Review for "Development and evaluation of a new 4DEnVar-based weakly coupled ocean data assimilation system in E3SMv2" by Shi *et al*

Shi and co-authors evaluate an assimilation run produced with the 4DEnVar-based weakly coupled *ocean* assimilation method applied to the Earth system model E3SMv2. The 4DEnVar method follows the dimension-reduced projection 4DVar method of Wang et al (2010). The present study, together with a recently published work also by Shi et al (2024) [`https://doi.org/10.5194/gmd-17-3025-2024`] on a weakly coupled *land* data assimilation for E3SMv2 using the same methodology, appear to be part of the authors effort to produce realistic initial conditions for decadal climate predictions with E3SMv2. This is welcome news for the decadal climate prediction community and climate prediction users at large.

The present work briefly describes the authors implementation of the 4DEnVar method, discussed at length by Shi et al (2024), and presents evaluations of an E3SMv2 assimilation run. These include evaluations of the model 3D ocean temperature and salinity, which are constrained directly with the EN4.2.1 observational dataset, and air temperature and precipitation over the contiguous United States, which are constrained indirectly from the effects of the constrained ocean on the atmosphere. Since the methodology has already been discussed and tested elsewhere, and the results discussed here appear to be robust, I do not have major critical concerns. I do however have several suggestions/comments that I exhort the authors to address to hopefully improve the presentation and enhance the relevance of their work:

1. The evaluation of the E3SMv2 assimilation run only uses a control simulation that does not ingest observational data (other than the external forcing, common to both runs). That is, there is no assimilation method to benchmark against (e.g., simple nudging, or just imposing the observed temperature and salinity fields) so as to assess the effectiveness of the 4DEnVar-based methodology. For example, while Fig. 3 shows that the assimilation run outperforms the pre-assimilation background, it is unclear whether the 4.20 % error reduction (L239-240, Section 3.1.) can be considered "good" enough to justify the complexity of 4DEnVar. I do not suggest to produce a benchmark simulation using a different assimilation method, but I wonder whether the authors (or someone else) have tried a simpler approach on the E3SM model, or whether the authors can comment on previous work showing comparisons between different initialization methods that can shed light on this.

2. Section 2.4. In the experiment design, did the authors used a spinup run for equilibration before performing the assimilation, or is the assimilation applied directly from a piControl (L167) or historical (L205-207) run? Can the authors clarify and expand on this?

3. L222-224 Do "cost function", "cost function reduction" and "reduction rate of the cost function" refer to the same quantity? Please clearly name the quantity in Eq. 1 and use the same terminology subsequently.

4. L227 How is the observation error covariance matrix $\boldsymbol{R}$ computed? How is this matrix for EN4.2.1 ocean temperature and salinity? e.g., Is $\boldsymbol{R}$ diagonal or quasi-diagonal? If not, any insight on its spectral properties? How the characteristics of $\boldsymbol{R}$ are expected to impact the assimilation process?

5. L234-242, Section 3.1. While the authors' message is clear, the use of negative percents is odd. Consider showing positive percents specifying that they correspond to improvements due to the assimilation method.

6. L255 Can the authors expand on the two suggested reasons for the performance degradation in the deep ocean? If possible, can the authors provide some comments specific to the E3SMv2 model and the EN4.2.1 observational data?

7. L268-269 Is the seasonal cycle removed from the time series before computing the correlations? Please specify. And, is the linear trend removed?

8. L277-278 The authors suggest that the degradation in performance is "possibly due to sparse observational data or complex ocean dynamics". Can the authors expand on this? In particular, if the control run does not use observations (except for the external forcing, as it is the case for the assimilation run), how/why the sparse temperature and salinity observations would degrade the performance of the assimilation run relative to that of the control run?

9. L305. According to the text, Fig. 9 shows global mean RMSE of vertically averaged temperature and salinity. From the caption to Fig. 9, it shows RMSE of the vertically averaged global mean ocean temperature and salinity. As these are two different quantities, please correct and clarify which one is shown.

10. Figure 10. Panels (a) and (c) show CTRL minus OBS. However, from the caption and panels titles, (b) and (d) show ASSIM minus CTRL. Why? I would expect to see ASSIM minus OBS to assess the biases of the assimilating runs relative to those of the control. Please clarify, otherwise I would suggest to show and discuss the results for ASSIM minus OBS. This would imply changes to the discussion in L318-332.

11. L318. What "mean differences"? Please specify in the text. See previous comment.

12. L322. From Fig. 10b it is unclear whether the "SST biases found in CTRL are substantially reduced by ASSIM". See comment above on Figure 10.

13. In addition to Fig. 11, can the authors show the correlation and RMSE maps for both temperature and precipitation over the contiguous US (including statistical significance)? This will be useful to assess the regional impacts of the assimilated ocean. Perhaps the authors could show results for seasonal averages instead of annual means, choosing the seasons of strongest ENSO influence on US temperature and precipitation.

14. While Fig. 6 shows improved ocean temperature for the assimilation run, this result uses difference of correlations relative to control. As part of the analysis in section 3.6, it would be useful to directly assess the SST variability of the assimilation run in the tropical Pacific, which is expected to influence the simulated climate over land. For example, consider showing time series of the seasonal averaged (e.g., DJF) Niño 3.4 index for the CTRL, ASSIM and OBS, and their correlation with OBS.

15. EN products have data-sparse periods and regions. Salinity in particular is sparsely observed and could lead to spurious static instability in the absence of dynamical constrains. Can the authors expand on the potential limitations/advantages of using EN.4.2.1 for the assimilation process instead of reanalysis products such as ORAS5 [`https://cds.climate.copernicus.eu/datasets/reanalysis-oras5?tab=overview`] or GLORYS [`https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_PHY_001_030/description`]? This would be useful information in particular for producing centers of decadal predictions. Such discussion could be added to the concluding remarks, aligned with the authors "aim to advance the predictive capabilities of E3SM for decadal predictions".

Minor edits, etc

1. L109 Change "employs sophisticated representations of" with "represents"
2. L292 Change "variations in" to "of"
3. L306 Add "over the top 1000 meters"
4. L341 What "multiple US regions"? The contiguous US?
5. L633 Change "in the" to "averaged over".
6. L382 Consider changing "challenges" with "limitations"