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:

This manuscript evaluates the influence of a weakly coupled ocean data assimilation (WCODA) system implemented in E3SMv2 on the simulation of climate variability at global and regional scales. This manuscript primarily presents evaluation results rather than methodological or model-development advances. The WCODA system itself has already been fully described and evaluated in Shi et al. (2025, GMD). My major comment is that, in the present manuscript, it is not clear if there is any new algorithmic development, implementation detail, sensitivity analysis, or methodological innovation beyond what has already been published. Instead, the paper focuses on the climate pattern evaluation (e.g., ENSO, PDO, IOD, U.S. climate impacts), which seems to align with the scope of Journal of Climate or JGR-Atmospheres/Oceans more than the GMD. If the authors intend this work to be published in GMD, they must explicitly justify how this manuscript advances model development. At present, the manuscript reads as a results paper, not a model-development paper. Some major comments are listed as follows.

Response:

Thank you very much for providing us with very useful comments. We sincerely thank the reviewer for the constructive and thoughtful feedback. We would like to note that this study was submitted under the manuscript type "Model Evaluation paper", rather than "Development and Technical paper". As the reviewer noted, the development of the WCODA system is documented in Shi et al. (2025 GMD), which provided some verification of the DA system (i.e., error characteristics comparing the simulated fields that are constrained by DA with the observations). The present work focuses on the evaluation of the WCODA system, which is currently being used to generate realistic initial conditions for seasonal-to-decadal (S2D) hindcast experiments using the E3SM model. The main purpose of this study is to document the general features, strengths, and limitations of this newly developed WCODA system for the broader modeling community, with particular emphasis on large-scale climate modes and their remote impacts relevant to S2D prediction. We consider that documenting the performance improvements and inherent limitations of this system is critical for future modeling efforts and aligns well with the aims and scope of a "Model Evaluation paper" in GMD. We have explicitly clarified this evaluation focus in both the Introduction (L81-87) and the Conclusions (L455-457) of the revised manuscript.

L81-87: This 4DEnVar-based WCODA system is intended to provide realistic initial conditions for developing seasonal-to-decadal (S2D) hindcast experiments using the E3SM model. The primary objective of this study is to evaluate the strengths and limitations of this newly developed WCODA system and to document its capability in capturing global and regional climate variability for the broader modeling community, with particular emphasis on large-scale climate modes and their remote impacts that are essential for generating realistic initial conditions for S2D hindcast experiments.

L455-457: This study aims to thoroughly document the capabilities and limitations of this WCODA system in simulating global and regional climate variability.

In response to your suggestions, we have carefully revised the manuscript with the following clarifications and improvements. Specifically, we have added several new analyses and figures (e.g., **Figures A1, A2, and A3**) to enhance the dynamical insight of the results. For example, we have included global stability diagnostics in the **new Figure A1** to ensure that the assimilation does not introduce artificial imbalances (in response to **Comment #5**). We have also added the **new Figure A2** to explicitly link tropical precipitation improvements to adjustments in the large-scale Hadley circulation (in response to **Comments #8 and #9**). Furthermore, a one-month lagged regression analysis in the **new Figure A3** has been included to account for the potential delayed impact of ENSO (in response to **Comment #14**).

In addition to these dynamical analyses, we have revised and expanded key methodological details. These include providing a more detailed description of the experimental initialization (in response to **Comment #1**), clarifying the linear detrending methods and EOF analysis procedures (in response to **Comments #3 and #12**), and standardizing the significance testing to the 95% confidence level across all figures (in response to **Comment #13**). We have also specified that all climate indices, such as Niño 3.4, are calculated independently for each corresponding dataset (in response to **Comments #12 and #14**). In addition, we have provided a detailed justification for the Arctic exclusion in the assimilation process (in response to **Comment #2**).

We hope that these revisions address your concerns and clarify the contributions of our work. Please find our point-by-point responses to each of your specific comments below.

Comment#1:
Section 2.4: Experimental Design and Initialization: How the control experiment is initialized should be described carefully. Whether CTRL is spun up from a long control integration or initialized from observations? Do the CTRL and ASSIM use the same initial condition integrated every month? Or the CTRL does not change any initial condition. If so, this CTRL is fundamentally different from ASSIM simulation.

Response:
Thank you for your insightful comment. In our study, the CTRL experiment is initialized from a fully spun-up long-term integration of the E3SMv2 model (Golaz et al., 2022). The CTRL and ASSIM experiments are initialized from the same initial condition on January 1, 1950. After initialization, the CTRL experiment is integrated freely without re-initialization throughout the entire simulation period. This design allows the CTRL simulation to serve as a free-running baseline for evaluating the impacts of the assimilation system.

Based on your suggestion, we have revised the description in Section 2.4 accordingly to clarify the initialization of the control experiment (L158-165).

L158-165: Two numerical experiments are conducted to assess the impact of the WCODA system on climate variability: a control experiment (CTRL) and an assimilation experiment (ASSIM). Both experiments are initialized from the same initial condition on January 1, 1950, which is derived from a fully spun-up long-term integration of E3SMv2 (Golaz et al., 2022). Following initialization, the CTRL experiment is integrated freely without re-initialization or further modifications to the model state from 1950 to 2021 under observed historical external forcings. This experiment serves as the reference simulation for evaluating the influence of the assimilation system.

