

Authors' Response to Reviews of

Enhancing sea ice knowledge through assimilation of sea ice thickness from ENVISAT and CS2SMOS

Nicholas Williams, Yiguo Wang and François Counillon
The Cryosphere,

RC: Reviewers' Comment, AR: Authors' Response, Manuscript Text

Dear Editor, Mitchell Bushuk and Anonymous Reviewers

Firstly, we would like to thank you all very much for the constructive comments and suggestions for the manuscript "Enhanced Predictability of Antarctic Sea Ice through Sea Ice Thickness Assimilation". Your insights are very useful in enhancing the quality of our work. Based on the comments and suggestions, we will revise the manuscript accordingly.

Please find our detailed point-by-point responses to the reviewers' comments in the following sections. Below, we list each comment (Reviewer Comment, **RC**) and insert our response (Authors' Response, **AR**) along with the corresponding revisions of the manuscript (inside the **black box**).

Sincerely,

Nicholas Williams
On behalf of all the authors

Anonymous Reviewer 2

RC: *This manuscript presents a clear and well-designed assessment of the impact of assimilating sea ice thickness observations on the reanalysis and seasonal prediction of Antarctic sea ice using the Norwegian Climate Prediction Model. The results demonstrate meaningful improvements in sea ice extent and thickness prediction skill, particularly in key regions such as the Ross, Weddell, and Amundsen-Bellingshausen Seas. The study is timely and the experimental framework is generally sound. A few issues and clarifications are needed before the manuscript can be considered for publication.*

AR: We would like to thank the reviewer for their thorough review, and constructive, insightful and positive comments on our manuscript.

0.1. Major points

RC: *The manuscript shows that SIT assimilation extends SIE predictability up to 12 months for October initialisation. The interpretation of this “memory effect” remains mostly statistical. Please discuss whether this memory primarily arises from thermodynamic inertia of thick ice, dynamic processes (e.g., advection and ridging), or oceanic heat storage, and whether these mechanisms differ across regions (e.g., Weddell vs Amundsen-Bellingshausen Seas).*

AR: We propose that this primarily arises from thermodynamic inertia of thick ice due to how the SIT assimilation changes the SIT and SIV in EXP-OCT, thinning the sea ice in autumn, thus leading to additional melt in spring. There are also smaller contributions from the dynamic processes due to changes in the distribution of the sea ice cover from the assimilation (Figure 7), which can affect how ice advects and ridges. We are happy to add a paragraph in the Discussions and Conclusions section which is a further interpretation of these mechanisms and across the regions.

The increased prediction skill associated with the October initialisation likely reflects the thermodynamic inertia of the thicker sea ice. Through the assimilation of SIT observations, EXP-OCT reduces SIT during autumn. Thinner ice is more susceptible to melt during the following melt season, which can influence SIC several months later and thus contributes to the apparent memory in the system. Xiu et al. (2025) also showed that realistic SIT can contribute to SIT in the Weddell Sea. In addition to this thermodynamic mechanism, changes in the spatial distribution of sea ice introduced by the assimilation may also play a role. As shown in Figure 7, the SIT assimilation modifies the regional distribution of Antarctic sea-ice thickness, whereby a better initial SIT estimate will lead to more accurately estimated advection and ridging due to a better distribution of the SIT, thereby affecting the evolution of the ice cover. These effects are most important in the Weddell Sea - where ice is thickest - and the Amundsen-Bellingshausen Seas where ice is also thick and dynamics are important.

0.2. Specific Comments

RC: *Lines 205–208: The hindcasts use 10 ensemble members selected as the first 10 members of the reanalysis ensemble, with all members assumed to be equally likely under the EnKF framework. Please justify whether this ensemble size is sufficient to robustly estimate skill metrics such as RMSE and HEE, particularly for regional and seasonal stratifications. In addition, please clarify whether the results are sensitive to the specific choice of ensemble members (e.g., first 10 vs. alternative random subsets), and whether any tests were performed to assess sampling uncertainty.*

AR: Thanks to the reviewer for their insightful comment. Another reviewer also highlighted this. In this study,

hindcasts use 10 ensemble members, selected as the first 10 ensemble members from the reanalysis. In the EnKF, all members are equally likely and identically distributed, so that it would make no difference if we pick the first 10 or randomly 10 other members. Our choice to limit our ensemble size is driven primarily by computational constraints, as the seasonal prediction experiments require a large number of integrations across multiple start dates and lead times (accounting for 1200 model years). A larger sample size would reduce sampling uncertainty. Sampling error with a random process reduces as the square root of the ensemble size. For the ensemble mean (used for validation), it is usually considered that 10 members are sufficient. The hindcast validation metrics used are computed over multiple years and from the ensemble mean, which mitigates the effect of sampling error. We also do not see large variations in the metric, suggesting that this pragmatic choice is reasonable. In the revised manuscript, we clarify and justify this choice following lines 205-208:

This choice represents a compromise between computational cost and ensemble size, as the seasonal prediction experiments require a large number of integrations from 1995-2022 across multiple initialisation dates and lead times. While a larger ensemble could further reduce sampling uncertainty, the skill metrics presented in this study are computed from the ensemble mean over multiple years and 10 ensemble members, which helps mitigate the impact of ensemble sampling variability. We also see no large discontinuity in the validation metric, suggesting that this ensemble size is sufficient to extract the main patterns in the ensemble mean. Note that while we pick the first 10 members, this choice does not matter because, with the Ensemble Kalman Filter, all members are equally likely to yield the best results.

