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, Alison Delhasse, and François Massonnet

Firstly, we would like to thank you all very much for the constructive comments and suggestions for the manuscript "Enhancing sea ice knowledge through assimilation of sea ice thickness from ENVISAT and CS2SMOS". Your insights are very useful in enhancing the quality of our work. Based on the comments and suggestions, we have revised the manuscript.

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

1. Reviewers: Alison Delhasse and François Massonnet

RC: ***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).***

AR: The overview of the manuscript provided by the reviewers is precise. We thank the reviewers for their time and effort in carefully reviewing the manuscript.

1.1. General comments

RC: ***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.***

AR: We appreciate that the reviewers found this study insightful, interesting, and novel.

RC: ***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.***

AR: Thanks.

RC: ***Methodology. We have several comments/pieces of advice for improvement.***

AR: We appreciate the reviewers' thoughtful and insightful feedback and suggestions, which helped us to improve the manuscript. Please find our detailed point-by-point responses to the reviewers' comments in the following text.

RC: ***Sometimes, the interpretation of statistics is not clear to readers who are not in the data assimilation community. A few examples: 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?***

AR: We agree with the reviewer that the degree of freedom for signal (DFS) is a challenging metric and that our explanations were too succinct. We have added further explanation about DFS in the section "Validation Metrics". We believe that with the current modifications, readers who are not experts in data assimilation will be able to understand the purpose of the metric better.

> The DFS metric quantifies how the assimilation of observations has reduced the dimension or rank of the ensemble (Sakov et al., 2012). A larger DFS value implies that the assimilation has more change into the system, i.e., reducing the number of degrees of freedom (the unit of the metric). The DFS can be between 0 (no impact) and the total number of degrees of freedom minus one (where all members collapse to a single member)[a]. A well-balanced data assimilation system aims to make minimal changes necessary to comply with observations, and as such, one should reach neither the lower nor the upper DFS value. As a consequence, DFS is often used to calibrate the strength of the data assimilation system when observation error is poorly known (Sakov et al., 2012) -prevent too strong or too weak assimilation. DFS can also be used to isolate the relative influence of each observation on the total impact of the assimilation, which is our aim in this study. More specifically, DFS is used here to diagnose the relative influence of the SIT assimilation compared to other datasets, for instance, where it is most beneficial, as well as to quantify the impact of ENVISAT SIT versus C2SMOS.
>
> [a]Note that the total number of degrees of freedom is the minimum between the ensemble size and the number of observations used in the local analysis.

RC: ***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***

AR: We would like to thank the reviewer for this nice suggestion that aligned well with our previous analysis for SIT and SIC – decomposing the error as mean absolute error and displacement. We have now included the decomposition (L378) and Figure R1 in the revised manuscript.

> The IIEE can also be decomposed in a different way, using a mean absolute error (MAE) and a displacement (DISP). This is formulated in Goessling et al. (2016) as
>
> $$\mathrm{IIEE} = \mathrm{MAE} + \mathrm{DISP} \quad (1)$$
> $$\mathrm{MAE} = |\mathrm{O} - \mathrm{U}| \quad (2)$$
> $$\mathrm{DISP} = 2 \cdot min(\mathrm{O}, \mathrm{U}), \quad (3)$$
>
> where O is the area where SIE is overestimated, and U is the area where SIE is underestimated.

It is in quite good agreement with the rest of our analysis for SIC and SIT. Until summer, the MAE is well constrained, which yields an improvement over both CTRL and FREE. The displacement IIEE error is also reduced. During the summer, the MAE is still reduced, but the displacement of IIEE in +SIT has increased compared to CTRL between June and October.

RC: ***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.***

AR: We thank the authors for this insightful comment and agree that showing the spread of the ensemble would be important. We have adjusted Figure 6 to additionally show the ensemble spread using shaded colours, modified the caption as appropriate, and added some discussion of this in the text. We also have an analysis

