We thank Reviewer #2 for the constructive comments and suggestions, which greatly help to improve the quality of our manuscript. We have made revisions and replied to all the comments. Please find the point-by-point responses to the comments below. Our responses are shown in "Blue" and the changes in the manuscript are shown in "Red". The line numbers correspond to those in the clean version of our revised manuscript.

Response to the comments from Reviewer #2

General Comment:

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.

Response:

We would like to express our sincere gratitude for your time and effort in reviewing our manuscript. We truly appreciate your constructive comments and suggestions, which have significantly contributed to enhancing the quality of our work. We have carefully addressed each comment, as outlined below, and have made the necessary revisions to our manuscript.

Comment#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.

Response:

Thank you for your insightful comments. We agree that comparisons with simpler approaches could strengthen the effectiveness of the 4DEnVar method. Unfortunately, to the best of our knowledge, no previous studies have applied a simpler method (e.g., nudging) for ocean data assimilation in the E3SM model, and implementing such methods for comparison with 4DEnVar would require substantial effort that goes beyond the current scope of this study. However, He et al. (2020) implemented a similar 4DEnVar-based ocean data assimilation system for the FGOALS-g2 model and demonstrated improvements over the previous assimilation system in that model. Specifically, the 4DEnVar method in their study showed an average monthly reduction rate of the cost function of 4.4%, better than 3.0% from the previous system. In our study, we observed a comparable 4.20% reduction, further supporting the effectiveness of the 4DEnVar approach. Furthermore, previous studies have shown that 4DVar-based methods outperform simpler methods, such as nudging and 3DVar, by maintaining dynamical consistency with the model and minimizing initial shocks in the forecasts (Sugiura et al., 2008; Zhang et al., 2020).

Based on your suggestions, we have revised the manuscript to include the advantages of using 4DVar-based methods over simpler assimilation techniques (L87-89) and note that the 4.20% reduction rate in our study is comparable to the 4.4% reported by He et al. (2020a) using a similar 4DEnVar-based assimilation system (L261-263), further supporting the effectiveness of this methodology.

L87-89: Previous studies have shown that 4DVar-based methods outperform simpler schemes (e.g., nudging or 3DVar) by maintaining dynamical consistency with the model and minimizing initial shocks in the forecasts (Sugiura et al., 2008; Zhang et al., 2020).

L261-263: This average reduction rate of 4.20% is comparable to the 4.4% reduction rate reported by He et al. (2020a), who used a similar 4DEnVar-based assimilation system in a different climate model, further supporting the effectiveness of the 4DEnVar approach.

Comment#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?

Response:

To clarify, the assimilation experiment (ASSIM) is initialized directly from the historical run in 1950, with no additional spin-up performed. We have added the following sentence (L228-230) to make this clear.

L228-230: The assimilation run is initialized directly from the historical run in 1950, using the fully coupled state at the start of the simulation.

Comment#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.

Response:

Thank you for pointing out this inconsistency. To clarify, "cost function reduction" and "reduction rate of the cost function" refer to the same concept. The reduction rate of the cost function represents the percentage decrease in the cost function. The formula presented in Eq. 1 represents the reduction rate of the cost function.

To ensure consistency, we have revised the manuscript to consistently use the term "reduction rate of the cost function" throughout the manuscript, including the description of Eq. 1 (L240-241) and all subsequent mentions.

L240-241: The reduction rate of the cost function serves as a fundamental measure to assess the assimilation system's accuracy, calculated using the formula:

Comment#4:

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

Response:

The observation error covariance matrix ***R*** is determined statistically by estimating the variance of the EN4.2.1 ocean temperature and salinity data. In this study, ***R*** is assumed to be diagonal. The characteristics of ***R*** directly influence the weighting of observations in the assimilation process: larger values of ***R*** result in smaller weights for observations, whereas smaller values increase the weight of observations.

We have added this clarification (L245-248) to explain the computation of ***R***, its diagonal assumption, and its implications for the assimilation process.

L245-248: In this study, ***R*** is assumed to be diagonal and its diagonal elements are statistically computed based on the variance of the EN4.2.1 ocean temperature and salinity data. The characteristics of ***R*** directly influence the assimilation process, where larger values reduce the relative weight of the EN4.2.1 reanalysis and smaller values increase it.

Comment#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.

Response:

Thank you for your suggestion. We have revised the manuscript to specify that positive percentages represent improvements due to the assimilation and to present the reduction rate of the cost function as positive percentages (L256-261).

