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 1

RC: ***Overall, this manuscript makes a timely and valuable contribution to seasonal-to-interannual sea ice prediction and will be of interest to the cryosphere communities. While the study presents promising results, several aspects require further clarification and improvement. I therefore recommend minor revisions and believe that addressing these points would enhance its suitability for publication.***

AR: We thank the reviewer for their constructive comments, insight and positive feedback for the manuscript.

RC: ***The manuscript frequently uses the term "enhanced predictability", while the analyses mainly demonstrate improved prediction skill resulting from better initialization through SIT assimilation. Since predictability refers to the intrinsic limits of the climate system, whereas prediction skill reflects model performance, the authors should clarify which aspect is improved. If the results primarily indicate enhanced prediction skill in NorCPM, the terminology should be revised accordingly throughout the manuscript, including the title.***

AR: We thank the reviewer for this important clarification. As suggested, we revise accordingly throughout the manuscript. For example, we change the title to "Enhanced Prediction Skill of Antarctic Sea Ice through Sea Ice Thickness Assimilation".

RC: ***Line 111–112: NorESM is run with CMIP5 historical forcings and RCP8.5 beyond 2005. Given the citation of Xiu et al. (2025), which shows that improved atmospheric state representation can enhance sea ice prediction skill, the authors are encouraged to discuss whether the use of RCP8.5 forcing may introduce systematic atmospheric biases, and whether improvements in atmospheric forcing (rather than nudging) could lead to more realistic atmospheric states and improved sea ice prediction skill.***

AR: Please note that Xiu et al. (2025) did not address the issue of external forcings, but year-to-year variability. Indeed, improvements in atmospheric states are a source of skill for springtime sea ice predictions (Xiu et al., 2025). The motivation to stick to CMIP5 forcing was primarily due to a bug in our external forcing implementation with CMIP6 (see our response to another reviewer who commented on that). We add the following statement to the revised manuscript:

> It is worth noting that using a more realistic external forcing would improve the model representation of the atmospheric trends and improve sea ice prediction skill of NorCPM.

RC: ***The assimilation implementation requires clarification. In Lines 128–131, it is unclear whether SIC is updated twice within the same assimilation cycle, and if so, how these updates are implemented and treated differently. In addition, both the Rfactor and K-factor are used to inflate the observation error; please clarify their respective roles and, if possible, quantify their effects, for example by indicating how much the ensemble spread or observation error is inflated.***

AR: SIC is updated twice within the same assimilation cycle. The observations of SIC and SIT are independent, therefore there is no overfitting issue that could occur within the system. In data assimilation theory, assimilating sequentially two independent observation data sets or together is equivalent for a linear system (it differs slightly for non-linear cases, but with no clear conclusions). Some data assimilation schemes (such as the EAKF used for sea ice prediction at GFDL) assimilate all observations sequentially. The reason why we decided to assimilate separately these two data set in our system is that for SIC and ocean observations, we see that it can influence all ocean and sea ice variables, while for SIT, we do not expect a strong relationship with ocean state variables in the location where LEGOS data is available (in the pack ice).

The R-factor belongs to the category of ensemble inflation methods (multiplicative inflation, additive inflation, relaxation to the prior, EnKF-N) that counteract the spurious deflation of ensemble spread caused by the finite ensemble size and model error (Raanes et al., 2019). The specificity of the R-Factor is that it uses the standard observation error to update the ensemble mean, but an inflated observation error (e.g., by a factor of 4) to update the ensemble anomalies. This ensures a good accuracy of the ensemble mean but limits the spurious excessive deflation. It also acts only in locations with observations (unlike multiplicative inflation, which can cause issues with a varying observation network, as in ice-covered regions).

The K-factor is used to ensure a sound dynamical consistency of the analysis. It artificially inflates the observational uncertainty to keep the analysis within 2 times the ensemble standard deviation (with K-factor=2). The reason is that the EnKF performs a linear analysis update, and if the analysis exceeds 2 times the standard deviation, it produces a very unlikely state (more than 95%). An unrealistic state can cause a dynamical imbalance and a degradation afterwards. The rationale of the K-factor is thus to push more softly the model towards the data (if it is not just an outlier) so that the model and data get synchronised after several assimilations.