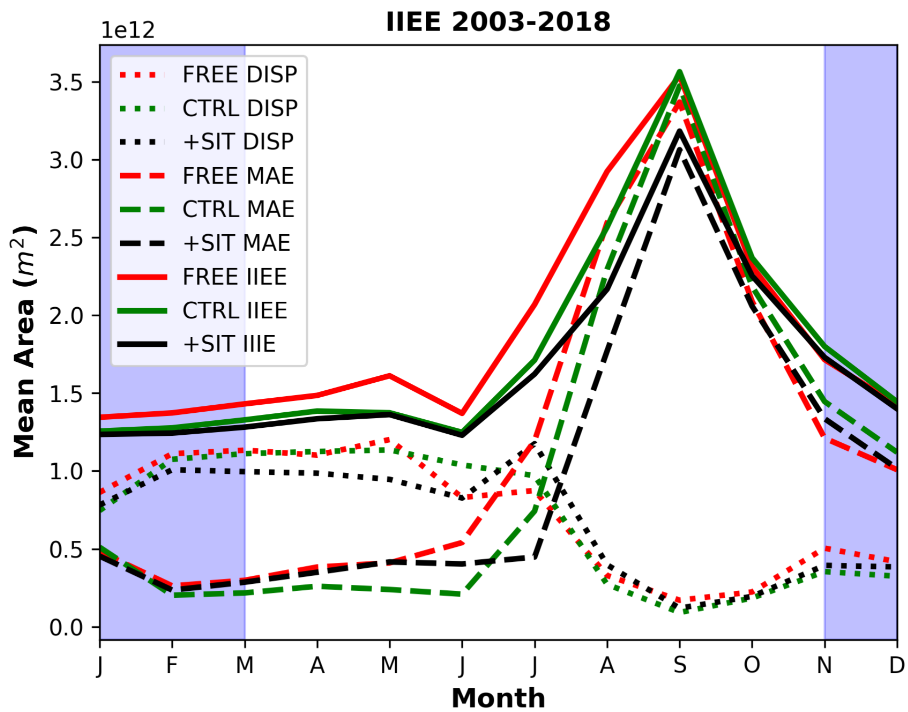

Figure R1: Climatology of the displacement (DISP), mean absolute error (MAE) and total IIEE of the FREE, CTRL and +SIT reanalyses.

on the reliability of the system (cf our answer to the next comment)

> The ensemble spread in extent and volume is shown in all experiments (Figure **??**). The spread in FREE is the largest of the three experiments by far, particularly for the SIE. This is expected as FREE does not assimilate any observations, so the spread is less constrained. The ensemble spreads in CTRL and +SIT are comparable for the SIE in March and September, whereas +SIT has a lower spread for the SIV, particularly in March. Again, this is no surprise, because SIT observations are not assimilated in summer, so we expect +SIT to have a larger spread at the end of summer in September.

RC: ***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.***

AR: We indeed investigated if the filter was working as expected for the SIT assimilation before we began the work that appears in this study. We include in the manuscript the plot shown below that shows the changes in ensemble spread, bias and RMSE of the sea ice thickness over time.

We also added the following to the manuscript (L281)

> In Figure 1, we present the time evolution of the assimilation diagnostic. We can first notice that the bias has a seasonal signal that relates to the lack of observation during summer, when the bias increases. For ENVISAT, the system is too thick at the start but gets too thin at the end of the season, while with C2S the too thick bias remains positive until the end of the seasonal observation period. In an ensemble data assimilation system, one can use the ensemble spread as a measure of the system's error. A first check to assess the reliability of the system is to ensure that the quadratic sum of the ensemble standard deviation and observation error (here denoted as total error) matches the bias-free error of the ensemble mean (RMSE, Rodwell et al., 2016). We can notice that our system exhibits too high dispersion during the ENVISAT period, but that the reliability is very good in the C2S period. The overdispersion during the ENVISAT may relate to the observation error, which is very high. We can also notice that the ensemble spread and RMSE covary in time very well (both seasonally and interannually). There is a discrepancy at the start of the assimilation season that relates to the bias being large (not to be accounted for in the reliability budget analysis).

RC: ***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.***

AR: The reviewers are correct that the SST and SIC from the OISSTV2 dataset are highly correlated due to the derivation of SST from SIC. In our studies, SST under sea ice is not assimilated. Note also that we update SST when assimilating SIC (strongly coupled data assimilation), an approach that performs better than letting the SST adjust (Lisæter et al., 2007; Sakov et al., 2012; Kimmritz et al., 2018). For clarity, we have modified the text (L167-168) as follows:

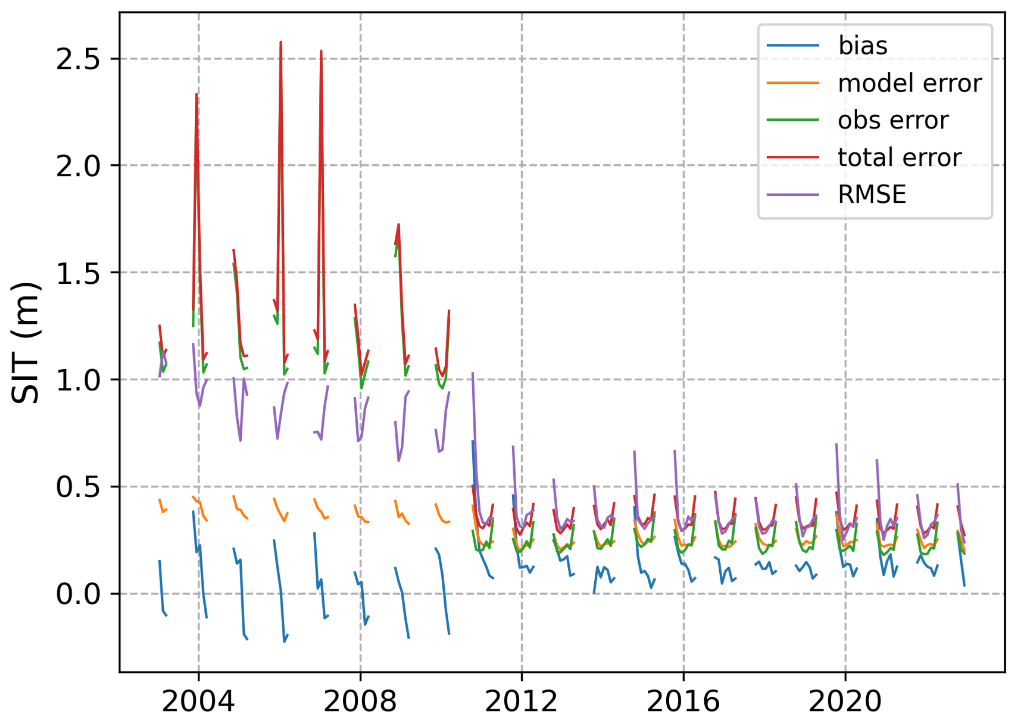

Figure R2: The time evolution of five key metrics of the EnKF for SIT assimilation (computed on the innovation vector – the difference between the observation and the model at the observation location): which is the bias, model error (ensemble standard deviation), observation error (standard deviation), total error (quadratic sum of model error and observation error) and RMSE. Note that there are gaps in the figure due to the lack of SIT observations outside winter.

> The SST data are produced by combining both in-situ and satellite observations and SST's simulated by sea ice cover (Reynolds et al., 2007). Note that SST data in the regions covered by sea ice are not assimilated in this study (Kimmritz et al., 2019; Wang et al., 2019). However, the ocean underneath the sea ice is updated based on the ensemble covariance.

RC: ***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.***

AR: As the reviewers state, it is a tradeoff in these qualities. A sample size of 30 is typically the size needed to have a reasonable estimation of the ensemble variance of a random variable. An ensemble of 30 members is sufficiently large to represent the subspace of a water column in the local analysis and has been found large enough to produce robust results for ocean and sea ice update Counillon et al. (2014); Kimmritz et al. (2018). Many parameters of the assimilation system have been tuned to work with the ensemble size of 30 (e.g., localisation and inflation). For clarity, we have added some statements to justify the use of 30 ensemble members as follows:

> FREE allows us to estimate the skill related to external forcings (Kimmritz et al., 2019). We use CTRL to compare the model with and without the assimilation of SIT data, while still assimilating the ocean and SIC observations. We use an ensemble size of 30 members primarily due to limited computational resources. In addition, many of the parameters in NorCPM (localisation, inflation) have been tuned to work ideally with an ensemble size of 30 which we found large enough to provide robust results for ocean and sea ice update with NorCPM (Counillon et al., 2014; Kimmritz et al., 2018).

RC: ***In the same vein, why do the authors do monthly assimilation (why this frequency)?***

AR: This is a choice driven primarily by computational limit - the assimilation step in NorCPM requires stopping and restarting the model. At start, NorESM reinitialises all the components of the fully coupled Earth system model NorESM, which takes a substantial amount of time. For this reason, using a frequency of assimilation higher than monthly (e.g., daily assimilation) was making our system too slow. In the latest version of NorCPM, we have implemented the capability to pause and resume the model allowing us to perform the assimilation online. This solves the issue and we can now run the system with a daily assimilation frequency. We are currently testing the added value of a higher frequency in preparation for the CMIP7 Decadal Climate Prediction Project.

RC: ***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 fullfield?***

AR: We have not tried the full-field assimilation for both SIC and SIT. The anomaly-field assimilation (updating only the model variability) has been a standard setting of NorCPM for ocean observations, as the Earth system models often have large model biases. Correction for model bias in the ocean has a detrimental impact on the ocean when there is limited observation (transfer of bias in the ocean interior, instability that takes several decades to dissipate; see discussion in Counillon et al. (2016) and degrades the prediction skill (Garcia-Oliva et al., 2024). In the context of sea ice, the ocean has a much larger heat capacity, so it is essential that the mask of sea ice and freezing surface temperature are updated in agreement. It is why we decided to update the SIC and ocean data both in anomaly, as the two are tightly correlated.

Initially, we attempted to update SIT in the anomaly field to be fully consistent. However, it had no real influence on our prediction (either improvement or degradation). We noticed that in our system, SIT bias correction was persisting for a very long time (10 years). We decided to test the current setup.

We are unsure of understanding the challenges with the example mentioned. In our system, the assimilation takes place in two distinct steps: 1) assimilate the oceanic and SIC data (in anomaly field) and 2) assimilate SIT data (full field). In the example provided, assimilation in step 1 increases the overall aggregated concentration, and the thickness will adjust the individual thickness concentration so that the ice is primarily in the thinner class. Note that assimilation will deplete the spread by construction.