RC: *Lines 224–225; 272–274: The ICESat, SMOS, and LEGOS datasets cover different time periods and seasons. Please clarify how temporal mismatches are handled in the evaluation and whether all metrics are computed over identical periods for fair comparison.*

AR: All metrics are computed only over periods where the data are available both spatially - masking out grid points where no data exists in one of the datasets - and temporally - e.g. if we did not have sufficient daily data points for a monthly average, then we discard the data for that month. We clarify this in the manuscript by adding to lines 224-225:

When comparing and validating against the ICESat-2, SMOS, and LEGOS datasets, we restrict the analysis to grid cells and time periods for which both the model output and the observational data are available. Monthly values are therefore computed only when a sufficient number of daily observations are available to form a representative monthly mean, ensuring that all metrics are evaluated over consistent spatiotemporal subsets of the data.

RC: *Lines 262–265: The improvements appear strongest in regions with thick ice (Weddell and Ross Seas). Please briefly discuss whether the weaker improvements in thinner-ice regions are due to observational uncertainty, model deficiencies, or SIT-SIE coupling.*

AR: We thank the reviewer for highlighting this point. The weaker improvements in thinner ice regions result primarily from observational uncertainty. In thinner-ice regions, LEGOS SIT observations have much larger uncertainty, implying that they will have nearly no impact during assimilation. To confirm that, we have analysed data assimilation metrics (degree of freedom of the signal and spread reduction factor, see Sakov et al. (2012)) that quantify the influence of each observation (not shown). Additionally, improvements are stronger in regions with thick ice because the model largely overestimated ice thicknesses in these regions (due to an underestimation of melting), and as such, the misfits between the model and observation (called

innovation) are large, which often result in larger corrections. SMOS SIT could potentially be used to improve SIT forecasts at the ice edge. However, it covers a much shorter time period (15 years less than LEGOS), which is too short to robustly demonstrate its added value for seasonal time scale. We add the following to the manuscript to clarify this:

The weaker improvements in thinner-ice regions are primarily attributed to larger observational uncertainties in the assimilated sea ice thickness (SIT) products. In contrast, thicker-ice regions exhibit stronger improvements partly because the model tends to overestimate ice thickness there, likely due to underestimated melting processes, leading to larger innovations (i.e., model and observation mismatches) and thus stronger corrections during assimilation. While alternative datasets such as SMOS SIT may offer potential improvements near the ice edge, their shorter temporal coverage limits their suitability for robust assessment at seasonal timescales.

RC: *Lines 265–269: The authors note that positive SIT biases near the ice edge may partly originate from SIE mismatches rather than intrinsic thickness errors. This is an important point. Could the authors provide a more quantitative separation of these two contributions, for instance by masking regions with consistent ice cover or by analysing thickness biases conditional on correct ice presence?*

AR: We thank the reviewer for highlighting this important point, and we agree that some of the positive SIT biases are from mismatched SIE, and that this can be clarified better in the manuscript with further analysis. The areas where both model and observations have consistent ice cover are relatively few in two of our five regions. Only thicker sea ice in the Weddell, Ross, Amundsen and Bellingshausen Seas persists through summer - and only in the Weddell Sea are there significant areas of sea ice year-round. However, we believe it could still add some value to this study. To better evaluate this issue, we add analysis in the revised manuscript analysing year-round thickness bias averages between LEGOS, SMOS, ICESat-2, EXP-OC and EXP-OCT (where we have already plotted SIT and differences in Figures 3, 6 and 7), masking regions where the comparisons do not not have consistent ice cover at this time of the year over the reanalysis period 1994-2003.

RC: *Lines 270–276: The manuscript compares EXP-OCT against SMOS and ICESat-2, which have very different sensor characteristics and thickness sensitivity ranges. Please clarify which product is considered more reliable for thin ice and thick ice, respectively, and how this affects the interpretation of RMSE differences.*

AR: SMOS is generally considered to only be reliable for the thinnest ice between 0 and 1 m, and observations are more reliable up to 0.5 m, with uncertainty increasing as thickness increases above that threshold (Kaleschke et al., 2024). ICESat-2 uses a Lidar instrument, which has higher uncertainty at lower thicknesses with uncertainty decreasing as thickness increases, but there is further uncertainty due to the poorly constrained snow depths on Antarctic sea ice. ICESat-2 appears more reliable than LEGOS for thick ice, but the time coverage is too short to robustly demonstrate added value for seasonal predictions. We will add the following clarification in the paper:

SMOS is generally considered reliable primarily for thin sea ice (0–1 m), with the highest accuracy for ice thicknesses up to approximately 0.5 m; uncertainty increases progressively beyond this range (Kaleschke et al., 2024). In contrast, ICESat-2 employs a lidar instrument, which exhibits greater uncertainty at lower ice thicknesses, with accuracy improving as thickness increases. However, its estimates are additionally affected by uncertainties in snow depth, which remain poorly constrained over Antarctic sea ice. While ICESat-2 appears to outperform LEGOS for thicker ice conditions, its relatively short temporal coverage limits the ability to robustly assess its added value for seasonal

predictions.

