Response to reviewer 1

December 1, 2025

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

Major Comments

1. Conceptual and Methodological Clarity

- Lines 15–40+: Please define all acronyms (ECMWF, ORAS6, LIM2, SI3, NEMOVAR, etc.) upon first mention. The introduction assumes strong familiarity with ECMWF's system architecture, which may limit accessibility for broader readership.

  Indeed there are some missing definitions which we will gladly expand.

- Lines 120–135:

  The statement in Line 126 that "sea ice concentration is orthogonal to sea ice thickness" requires careful revision. These quantities are not strictly orthogonal in the linear algebraic sense; rather, they are inherently linked through the ice volume relationship SIVn = SICn x SITn. For instance, during the melt season, increased atmospheric and oceanic temperatures cause thinner ice to melt more rapidly, leading to a decrease in SIC. Meanwhile, thicker ice (i.e., higher SIT) tends to persist longer, resulting in spatial patterns where SIC and SIT covary—contradicting the claim of orthogonality. During the freeze-up period, SIC and SIT may indeed covary less, but as illustrated above, this relationship is not universally negligible. Assuming zero covariance between the two may therefore misrepresent coupled ice processes and potentially lead to unphysical SIT (and consequently SIC) states. While SIC and SIT are often treated as distinct state variables, it is inaccurate to describe them as orthogonal. This assumption should be revisited, and the implications for the chosen increment distribution scheme should be discussed. If SIC is updated independently across categories primarily for practical reasons, that rationale should be made explicit.

  We understand the reviewer's strong view here, and agree with all their scientific points. However we think this is partially a miscommunication on our part, as we have used the term "orthogonal" in the geometric sense to mean perpendicular. We absolutely agree that the quantities are related and correlated, especially with a multicategory model, but they remain representative of spatial extent (SIC) and vertical extent (SIT). This we attempt convey in Figure 1.

  We would propose to change the wording such that our statement "Noting that sea ice concentration is orthogonal to sea ice thickness" would become "Noting that sea ice concentration represents the spatial extent and sea ice thickness represents the vertical extent"

- Following up on this point (and similar to the one above), since sea ice volume is an extensive variable and is updated as such in Equation 4 (Line 158), have you considered how restricting changes in the ITD impacts your updates? For example, imagine an update in the Central Arctic, where ice is relatively thick (< 1 m). Suppose a ridging event occurs that leaves open water, and the ice then refreezes according to observations and the prevailing cold atmospheric state, warranting a positive SIC increment. This newly formed ice is likely thin to start, according to well-established theory, but under your current update scheme, it would be constrained to adhere to the original thickness distribution—potentially overcompensating and producing too much ice volume, leading to an unphysical amount of ice. Have you considered this or similar scenarios? the ice volume increment could be scaled more strongly toward thinner ice categories in both positive and negative increments to better represent the persistence of thicker ice, as widely documented in the literature.

This is a very clear example of a limitation of our system. In this case we rely completely on the model dynamics to simulate such events, and if they are missed then the resulting analysis will become biased. What we need in order to constrain our system are observations of the sea ice thickness. We lack the capacity to assimilate available sea ice thickness observations and are working actively on their inclusion in a future system. (See for example, the WMO WWRP PCAPS task team on sea ice thickness data assimilation `https://www.wwrp-pcaps.net/what-we-do#task-teams`, ESA funded DANTEX project `https://www.ecmwf.int/en/elibrary/81674-earth-system-assimilation-cryosphere-status-and-way-forward-td-01`). We assimilate SIC observations in our multicategory model so that the category-average thickness is left unaltered, partly because we did not want SIC observations to bear an impact on effective thickness. Our hope is that these future SIT related observations will then be able to constrain that other dimension to the representation of the sea ice. We also note that the scheme of Peterson et al. (2015) goes some way to implement the methodology you suggest, which we tested (line 230-245) and showed to perform less well than Gamma or background splitting.

