We thank Reviewer #1 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 #1

General Comment:

The manuscript describes the implementation of a new weakly coupled ocean data assimilation in the E3SMv2. The authors show that using this assimilation method helps reduce temperature and salinity bias and RMSE in general in both surface and deeper ocean layers. The results of the manuscript are promising and will be valuable for the community. It is overall well-written and easy to follow. That being said I believe that more work must be done on this manuscript before being published. A lot of things need to be clarified or added to improve the clarity and robustness of the results.

As a general comment here, before going into details, the authors need to be more quantitative, in the text and in the abstract, when describing results in addition to saying increase/decrease or cold/warm biases please indicate some values. Also, units of temperature and salinity are missing on every figure, please add units in both figures and captions as otherwise, it makes things hard to understand, i.e. figure 10 see comment below.

One other main concern here is about section 3.6 on the influence of ocean data assimilation on the regional climate over land. Although this is an interesting topic, it is most likely a manuscript in itself. The results presented in this section are highly preliminary, no details on observations (definitely not enough to only briefly mention it in the Figure 11's caption) or methodology used are given. Very little can be said with certainty on the influence of ocean data assimilation on regional climate over land with only the figure and analysis presented (fig 11). This needs a much more rigorous analysis to be able to draw robust conclusions. In my opinion, this section should be removed and can't be published as is in this manuscript.

Response:

Thank you very much for taking the time to review our manuscript and providing us with very useful comments. We have revised the abstract (L28-31), results (L373-376, L379-380, and L381-383), and conclusion (L404-408) sections to include numerical values that quantify improvements in temperature and salinity biases. We have also added units to all relevant figures and captions (Figs 3-5, 8-9) and explicitly updated Fig. 10 to include missing units for better clarity.

After careful consideration, we have decided to remove section 3.6 and Figure 11 from the manuscript. We agree that the current results are preliminary and require further analysis in future work. The removal of this section 3.6 has been reflected in the revised manuscript, and all references to this section have been updated accordingly.

L28-31: For sea surface temperature, cold biases in the North Pacific and North Atlantic are diminished by about 1-2 °C, and warm biases in the Southern Ocean are corrected by approximately 1.5-2.5 °C. In terms of salinity, improvements are observed with bias reductions of about 0.5-1 psu in the North Atlantic and North Pacific and up to 1.5 psu in parts of the Southern Ocean.

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.

L379-380: Notably, in the Mediterranean Sea, CTRL exhibits a large positive salinity bias exceeding 2.5 psu.

L381-383: 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.

L404-408: ASSIM substantially mitigates model biases for SST and SSS observed in CTRL, particularly reducing cold biases in the North Pacific and North Atlantic by approximately 1-2 °C, correcting warm biases in the Southern Ocean by about 1.5-2.5 °C, and significantly increasing salinity estimates to reduce the model fresh biases by approximately 0.5-1 psu in the North Atlantic and North Pacific, and up to 1.5 psu in parts of the Southern Ocean.

Comment#1:

Line 114-116: As the manuscript focuses on ocean, more info is needed on the ocean model set-up here. Does the ocean model has the same horizontal resolution as the atmosphere? how many vertical layers, vertical resolution? What is the vertical mixing scheme used?

Response:

Thank you for your constructive comment to clarify the ocean model configuration in our manuscript. The ocean model in E3SMv2 employs the Model for Prediction Across Scales-Ocean (MPAS-O), which has a horizontal resolution of ~60 km in the midlatitudes and ~30 km at the equator and poles, differing from the atmospheric model's resolution of 110 km. MPAS-O is configured with 60 vertical layers, with finer resolution (~10 m) near the surface and coarser resolution (~200 m) at depth. The vertical mixing scheme employed is the K-profile parameterization, as described by Van Roekel et al. (2018).

Based on your suggestion, we have revised the manuscript to provide a more detailed description of the ocean model configuration (L120-124).

L120-124: MPAS-O operates at a horizontal resolution of ~60 km in the midlatitudes and ~30 km at the equator and poles, differing from the atmospheric model's resolution of 110 km. It is configured with 60 vertical layers, with finer resolution (~10 m) near the surface and coarser resolution (~200 m) at depth. The vertical mixing scheme employed is the K-profile parameterization, as described by Van Roekel et al. (2018).

Comment#2:

On section 2.2: Just by reading this, for an external reader, it is not very clear what is this product (EN4.2.1) and thus what is assimilated. Those are profiles with spatiotemporal variability (right?), what are the typical depths where observations are available? Typically, what are the regions where we have a lot or few observations, and thus can/should we expect improvement in these areas or not? In summary, a better description of this product here would help us understand better the results presented after.

Response:

Thank you for pointing out the need for more clarity regarding the EN4.2.1 dataset. The EN4.2.1 product is developed based on quality-controlled ocean temperature and salinity profiles from four input sources: Argo, ASBO (Arctic Synoptic Basin Wide Oceanography), GTSPP (Global Temperature and Salinity Profile Program), and WOD09 (World Ocean Database). The EN4.2.1 dataset includes observations from a wide range of profiling instruments, such as Argo floats, expendable bathythermographs (XBTs), and mechanical bathythermographs (MBTs) (Chen et al., 2020).

According to Good et al. (2013), observations in EN4.2.1 are most abundant in the upper 100 meters, with vertical resolution refined to ~1 m in the top 100 m. Spatially, data density is high in regions such as the North Atlantic and western Pacific but decreases significantly in high-latitude and deep ocean regions. Areas with higher observational density, such as the upper North Atlantic, are expected to show greater improvement in the assimilation results, while regions with sparse observations may exhibit limited improvements.

Following your advice, we have revised the manuscript to include a more detailed description of the EN4.2.1 dataset, emphasizing the spatiotemporal distribution of observations and their implications for data assimilation (L135-146).