In a post-processing, which takes place at the end of the assimilation step, the sea ice volume (vice) in each thickness category is changed proportionally so that the thickness of each thickness category remains identical to that of the prior. For clarity, we have modified the relevant text (L141) as follows:

> NorESM has a large SIT bias (Bentsen et al., 2012), and while assimilation of ocean observation reduces it partially, some of the bias remains. Bethke et al. (2021), compared two versions of NorCPM assimilating ocean observations, one that updates only the ocean component and one that updates the ocean and sea ice components. The latter yields a strong reduction of the bias of SIT and provides enhanced predictions. Note also that it takes about ten years for the model to rebuild the SIT bias once assimilation is stopped (their Figure S15). We, therefore, use full-field assimilation to correct the SIT bias that can influence the variability. In the first attempt, we used anomaly-field assimilation. However, the assimilation impact of SIT anomalies was inconclusive, with no added skill for predictions (not shown). Please also note that full field assimilation for the ocean is preferred in NorCPM (Counillon et al., 2016), because with the imbalanced observation data set (i.e. observation nearly only at the surface prior to Argo), full field assimilation produces large drift (Garcia-Oliva et al., 2024). As the ocean has a much larger heat capacity than sea ice, we prefer to update the ocean and the sea ice mask in agreement. This explained why we did not attempt to perform a full-field assimilation of ocean, sea ice concentration, and sea ice thickness.
>
> When assimilating SIT observations, we only update the individual category sea ice fraction, which can change the sum of the ice fraction. In the post-processing of the assimilation, the sea ice volume in each thickness category is changed proportionally so that the thickness of each thickness category remains identical to that of the prior. We do not update the ocean component, as the covariances between SIT and the ocean are very small, and may cause more harm than benefit because of sampling error.

RC: ***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.***

AR: There is a post-processing applied to the sea ice and ocean components. This post-processing is detailed in Kimmritz et al. (2018), and there is a reference to this paper in methods. However we have also clarified some of the post-processing as explained in the methods, which the reviewers also noted further down in this review.

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

AR: As suggested, we have added a map as a new figure (Figure 8) in the study in the prediction section.

RC: ***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/marinescience/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.***

AR: We were also not aware of this treatment. We re-plotted the IIEE results using this ensemble treatment. However, there was no noticeable difference in any of the results using this treatment of ensemble information. The reviewers can compare the new IIEE plot in the revised manuscript with the originally plotted IIEE (Figure R3).

1.2. Presentation

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

AR: The crosses are to show periods when the validation is not possible due to the lack of SIT observations from ENVISAT/CS2. The dots represent the ACC values that are not statistically significant. We have added this explanation to the captions for these figures (note that these are now figures 9 and 10 due to the addition of new figures).

1.3. Minor comments

RC: ***Line 1: the decline extends to at least four decades, so better write "four" here.***

AR: Thanks. We have modified the text (L1-2) as

> Arctic sea ice extent has declined significantly over the past four decades, opening up the Arctic to shipping and resource extraction while also impacting wildlife and local communities.

RC: ***Line 5: Specify the spatial resolution of NorCPM in your abstract.***

AR: As suggested, we have revised the relevant text (L5) as follows:

> We use the Norwegian Climate Prediction Model (NorCPM) with 1° horizontal resolution for the ocean and sea ice components and approximately 2° for the atmosphere and land components, which has previously assimilated ocean and sea ice concentration observations.

RC: ***Line 7: Specify which two decades.***

AR: We have added the exact period of the reanalysis as follows (L8-9):

> This allows us to produce a two-decade (2003-2023) reanalysis with sea ice thickness assimilation focusing on the Arctic Ocean.

RC: ***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.***

AR: We agree with the reviewers on this comment and have reworded this sentence to the following (L24-25):