- Lines 165–175: The approach of assigning a fixed thickness of 45 cm to newly formed ice is insufficiently justified. The sensitivity tests in Section 6.2, which explore the additional fixed thicknesses of 22.5 cm and 10 cm, are valuable but not comprehensive. A more physically consistent approach would be to assign thickness proportionally to the magnitude of the SIC increment (e.g., thinner ice—approximately 10 cm—for small increments and thicker ice—up to 45 cm—for larger increments). Please also provide the specific reference supporting the 0.5 m threshold used in ORAS5.

We cannot go any thicker than 0.45m as adding ice into categories that are not the thinnest poses technical challenges which we were unable to overcome. What we have shown with the sensitivity tests is that we need to add ice with as much enthalpy as possible to stop the model from melting out the new ice. Thus any weighting by SIC increment would act to give less enthalpy to the new ice, and thus the preexisting biases in the ocean state and atmospheric forcing would more easily remove the ice that the observations are informing should be there. The 0.5m minimum SIT was not documented in the ORAS5 paper. This is the same value used by LIM2 model when forming new sea-ice in the open water. We understand that the same 0.5m of SIT was used in the UK Met Office system when new ice is added.

We are happy to add a reference to Fichefet and Maqueda [1997] where they also use 0.5m as the thickness of new ice.

T. Fichefet and M. M. Maqueda. Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *Journal of Geophysical Research: Oceans*, 102(C6):12609–12646, 1997

- Lines 180–195: The implementation of an "ice-induced temperature increment" is a creative solution to maintain thermodynamic balance, but its physical basis and vertical extent (to 19.5 m) are questionable. Please provide justification for the chosen α values and the depth profile f(z), as decreasing ocean temperature at such depth could unintentionally alter stratification or vertical mixing. It is unclear whether this adjustment differentiates between increments arising from advection versus thermodynamics—this distinction should be addressed. For example, in the case of advection, atmospheric forcing from cyclones may move ice equatorward, which most likely does not alter the near-surface ocean profile in the same way that the thermodynamic growth of ice would.

  We agree that the physical basis for this mechanism is not carefully considered . Indeed, it should not be needed if the ocean temperatures were perfect and conducive to maintaining ice of the required concentration at the surface. Ideally further developments could be made in this area, such as ensuring the energy content of the temperature increment is somehow related to the latent heat of fusion. As explained, ideally such an approach should be achieved within the balance operator of data assimilation scheme, which is not yet implemented in our system. Various options of alpha have been tested (see Chapter 6.3). However, we do acknowledge that the final choice of this vertical profile f(z) and value of alpha could be refined further, based on physical processing and exchange of energy between liquid water and sea-ice.

  There is no case where we differentiate between increments arising due to model deficiencies in either advection or thermodynamics. Untangling which scenario we are in from observations of SIC alone seems like a challenging problem to automate for every cycle of a reanalysis.

- Lines 220–270: The experimental design is clearly described; however, several key methodological details are missing and appear to rely on the reader's prior familiarity with ORAS6. Please specify the number of ensemble members at the first mention (e.g., near Line 100, rather than only at Line 216). Further clarification is also needed regarding the "deterministic (single-member) experiments" described at Line 220—specifically, how they were implemented, why this approach was chosen, and whether any statistical significance testing was performed to account for potential control-member dependence (e.g., rerunning with five different control members to assess robustness). Additionally, the presentation of normalized standard deviations in Tables 2– 5 would be clearer if accompanied by absolute RMSE values and the sample size used for each experiment (e.g., n = ...).

  We are happy to provide these clarifications. We would add the deterministic method is chosen for computational cost reasons given the high expense of running a global ocean/sea ice reanalysis for the 5 year period. We are unsure how to describe "how they were implemented" in a way that adds any meaningful insight. No significance testing was performed to account for potential control-member dependence.