L135-146: Produced by the Met Office Hadley Centre, the EN4.2.1 product is developed based on quality-controlled ocean temperature and salinity profiles from four input sources: Argo, ASBO (Arctic Synoptic Basin Wide Oceanography), GTSPP (Global Temperature and Salinity Profile Program), and WOD09 (World Ocean Database) (Good et al., 2013). The EN4.2.1 dataset includes observations from a wide range of profiling instruments, such as Argo floats, expendable bathythermographs (XBTs), and mechanical bathythermographs (MBTs) (Chen et al., 2020). According to Good et al. (2013), observations in EN4.2.1 are most abundant in the upper 100 meters, with vertical resolution refined to ~1 m in the top 100 m. Spatially, data density is high in regions such as the North Atlantic and western Pacific but decreases significantly in high-latitude and deep ocean regions. This distribution in data availability influences the assimilation results. Areas with denser observational coverage, such as the upper North Atlantic, are expected to show greater improvements through assimilation, while regions with sparse observations may exhibit limited improvements.

Comment#3:

Line 150: “significantly reducing”: by how much the computational resources are reduced?

Response:

To clarify, the computational efficiency of the DRP-4DVar system used in this study has been demonstrated in previous studies. Based on Zhu et al. (2022), the DRP-4DVar system reduces computational time by approximately 50% compared to traditional 4DVar systems. For example, in a 6-hour assimilation window, the 4DVar system required 25 minutes, while the DRP-4DVar system only took 13 minutes due to its efficient computation.

In response to this comment, we have revised this sentence (L162-163) for clarification in the revised manuscript.

L162-163: Zhu et al. (2022) demonstrated that the DRP-4DVar method significantly reduces computational time by approximately 50% compared to traditional 4DVar systems.

Comment#4:

Line 207: what do you mean by “observed external forcing”? what is this?

Response:

The “observed external forcing” refers to external factors influencing the climate system, such as solar radiation and greenhouse gas and aerosol concentrations. These transient historical external forcings are prescribed following the CMIP6 protocol (Eyring et al., 2016). These observed external forcings directly influence the atmospheric component of the model and subsequently influence other components (e.g., land and ocean) through their coupling with the atmosphere.

We have revised this sentence (L218-222) to include a more precise explanation of “observed external forcing” for better clarity.

L218-222: driven exclusively by observed external forcings (e.g., solar radiation and greenhouse gas and aerosol concentrations). The observed external forcings, prescribed according to the CMIP6 protocol (Eyring et al., 2016), directly influence the atmospheric component and subsequently affect other components (e.g., land and ocean) through their coupling with the atmosphere.

Comment#5:

Line 211: “across sixty ocean layers”. Is this all the model ocean layers? if not what’s the depth of the 60 layer and why only 60. Please clarify this.

Response:

Yes, the referenced "sixty ocean layers" corresponds to all the vertical layers in the ocean component of the E3SMv2 model. We have revised this sentence (L226-227) for better clarity.

L226-227: across all sixty ocean layers spanning the entire ocean depth

Comment#6:

Line 237-238: what is the reason for the initial jump in the cost function reduction from -12% to -4% ?

Response:

The observed initial jump from -12% to -4% can be attributed to the rapid adjustments of the model during the first two years of assimilation. At the start of the assimilation cycle, the model state undergoes rapid adjustments to align with the reanalysis data, leading to a sharp reduction in the cost function. As the assimilation progresses, subsequent iterations refine these adjustments, resulting in a slower rate of reduction

To clarify this point, we have provided a brief explanation of this phenomenon (L263-266).

L263-266: The initial sharp reduction rate of the cost function reflects the rapid adjustments made by the model to align with the reanalysis data. As the assimilation progresses, subsequent iterations refine these adjustments, resulting in a slower rate of reduction.

Comment#7:

Line 246: “nine ocean layers”: why did you choose to show specifically these 9 layers? Is 85m the last layer where observations are assimilated ? if not why not showing anything below? This need to be clarified.

Moreover, is it necessary to show these nine as they relatively show the same results, maybe you could only show 5m, 45 and 85m ..? Reading further, you’re showing profiles up to 1000m so why not showing maps of the deeper ocean here as well? 85m is not ‘deep ocean’, depending on region and/or season this is still in the ocean mixed layer.

On that note, it could be interesting to show seasonal maps as well. Is there any seasonal variability on these results, i.e. if maybe there are fewer observations during winter months does the ASSIM still perform better?

Response:

Thank you for your insightful comments. The original nine layers were simply chosen as the first nine layers of the ocean model, ranging from 5 m to 85 m. However, we recognize that this approach limited the depth representation and did not fully capture the assimilation impacts across deeper layers. Based on your suggestion, we have revised the depth of the nine layers to span a broader depth range. The updated layers are at depths of 5 m, 45 m, 85 m, 135 m, 327 m, 528 m, 708 m, 879 m, and 1106 m. We have revised Figure 4 to include the nine newly selected ocean layers from 5 m to 1106 m and updated the description of assimilation performance (L285-287) in the deeper ocean layers (879 m and 1106 m).

L285-287: In the deeper layers, the assimilation still shows notable improvements in regions such as the North Pacific and parts of the Southern Ocean, though with more pronounced degradation observed in the equatorial Atlantic and parts of the Indian Ocean.

We have also added supplementary figures (Figs. A1 & A2) and the corresponding sentences (L283-285) to show the seasonal performance of RMSE differences for summer (Fig. A1) and winter (Fig. A2) across the same nine depths. In the upper ocean layers, we observe better RMSE performance during winter compared to summer in some regions, such as the tropical Pacific.

L283-285: In the upper ocean layers, RMSE performance is better during winter compared to summer in some regions, such as the tropical Pacific (Figs. A1 & A2).