Comment#2:

Whether ocean temperature and salinity drift exists prior to the analysis period. Also, why Arctic Ocean is excluded from the assimilation process? This should be emphasized more. The exclusion of the Arctic Ocean from assimilation requires a much stronger justification. The manuscript states that this is due to "sparse observational coverage," yet EN4.2.1 does include Arctic observations. Sparse coverage alone is not a convincing argument, as the Southern Ocean suffers from similar limitations but is still assimilated. Potential dynamical consequences of excluding the Arctic, especially for high-latitude biases and global energy balance.

Response:

We thank the reviewer for raising these important points. Regarding the first question, no substantial drift in ocean temperature or salinity is present. As documented in Golaz et al. (2022), the E3SMv2 model underwent a rigorous 500-year pre-industrial spin-up followed by a historical integration from 1850 to 2014, ensuring that the model had reached a stable quasi-equilibrium state well before the start of our analysis period in 1950. The initial conditions for both CTRL and ASSIM were therefore derived from a fully equilibrated model state.

Regarding the exclusion of the Arctic Ocean from assimilation, we agree that this point requires clearer justification. Unlike the Southern Ocean, much of the Arctic Ocean is characterized by year-round sea-ice cover, and the absence of coupled ice-ocean assimilation presents a primary technical challenge. Without simultaneously updating sea ice variables, assimilating only ocean temperature and salinity could introduce physical inconsistencies at the ice-ocean interface. We therefore chose to exclude the Arctic Ocean from the assimilation domain in the current system. Addressing this limitation remains a key priority for future system development.

In response to this comment, we have clarified the quasi-equilibrium state of the model before the analysis period (L166-168) and added a detailed justification for excluding the Arctic Ocean from assimilation (L184-190).

L166-168: The E3SMv2 model had already reached a stable quasi-equilibrium state through a long spin-up and historical integration (Golaz et al., 2022).

L184-190: However, the Arctic Ocean is excluded from the assimilation domain due to sparse observational coverage and the current absence of a coupled ice-ocean assimilation scheme. Much of the Arctic Ocean is characterized by pervasive year-round sea ice cover. Without simultaneously updating sea ice variables, assimilating ocean temperature and salinity alone could introduce physical inconsistencies at the ice-ocean interface. Addressing this technical limitation remains a high priority for future system development.

Comment#3:

Fig. 1 shows detrended global, ocean, and land surface air temperature anomalies. However, the exact detrending method is unclear (linear? piecewise? global-mean vs grid-point detrending). Different detrending approaches can substantially alter correlation metrics in the following analysis. This figure may include both detrended and non-detrended time series. I also suggest to discuss with the results with the pacemaker experiments which only nudges the SST (e.g., Kosaka and Xie, 2013, 2016; Douville et al., 2015). The significance of assimilating ocean subsurface data should be emphasized here.

Douville, H., A. Voldoire, and O. Geoffroy (2015), The recent global warming hiatus: What is the role of Pacific variability? Geophys. Res. Lett., 42, 880–888.

Kosaka, Y., and S.-P. Xie (2013), Recent global-warming hiatus tied to equatorial Pacific surface cooling. Nature, 501, 403–407

Kosaka, Y. and S.-P. Xie (2016), The tropical Pacific as a key pacemaker of the variable rates of global warming. Nature Geo., 9, 669-673.

These papers demonstrate that nudging SST only in the tropical Pacific already reproduces much of the observed global temperature variability. The authors must therefore clarify: what additional value does subsurface ocean assimilation provide beyond SST nudging (impacts on the atmospheric and climate variability)? This is a critical scientific question that is not addressed.

