Review of "*Enhancing sea ice knowledge through assimilation of sea ice thickness from ENVISAT and CS2SMOS »* by Nicholas Williams, Yiguo Wang and François Counillon.

This review is co-signed by François Massonnet and Alison Delhasse (UCLouvain)

This manuscript focuses on the role that sea ice thickness (SIT) knowledge plays on the estimation of the Arctic sea ice state as well as on the skill of Arctic sea ice predictions. The authors use the Norwegian Climate Prediction Model (NorCPM) to conduct a series of sensitivity experiments : a FREE ensemble run where no data is assimilated at all, a CTRL run where sea ice concentration (SIC), sea surface temperature (SST) and hydrographic profiles data are assimilated, and finally a +SIT run where SIT data (the CS2SMOS product and the ENVISAT product) is assimilated in addition to the data considered in CTRL. The authors use several standard metrics to first evaluate the impact of SIT assimilation on the mean state, which is found to be improved against independent observational datasets. Then, the authors look at the skill of seasonal predictions that are initialized from CTRL and SIT and show clear improvement for the prediction of the sea ice edge (the improvement for sea ice extent, SIE, is a bit less clear).

General comments

Novelty. The study clearly adds to the body of literature, as it is the first time that ENVISAT data are assimilated in a coupled model, to our knowledge. This is also an insightful study to better understand the role that sea ice thickness plays on the skill of seasonal predictions, even though there is already some evidence from previous studies that this is the case. The study is also insightful from methodological aspects, since doing strongly coupled assimilation with a coupled model is a real challenge. There are interesting aspects of full-field *vs.* anomaly initialization that can be useful for practitioners.

Positioning. with respect to previous works. The authors seem to cite all relevant literature for this work and the work is well positioned with respect to the existing body of knowledge.

Methodology. We have several comments / pieces of advice for improvement.
1. Sometimes, the interpretation of statistics is not clear to readers who are not in the data assimilation community. A few examples:
    o Degrees of Freedom for Signal (DFS) : We have the definition at line 227 but we would like to have more guidance about the interpretation of this statistic. What are the units? Is DFS supposed to be large, low? What is the desired look of maps like in Fig. 1? Somehow, we understand that the maps should show some form of complementarity (i.e., DFS associated with one variable should not be redundant with the DFS associated with another)? Should the maps be orthogonal to each other?

- The Integrated Ice Edge Error (IIEE) can be decomposed as in Eq. 5 as a contribution from overestimation and underestimation, but also as the sum of a mean absolute error and a displacement error. Have the authors tried to produce the timeseries following the latter decomposition? This would help understand what type of error is driving the IIEE.

2. We are wondering why the authors focus all the analyses on the ensemble mean and never display the spread/variability of the ensemble. For example, in Fig. 5, if we had access to the range (min-max) of the FREE, CTRL and +SIT ensembles, we could assess whether the change in the mean is significantly larger than the intra-member spread. To us, it is an important information because if the changes brought by the data assimilation are lying within a lot of background noise, then the information is not the same as if the changes clearly emerging from the background noise.

3. It would be good to know if the filter works as expected, by producing sufficiently dispersed forecasts before the analysis time steps. Could the authors plot the time series of the ensemble standard deviation of some variable (e.g. SIE) to see if this spread is commensurate with the bias that is expected to be fixed by the filter? Ideally, the same analysis for other variables (volume, spatial sea ice thickness) would be welcome. Maybe the authors have already done this and have a figure that they can pull up for this review. It does not need to be an extra figure in the manuscript, but that point can be mentioned in the manuscript.

4. The model assimilates SIC, but also SST from the OISSTV2 dataset, which is derived (line 167) using a quadratic function of SIC itself. Isn't there some unnecessary redundancy here? How does the filter behave when it assimilates both x and x^2? The filter works on linear assumptions for the covariances between variables, so our initial guess would have been that it is better to let the SST adjust to the SIC when the latter is assimilated, but it would be interesting to know more if the authors have some thoughts here.

5. Can the authors justify the use of $N = 30$ members ? This choice is certainly guided by a tradeoff between practical experience, computational limitations, and statistical power, but having a sense of the history behind that choice can be useful for other teams.

6. In the same vein, why do the authors do monthly assimilation (why this frequency)?

7. Line 137: It appeared strange to us to have SIC assimilated in anomaly while SIT is assimilated in full field. Now, we understand that this is the result of a few attempts, and this is also where the paper has a real value. Nevertheless, we can imagine a few cases where this methodology makes things complicated. Suppose a grid cell where, at one assimilation step, there is much more ice in satellite observations for that year than in the satellite climatology. The SIC anomaly initialization will drive a strong increase in SIC. But suppose that, at the same grid cell, the model as a thick bias, so that the (full field) assimilation of SIT drives a negative update of SIT. The filter will thus produce very spread and thin ice in that grid cell, which results from two different causes: one is a climatic reason and one is a model issue. Isn't the filter producing strange-looking updates in some cases if SIC and SIT are assimilated in anomaly and full-field mode, respectively? The authors say that they tried SIT anomaly initialization too,

but that this was inconclusive. Did they try SIC full-field together with SIT full-field?