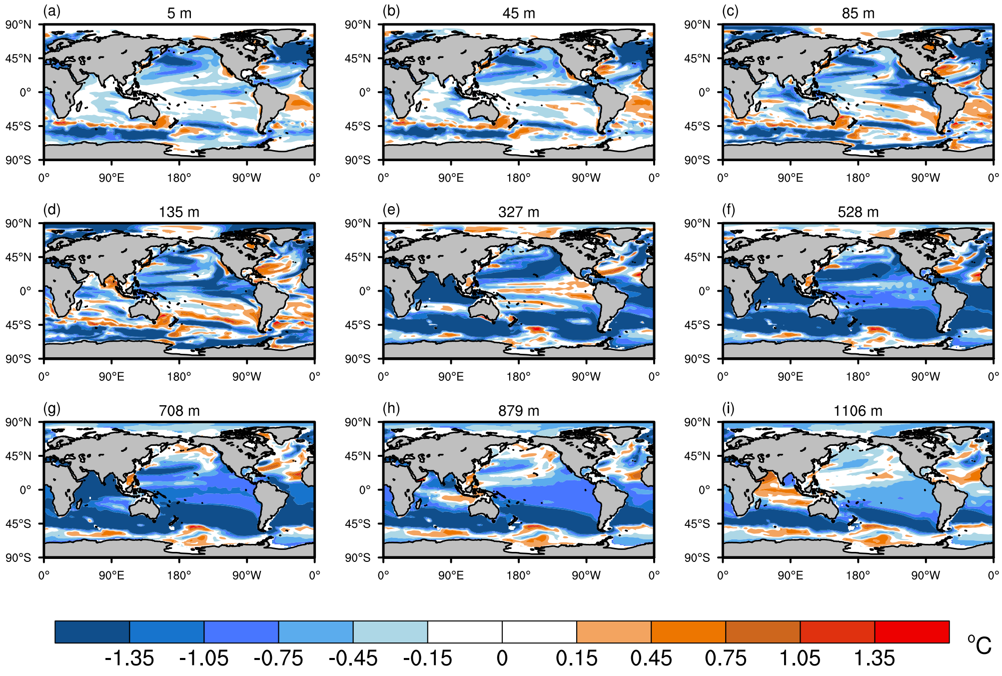


Figure 4. Spatial patterns of root mean square error (RMSE) differences in ocean temperature (unit: °C) between ASSIM and CTRL across nine ocean layers from 1950 to 2021. The RMSE differences are shown for nine different ocean depths: (a) 5 m, (b) 45 m, (c) 85 m, (d) 135 m, (e) 327 m, (f) 528 m, (g) 708 m, (h) 879 m, and (i) 1106 m.

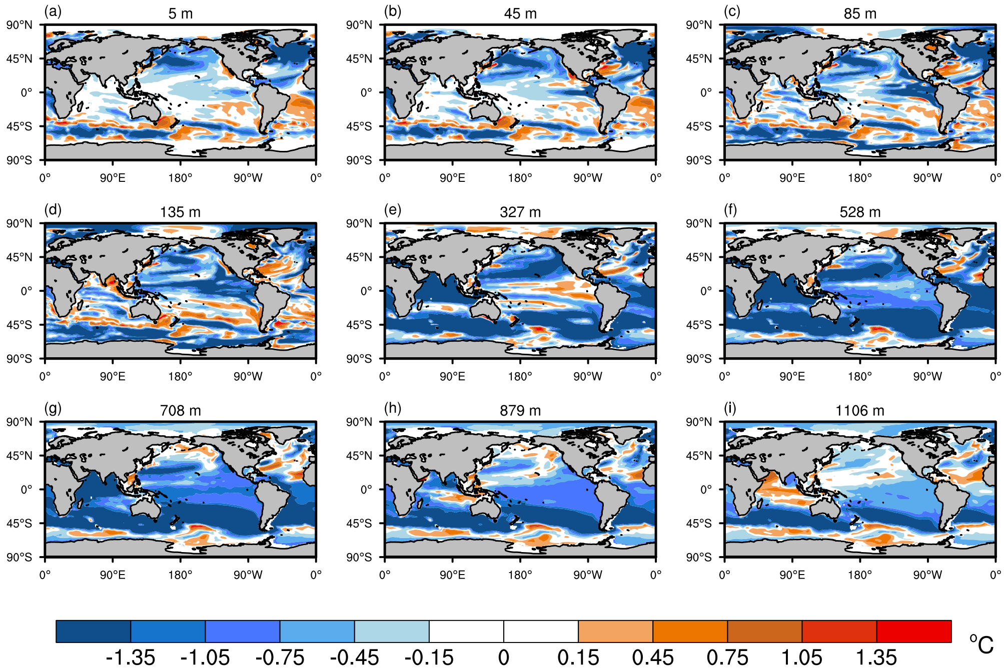


Figure A1. Similar to Figure 4 but during summer.

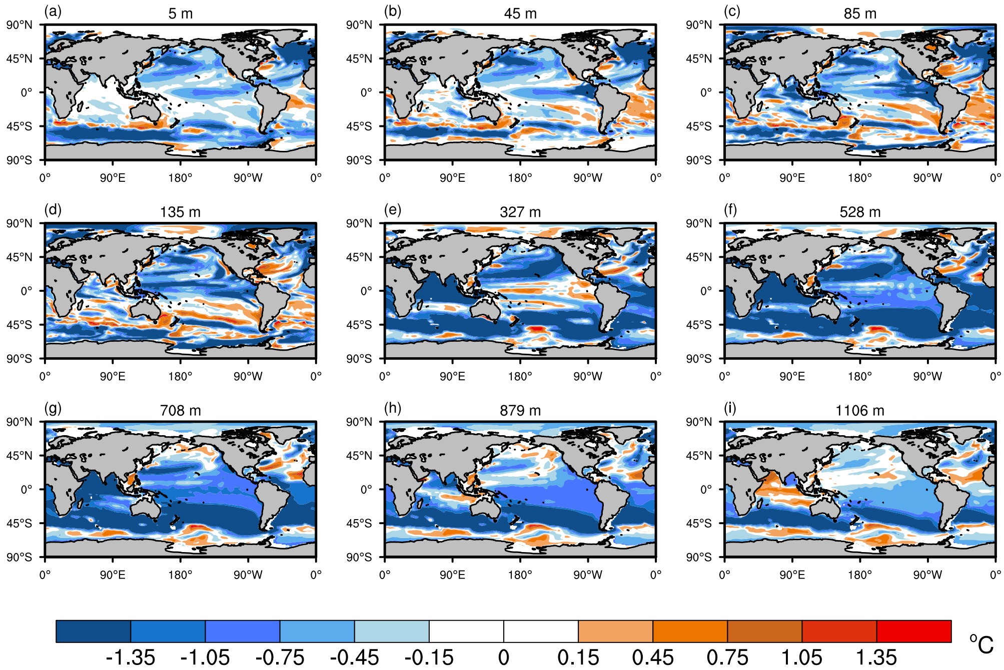


Figure A2. Similar to Figure 4 but during winter.