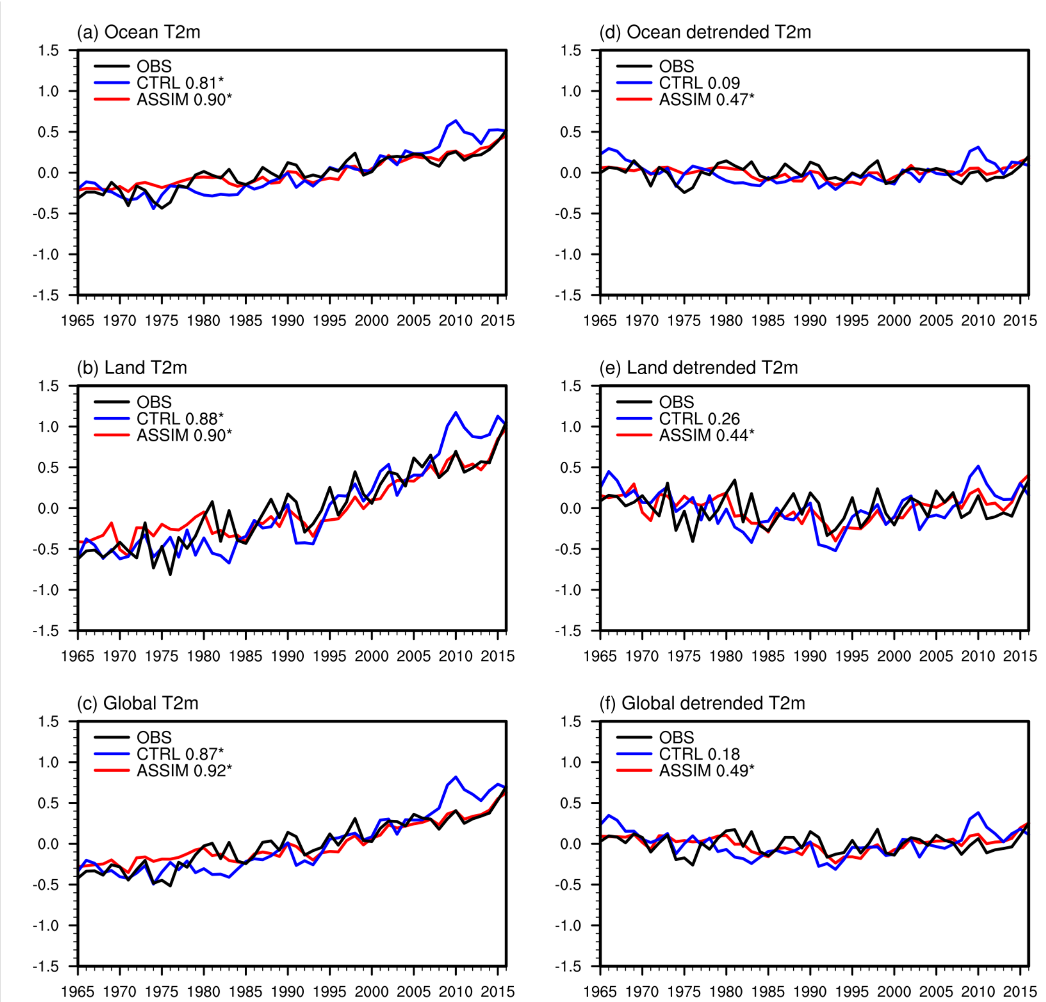
Response:

In our analysis, a linear detrending method was applied to the time series of annual mean surface air temperature anomalies. The detrending was performed on the domain-averaged time series over the global, ocean, and land regions. We have revised Figure 1 to include both non-detrended (Figs. 1a-c) and linearly detrended (Figs. 1d-f) time series of annual mean surface air temperature anomalies over ocean, land, and global domains.

We thank the reviewer for pointing out the relevance of SST pacemaker experiments (e.g., Kosaka and Xie, 2013, 2016; Douville et al., 2015). We agree that such experiments have been instrumental in isolating the influence of tropical Pacific SST variability on global and regional climate variability. However, data assimilation and pacemaker experiments are designed for different scientific purposes. Pacemaker experiments are typically used to isolate the impact of prescribed SST anomalies on atmospheric variability, whereas the primary goal of data assimilation is to generate realistic initial conditions for hindcasts and forecasts.

The SST-based nudging may potentially introduce physical inconsistencies between surface and subsurface layers. In contrast, our data assimilation system constrains the full three-dimensional ocean state, thereby ensuring vertical dynamical consistency. This feature has been shown to be essential for capturing decadal-to-multidecadal variability from previous studies (Morioka et al., 2018; Chikamoto et al., 2019). For instance, Morioka et al. (2018) demonstrated that the skillful prediction of decadal variability in the South Atlantic requires the assimilation of subsurface ocean temperature and salinity, as SST-only assimilation fails to accurately represent zonal heat transport. Chikamoto et al. (2019) also highlighted that data assimilation below 300 m in the ocean is necessary to improve decadal climate predictability. While we agree that a systematic comparison between subsurface ocean data assimilation and SST-based pacemaker approaches could provide valuable insights and represent a promising direction for future research, such a comparison is beyond the scope of the present study because these experiments are performed for different scientific objectives. The relevant discussion has been added to the revised manuscript.

In response to this comment, we have updated Figure 1 in the revised manuscript to include both non-detrended and linearly detrended time series, and revised the caption of Figure 1 (L757-763) to clarify the detrending methodology and added the corresponding text (L210-214) to describe the updated figure. Additionally, we have added further clarification and cited relevant studies (L176-184) to emphasize the additional value of assimilating subsurface ocean temperature and salinity, and also noted that comparing with SST-only pacemaker experiments remains an important direction for future research (L494-500).



L757-763: **Figure 1.** Time series of non-detrended (left column) and linearly detrended (right column) annual mean surface air temperature anomalies (units: °C) over (a, d) ocean, (b, e) land, and (c, f) global domains from 1965 to 2016. Black line: observation; blue line: CTRL; red line: ASSIM. Temperature anomalies are computed by removing both the climatological mean and long-term trend. For the detrended series, a linear trend is removed from the domain-averaged time series. The correlation coefficients of CTRL and ASSIM with the observation are also shown. The asterisk denotes statistically significant correlation at the 95% confidence level.

L210-214: The non-detrended time series (Figs. 1a-c) highlight the long-term warming trend, while the detrended series (Figs. 1d-f) isolate interannual variability by removing the linear trend from each domain-averaged time series. In both cases, ASSIM better captures the observed interannual surface air temperature variability compared to CTRL across all domains.