8. After an analysis by the filter, is the model immediately restartable? Or do the authors apply some post-processing / regularization / sanity check to avoid unphysical initial states? There may be cases where, for example, sea ice volume is not zero but sea ice concentration is zero, which can cause a model crash.

9. It would be helpful to add a map of the regions used for the analyses (i.e. the Bushuk et al's regions).

10. Line 237: the treatment of ensemble information when using the IIEE is actually quite subtle, as one of the reviewers learned when interacting with Helge Goessling in another study (Massonnet et al. 2023, https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2023.1148899/full). The spirit of the IIEE is not to be applied to all members individually, to be then averaged across members. Instead (quoting that study):

> "*For one-member forecasts, the calculation of the IIEE is straightforward: the spatial fields of SIC are converted to 1 or 0 based on the 15% SIC threshold and the resulting binary field is compared to the observed binary field of ice presence. The areas of grid cells where sea ice is present in observations but absent in the forecasts, or absent in observations but present in the forecasts are then summed. For multi-member forecasts, the calculation is slightly different: binary fields of sea ice presence are defined for each member individually [based on the 15% SIC threshold], and a probability of sea ice presence is calculated by averaging the binary fields across the ensemble. The areas of grid cells where sea ice present in observations but present with < 50% probability in the ensemble, or absent in observations but present with > 50% in the ensemble, are then summed.* »

Presentation.

1. Fig. 7 and similar: please explain the meaning of dots / crosses in the caption.

Minor comments

- Line 1: the decline extends to at least four decades, so better write "four" here.
- Line 5: Specify the spatial resolution of NorCPM in your abstract.
- Line 7: Specify which two decades.
- Line 23: Seasonal predictions do not really aim to predict the response of the sea ice to climate change (climate projections do). Seasonal predictions are more an initial-value problem, and the changes in forcings play little role at these time scales.
- Line 24: The Arctic basin itself has not warmed, it is the atmosphere or the oceans that have warmed.
- Line 26: remove "ice"

- Line 26: is it the melting season that has lengthened, or the open-water season? We think the latter.
- Line 26-27: It is rather unclear what "such a radical change influences internal variability". The internal variability is the manifestation of internal processes within the climate system, how do sea ice changes modify this internal variability?
- Line 39: We would also cite: Koenigk, T., König Beatty, C., Caian, M., Döscher, R., & Wyser, K. (2012). Potential decadal predictability and its sensitivity to sea ice albedo parameterization in a global coupled model. *Climate Dynamics*, *38*(11–12), 2389–2408. https://doi.org/10.1007/s00382-011-1132-z
- Line 43: "comprise of melting" is a strange sentence.
- Line 56: "the atmosphere, particularly wind" is not really meaningful. Wind is not a part of the atmosphere; it is a process that occurs within the atmosphere.
- Line 62: To our knowledge CS2 has not been defined yet.
- Line 106: you may also want to cite Massonnet et al. 2019 where we also recommended five categories for climate studies (https://doi.org/10.5194/gmd-12-3745-2019)
- Line 127: What do you mean by strong and weak data assimilation method? Please clarify.
- Line 129-130: "so that the individual thickness category remains identical". This is strange, the category (limits) are fixed, aren't they? Do the authors mean that the thickness in each category is constrained to not change? Maybe a schematic would help here to explain to the reader how the redistribution of volume works.
- Line 140: a large fraction of what?
- Line 144-145: "we update the individual multi-category sea ice fraction" → but not volume then? Maybe a schematic would help.
- Line 158: error → errors
- Line 183: "The observation error is provided by the datasets". We assume you mean that the error *statistics* is provided (otherwise the true state would be known ☺)
- L188: Could you be a bit more specific about your random states: are they picked up between 1850-2005? Which months exactly? Perhaps a list of these random state years/months would be appreciated in the supplement.
- L190: Could you precise how your ensemble is generated?
- Line 195: disentangle from what?
- Line 205: the area *of grid cells*
- Figure 1:
  - Is OCN DFS limited to 6 or could be higher value? In this last case, could you adapt your color scale for subplot 1a)?
  - Can you refer to a) to e) in your caption instead of/or added to your location indications (top right, left,…)?
- Figure 2:
  - we would propose that there is an extra row showing the mean SIT just to have a sense of the spatial distribution of that variable.
  - change the color of the localization of your observation points as there are indistinguishable from islands for instance.

- How do explain the negative bias of +SIT compared to observation for SIT (Figure 3 & Table 1)?
- Figure 4b: Do you have any idea to explain this minimum in bfRMSE in June for FREE experiment?
- Line 339: "we remove the trend": is it the linear trend?
- Line350: (Figure 7) → (Figure 7 &8)?
- Line 360: "but it is only significant in November-December, for hindcast initialized in January and March." If significance is marked without a point, I missed the significance for prediction of nov-dec initialized in March.
- L447: in the Beaufort *Sea*