Response to reviewer 3

December 12, 2025

Thank you to the reviewer for taking the time to provide excellent constructive comments which should help improve the manuscript.

Major Comments

- The results from the manuscript are based on the non-perturbed control member of the ensemble ORAS6 system. Therefore the short existing description describing the perturbations to the sea ice concentration observations could be removed. As it stands, I found the ensemble perturbation description just informative enough to not understand what has been instituted. However, for selfish reasons – I really wish to understand this – I would prefer they firm up the description somewhat, even if it does not have direct bearing on the manuscript. The alternative might be to refer the reader to some other reference, either past or future (tech documents being acceptable). The tech document [Zuo et al., 2017] gives a detailed description of the random superobbing of OSISAF sea ice concentration used in ORAS5. I assume this corresponds to the one sentence statement (l.100) "The data is randomly thinned to boxes of 0.25/0.25 degrees." I have many questions regarding the further description detailing randomly sampling both the structural error determined by differences between OSTIA and ESA SST CCI v2, and the analysis error determined by variance in ERA5.
  - Are the structural and analysis errors actually independent. Doesn't ERA5 use OSTIA as the lower boundary condition?

    ERA5 uses HadISST2 until 2008 and then OSTIA NRT
  - Are the differences between OSTIA and ESA SST CCI v2 a good representation of the uncertainty in observed sea ice concentration. Do they not use the same OSI-SAF sea ice product, or at least similar OSI-SAF products? I researched this, OSTIA uses the near realtime OSI-401-b and ESA SST CCI v2 presumably (no clear link to which product used) uses one of the climate reanalysis products of OSI-450-a1 or OSI-430-a. These will differ, but only by the choice of the long term anchor mechanism use to produce a self-consistent climate product, but not in the actual retrieval method.

    This is true. See below for more on this.
  - How do you build a database from each of these? I envision this could be as small a database as two numbers – a difference, and a variance. However, I suspect it must be a much fuller database that somehow incorporates temporal and spatial covariances into it. A more detailed explanation, or reference to a more detailed explanation of what is done here would be beneficial.

    See below.

  Why this might be important: There is a large uncertainty between the sea ice analysis products [e.g. Peterson et al., 2022, Niraula, 2023]. This large uncertainty, and in particular the fact that the sea ice edge from the OSI-SAF products appears to be too diffuse Renfrew et al. [2021], could have an impact on your results.

We understand your frustrations here and think some more context would be useful. The full ensemble perturbation strategy will be part of the up-coming ORAS6 reference paper. The short description we include here to keep such sea ice concentration details together. So we will ensure that a new version includes clear direction to either the new ORAS6 reference if it is submitted in time, or to the ORAS5 paper as you noted.

More importantly we already know that our perturbation strategy is insufficient to achieve realistic spread. In no large part this is because, as you have pointed out, OSTIA and ESA SST CCI v2 are based on similar underlying OSISAF product. This feature of the system was identified after production was completed and so we had no time to correct it with any other methodologies.

We wanted to keep this paper focused on the assimilation methodology which we would apply regardless of observation perturbation method. So we should acknowledge that the uncertainty (and how to account for it) in SIC observations could affect our assessments, and they be viewed with this in mind.

And as a further note, the lack of spread does not directly influence the data assimilation methodology we are using, as the spread is not used to estimate any flow dependent background errors for the sea ice. This is different to the 3D ocean, where background errors are derived from the ensemble spread. Work is planned for future systems to improve ensemble reliability in this area, both with new observation perturbation strategies, but also with stochastic physics developments in the model itself.

The following is a short, but more detailed description of the construction of the perturbation database. We prefer to leave such details out of this paper, and leave them for the overarching ORAS6 reference paper, but does no harm to share the details with you here.

We are building a SIC perturbation repository for both analysis and structural errors. Analysis errors are computed as differences between monthly individual ERA5 members and ERA5 ensemble mean over 1979-2020. Structural errors are computed as differences between OSTIA REP and ESA-CCIv2 monthly fields over 1982-2020. Hence we have *spatial maps* of samples of these errors. All fields are interpolated on a 1x1 grid and differences in the mean state removed.

The perturbation repository is stratified by calendar month and date range to capture uncertainties that correspond to the season and the level of sampling of the ocean by the observing system. For any given date, the perturbation patterns are randomly selected within the corresponding period of the repository and added to the corresponding field (i.e. done consistently for both sea ice and ocean perturbations).