L176-184: Unlike the sea surface temperature (SST) relaxation approach used in pacemaker experiments to isolate the ocean influence on climate variability (e.g., Kosaka and Xie, 2016), our assimilation system is implemented to generate initial conditions for hindcast experiments by incorporating full-depth ocean temperature and salinity reanalysis to constrain the complete three-dimensional ocean state. This approach preserves vertical dynamical consistency and retains the subsurface ocean memory necessary for capturing decadal-to-multidecadal variability. For example, Morioka et al. (2018) demonstrated that skillful prediction of decadal variability in the South Atlantic requires the assimilation of subsurface ocean temperature and salinity, as SST-only assimilation fails to accurately represent zonal heat transport.

L494-500: Given that the assimilation of subsurface ocean temperature and salinity is essential for skillful decadal prediction (Morioka et al., 2018; Chikamoto et al., 2019), future efforts could explore systematic comparisons with SST-based pacemaker experiments to elucidate the distinct roles of surface and subsurface ocean constraints. Such comparisons would provide valuable insights into the mechanisms governing decadal predictability and inform the design of more effective initialization strategies.

Comment#4:

Line 174: I am surprised that assimilating full-depth ocean temperature and salinity increases the global mean temperature correlation only to 0.47. This seems to be weaker than correlations achieved by simple SST pacemaker experiments. Even more puzzling is that ASSIM fails to capture major ENSO impacts on the global 2m temperature (e.g., 1982/83, 1986/87, 1997/98) and strong La Niña events, despite assimilating ocean data (compared to Fig. 2 in Douville et al., 2015). This raises some issues why subsurface assimilation does not outperform SST-only nudging. Does the DA system overly smooth variability? Is the monthly assimilation window too long to retain ENSO phase locking?

Response:

We would like to clarify that the reported correlation of 0.85 in Douville et al. (2015) for the global annual mean surface air temperature is based on non-detrended time series. To ensure a consistent comparison, the corresponding non-detrended correlation in our ASSIM experiment is 0.92 (as shown in Fig. 1c), which is slightly higher than the 0.85 in Douville et al. (2015). The value of 0.47 refers to the detrended correlation, highlighting how well ASSIM captures interannual variability.

We agree that the ASSIM simulation underrepresents the atmospheric response to several ENSO events. This may be due to fundamental differences in experimental design for different purposes. Pacemaker experiments apply strong and high-frequency nudging to force the tropical Pacific SST towards observations, which directly enhances ENSO-related atmospheric responses to isolate the ENSO influence. In contrast, our assimilation system imposes global ocean constraints rather than targeting the tropical Pacific alone to produce globally balanced coupled states for initializing hindcasts/forecasts. The broader assimilation domain may introduce competing influences from other regions that may partially offset the atmospheric imprint of ENSO. In addition, the use of a monthly assimilation window may smooth the rapid phase transitions of ENSO, reducing the amplitude of the associated atmospheric signals. These limitations are now acknowledged in the revised manuscript.

In response to this comment, we have revised the manuscript to compare with the results of pacemaker experiments in Douville et al. (2015) (L214-217), and to briefly explain the weaker atmospheric response to certain ENSO events (L226-230).

L214-217: For the non-detrended global annual mean surface air temperature, the correlation of ASSIM with observations is 0.92, slightly higher than the value of 0.85 reported by Douville et al. (2015) in their tropical Pacific pacemaker experiment.

L226-230: However, compared with pacemaker experiments in which SST is directly nudged within the tropical Pacific, the relatively weaker atmospheric response to certain ENSO events in ASSIM may be related to the competing influence of global ocean constraints beyond the tropical Pacific, as well as the smoothing effect introduced by the monthly assimilation window.

Comment#5:

Fig.2 I do agree that errors are reduced mostly in ASSIM, however, I notice red shading in Arctic is increased in ASSIM than CTRL. Is it right? That implies assimilation enhances the biases over the regions where no assimilation is made. Can you comment on this issue? Can you ensure the assimilation provides a reasonable dynamic is still hold? What's the imbalance of this Earth System Model after the assimilation (global TOA imbalance, global atmosphere temperature, global ocean temperature and salinity etc). Without such diagnostics, it is difficult to assess whether the assimilation introduces artificial imbalances.

Response:

It is correct that the climatological warm biases over the Arctic appear slightly enhanced in ASSIM compared to CTRL. This feature reflects the fact that no ocean data assimilation is applied over the Arctic Ocean. In this region, the ocean state is not constrained by assimilation and may respond indirectly to adjustments introduced by assimilation in other regions. Thus, the enhanced Arctic bias may reflect a redistribution of temperature biases, rather than a breakdown of the model's dynamical consistency. Despite this localized increase, the overall bias patterns show broad improvements across most regions in ASSIM. We acknowledge this limitation in the revised manuscript.

Regarding the concern about possible model imbalances, we have added a new Figure A1 to diagnose the temporal evolution of global mean top-of-atmosphere net energy flux, surface air temperature, ocean temperature and salinity averaged over the upper 1000 m. These results confirm that the ASSIM experiment maintains the overall stability of the coupled system, indicating that the data assimilation does not introduce artificial imbalances.