Please refer to (Sakov et al., 2012) for further information on those implementations. Please note that these settings are commonly used in all NorCPM simulations and have been robustly tested. We clarify this in the manuscript in lines 147-153 as follows:

> We use the R-factor inflation method, where the observation error is inflated by a factor of 4 (i.e., R-factor=4) for the update of the ensemble anomaly and the k-factor formulation, in which observational error is artificially inflated if the assimilation pushes the update beyond two times the ensemble spread (K-factor=2) (Sakov et al., 2012).

RC: ***Figure 4: Compared with SMOS and ICESat-2, the assimilated LEGOS SIT appears to have a larger RMSE than EXP-OC. If so, this may indicate better agreement of EXP-OC with these independent datasets. The authors are encouraged to clarify this by directly comparing both EXP-OC and LEGOS SIT against SMOS and ICESat-2, and to explain why further assimilating LEGOS SIT nonetheless leads to improved sea ice estimates and prediction skill.***

AR: We thank the reviewer for highlighting this point and agree that it requires a clearer explanation in the manuscript. Figure 4 shows that the standalone LEGOS dataset has larger RMSE than both EXP-OC and EXP-OCT when evaluated against SMOS and ICESat-2. However, when LEGOS SIT is assimilated (EXP-OCT), the resulting reanalysis achieves the lowest RMSE overall, particularly against ICESat-2, where improvements are consistent across all months. Improvements relative to SMOS are more mixed, with clear reductions in RMSE in austral summer, with no differences at the end of austral winter.

This apparent discrepancy arises because data assimilation does not aim to reproduce the assimilated observations, but instead estimates the most likely sea ice state by optimally combining model states with multiple observational constraints, each weighted by its respective error statistics. Although the LEGOS product has larger errors relative to independent datasets, it can still provide complementary spatial and temporal information that helps correct model biases when appropriately weighted in the assimilation system. As a result, even a noisier dataset can improve the posterior estimate if it contributes independent information that is not fully captured by the other observations.

To clarify this point, we revise the text at lines 276 as follows:

> Although the standalone LEGOS product can exhibit larger errors than EXP-OC when evaluated against independent datasets, its assimilation in EXP-OCT nevertheless leads to a reduction in RMSE overall.

> This reflects that the data assimilation framework estimates the most likely sea ice state by optimally combining model states with multiple observational constraints, rather than directly reproducing any individual dataset. Differences between LEGOS, ICESat-2, and SMOS are also expected due to their distinct retrieval approaches and differing sensitivities to sea ice and snow properties. EXP-OCT achieves the lowest RMSEs in austral spring and austral autumn, while errors are higher in February, at the end of the melt season, and during winter, when sea ice is thickest and most spatially variable. The largest RMSE reductions relative to ICESat-2 occur outside the winter period. This seasonal dependence likely reflects variations in both model and observational uncertainty. During transition seasons, larger model biases in sea ice thickness provide greater scope for improvement through assimilation. In contrast, during winter, increased variability in snow depth and density introduces larger uncertainties in the conversion from freeboard to thickness, reducing the effectiveness of assimilation increments due to higher observational noise.

0.1. Minor Comments

RC: ***Line 47: Please clarify what "OISST" refers to at first mention?***

AR: OISST refers to Optimal Interpolation Sea Surface Temperature, a NOAA SST Dataset. We clarify this in the revised manuscript.

RC: ***Line 80: "ICESAT-2" should be corrected to "ICESat-2."***

AR: Thank you, we correct this in the revised manuscript.

RC: ***Only one-third of the available ensemble members is used for prediction. Please clarify the rationale for this choice and discuss whether the reduced ensemble size could affect prediction skill.***

AR: Thanks to the reviewer for noticing this, this has also been highlighted by another reviewer. In this study, hindcasts use 10 ensemble members, selected as the first 10 ensemble members from the reanalysis. This choice 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, it is usually considered that 10 members are sufficient. However, because the metrics used are computed over multiple years and the validation primarily assesses the ensemble mean, this already mitigates the effect of the sampling uncertainty sufficiently. We can notice that as well from the fact that the metrics (correlation and RMSE) are relatively continuous. In the revised manuscript, we clarify and justify this choice following lines 205-208. As the skill metrics are computed over a long period and 10 ensemble members, while we have not systematically tested alternative ensemble subsets, the influence of the sampling variability is reduced. We will add additional sentences clarifying this in the experiment setup section:

> 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: ***Section 3.2: This section requires clarification. bfRMSE is mentioned but not used elsewhere; please clarify or remove it. In addition, define MSE and explain how MSEforecast and MSEreference are computed. In Equation (4), the variables appear to represent anomalies rather than the raw model and observation values described in Line 241.***