L256-261: As noted earlier, negative values of the reduction rate of the cost function indicate the successful incorporation of reanalysis data into the coupled model. However, the reduction rate is presented here as positive percentages to represent improvements due to the assimilation. The reduction rate of the cost function reaches 12.03% in the first month. Over the entire 72-year period from 1950 to 2021, the average monthly reduction rate of the cost function is 4.20% for all months in ASSIM.

Comment#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?

Response:

We have expanded on the two suggested reasons for the performance degradation in the deep ocean, providing additional context specific to the EN4.2.1 observational data and the E3SMv2 model. For the EN4.2.1 reanalysis, the coverage and quality of observations tend to decrease with depth, which may lead to higher uncertainties in the deep ocean. This sparse observational coverage limits the constraints that data assimilation can impose on the model state in the deep ocean. Furthermore, in the E3SMv2 model, the complexity of simulating deep-ocean processes, such as vertical mixing and bottom water formation, may contribute to biases that are difficult to correct through data assimilation.

In response to this comment, we have incorporated this discussion into the revised manuscript (L298-305) to clarify the potential reasons for the performance degradation in the deep ocean.

L298-305: The degradation in the deeper ocean layers can be attributed to two main factors: observational data limitations and challenges in representing deep-ocean processes in the model. For the EN4.2.1 reanalysis, the coverage and quality of observations tend to decrease with depth, potentially resulting in greater uncertainties in the deep ocean. This sparse observational coverage limits the constraints that assimilation can impose on the model state. Furthermore, in the E3SMv2 model, the complexity of simulating deep-ocean processes, such as vertical mixing and bottom water formation, may contribute to biases that are difficult to correct through assimilation.

Comment#7:

L268-269 Is the seasonal cycle removed from the time series before computing the correlations?

Please specify. And is the linear trend removed?

Response:
Yes, we have removed the seasonal cycle and linear trend before computing the correlations. This clarification has been added to the revised manuscript in Lines 309-310.

L309-310: The seasonal cycle and linear trend have been removed before computing the correlations.

Comment#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?

Response:
Thank you for your insightful comment. Sparse temperature and salinity observations may introduce higher uncertainty into the assimilation process. In regions with sparse observations, the assimilation process may introduce biases or errors when attempting to fit the model to incomplete or uncertain data, leading to localized performance degradation relative to the control run. Additionally, possible imbalances between ocean state variables during the assimilation process may also degrade the assimilation performance in certain areas.

In response to this comment, we have revised this sentence (L319-321) to include more specific factors contributing to the diminished performance.

L319-321: However, certain areas exhibit diminished performance, possibly due to sparse observational coverage introducing higher uncertainty into the assimilation process or imbalances between ocean state variables during the assimilation (Edwards et al., 2015; He et al., 2020b).

Comment#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.

Response:
Thank you for pointing out this inconsistency. We have corrected the caption of Figure 9 (L661-662) to accurately describe the global mean RMSE of vertically averaged ocean temperature and salinity.

L661-662: **Figure 9.** Temporal variations of the global mean bias (a, c) and RMSE (b, d) for ocean temperature (unit: °C) and salinity (unit: psu) averaged over the upper 1000 meters from 1950 to 2021.

Comment#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.

Response:

Based on your suggestion, we have updated panels (b) and (d) in Figure 10 to display "ASSIM minus OBS" instead of "ASSIM minus CTRL". In addition, we have revised the discussion accordingly (L373-376, and L380-383) to reflect the updated figure and results.

L373-376: In contrast, these SST biases found in CTRL are substantially reduced by ASSIM (Fig. 10b), with cold biases in the North Pacific and North Atlantic diminished by approximately 1-2 °C, and warm biases in the Southern Ocean corrected by about 1.5-2.5 °C.

L380-383: Compared with CTRL, ASSIM significantly reduces the overall fresh biases in CTRL (Fig. 10d). Notable improvements are observed in the North Atlantic and North Pacific, where salinity biases are reduced by approximately 0.5-1 psu, and in parts of the Southern Ocean, where reductions reach up to 1.5 psu.