Based on your suggestions, we have added clarification on the enhanced Arctic warm bias (L245-248) and included the new Figure A1 to demonstrate that the data assimilation does not introduce artificial imbalances (L248-251).

L245-248: The climatological warm biases over the Arctic Ocean appear slightly enhanced in ASSIM compared with CTRL. Since the Arctic Ocean state is not directly constrained by assimilation, it may respond indirectly to remote adjustments introduced elsewhere, potentially leading to a redistribution of temperature biases.

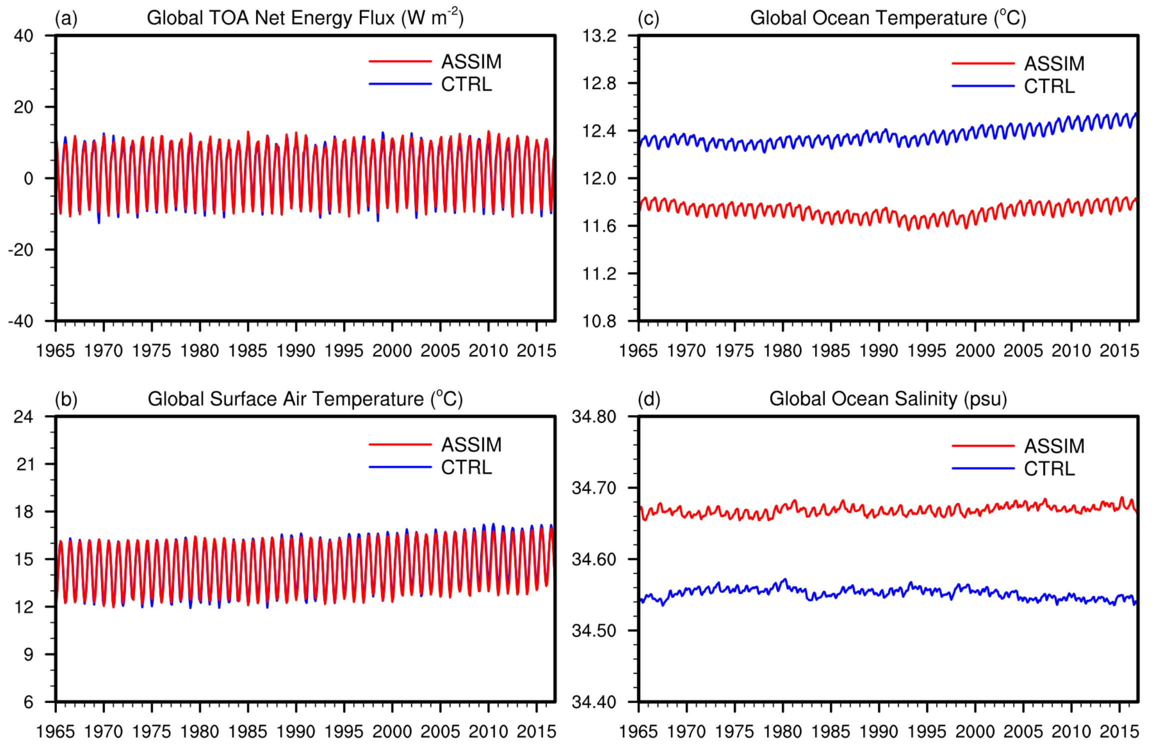


Figure A1. Time series of (a) global mean top-of-atmosphere (TOA) net energy flux (W m^{-2}), (b) global mean surface air temperature (°C), (c) global mean ocean temperature averaged over the upper 1000 m (°C), and (d) global mean ocean salinity averaged over the upper 1000 m (psu). The blue and red lines denote the CTRL and ASSIM experiments, respectively.

L248-251: Despite this localized degradation, the assimilation maintains the overall stability of the coupled system. As shown in Fig. A1, key global mean diagnostics, including top-of-atmosphere net energy flux, surface air temperature, ocean temperature and salinity, remain stable throughout the ASSIM integration.

Comment#6:

Line 195-196. It seems the overall spatial pattern of the biases is reduced but the shape is similar. Are you sure the realistic ocean state play a key role? Maybe assimilating the SST is sufficient. The authors may disentangle SST-only effects and Subsurface temperature/salinity impacts.

Response:

We agree that the similar spatial pattern of bias reduction between surface air temperature and SST highlights the strong coupling between the ocean surface and the atmosphere. It is possible that SST assimilation alone could account for a substantial fraction of the surface air temperature improvement. However, the current assimilation experiment involves the assimilation of full-depth ocean temperature and salinity, making it difficult to isolate the individual contributions of SST and subsurface ocean constraints. We acknowledge that a dedicated SST-only assimilation experiment would be required in future work to disentangle these respective effects.

Following your suggestions, we have revised the manuscript (L239-244) to highlight the need for future targeted SST-only assimilation to distinguish the individual contributions of SST and subsurface ocean constraints.

L239-244: The spatial pattern of reduced surface air temperature bias closely aligns with the reduction in SST bias reported by Shi et al. (2025), highlighting the strong coupling between the ocean surface and the atmosphere. However, the current assimilation experiment does not explicitly isolate the respective contributions of SST and subsurface ocean states. Future SST-only assimilation experiments would be needed to disentangle their individual effects on surface air temperature.