Comment#8:

On Figure 4 : Is there any ocean current-related impact? As it seems that the RMSE is somewhat increased near strong ocean currents (i.e. Gulf Stream separation, ACC from Agulhas to East Australian current with some sort of dipole increased/decreased RMSE)? and/or near upwelling regions California coast, Northwest Africa? any reasons on why the assimilation would not do better there? As in Figure 6 and 7 statistical significance should be added on both Figure 4 and

5, this will make the result more robust and these increased RMSE might not turn out to be significant (?). Also please indicate how the significance is calculated.

Response:

Thank you for highlighting the potential impact of ocean currents and upwelling regions on RMSE differences in Figure 4. We agree that the assimilation results show degradation near strong ocean currents and upwelling regions. These regions are characterized by strong horizontal gradients and mesoscale variability, which are not well captured by MPAS-O at relatively low resolution and hence pose challenges for the assimilation system and likely contribute to increased RMSE.

We have revised the manuscript to include additional discussion (L279-283) to highlight the impact of strong ocean currents and upwelling regions on assimilation performance.

L279-283: However, increased RMSE values are observed near strong ocean currents and upwelling regions, such as the Gulf Stream, Agulhas Current, and the California coast. These regions are characterized by strong horizontal gradients and mesoscale variability, which are not well captured by MPAS-O at relatively coarse resolution and hence present challenges for the assimilation system and likely contribute to diminished performance.

In response to your suggestions, we have applied the paired t-test to assess the statistical significance of the RMSE differences in Figures 4 and 5. Specifically, the mean and standard deviation of the square error differences were calculated at each grid point. The t-statistic was computed as the ratio of the mean difference to the standard error to determine statistical significance. A confidence level of 95% ($p < 0.05$) was used to identify regions where the RMSE differences are statistically significant. Our results show that some regions of increased RMSE also pass the significance test.

We have updated Figures 4 and 5 to use dotted regions to illustrate statistical significance. Regions where the RMSE differences are statistically significant at the 95% confidence level are now marked with dots in the revised figures.

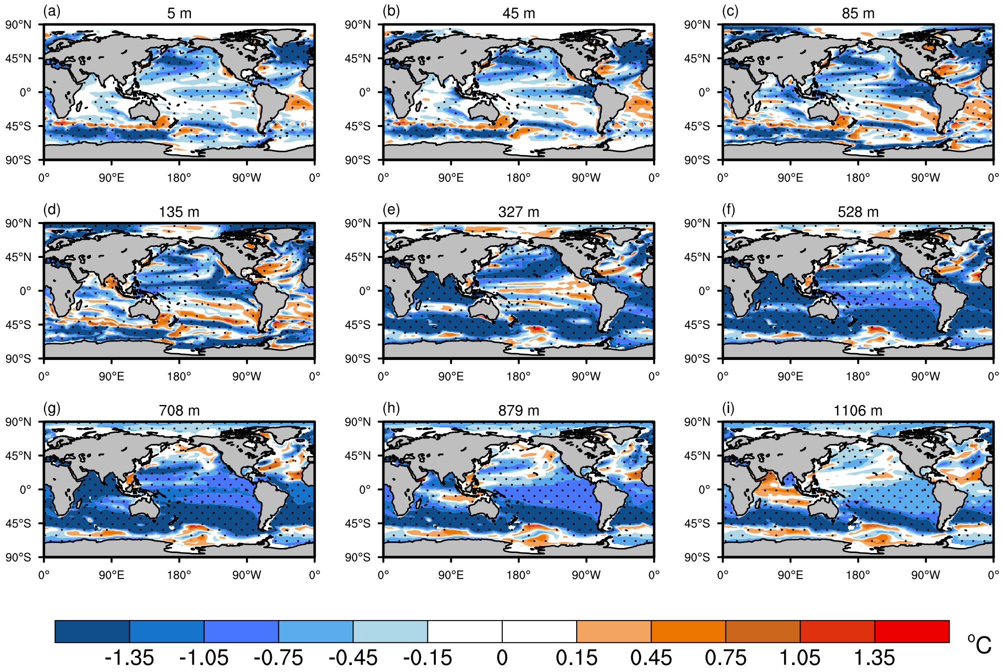


Figure 4. Spatial patterns of root mean square error (RMSE) differences in ocean temperature (unit: °C) between ASSIM and CTRL across nine ocean layers from 1950 to 2021. The RMSE differences are shown for nine different ocean depths: (a) 5 m, (b) 45 m, (c) 85 m, (d) 135 m, (e) 327 m, (f) 528 m, (g) 708 m, (h) 879 m, and (i) 1106 m. Dotted areas represent statistical significance at the 95% confidence level.