  We have approximately 250,000 observations per cycle, and the experiments run for 360 cycles over the 5 year period, leading to a total number of observations around n = 90,000,000. We can add that the control standard deviations are globally: 0.098 m^2/m^2 with a sample size of 9.15×10^7, northern hemisphere 0.130 m^2/m^2 with a sample size of 3.82×10^7, southern hemisphere 0.095 m^2/m^2 with a sample size of 5.32×10^7,

2. Physical Justification and Link to Observations

- Lines 50–90:

  The discussion of SI3 physics is thorough but tends to conflate "detail" with "accuracy." Please avoid language such as "more accurate" when the evidence shown only supports "more detailed" physical representation of the ITD. Claims of improved accuracy should be supported by comparisons with independent observational datasets or independent models.

We are happy to remove statements about accuracy here, we do not want this paper to be a reference for the model performance comparing SI3 with LIM so it is right that we reword things here.

- Lines 90–120: Please clarify the rationale for the choice of observation products (OSI SAF datasets, Table 1). This could be as simple as noting that these datasets were considered state-of-the-art for their respective time periods. For example, while the text indicates that OSI-450-a, OSI-430-a, and OSI-438 were used sequentially, it does not discuss potential cross-calibration issues or bias transitions between these products. If such effects were evaluated and found negligible—or were deemed unquantifiable—please state this explicitly.

  As in response to RC2:

  We move to L3 because the assimilation system can cope with missing data, something the old system could not do and therefore required L4 data (gap filled). OSISAF provides a climate data record from 1978 onwards which seamlessly transitions into a near real time product. This near real time product is suitable for the operational schedule for NWP production. Moreover the data is available with operational levels of support. We have decided to move the contexts about OSI-SAF data transition to fast track data and OSI-438 product to the ORAS6 reference (currently in preparation), including corresponding discussion on the cross-calibration and continuation of sea-ice states during these transition.

- Lines 110–115: Please improve the justification of the chosen NEMOVAR settings. Further explain why only a few parameters were modified, and reference relevant literature to support the selected background error covariance matrix configuration. What is the rationale for using a constant value of 0.2 for the observation operators? It is likely that a parabolic error distribution would be more appropriate for SIC, as it better represents lower uncertainty at the physical bounds (0 and 1) and higher uncertainty at intermediate concentrations. As noted we have kept all the settings listed in this section the same as implemented in ORAS5. We agree with the reviewer that these settings, particularly, the specification of observation errors for sea-ice concentration, could be further refined, e.g. using the parabolic distribution as suggested. We will add a sentence to clarify this point. All of these could be optimised but have not been.

3. Figures and Visualization Figures 1–8: Figures would benefit from clearer visual presentation. Specific points:

- Figure 1: Use darker ocean coloring and simplified snow overlays to better illustrate ITD structure and surface layering.

  Figures 4–8: Label all color bars and ensure aspect ratios are not distorted. Several figures appear stretched vertically.

  Figure 7: Avoid duplicate words (e.g., "sea sea")

  Figure 8: Clarify whether this map represents a single cycle or an average over multiple cycles.

  Please include the number of experiments or assimilation cycles used to produce each composite figure, either within the figure itself or in its caption. Without this information, it is difficult to assess the reproducibility and robustness of the presented results.

We are happy to correct and/or clarify these. Unless otherwise stated (figures 6 & 7), the plots are from the entire 5 year experimental period. We can add that Figure 2 is an instantaneous look at the model field.

4. Results and Interpretation

- Lines 240–260 (Gamma and Peterson splitting tests): The text concludes that the proportional ("background") increment method performs best, due to worse performance from the gamma-based distribution in the Fram Strait (Fig. 4), which is important for the ECMWF NWP system. Was it ever considered to use the gamma-based distribution within the Central Arctic, which yields better SIC scores, and the "background" increment scheme near the ice edge? Also consider whether observational uncertainty or regional biases in the Level 3 product may have influenced these comparisons (especially in the marginal ice zone).