Comment#7:

Line 206-207, why do you think surface air temperature and precipitation correlation over land in CTRL is “relatively high” (0.26 and 0.27)? Have you compared this with the case without external forcing on land?

Response:

Since we have not conducted a specific control experiment without external forcing over land, referring to these correlation values as "relatively high" may overstate the result for this study.

In response to this comment, we have removed this sentence from the revised manuscript to avoid overstatement.

Comment#8:

Section 3.1: this section is largely descriptive, resembling a student report rather than a scientific analysis. The manuscript does not explain why precipitation biases are reduced, which circulation or moisture processes are improved, whether Walker/Hadley circulation changes are involved. I couldn’t see any scientific insight. I suggest the authors enhance the dynamics behind the enhanced precipitation bias pattern.

Response:

We thank the reviewer for this constructive comment. We agree that the previous analysis lacked sufficient dynamical interpretation. In the revised manuscript, we have added a new Figure A2 that presents the zonal-mean meridional streamfunction to evaluate assimilation-induced changes in the Hadley circulation.

This analysis provides dynamical insight into the precipitation improvements. As shown in Figure A2, the CTRL simulation exhibits a pronounced positive streamfunction anomaly between 5°N and 15°N, indicating excessive ascending motion in the northern branch of the Hadley cell. This feature is consistent with the distinct wet bias over the tropical Pacific. In contrast, the ASSIM experiment substantially mitigates this positive anomaly, suggesting that the overactive Hadley cell is effectively constrained. As a result, the precipitation distribution in ASSIM becomes more realistic. We have also examined changes in the Walker circulation (not shown) but found no substantial differences between CTRL and ASSIM. Thus, the precipitation improvements are more closely associated with changes in the Hadley circulation.

In response to this comment, we have included the new Figure A2 and incorporated this dynamical interpretation (L277-282) in the revised manuscript to explicitly link precipitation bias reduction to large-scale circulation changes.

L277-282: The improvement in tropical precipitation is closely linked to changes in the large-scale meridional circulation. As illustrated in Fig. A2, CTRL exhibits an anomalously strong ascending branch of the Hadley circulation in the northern tropics, which sustains a pronounced wet bias over the tropical Pacific. In contrast, this anomalous ascent is effectively suppressed in ASSIM, contributing to a more realistic precipitation distribution across the tropical and subtropical regions.

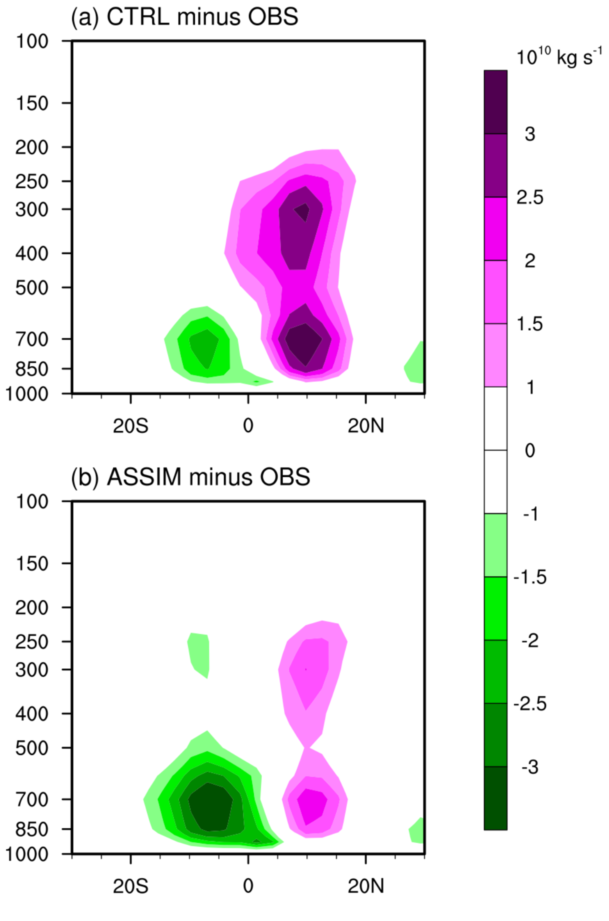


Figure A2. Zonal-mean meridional streamfunction differences (units: 10^{10} kg s^{-1}) between model simulations and observations averaged over the period 1980–2016. Panel (a) shows the difference between CTRL and observations, and panel (b) shows the difference between ASSIM and observations. The vertical axis represents pressure levels (hPa), and the horizontal axis denotes latitude.

<u>Comment#9:</u>

Fig. 4: it is very interesting to see that tropical-subtropical precipitation difference is not statistically significant in ASSIM. This is the region where we have the largest precipitation biases. It is very interesting to understand if assimilation in ocean suppress variability? Or is this a sign of over-constraint?

Response:
We do not interpret the lack of statistically significant differences in ASSIM as evidence of suppressed variability or over-constraint. Instead, it primarily reflects a reduction in systematic mean-state biases due to the assimilation. In CTRL, large precipitation biases lead to pronounced climatological differences relative to observations. In contrast, these biases are substantially reduced in ASSIM, bringing the simulated mean state closer to observations and thereby reducing the residual differences below the threshold of statistical significance at the 95% confidence level. This interpretation is further supported by the weakened anomalous ascent in the northern branch of the Hadley circulation shown in the new Figure A2, which contributes to the more realistic precipitation distribution.