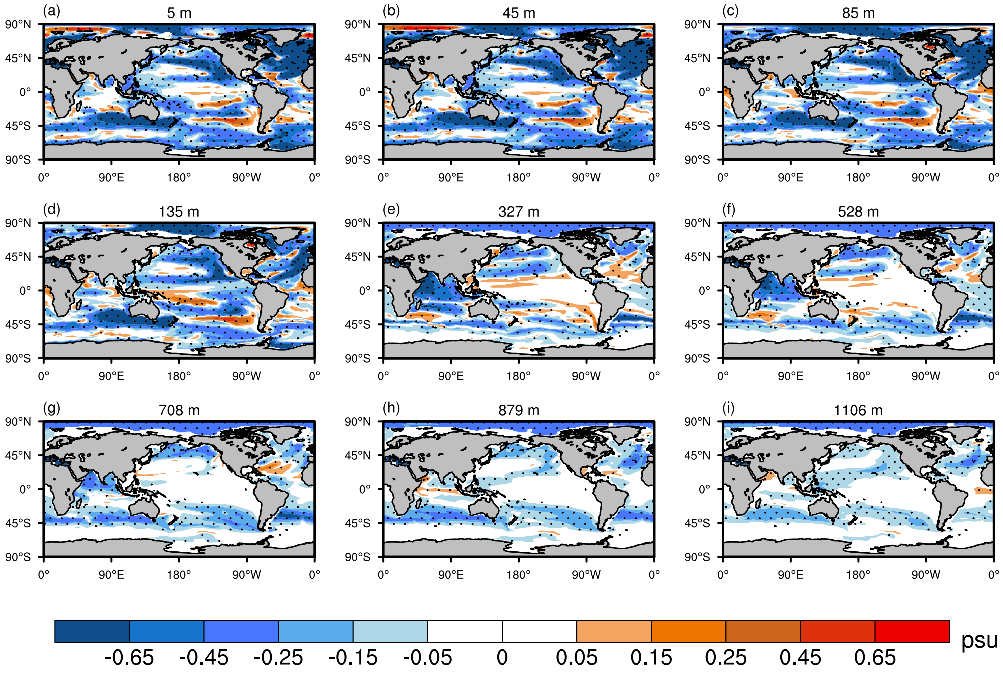


Figure 5. Similar to Figure 4 but for ocean salinity (unit: psu).

Comment#9:

Line 256: Again, 85m is not "deep-ocean dynamics"

Response:

Yes, we have included deeper ocean layers at 879 m and 1106 m in the updated Figure 4, and replaced the term "deep-ocean dynamics" with "deeper ocean processes" (L289).

L289: deeper ocean processes

Comment#10:

Line 274-276: It seems a good hypothesis with maybe hints of this in Figure4 as well showing reduced RMSE on temperature in this region. Since you have both the observations and simulations it would be interesting to confirm and show this, if assimilation actually helps representing El Nino/La Nina better. Otherwise one can wonder why the correlation is largely increase in the Pacific compare to, i.e, the Atlantic.

Response:

Thank you for your insightful comment. We have conducted an additional analysis of the Niño 3.4 index and included a new supplementary figure (Figure A3). Figure A3 compares the time series of the winter Niño 3.4 index from the observation, ASSIM, and CTRL. Our analysis shows that the correlation coefficient between the time series of the winter Niño 3.4 index and observation is improved from 0.06 in CTRL to 0.79 in ASSIM, confirming that the assimilation enhances the representation of El Niño/La Niña variability.

We have included this new Figure A3 and the associated results (L315-317) in the revised manuscript to support the better representation of El Niño/La Niña variability from ASSIM.

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.

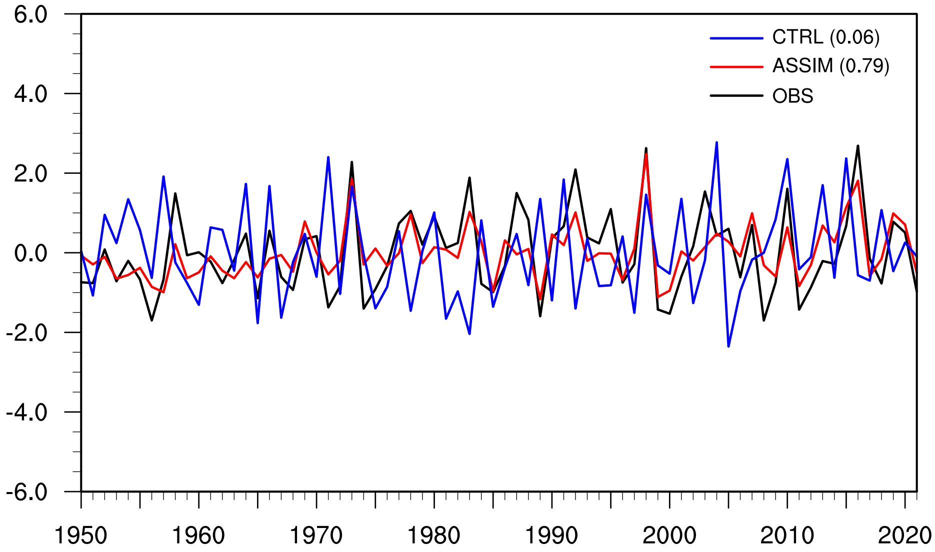


Figure A3. Time series of the winter Niño 3.4 index from 1950 to 2021 for the observation (black line), ASSIM (red line), and CTRL (blue line). The correlations of the Niño 3.4 index with the observation in ASSIM and CTRL are also shown.

Comment#11:

Line 276: “considerable”: considering the values in the Indian Ocean this vocabulary might be too strong.. otherwise the Southern Ocean have similar magnitude, is that also “considerable improvements”..?

Response:

We agree that the term "considerable" might overstate the magnitude of the improvements in the Indian Ocean. We have revised this sentence to remove the mention of the Indian Ocean and replaced "considerable" with "noticeable" (L317).

L317: Moreover, parts of the North Pacific also exhibit noticeable improvements.

Comment#12:

Line 277-278: “complex ocean dynamics” This term was used before in the text, but this is a bit too generic to explain differences and especially to explain the diminished performance, you have to be more specific or don't use this. what is "complex ocean dynamics"? In that regard, then it means that the simulation without assimilation (CTRL) is doing better at these "complex ocean dynamics", then why? I believe here, from the stippling on the figure, it doesn't seem that these reduced correlations are statistically significant (whereas all the increased ones are) and thus, this might just be due to internal variability of the ocean model?

Response:

Thank you for your insightful suggestion. We agree that the term "complex ocean dynamics" is too generic and does not adequately explain the observed diminished performance. In response to your feedback, we have removed the term "complex ocean dynamics" and added more specific factors. Additionally, when deeper layers (e.g., 708 m, 879 m, and 1106 m) are included, we observe that some regions of reduced correlations also pass the significance test. For these areas, the diminished performance may result from 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., 2020).

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#13:

Line 286: Again, this doesn't seem to be statistically significant

Response:

Yes, we have revised this sentence (L327-328) to remove the mention of the Indian Ocean and focus only on regions with significant results.

L327-328: Noteworthy improvements are evident in the tropical Pacific, North Pacific, and parts of the North Atlantic.