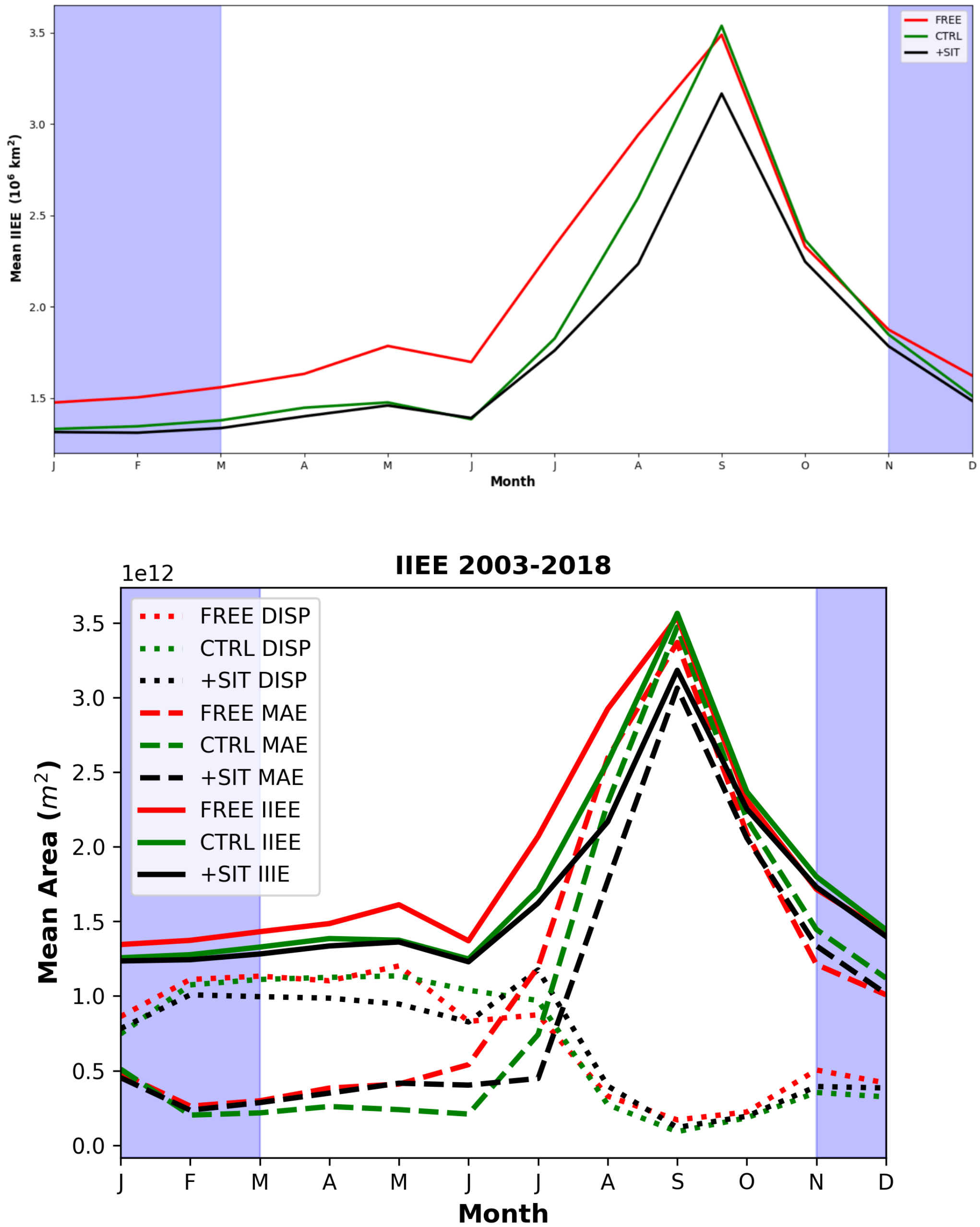

Figure R3: Top: Original figure from the paper, with the incorrect ensemble treatment of the climatological IIEE. Bottom: Climatology of the displacement (DISP), mean absolute error (MAE) and total IIEE of the FREE, CTRL and +SIT reanalyses.

> This has led to increasing interest in the study of Arctic sea ice and, in particular, of seasonal predictions, which aim to predict the sea ice on seasonal time scales.

RC: ***Line 24: The Arctic basin itself has not warmed, it is the atmosphere or the oceans that have warmed.***

AR: We have changed from "Arctic basin" to "Arctic climate" (L25).

RC: ***Line 26: remove "ice"***

AR: Removed, thanks.

RC: ***Line 26: is it the melting season that has lengthened, or the open-water season? We think the latter.***

AR: We have revised the sentence as follows (L27-28):

> Much of the Arctic sea ice has gone from being perennial year-round to seasonal, with an increase in the length of the open-water season.

RC: ***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?***

AR: We agree with the reviewers and have revised the sentence (L28-29) to

> Such radical change also makes seasonal Arctic sea ice prediction more challenging.

RC: ***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***

AR: Added, thanks (L40-41).

> Sea ice thickness (SIT) may be predictable up to 2 years in advance (Holland et al., 2011; Koenigk et al., 2012), and can be important for SIE predictions up to 2 years ahead (Tietsche et al., 2014) in an idealised framework.

RC: ***Line 43: "comprise of melting" is a strange sentence.***

AR: We have changed this text to

> Thermodynamic changes in the sea ice are composed of melting (lateral, bottom and top melt) and freezing (congelation and frazil ice formation).

RC: ***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.***

AR: We agree with the reviewers and have changed this to (L58-59)

> Atmospheric processes, particularly wind, which determine the dominant Arctic ocean currents of the Transpolar drift stream and the Beaufort Gyre, are also important but have a short memory.