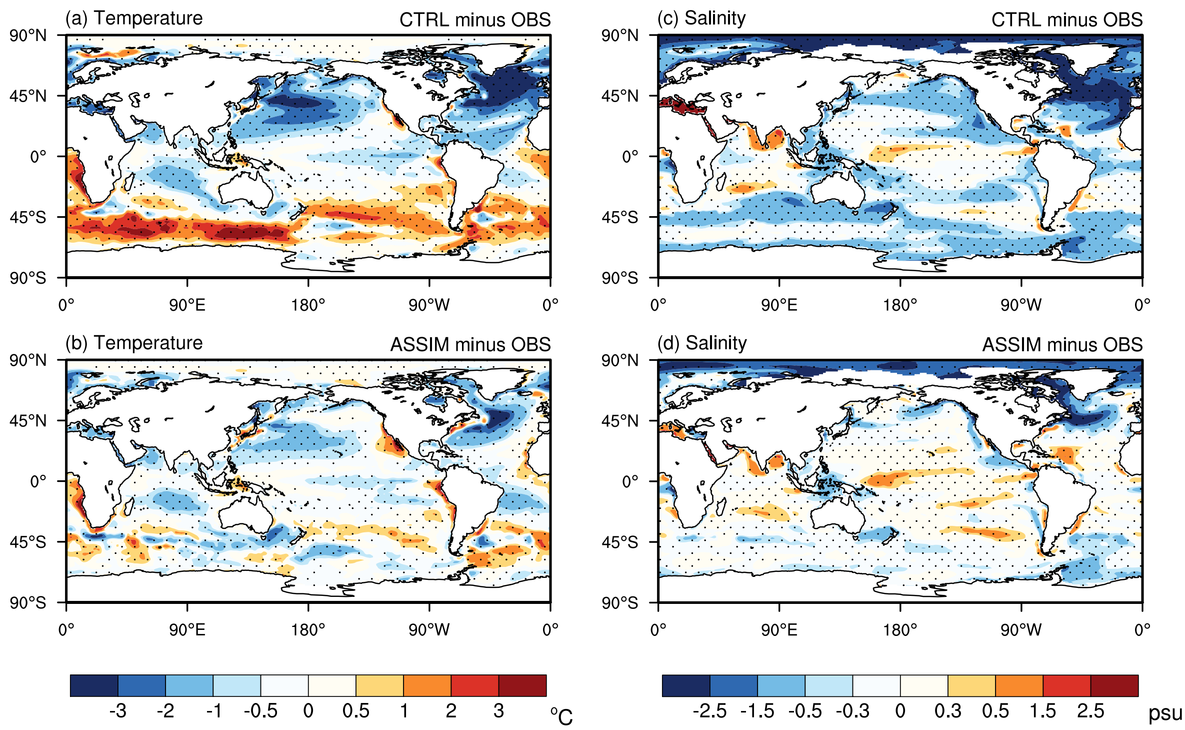

Figure 10. Climatological mean differences in sea surface temperature (left, unit: °C) and

salinity (right, unit: psu) from 1950 to 2021. The top panels show the differences between CTRL and observation, while the bottom panels show the differences between ASSIM and observation. Dotted areas indicate regions where the differences are statistically significant at the 95% confidence level.

Comment#11:

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

Response:

We have revised this sentence (L369-370) to show the climatological mean differences between CTRL and observation, as well as between ASSIM and observation.

L369-370: Figure 10 presents the climatological mean differences between CTRL and observation, as well as between ASSIM and observation, for both sea surface temperature (SST) and salinity (SSS).

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.

Response:

We have revised Figure 10 to replace "ASSIM minus CTRL" with "ASSIM minus OBS" in panels (b) and (d). Additionally, we have incorporated quantitative descriptions (L373-376) to better illustrate the magnitude of bias reduction.

L373-376: In contrast, these SST biases found in CTRL are substantially reduced by ASSIM (Fig. 10b), with cold biases in the North Pacific and North Atlantic diminished by approximately 1-2 °C, and warm biases in the Southern Ocean corrected by about 1.5-2.5 °C.

Comment#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.

Response:

Thank you for your valuable suggestion. We have extended our analysis to include the correlation and RMSE maps for temperature (Fig. R1) and precipitation (Fig. R2) during boreal winter, when ENSO exerts its strongest influence on US climate variability. For winter temperature (Fig. R1), correlation improvements are observed across the northern and central US, while reductions in RMSE are primarily concentrated in the central and northeastern US. For winter precipitation (Fig. R2), correlation improvements are evident over the central and southern US, with RMSE reductions prominently observed in the southern US.

However, based on Reviewer #1's comment that the current results in Figure 11 are highly preliminary and require more rigorous analysis to draw robust conclusions, we have followed Reviewer#1's suggestion to remove Figure 11 from the revised manuscript. We sincerely appreciate your constructive feedback, which has been invaluable in refining our analysis. In future research, we plan to integrate Figure 11 and this analysis of correlation and RMSE maps into a separate study, with more comprehensive analyses to enhance the robustness and clarify our findings.