Comment#14:

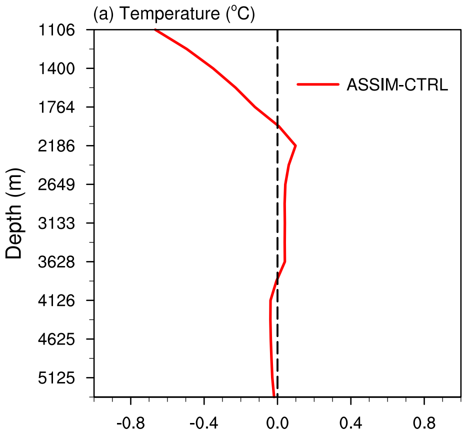
Figure 8: Related to previous comment on ocean model depth. Here, why stop the profiles at 1106m, nothing below? 1106m which is rather a very specific depth..? and thus why showing only up to 85 m on the previous panels, Figure 4 to 7? As previously mentioned, it would be interesting to see maps at deeper ocean levels as well as it seems here that the assimilation is improving significantly deeper in the ocean? Instead of showing 9 layers in the first 90m which are showing very similar results.

Response:

Thank you for pointing this out. We appreciate your suggestion to extend the depth range and improve the layer selection in the figures. The ocean model consists of sixty ocean layers from 5 m to 5375 m. The original profiles stopped at 1106 m because this depth corresponds to the 41st layer of the 60-layer ocean model, which is the closest to the upper 1000 meters of the ocean. Based on your feedback, we have added Figure A4 to extend the vertical profiles to cover all layers from 1106 m to the deepest layer at 5375 m. The results in Figure A4 show that the RMSE differences between ASSIM and CTRL become notably smaller for both ocean temperature and salinity, with temperature differences generally within 0.10 °C and salinity differences within 0.02 psu below 2000 m. This suggests the limited impact of assimilation in the deeper ocean.

We have included this new Figure A4 to extend the vertical profiles from 1106 m to 5375 m, and added the relevant results (L347-349) to the revised manuscript.

L347-349: The extended profiles in Figure A4 indicate that below 1106 meters, the RMSE differences between ASSIM and CTRL gradually decrease for both ocean temperature and salinity, suggesting the limited impact of assimilation in the deeper layers.


(a) Temperature (°C)

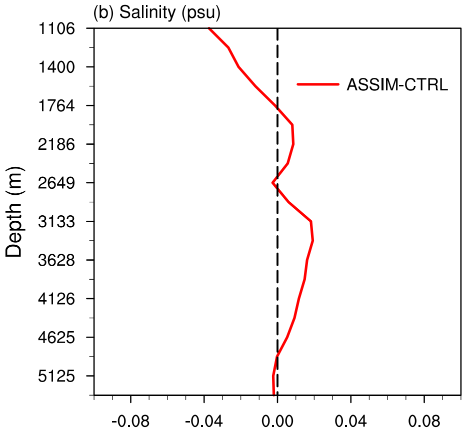

(b) Salinity (psu)

Figure A4. Vertical profiles of the globally averaged RMSE differences between ASSIM and CTRL for (a) ocean temperature (unit: °C) and (b) ocean salinity (unit: psu) with depths from 1106 m to 5375 m.

In response to this comment, we have revised Figures 4 to 7 to include nine different ocean layers: 5 m, 45 m, 85 m, 135 m, 327 m, 528 m, 708 m, 879 m, and 1106 m. Additionally, we have updated the corresponding descriptions of the assimilation results in the deeper layers (879 m and 1106 m) for Figure 4 (L285-287), Figure 5 (L295-298), Figure 6 (L317-318) and Figure 7 (L328-331).

L285-287: In the deeper layers, the assimilation still shows notable improvements in regions such as the North Pacific and parts of the Southern Ocean, though with more pronounced degradation observed in the equatorial Atlantic and parts of the Indian Ocean.

L295-298: In the deeper layers, the improvements are less extensive but remain evident in regions such as parts of the North Atlantic and North Pacific. However, RMSE degradation becomes notable in the equatorial Atlantic and parts of the Indian Ocean, highlighting the need for further improvements in these regions.

L317-318: In the deeper layers, improvements are observed in the western Pacific and parts of the Southern Ocean.

L328-331: In the deeper layers, the improvements in correlation become more localized, primarily concentrated in the western Pacific and parts of the Southern Ocean. Meanwhile, reductions in correlations are observed in parts of the equatorial Pacific and the South Atlantic, indicating the need for further improvements.

Comment#15:

Line 299: "gradually decrease as depth increases" but it does increase again below 300m? Is it not significant ?

Response:

Thank you for your comment. We agree that the phrase "gradually decrease as depth increases" does not fully capture the observed trend. The RMSE differences for salinity gradually decrease from 155 meters to 305 meters, but a slight increase is observed between 305 meters and 1106 meters. We have revised this sentence (L343-345) to clarify this result.

L343-345: The RMSE differences gradually decrease as depth increases from 155 meters to 305 meters, but a slight increase is observed between 305 meters and 1106 meters.

Comment#16:

Line 306-307 and Figure 9: it could be useful to show the average over different depths not only 0-1000m. i.e. 0-300m, 0-700m, 0-1000m. From fig 8, doesn't this "systematic overestimation of temperature" come from 300m and below? if looked at the surface or over different depths it may not systematically overestimate the temperature? This would give more insight into what is improved or not depending on the depth, as in Fig8. Similar thing could be done for salinity.

Also, is there a spin-up period to take into account to analyze the result or when doing this kind of assimilation? it seems to take 10~15 years for the bias to approach the 0 line (Fig 9a,c). Maybe a comment on that would be useful.

Response:

Thank you for these insightful suggestions. We have included additional analyses showing the temporal variations of bias and RMSE averaged over different depth ranges: 0-300 m (Figure A5) and 0-700 m (Figure A6). The results in Figures A5 and A6 show that the systematic overestimation of temperature in CTRL primarily originates from depths below 300 meters, consistent with the vertical profiles shown in Figure 8. In contrast, the salinity bias in CTRL is already prominent in the upper 300 meters.

We have added Figures A5 and A6 to include these depth-specific averages and revised the corresponding text (L354-355 and L360) to reflect these findings.