I want to be clear: I completely agree that it is best to test the methodologies for distributing the sea ice into thickness categories in the control, unperturbed, deterministic simulation, and then validating the sea ice concentration fit against the assimilated observations. (I.e. I am not suggesting any changes to your methodology.) However, I believe results such as the degraded (Atlantic) sea ice edge thickness (Figure 4) when employing the gamma thickness distribution are likely related to uncertainties in the ice concentration observations – and this should probably be acknowledged. Similarly, the need to suppress temperature increments at the expense of sea ice increments (Section 6.3) is again likely a result of uncertainties in both the sea ice observations and SST analysis close to the sea ice edge as covered in my next point.

- Section 6.3: I have a small issue with referring to the suppression of temperature increments as a balance relationship. For a balance constraint, typically one would have a balanced temperature increment dependent on the sea ice, but then an additional unbalanced term and unbalanced increments to temperature that could still improve the fit to temperature observations. But more philosophically, here you are literally placing your finger on the balance scale by tipping the resulting analysis fit towards the sea ice observations and away from the temperature observations. If anything, it would be more akin to introducing bogus (zero innovation) observations in the presense of sea ice. All these methods, along with still others are employed to improve retention of increments in the analysis, so I have no issue with the method, just perhaps the terminology.

  This is the main issue at hand, so we are very glad to see it is clear until this point. We are happy to replace the term "balance" with "physical relationship" to improve communication around this terminology.

  However, my understanding is not complete, and maybe they are all interconnected. When first I looked at the results it is clear that the fit to sea ice does improve with increasing alpha – but it was not clear that your choice of alpha was actually maximizing that fit (i.e. would it be a good idea to test a higher value of alpha). With better understanding, I now believe there would be no maximization (although perhaps saturation); continuing to increase alpha should continue to increase the fit to the sea ice observations, although presumably in detriment to the fit to temperature observations and presumably with lowering levels of improvement as alpha is further increased. What is shown (Tables 4 &5; Figures 6,7 &8) is only half the story as presumably this is all at the cost of sea surface and sub-surface temperature statistics (but only in the vicinity of the sea ice edge). This then connects with item 1: How does this connect with uncertainties in the sea ice concentration and temperature observations? Are you improving a fit to observation (in reality, that "observation" is actually an analysis) to a point the fit is well below the significant uncertainty in the observation? Figure 7, 8 & 9 of Renfrew et al. [2021] would seem to suggest this might be so. Is there a need to re-address this issue with post 2025-07-06 results and the introduction of the OSI-438 (AMSR2) product (Table 1)?

  This is an excellent summary which hopefully we can help expand on to make it clearer.

  Please see our responses to RC1 and RC2 who both suggested increasing α further to partially answer your question here.

  Ocean temperature observations are few and far between near the ice edge or under the ice pack, so when we look at observation fits to temperatures we do not see degradations (these are not shown as neutrality seemed not very interesting for the reader). So we could continue to increase α and not see degradations, all the while we are making larger and potentially unconstrained changes to the ocean below the ice. It might be that instead of the ocean temperatures being the reason for ice increments not being retained, it might be instead the atmospheric forcing that means ice increments are rejected.

Minor Comments

1. I think it should be pointed out other methods of distributing sea ice concentration amongst the thickness categories do exist. For example, Smith et al. [2016] propose the rescaled forecast tendencies (RFT) method that is largely identical to your method. For instance their Eqn. 1 is identical to your equation 3. Please cite.

Thank you very much for pointing this out. We were unaware of this work and it clearly needs to be cited and discussed. We think our method is the same as their Rescale the Existing ice thickness Distribution (RED) method and not their Rescaled Forecast Tendencies (RFT) method. "In effect means that the increment is distributed in proportion to the existing ice thickness distribution as in the *Rescale the Existing ice thickness Distribution (RED)* method of Smith et al. [2016]".

We will further add to the conclusions some discussion on their RFT methodology:

"We have shown that results are highly sensitive to the method used to distribute a single category increment across the multiple prognostic thickness categories of the model. Our choice to keep the increments in proportion to the background profile of ice distribution was shown to perform well, and has the benefit of leaving the sea ice thickness unaffected, by design. There are other methods in the literature for distributing ice increments across thickness categories, such as the Rescaled Forecast Tendencies (RFT) method of Smith et al. [2016]. Such methods that estimate what type of model deficiency led to an increment to the sea ice concentration, and apply concentration increments in such a way to counteract them. We leave comparisons against such methods for future investigations."