Based on this comment, we have revised the manuscript (L274-282) to clarify this interpretation, with the newly added Figure A2 providing supporting evidence of the improved meridional circulation.

L274-282: These reductions in systematic biases bring the climatological precipitation in ASSIM into better agreement with observations, with residual differences over some regions, particularly the tropical Pacific, no longer reaching statistical significance. The improvement in tropical precipitation is closely linked to changes in the large-scale meridional circulation. As illustrated in Fig. A2, CTRL exhibits an anomalously strong ascending branch of the Hadley circulation in the northern tropics, which sustains a pronounced wet bias over the tropical Pacific. In contrast, this anomalous ascent is effectively suppressed in ASSIM, contributing to a more realistic precipitation distribution across the tropical and subtropical regions.

Comment#10:
Line 231-232, it needs to be very careful to say the model's inability to reproduce the tropical SST variability. Normally, we never expect the climate model to reproduce the exact timing or peaks of ENSO or other tropical climate pattern. Instead, if the ENSO spectrum and strength are well compared with the observation, we believe this model can capture the dynamics related to ENSO and is capable of reproducing tropical SST variability. However, if the model cannot well simulate the tropical dynamics while we try to nudge the simulation using data assimilation, this is very dangerous. This can be seen in Fig. 6. The discussion needs careful, cautious wording and deeper discussion.

Response:
We agree that the original wording "the model's inability" is inappropriate. In the revised manuscript, we have removed this phrasing to avoid implying an intrinsic deficiency of the model. Instead, we have revised this sentence (L292-293) to clarify that the low correlation in CTRL primarily reflects poor phase agreement with the observed ENSO indices. As noted

earlier, an important purpose of data assimilation is to generate initial conditions for hindcasts/forecast. Hence correctly capturing the phasing of ENSO and other modes of variability is critical for providing realistic initial conditions for predictions.

L292-293: indicating poor phase agreement with the observed Niño 3.4 index

Comment#11:

Fig. 8 I am surprised that the DMI correlation in ASSIM is only 0.56, given that SST is assimilated. I thought the correlation should be very similar to the Nino3.4 index while you assimilate the SST. Can you explain why? This implies your SST in Indian Ocean is not quite close to the observation. Why? What's the correlation of DMI between EN4 SST and HadiSST you used as the validation? Is Indian Ocean SST poorly constrained?

Response:

The relatively lower DMI correlation in ASSIM is related to known limitations of the EN4.2.1 dataset in the Indian Ocean. As documented by Good et al. (2013), the EN4.2.1 product exhibits regional deficiencies in the Indian Ocean due to the lack of near-surface observations that pass quality control. We further computed the correlation of the DMI index between the EN4.2.1 and HadISST dataset (Figure R1). The resulting correlation is only 0.75, indicating substantial discrepancies between these two observational products in the Indian Ocean. Therefore, the limited quality of the EN4.2.1 dataset likely limits the effectiveness of the assimilation in the Indian Ocean, leading to the relatively lower DMI correlation in ASSIM.

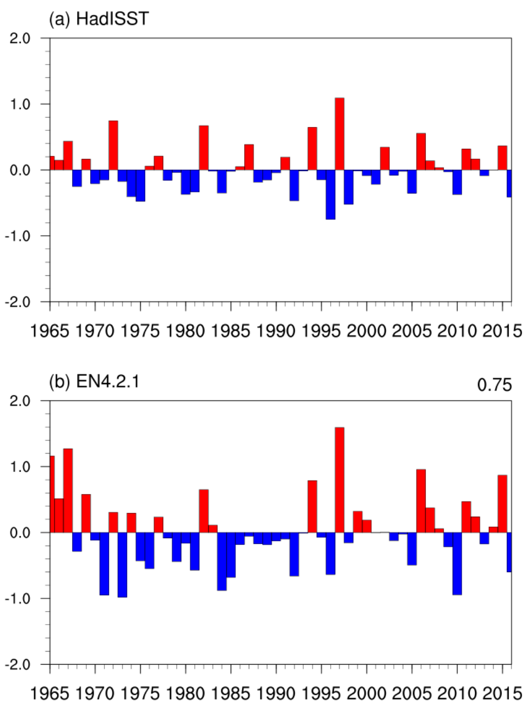


Figure R1. Time series of the autumn Dipole Mode Index (DMI) for (a) the HadISST dataset and (b) the EN4.2.1 dataset. The correlation coefficient between the EN4.2.1 and the HadISST dataset is shown in the upper-right corner of panel (b).

In response to this comment, we have added a clarification and cited Good et al. (2013) (L338-342) to note that the lower DMI correlation in ASSIM is related to known limitations of the EN4.2.1 dataset in the Indian Ocean.

L338-342: Nevertheless, the lower DMI correlation in ASSIM relative to Niño3.4 is partly attributable to known limitations of the assimilated EN4.2.1 dataset in the Indian Ocean. As noted by Good et al. (2013), the EN4.2.1 reanalysis exhibits regional deficiencies in this basin due to the limited availability of near-surface observations that meet quality-control standards, thereby limiting the assimilation effectiveness in the Indian Ocean.

