This manuscript presents the development and evaluation of sea ice concentration (SIC) data assimilation within ECMWF's new Ocean Reanalysis System 6 (ORAS6), which incorporates the multicategory SI3 sea ice model. The authors describe modifications to the NEMOVAR data assimilation framework, including the distribution of SIC increments across thickness categories and an imposed ice–water temperature balance adjustment. The experiments are well-structured and represent a meaningful advancement toward more physically consistent SIC assimilation in a multicategory model.

That said, the paper requires substantial clarification and justification across multiple sections before it is ready for publication. Many of the presented choices (e.g., increment-splitting schemes, fixed new-ice thickness, and temperature adjustments) lack sufficient physical motivation or citation support. In addition, key assumptions—such as the claimed orthogonality between sea ice concentration and thickness—are questionable and should be more carefully examined or revised. Figures also require clearer labeling, scaling, and physical interpretation. Overall, the manuscript is promising but would benefit from a more transparent exposition of methodological decisions, improved figure presentation, and tighter linkage between the experimental setup and the stated objectives.

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.

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 $SIV_n = SIC_n \times SIT_n$. 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.

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.

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.

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.

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 = ...).

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.

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.

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.

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.

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).

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.

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.

Minor Comments:

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

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
Line 241: "The scores shown in Table 2 show that" → "Table 2 shows that" (remove redundancy)

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