L354-355: This overestimation in ocean temperature primarily originates from depths below 300 meters (Figs. A5 & A6).

L360: This salinity bias in CTRL is already prominent in the upper 300 meters (Figs. A5 & A6).

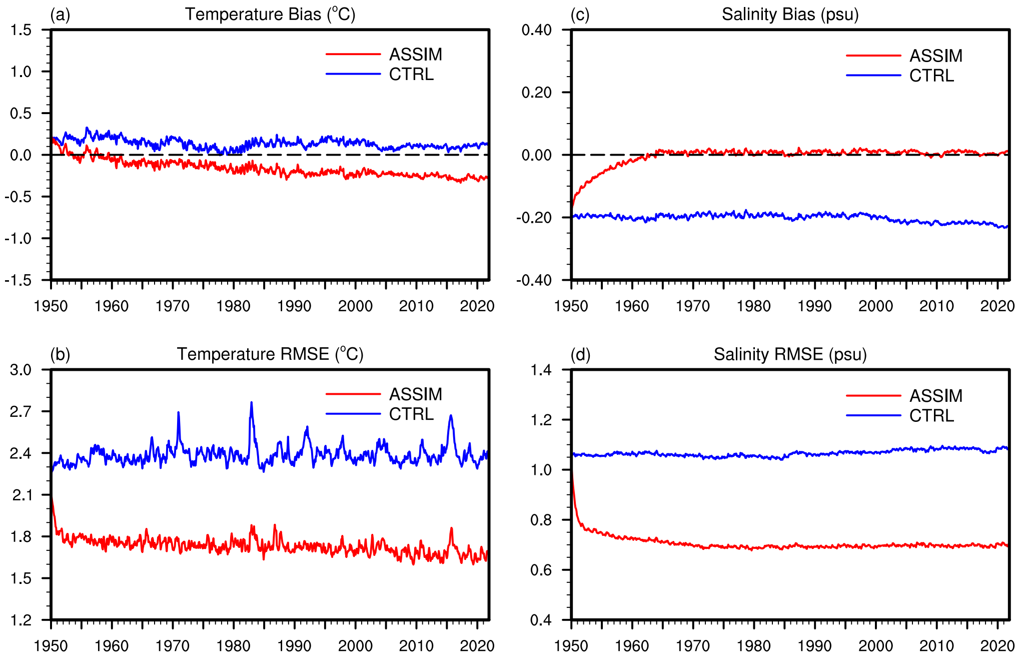


Figure A5. Temporal variations of the global mean bias (a, c) and RMSE (b, d) for ocean temperature (unit: °C) and salinity (unit: psu) averaged over the upper 300 meters. The red lines represent ASSIM, while the blue lines represent CTRL.

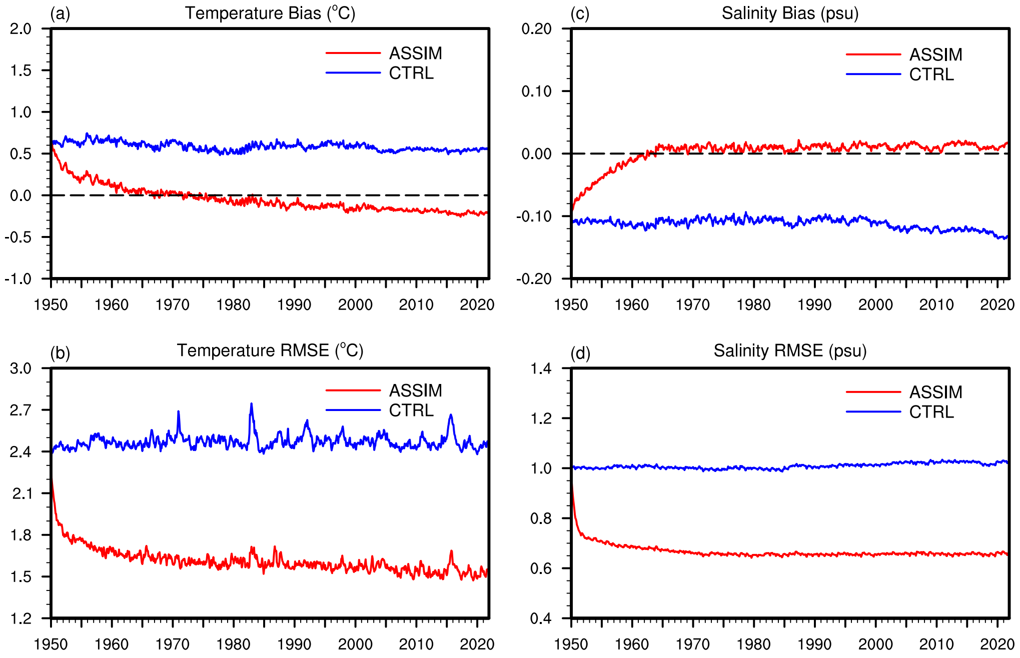


Figure A6. Similar to Figure A5 but averaged over the upper 700 meters.

Regarding the "spin-up" period for the assimilation system, we also acknowledge that it takes approximately 10-15 years for the biases in temperature and salinity to stabilize near the zero line. This behavior reflects the system's equilibration during the assimilation process, where the model takes time to adjust to the assimilated reanalysis.

We have included the following sentence (L363-365) to note that the first 10-15 years were considered as part of the adjustment period.

L363-365: Notably, it takes approximately 10-15 years for the biases in both temperature and salinity to stabilize near the zero line, reflecting an adjustment period where the assimilation system equilibrates.

Comment#17:

Figure 10: What are the units here? The colorbar is the same for both temperature and salinity? is it up to 3 Deg C difference and 3 psu difference or is it in %? This need to be clarified. if in % please quantify in the text how much in degC/psu.

Response:

The color bar in Figure 10 represents differences, not percentages. The units are °C for temperature, and psu for salinity, and the same color bar is used for both panels. To address this, we have added the units (°C for temperature and psu for salinity) to both the figure and its caption (L665-666) to ensure clarity.