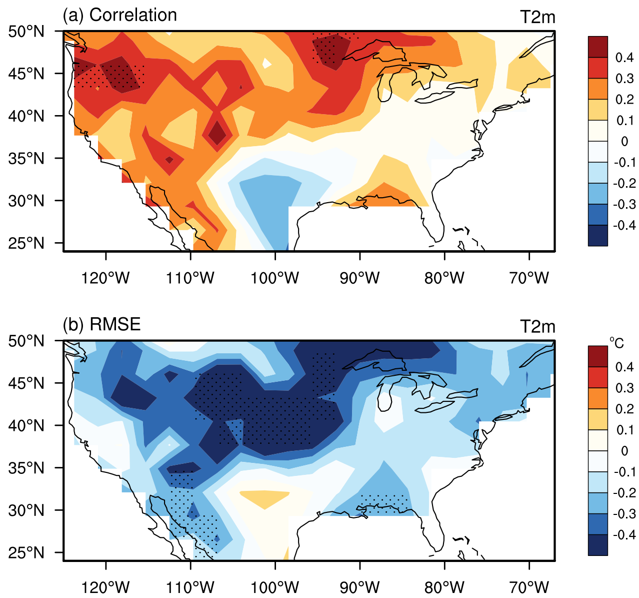

Figure R1. Spatial patterns of correlation and RMSE differences in surface air temperature during boreal winter between ASSIM and CTRL over the contiguous US from 1950 to 2021. Dotted areas represent statistical significance at the 90% confidence level.

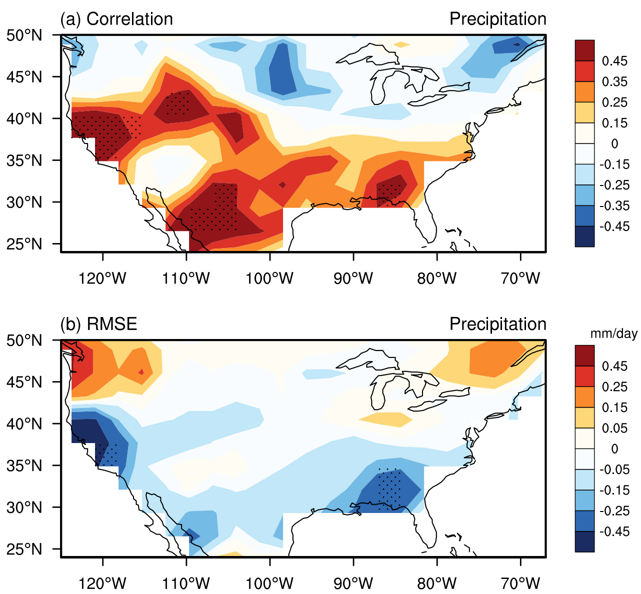

Figure R2. Similar to Figure R1 but for winter precipitation.

Comment#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.

Response:

We have added a new Figure A3 to show the time series of the winter Niño 3.4 index for CTRL, ASSIM, and OBS. The correlation coefficient of the winter Niño 3.4 index with observation increases from 0.06 in CTRL to 0.79 in ASSIM, highlighting the enhanced representation of ENSO variability in the assimilation run.

In response to this comment, we have included this analysis as Figure A3 and added the corresponding results (L315-317) in the revised manuscript to underscore the enhanced SST variability in the tropical Pacific.

L315-317: Further analysis of the winter Niño 3.4 index (Fig. A3) confirms that the assimilation improves the representation of ENSO variability, with the correlation coefficient increasing from 0.06 in CTRL to 0.79 in ASSIM.