However reading their paper it is not clear which forecast lead time to use to appropriately compute such forecast tendencies given the First Guess at Appropriate Time (FGAT) methodology we use. We will need to discuss with Smith et al. in this regard.

2. Joint Minimization (l. 114): It has been my understanding that joint minimimization of the univariate sea ice and multivariate ocean has always been a problem that leads to degraded fits, so I am happy that you have managed to achieve this. However, will this not be a function of the observing system. Could improved numbers of near ice temperature observations change this dynamic? I suppose the hope is to go toward a more balanced, joint multi-variate ocean sea-ice assimilation before any substantial changes occur in the observing network.

   That was our previous experience, when assimilating level 4 sea ice concentration products. It was implemented somewhat naively, sampling sea ice observations globally where clearly the data is of no use in regions such as the tropics. This led to an imbalance which observations the minimsation targetted with its gradient descent method. The two key developments were moving to L3 sea ice observations where missing data is accounted for appropriately and we do not sample regions far from sea ice, and the modifications to the oceanic component of the background error covariance matrix as documented in Chrust et al. [2024].

3. Figure 4 (Section 6.1). As stated above in Major Item #1, I believe the degradation in sea ice thickness when using the gamma distribution is ultimately due to the uncertainties in the sea ice concentration and location of the ice edge. Although we are given insufficient information to confirm this, I suspect the gamma distribution has a bias towards thicker ice, at least at the ice edge (what is the hemisphere change in bias?) – which then shows up as increased RMSE there as the thickness observations would be sensing zero thickness ice. (I.e. Your decision – I believe correct decision – to not utilize the slightly better in terms of ice concentration, gamma distribution, is likely due to offsetting biases.) This might be a result that could change in the perturbed solution. It also again would likely change post 2025-07-06 with the introduction of OSI-438. We agree. Hopefully the new figure we propose as a response to Reviewer 2's comments will also address this point - it is indeed a bias in the marginal ice zone which we see.

4. Open Water (Section 6.2) I could not help but think your open water fit results were one-sided. The results continue to get better as you increase the thickness ice increments in open water. While I would agree adding ice to open water into the 2nd category would seem weird – and would disconnect your data assimilation new ice thickness to your thermodynamic new ice thickness. (Actually that wording is a little ambigouous in Section 4.3.2 – the new ice thickness in the DA is or is not identical to the thermodynamic new (frazil) ice – or just the other properties?) At any rate, the same arguments that the data assimilation process is correcting for errors in dynamical movement of ice as much as errors in thermodynamics as justification for adding ice across all categories could presumably be used to justify adding higher category (relocated ice) for positive increments in open ice. Sorry: I seem to recall a statement is made along these lines, but I could not track it down in the manuscript. It might have been interesting to test adding 2nd category thickness ice to see where the fit to observations begins to decrease.

   We understand this point of view. It has certainly been a technical limitation that stopped us from increasing the thickness of new ice to open water beyond the maximum allowed in the thinnest category. As for the choices we made to distribute increments to the existing ice pack, we are very much hoping that (passive microwave) observations of sea ice thickness will ultimately tell us what we should be doing in these areas.

   Until that point we want to keep the system as simple as possible. Once we start adding thicker ice, would we add distribute across the first 2 ice thickness categories? Or put it all in the second? These would be interesting to investigate, but will have to be left for future work.

   "The other sea ice properties are set so that the sea ice created by the DA is identical to sea ice that would have thermodynamically formed into the open water." we should reword this to read

   "The properties of this new sea ice are the same as frazil new ice formed themodynamically by the model from open water."

5. Table 4&5/Figure 8: If I have not missed something, α (Eqn. 5) has units of °C or K (it is a change in temperature – so your choice). The values of α in Tables 4&5 and Figure 8 – and in the text if they occur there – should be accompanied with that unit.

   Happy to add that.

6. I will forego my anonymity as penance for my tardiness in achieving this review

   Tardiness is welcome in this case, and we greatly appreciate your time and effort to help us improve this paper.