L665-666: **Figure 10.** Climatological mean differences in sea surface temperature (left, unit: °C) and salinity (right, unit: psu) from 1950 to 2021.

Comment#18:

Figure 10b,d Should be "ASSIM minus Obs", it would then be much easier to appreciate the improvements in ASSIM in comparison to panels (a,c). Right now it is not clear or difficult to see actually how much the biases are reduced (or not) in ASSIM.

Response:

Thank you for your valuable suggestion. We have modified panels (b, d) in Figure 10 to show "ASSIM minus OBS" instead of "ASSIM minus CTRL". The figure caption has also been updated to reflect this change.

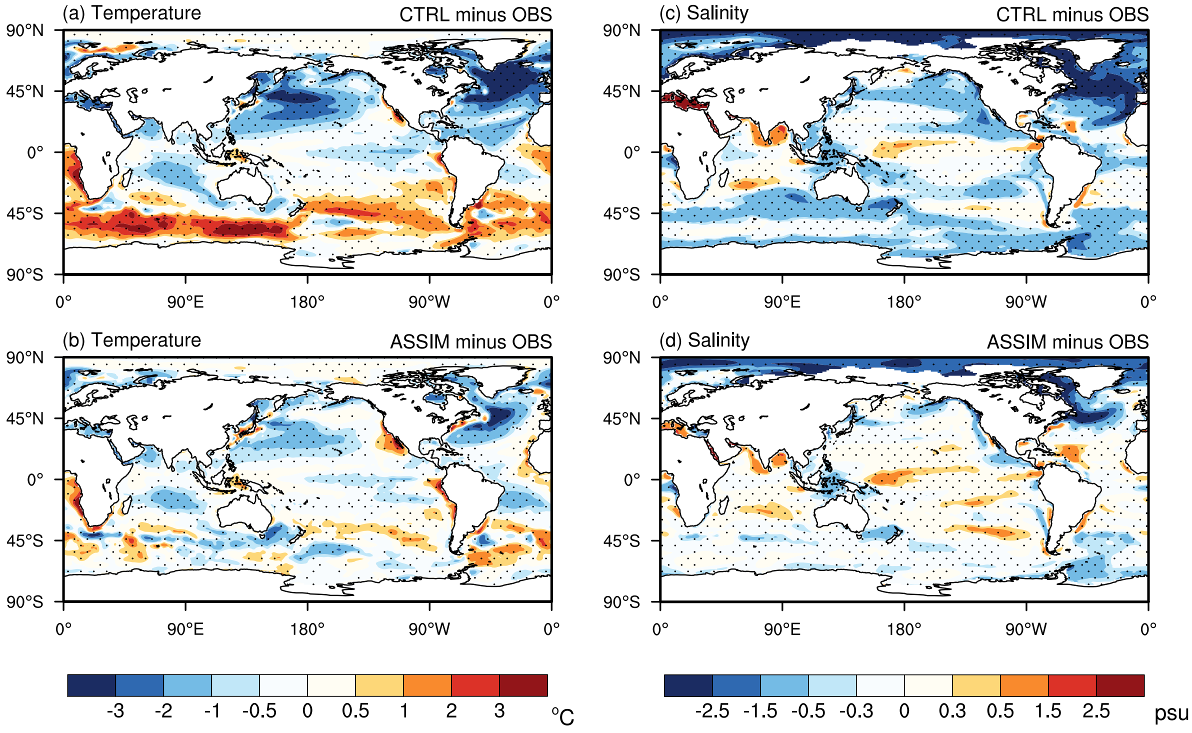


Figure 10. Climatological mean differences in sea surface temperature (left, unit: °C) and salinity (right, unit: psu) from 1950 to 2021. The top panels show the differences between CTRL and observation, while the bottom panels show the differences between ASSIM and observation. Dotted areas indicate regions where the differences are statistically significant at the 95% confidence level.

Comment#19:

Line 323-324: Again, quantify better how much.

Response:

We have revised the text (L373-376) to include specific numerical values to quantify the improvements. Specifically, in the North Pacific and North Atlantic, the cold biases in CTRL are reduced by approximately 1-2 °C. Similarly, in the Southern Ocean, the warm biases are corrected by about 1.5-2.5 °C.

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#20:

Line 325-326: could be worthwhile to note the large high bias in salinity in the Mediterranean Sea as the bias trend seems to be opposite to the global ocean, and it seems to be improved in the ASSIM as well.

Response:

Thank you for pointing this out. In response, we have revised the text (L379-380) to highlight the large positive salinity bias in the Mediterranean Sea from CTRL.

L379-380: Notably, in the Mediterranean Sea, CTRL exhibits a large positive salinity bias exceeding 2.5 psu.

Comment#21:

Line 378-380: In light of my previous comment, I believe this statement should be also removed.

Response:

We agree with your suggestion and have removed this statement from the manuscript.

References:

Chen, J., Liu, H., Bai, C., Yan, H., Lu, K., Bao, S., and Liu, K.: Identifying climate modes contributing to sea surface salinity decadal variation in the North Pacific Ocean, Journal of Geophysical Research: Oceans, 125(10), e2019JC016011, https://doi.org/10.1029/2019JC016011, 2020.

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.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, Journal of Geophysical Research: Oceans, 118(12), 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.

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, 2020.

Van Roekel, L., Adcroft, A., Danabasoglu, G., Griffies, S. M., Kauffman, B., Large, W., Levy, M., Reichl, B., Ringler, T., and Schmidt, M.: The KPP boundary layer scheme for the ocean: Revisiting its formulation and benchmarking one‐dimensional simulations relative to LES, Journal of Advances in Modeling Earth Systems, 10, 2647–2685, https://doi.org/10.1029/2018MS001336, 2018.

Zhu, S., Wang, B., Zhang, L., Liu, J., Liu, Y., Gong, J., Xu, S., Wang, Y., Huang, W., Liu, L., He, Y., and Wu, X.: A Four‐Dimensional Ensemble‐Variational (4DEnVar) Data Assimilation System Based on GRAPES‐GFS: System Description and Primary Tests, Journal of Advances in Modeling Earth Systems, 14(7), e2021MS002737, https://doi.org/10.1029/2021MS002737, 2022.