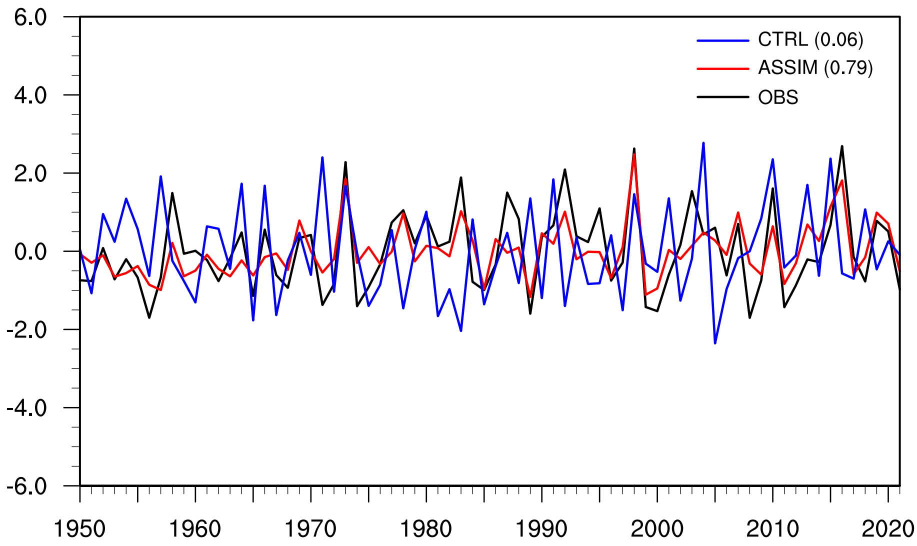

Figure A3. Time series of the winter Niño 3.4 index from 1950 to 2021 for the observation (black line), ASSIM (red line), and CTRL (blue line). The correlation of the Niño 3.4 index with the observation in ASSIM and CTRL are also shown.

Comment#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".

Response:
Thank you for your insightful comment. We agree that the sparse observational coverage in EN4.2.1, particularly for salinity, could pose limitations to the assimilation process, potentially introducing static instabilities in the absence of dynamical constraints. Reanalysis products such as ORAS5 and GLORYS offer promising alternatives for mitigating these limitations. Future efforts should explore incorporating these reanalysis products into the WCODA system to improve the assimilation performance in challenging areas.

In response to this comment, we have expanded the discussion (L410-415) in the concluding remarks to include the limitations of using the EN4.2.1 dataset and the potential benefits of employing alternative reanalysis products like ORAS5 and GLORYS.

L410-415: The reliance on the EN4.2.1 product could pose limitations to the assimilation process due to the sparse salinity observations and potential for static instabilities in data-sparse regions. Reanalysis products such as ORAS5 and GLORYS provide promising alternatives for mitigating these limitations. Future efforts should explore incorporating these reanalysis products into the WCODA system to improve the assimilation performance in challenging areas.

Comment#16:
L109 Change "employs sophisticated representations of" with "represents"

Response:
Done.

Comment#17:
L292 Change "variations in" to "of"

Response:
Done.

Comment#18:
L306 Add "over the top 1000 meters"

Response:
Done.

Comment#19:
L341 What “multiple US regions”? The contiguous US?

Response:
To clarify, "multiple US regions" refers to areas within the contiguous US. We have removed this term in the revised manuscript.

Comment#20:
L633 Change “in the” to “averaged over”.

Response:
Done.

Comment#21:
L382 Consider changing “challenges” with “limitations”

Response:
Done.

References:

Edwards, C. A., Moore, A. M., Hoteit, I., and Cornuelle, B. D.: Regional ocean data assimilation, Annual Review of Marine Science, 7(1), 21–42, https://doi.org/10.1146/annurev-marine-010814-015821, 2015.

He, Y., Wang, B., Huang, W., Xu, S., Wang, Y., Liu, L., Li, L., Liu, J., Yu, Y., Lin, Y., Huang, X., and Peng, Y.: A new DRP-4DVar-based coupled data assimilation system for decadal predictions using a fast online localization technique, Climate Dynamics, 54, 3541–3559, https://doi.org/10.1007/s00382-020-05190-w, 2020a.

He, Y., Wang, B., Liu, L., Huang, W., Xu, S., Liu, J., Wang, Y., Li, L., Huang, X., Peng, Y., Lin, Y., and Yu, Y.: A DRP-4DVar-based coupled data assimilation system with a simplified off-line localization technique for decadal predictions, Journal of Advances in Modeling Earth Systems, 12(4), e2019MS001768, https://doi.org/10.1029/2019MS001768, 2020b.

Sugiura, N., Awaji, T., Masuda, S., Mochizuki, T., Toyoda, T., Miyama, T., Igarashi, H. and Ishikawa, Y.: Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations, Journal of Geophysical Research: Oceans, 113, C10017, https://doi.org/10.1029/2008JC004741, 2008.

Zhang, S., Liu, Z., Zhang, X., Wu, X., Han, G., Zhao, Y., Yu, X., Liu, C., Liu, Y., Wu, S., Lu, F., Li, M., Deng, X.: Coupled data assimilation and parameter estimation in coupled ocean–atmosphere models: a review, Climate Dynamics, 54, 5127–5144, https://doi.org/10.1007/s00382-020-05275-6, 2020.