  No this was not considered. It is certainly an interesting suggestion which we can consider for future work.

- Lines 260–290 (α sensitivity tests): The selection of $\alpha = 5$ as optimal appears partially ad hoc. Please include the rationale behind only testing $\alpha \leq 5$. Mention whether $\alpha > 5$ yields further improvement or instability.

  As in response to RC2:

  This is a very fair question. We are conservative in the approach and do not want the ice induced temperature increments to be larger as they could be compensating for errors in the forcing and not in the ocean state. Following equation (5), if δa is, say, 0.5 (a 50% change in ice concentration - large but not unexpected near the ice edge), then with $\alpha = 5$ this gives a 2.5K temperature increment to the ocean. These are regions where there are limited in situ temperature measurements at depth, and so we fear making even larger temperature increments could have negative consequences on the 3D ocean state. We do hope to improve on the simplicity of this scheme in the future, and very much recognise the need for and welcome suggestions for future developments to this approach.

- Lines 275–285 (Conclusions): The discussion asserts that "ice-induced temperature increments give the largest impact of all developments" based on your results, but the causal physical explanation is missing. Please elaborate/explore the physics in greater detail why this term may dominate the performance gain. The conclusion is also much more of a discussion of future work as opposed to a wrap-up of the completed work. Perhaps a future work and/or discussion section focusing on this is desired.

  We can certainly expand on this. We will speculate that, due to the lack of observational temperature constraints below the sea ice, model biases exist which prevent the ice model from either forming or sustaining sea ice of the concentration detected by the observations. Hence the ice-induced temperature increments are going some way to counteract this model bias.

The following minor comments are all well noted and we are happy to address in a revised version of the manuscript.

Minor Comments:

- Line 18: "as a (thermo)dynamical isolating cover" → "as a (thermo)dynamic insulating barrier"

- Line 34: "Other developments on sea-ice assimilation including attempt" → "Other developments in sea-ice assimilation include attempts"

  Section 2

- Lines 59-60: ”uses enthalpy rather than temperature, as in LIM2, as a prognostic variable” → Reorder: ”uses enthalpy, as in LIM2, rather than temperature as a prognostic variable” (clearer)

  Section 3

- Line 77: Should equation (1) have a period after it?

  Section 4

- Line 124: ”This is not a well defined problem” → ”This is not a well-defined problem” (hyphenate)

- Line 126: ”it is desirable that” → ”ideally,” (more concise)

- Line 145: ”The net result of this is that” → ”Consequently,”

- Line 159: ”meltpond concentrations/volumes/lid volumes” → ”melt pond concentrations, volumes, and lid volumes” (consistency and clarity)

- Lines 160-161: ”potentially because sea ice areal age is proportional to concentration, not volume” → Please rephrase this statement for clarity and explain how SI3 calculates sea ice areal age.

  Section 5

- Line 212: ”Hourly data from ERAS5” → ”ERAS5” should be ”ERA5”

- Line 220: ”we show results from deterministic (single member) experiments, using the ORAS6 EDA” → Clarify: how do the single members relate to the EDA? Please be more specific.

  Section 6 redundancy)

- Line 241: ”The scores shown in Table 2 show that” → ”Table 2 shows that” (remove

- Line 252: ”Global (err)” in table headers → Define ”err” in caption – what does this error represent physically, and how does it relate to being a standard deviation?

  Section 7

- Line 271: ”Sea ice concentration is an important dataset to be assimilated” → ”Sea ice concentration is an important variable to assimilate” (datasets vs. variables)

- Line 296: ”In the near future” → Specify timeframe

  General observations:

- Inconsistent hyphenation: ”sea-ice” vs ”sea ice” - standardize throughout

- ”melt pond” vs ”meltpond” - be consistent

- Some sentences begin with ”This” or ”These” without clear antecedents - consider being more specific