RC: ***Line 62: To our knowledge CS2 has not been defined yet.***

AR: It has already been defined in L52 in the revised manuscript.

RC: ***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/gmd12-3745-2019)***

AR: Agreed and added, thanks.

RC: ***Line 127: What do you mean by strong and weak data assimilation method?***

AR: Strongly coupled DA refers to when model variables of a different component (ocean, atmosphere, sea ice) from the observed one are updated via cross-component covariance. (Stephen G. Penny, 2017), e.g., when ocean variables are updated when assimilating SIC data. If they are directly updated via the covariances during data assimilation, then it is strongly coupled; if not, then it is considered weakly coupled. For clarity, we have revised the text (L128-130) as follows:

> We update both the ocean and sea-ice components based on the observations from both components, an approach called strongly coupled ocean-sea ice DA (Laloyaux et al., 2016; Kimmritz et al., 2018; Stephen G. Penny, 2017). Strongly coupled ocean-sea ice DA in NorCPM was shown to be more effective than weakly coupled DA in which sea ice observations are used to only update the sea ice variables (Kimmritz et al., 2018).

RC: ***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.***

AR: It is correct that this was confusing. We do not think a schematic is needed, but we have revised the explanation (L132-134) as follows:

> The sea ice volume in each thickness category is changed proportionally in a post-processing of the assimilation so that the thickness of each thickness category for each member remains identical to its thickness category in the prior. As such, only the sea ice area of each thickness category is updated, which still yields an update of the total thickness but ensures that there is no need to re-allocate sea ice in a different thickness category post assimilation Kimmritz et al. (2018).

RC: ***Line 140: a large fraction of what?***

AR: We worded this poorly and meant to say that some of the sea ice thickness bias remains, even after assimilation. We have changed the text (L141-142) as follows:

> NorESM has a large SIT bias (Bentsen et al., 2012), and while assimilation of ocean observation reduces it partially, some of the bias remains.

RC: ***Line 144-145: "we update the individual multi-category sea ice fraction" but not volume then? Maybe a schematic would help.***

AR: We updated the sea ice fraction in each category during the assimilation and updated the sea ice volume in each category in the post-processing of the assimilation. For clarity, we have revised the text (L148) as follows:

> When assimilating SIT observations, we update the individual category sea ice fraction, which can change the sum of the ice fraction. However, as the prior assimilation of SIC observations has already constrained the ensemble close to the observed estimate, the subsequent SIT assimilation updates the individual fractions so that it complies with the observed ice thickness as well. The sea ice volume in each thickness category and for each member is changed proportionally so that the thickness of each thickness category remains identical to that of the prior.

RC: ***Line 158: error à errors***

AR: Modified, thanks.

RC: ***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).***

AR: Yes, this is what we meant. We have updated the manuscript (L187) to clarify as follows:

> The observation error statistics are provided by the datasets.

RC: ***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.***

AR: We have clarified this and the ensemble generation in the text (L190).

> a 30-member ensemble run without data assimilation, integrating from 1850 to December 2023 with CMIP5 historical forcings and with RCP8.5 beyond 2005. The ensemble is initialised on 01 January 1850 by randomly selecting 30 states on 01 January in different years from a stable pre-industrial run (with one single member) (Counillon et al., 2016).

The random states were picked over the last 100 years of a stable preindustrial run, and FREE is identical to the one used in Counillon et al. (2016).

RC: ***L190: Could you precise how your ensemble is generated?***

AR: Please refer to the previous reply.

RC: ***Line 195: disentangle from what?***

AR: For clarity, we have revised the text (L199) as follows:

> FREE allows us to estimate the skill related to external forcings (Kimmritz et al., 2019).

RC: ***Line 205: the area of grid cells***

AR: As suggested, we have modified the text (L212) to

We define the SIE as the total sum of the area of grid cells where the ensemble mean of SIC is at or above 15%.

RC: ***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)?***

AR: The DFS is limited by the minimum of the number of observations in the local analysis and the ensemble size (i.e., 30). It is correct that over the ocean domain, DFS is overly dominated by ocean observation. However, we aim to highlight the relative influence of different observations in the Arctic domain. If we adapt the different colour bars to individual panels, it would likely be confusing and not straightforward to compare the influence of different observations. If we keep using one colourbar for all panels and increase the range, then the DFS of SIT will become indistinguishable. Therefore, we have decided to keep the same colour scale as before.

RC: ***Figure 1: Can you refer to a) to e) in your caption instead of/or added to your location indications (top right, left,...)?***

AR: As suggested, we have modified the caption for the figure (Figure 2 in the revised manuscript).

RC: ***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.***

Figure 2: change the color of the localization of your observation points as there are indistinguishable from islands for instance.

