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.

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.

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.

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.

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?

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。

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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

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.