RC: *Lines 406–409: The “end-of-winter barrier” in SIT predictability is hypothesised to be partly due to limitations of the LEGOS SIT dataset. It would be useful to test whether this barrier is also evident when verifying against independent SIT datasets (e.g., ICESat-2) or whether it is specific to LEGOS-assimilated runs.*

AR: We hypothesise that the barrier in predictability may exist due to the high uncertainties at the ice edge in the LEGOS product because of uncertainties in the SIT product and differences between the available Antarctic SIT products. Due to the limited time period of ICESat-2 observations (2018-2022), we would not be able to robustly statistically test our predictions, as we would only have 5 samples for a specific month. We agree that this statement is speculative and we can remove it if the reviewer thinks it is preferable.

RC: *Lines 416–417: The study is framed as the “first long-term assessment of Antarctic SIT assimilation”. To substantiate this claim, the authors should explicitly contrast their experimental design, length, and findings with those of previous efforts and clarify what novel insight this work uniquely provides.*

AR: We expand in the conclusion to specifically state the novelty of our study:

This work represents the first long-term assessment of Antarctic SIT assimilation over a multi-decadal hindcast period, establishing SIT as a key mechanism for improving Antarctic sea ice prediction. Previous studies of Antarctic sea ice prediction (Guemas et al., 2016; Morioka et al., 2021; Bushuk et al., 2021; Xiu et al., 2025) have not been able to incorporate satellite SIT observations over sufficiently long periods to assess their impact on seasonal prediction skill, while studies that do assimilate satellite SIT have focused on reanalysis rather than prediction (Luo et al., 2021, 2023; Chenal et al., 2024). In contrast, this study assimilates a 30-year, year-round observational SIT record to produce a consistent reanalysis and initialise seasonal hindcasts, enabling a statistically robust assessment of how Antarctic SIT influences prediction skill across regions and lead times. The results in this study demonstrate that Antarctic SIT acts as a long-lived source of predictability for austral spring and summer initialisation, extending SIE and SIT prediction skill to longer lead times and across the Weddell, Ross, and Amundsen–Bellingshausen Seas, effects which were not demonstrated previously.

References

- Bushuk, M., Winton, M., Haumann, F. A., Delworth, T., Lu, F., Zhang, Y., Jia, L., Zhang, L., Cooke, W., Harrison, M., et al. (2021). Seasonal prediction and predictability of regional antarctic sea ice. *Journal of Climate*, 34(15):6207–6233.
- Chenal, A., Garric, G., Testut, C.-E., Hamon, M., Ruggiero, G., Garnier, F., and Le Traon, P.-Y. (2024). Assimilation of radar freeboard and snow altimetry observations in the arctic and antarctic with a coupled ocean/sea ice modelling system. *EGU sphere*, 2024:1–39.
- Guemas, V., Chevallier, M., Déqué, M., Bellprat, O., and Doblas-Reyes, F. (2016). Impact of sea ice initialization on sea ice and atmosphere prediction skill on seasonal timescales. *Geophysical Research Letters*, 43(8):3889–3896.
- Kaleschke, L., Tian-Kunze, X., Hendricks, S., and Ricker, R. (2024). Smos-derived antarctic thin sea ice thickness: data description and validation in the weddell sea. *Earth System Science Data*, 16(7):3149–3170.

- Luo, H., Yang, Q., Mazloff, M., Nerger, L., and Chen, D. (2023). The impacts of optimizing model-dependent parameters on the antarctic sea ice data assimilation. *Geophysical Research Letters*, 50(22):e2023GL105690.
- Luo, H., Yang, Q., Mu, L., Tian-Kunze, X., Nerger, L., Mazloff, M., Kaleschke, L., and Chen, D. (2021). Dasso: a data assimilation system for the southern ocean that utilizes both sea-ice concentration and thickness observations. *Journal of Glaciology*, 67(266):1235–1240.
- Morioka, Y., Iovino, D., Cipollone, A., Masina, S., and Behera, S. K. (2021). Summertime sea-ice prediction in the weddell sea improved by sea-ice thickness initialization. *Scientific reports*, 11(1):11475.
- Sakov, P., Counillon, F., Bertino, L., Lisæter, K., Oke, P., and Korablev, A. (2012). Topaz4: an ocean-sea ice data assimilation system for the north atlantic and arctic. *Ocean Science*, 8(4):633–656.
- Xiu, Y., Wang, Y., Luo, H., Garcia-Oliva, L., and Yang, Q. (2025). Impact of ocean, sea ice or atmosphere initialization on seasonal prediction of regional antarctic sea ice. *Journal of Advances in Modeling Earth Systems*, 17(2):e2024MS004382.