AR: Thank you to the reviewer for pointing out these issues. bfRMSE is meant to refer to RMSE in this equation and at one place in the text we use bfRMSE when we mean RMSE, we have now corrected this. Additionally we also add the definition for the MSE and clarify that the "reference" refers to the EXP-OC experiment, while the forecast refers to the EXP-OCT experiment. For equation (4), these do indeed represent anomalies rather than raw values which was not clearly stated in the original text. In the revised manuscript, we clarify that these are anomalies relative to the climatology in line 241 to ensure the formula and descriptive text are consistent. The text is revised as follows:

> We also use the mean squared skill score (MSSS), which quantifies the reduction in error relative to the free run. MSSS is computed using the mean squared error (MSE), and defined as
>
> $$\text{MSSS} = 1 - \frac{MSE_{forecast}}{MSE_{reference}}, \tag{1}$$
>
> where MSE is defined as
>
> $$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x-y)^2, \tag{2}$$
>
> where x and y are the model and observation values respectively. One benefit of MSSS is that it normalises errors by their pointwise variability and, as such, can help visualise improvements that would otherwise be masked when the amplitude of the error is smaller than in other regions.
>
> The anomaly correlation coefficient (ACC) is also used to test the skill in reproducing the observed variability in the reanalyses and hindcasts. The ACC is defined as
>
> $$\text{ACC} = \frac{\sum_{i=1}^{N} x_i' y_i'}{\sqrt{\sum_{i=1}^{N} x_i'^2}\sqrt{\sum_{i=1}^{N} y_i'^2}}, \tag{3}$$
>
> where x_i' and y_i' are model and observation values. In this study, we always analyse detrended ACC values, i.e., detrending the time series of the experiments and the observations before calculating the ACC. The trend is usually removed because it is considered to be easy to predict (Bushuk et al., 2017). We test the statistical significance of the ACC using a Student's t-test at a 5% significance level, with degrees of freedom calculated as in Von Storch and Zwiers (2002).

RC: ***Line 275 and Figure 4: EXP-OCT does not consistently exhibit lower RMSE than EXP-OC in July–September; please revise the corresponding text. In addition, the comparison and related discussion should refer to ICESat-2 rather than ICESat.***

AR: We have revised the manuscript accordingly by removing this statement. Thank you.

RC: ***Figure 3: Compared with LEGOS SIT, EXP-OC overestimates sea ice prior to SIT assimilation, while EXP-OCT shows an underestimation afterward, particularly in the Weddell Sea. Please discuss the potential mechanisms underlying this shift.***

AR: We thank the reviewer for highlighting this point. This point was also mentioned by another reviewer. In the revised manuscript, we expand the discussion and conclusions section to clarify the main mechanisms.

The reduction in SIE in EXP-OCT, primarily in areas of thicker ice, is associated with a reduction in SIT following SIT assimilation. In NorESM, the model tends to produce excessively thick ice in this region, which contributes to an overestimation of sea ice due to the intrinsic memory of SIT. When SIT observations are assimilated, the SIT is reduced toward observed values, which weakens this thermodynamic memory and leads to a corresponding reduction in SIE. In addition, the assimilation improves the spatial distribution of sea-ice thickness. Additionally, as shown in Figure 7, NorESM exhibits biases in the regional distribution of Antarctic SIT, while the assimilation helps redistribute ice thickness more realistically. These changes in the thickness distribution also contribute to the improved representation of SIE. In other words, while it is unexpected that assimilation changes the sign of the bias, the analysis falls within the uncertainty range, complying with the model's dynamical consistency.

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 the thickness of the sea ice cover 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. In addition to this thermodynamic mechanism, changes in the spatial distribution of sea ice introduced by the assimilation may also play a role.

References

Bushuk, M., Msadek, R., Winton, M., Vecchi, G. A., Gudgel, R., Rosati, A., and Yang, X. (2017). Skillful regional prediction of arctic sea ice on seasonal timescales. *Geophysical Research Letters*, 44(10):4953–4964.

Raanes, P. N., Bocquet, M., and Carrassi, A. (2019). Adaptive covariance inflation in the ensemble kalman filter by gaussian scale mixtures. *Quarterly Journal of the Royal Meteorological Society*, 145(718):53–75.

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.

Von Storch, H. and Zwiers, F. W. (2002). *Statistical analysis in climate research*. Cambridge university press.

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.