AR: As suggested, we have added an extra row showing the mean SIT. We have also changed the colour of the localisation of observation points to green squares, thanks for your suggestion. Please see the new figure below or Figure 3 one the revised manuscript.

RC: ***How do explain the negative bias of +SIT compared to observation for SIT (Figure 3 and Table 1)?***

AR: This can be caused by two factors. FREE has a very strong positive SIT bias, so when assimilating SIT, the assimilation removes a lot of this additional ice thickness. However, the analysis is a linear combination between the model and the observation and some bias remains. Additionally, the way the BGEP SIT data is derived from the ice drift can potentially lead to an overestimation of the SIT, compared to the observation product assimilated.

RC: ***Figure 4b: Do you have any idea to explain this minimum in bfRMSE in June for FREE experiment?***

AR: We agree that it is an interesting feature, and we confirm that there is nothing wrong with the plot or the experiment. We noticed a similar feature when using another observation product, such as OSTIA – albeit not as sharp as with NOAA OISSTV2. The reason for this is that the fraction of total variability explained by the trend (see Equation 1 in (Kimmritz et al., 2019)) is larger for June than for the surrounding months. Experiments with assimilation (CTRL and +SIT), which are built on the FREE run, retain similar features. However, while +SIT represent better the year-to-year variability than FREE, the bfRMSE is poorer for June because the discrepancies in the bias between ENVISAT and C2S introduce a spurious trend that degrades performance.

We have complemented the text by explaining this better.

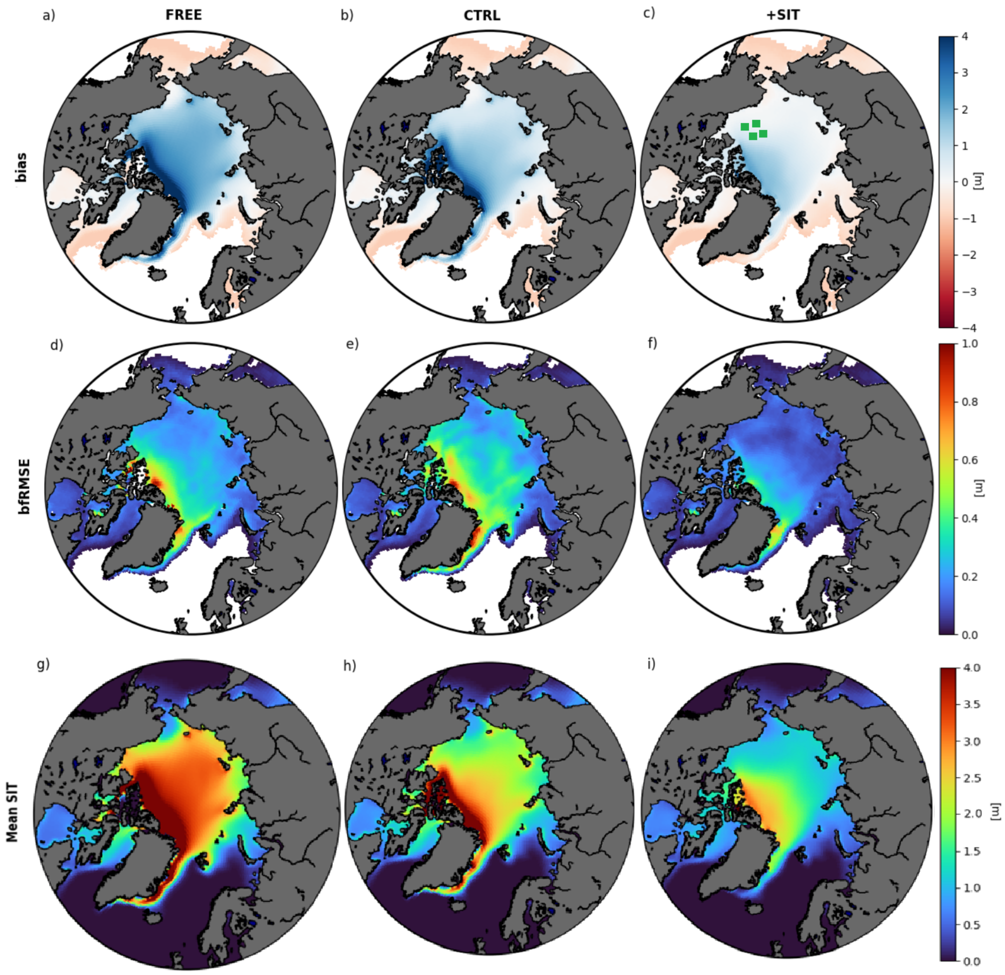

Figure R4: Bias (a - c), bfRMSE (d - f) and mean (g - i) of SIT in each of our NorCPM experiments in comparison to SIT observations from CS2SMOS between 2010 and 2023. In panel (c) we also show the locations of the BGEP ULS moorings as green squares.

