERA5, 42% for JRA-55, and 21% for MERRA-2. The coefficients of variation for the blowing snow parameters are 13% for ERA5, 38% for JRA-55, and 19% for MERRA-2.”

Line 325: Aren't spread and uncertainty the same thing?

We acknowledge that the phrasing was ambiguous, and will clarify by replacing mentions of “spread” here with “inter-product differences”. The point being discussed here was the impact of considering uncertainties from multiple products in aggregate.

Line 411: Suggest "following" instead of "consistent with".

Agreed, we will make the change at Line 411 as suggested.