For the bfRMSE of SIE (Figure 5b), +SIT and CTRL have lower errors than FREE, showing the positive impact of the assimilation of ocean and SIC data. However, the bfRMSE in SIT+ is larger than CTRL from January to August. In June, the bfRMSE for FREE reduces significantly, while there are smaller decreases in CTRL and +SIT. In that month, the fraction of total variability explained by the trend (see Equation 1 in (Kimmritz et al., 2019)) is substantially larger than for the surrounding months. While +SIT better represents the year-to-year variability than FREE, the bfRMSE is poorer for June because the discrepancies between ENVISAT and C2S introduce a spurious trend that degrades performance.

RC: ***Line 339: "we remove the trend": is it the linear trend?***

AR: Yes, it is the linear trend, we have added the word "linear" here to clarify (L375).

RC: ***Line350: (Figure 7) à (Figure 7 and 8)?***

AR: We have modified to "(Figures 9 and 10)" (L387).

RC: ***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.***

AR: Thank you for the correction, the "and March" was a mistake in the text. We have removed it.

RC: ***L447: in the Beaufort Sea***

AR: Modified. Thanks.

References

Bentsen, M., Bethke, I., Debernard, J., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I., Hoose, C., et al. (2012). The norwegian earth system model, noresm1-m–part 1: Description and basic evaluation. *Geosci. Model Dev. Discuss*, 5(3):2843–2931.

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H., Svendsen, L., Chiu, P.-G., et al. (2021). Norcpm1 and its contribution to cmip6 dcpp. *Geoscientific Model Development*, 14(11):7073–7116.

Counillon, F., Bethke, I., Keenlyside, N., Bentsen, M., Bertino, L., and Zheng, F. (2014). Seasonal-to-decadal predictions with the ensemble kalman filter and the norwegian earth system model: a twin experiment. *Tellus A: Dynamic Meteorology and Oceanography*, 66(1):21074.

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., and Bentsen, M. (2016). Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the norwegian climate prediction model. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1):32437.

Garcia-Oliva, L., Carrassi, A., and Counillon, F. (2024). Exploring the influence of spatio-temporal scale differences in coupled data assimilation. *EGUsphere*, 2024:1–26.

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T. (2016). Predictability of the arctic sea ice edge. *Geophysical Research Letters*, 43(4):1642–1650.

Holland, M. M., Bailey, D. A., and Vavrus, S. (2011). Inherent sea ice predictability in the rapidly changing arctic environment of the community climate system model, version 3. *Climate dynamics*, 36:1239–1253.

Kimmritz, M., Counillon, F., Bitz, C., Massonnet, F., Bethke, I., and Gao, Y. (2018). Optimising assimilation of sea ice concentration in an earth system model with a multicategory sea ice model. *Tellus A: Dynamic Meteorology and Oceanography*, 70(1):1–23.

Kimmritz, M., Counillon, F., Smedsrud, L. H., Bethke, I., Keenlyside, N., Ogawa, F., and Wang, Y. (2019). Impact of ocean and sea ice initialisation on seasonal prediction skill in the arctic. *Journal of Advances in Modeling Earth Systems*, 11(12):4147–4166.

Koenigk, T., König Beatty, C., Caian, M., Döscher, R., and Wyser, K. (2012). Potential decadal predictability and its sensitivity to sea ice albedo parameterization in a global coupled model. *Climate dynamics*, 38:2389–2408.

Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., and Janssen, P. (2016). A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142(694):65–78.

Lisæter, K., Evensen, G., and Laxon, S. (2007). Assimilating synthetic cryosat sea ice thickness in a coupled ice-ocean model. *Journal of Geophysical Research: Oceans*, 112(C7).

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of climate*, 20(22):5473–5496.

Rodwell, M., Lang, S. T., Ingleby, N., Bormann, N., Hólm, E., Rabier, F., Richardson, D., and Yamaguchi, M. (2016). Reliability in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 142(694):443–454.

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.

Stephen G. Penny, Santha Akella, O. A. C. B. M. B. M. C. F. C. C. D. S. F. Y. F. A. K. A. K. P. L. J.-F. M. M. M. M. P. X. W. (2017). Coupled Data Assimilation for Integrated Earth System Analysis and Prediction. Technical report.

Tietsche, S., Day, J. J., Guemas, V., Hurlin, W., Keeley, S., Matei, D., Msadek, R., Collins, M., and Hawkins, E. (2014). Seasonal to interannual arctic sea ice predictability in current global climate models. *Geophysical Research Letters*, 41(3):1035–1043.

Wang, Y., Counillon, F., Keenlyside, N., Svendsen, L., Gleixner, S., Kimmritz, M., Dai, P., and Gao, Y. (2019). Seasonal predictions initialised by assimilating sea surface temperature observations with the enkf. *Climate Dynamics*, 53:5777–5797.