Comment#12:

Similarly, I am also surprised that the PDO pattern and the time series may not be so close to the observation (may connected to my comment 3 above, how the EOF is performed). Same apply for IPO and AMO in Section 3.4. Can you clarify how you perform the EOF analysis for the observation and model? Do you project the observational data on to the model grid? Or do you perform the EOF analysis separately between the observation and model? Are they collocated within the same grid? Also, the linear trend you remove. Did you remove the global trend for all climate patterns or did you remove the linear trend for individual grid? These details should be clarified for a fair comparison among these climate modes.

Response:

The EOF analysis is performed separately for the observations and for each model experiment. The observational datasets and model outputs are first interpolated onto a common horizontal grid prior to the EOF analysis. No projection of observational data onto the model grid is performed. Regarding detrending, we remove the linear trend at each individual grid point.

Based on your suggestions, we have clarified the EOF method (L349-351) in the revised manuscript and added detailed descriptions to the captions of Figure 9 (PDO index) (L817-819), Figure 10 (IPO index) (L825-827), and Figure 11 (AMO index) (L833-834).

L349-351: The EOF analysis is performed separately for the observations and each model experiment. Prior to the EOF calculation, all observational datasets and model outputs are interpolated onto a common horizontal grid and linear trends are removed at each grid point.

L817-819: The EOF analysis is performed separately for the observations and model outputs after interpolating both onto a common horizontal grid and removing linear trends at each grid point.

L825-827: The IPO index is derived from EOF analysis after interpolating both the observations and model outputs onto a common horizontal grid and removing linear trends at each grid point.

L833-834: The AMO index is calculated by subtracting the global mean SST anomalies (60°S–60°N) from the North Atlantic SST anomalies (0–60°N, 0–80°W).

Comment#13:
Some figures use 95% significance (e.g, Fig. 4) while some use 90% (e.g., Fig. 12). Some figures do not include any statistical significance. This is inconsistent. A single confidence level should be used throughout unless explicitly justified.

Response:
In the revised manuscript, we have standardized the significance level across all relevant figures to the 95% confidence level to ensure consistency. Figures 12 and 13, which previously used 90% significance, have been updated accordingly. Figures 1, 2, 3, 4, 6, and 14 already applied the 95% confidence level. For other figures (Figures 5, 7, 8, 9, 10, and 11), we have added asterisks to indicate that the correlation coefficient is statistically significant at the 95% confidence level.

Based on your suggestions, we have updated the captions of Figure 12 (L838-839) and Figure 13 (L841-842) to clarify that the dotted areas denote regions where the differences pass the 95% confidence level. In addition, we have revised the captions of other figures (Figures 5, 7, 8, 9, 10, and 11) (L788-789, L802-803, L807-808, L814-815, L824-825, L832-833) by adding asterisks next to correlation values to indicate significance at the 95% confidence level.

L838-839: Dotted areas indicate regions where the differences are statistically significant at the 95% confidence level.

L841-842: Dotted areas denote regions where the differences pass the 95% confidence level.

L788-789, L802-803, L807-808, L814-815, L824-825, L832-833: The asterisk indicates that the correlation is statistically significant at the 95% confidence level.

Comment#14:
Fig. 14 is very important. However, the comparison may not be fair. Can you confirm if the Nino3.4 index in CTRL is based on the observed Nino3.4 or the modeled Nino3.4? Can you ensure the ASSIM still preserve the same atmosphere and ocean dynamics as the CTRL? Also, do you use the same time correlation or 1-2 month lag? It is common to see the impact of ENSO on the US winter climate after ~1 month lag.

Response:
We confirm that the regression analysis in each panel is separately performed using the respective Niño3.4 index calculated for each dataset. Specifically, the OBS regression uses the observed index, while the ASSIM and CTRL regressions use their own simulated Nino 3.4 indices. The ASSIM experiment uses the same model physics and dynamical core as the CTRL run. The only difference between ASSIM and CTRL is the assimilation of the ocean EN4.2.1 reanalysis.

All regressions are based on simultaneous DJF winter averages. The regression analysis is performed using the winter-mean (DJF) atmospheric variables against the simultaneous winter Nino 3.4 index. Following your suggestion, we have also added a new Figure A3 to regress the monthly atmospheric variables from December to February (DJF) onto the 1-month-leading monthly Niño 3.4 anomalies from November to January (NDJ). We found that the lagged regression patterns in the new Figure A3 are highly consistent with the simultaneous winter regression results shown in Figure 14.

In response to this comment, we have revised the caption of Figure 14 (L847-848) and added a clarification in the main text (L417-419) to specify that the Niño 3.4 index is calculated separately for each dataset and that all regressions are based on simultaneous DJF averages. Additionally, we have added the new Figure A3 and corresponding description (L438-442) to present a one-month lagged regression analysis.

L847-848: In each panel, the Niño 3.4 index is calculated separately from its corresponding dataset and all regressions are based on simultaneous winter (DJF) averages.

L417-419: The Niño 3.4 index is calculated separately from its corresponding dataset, and all regressions are performed using simultaneous winter averages.

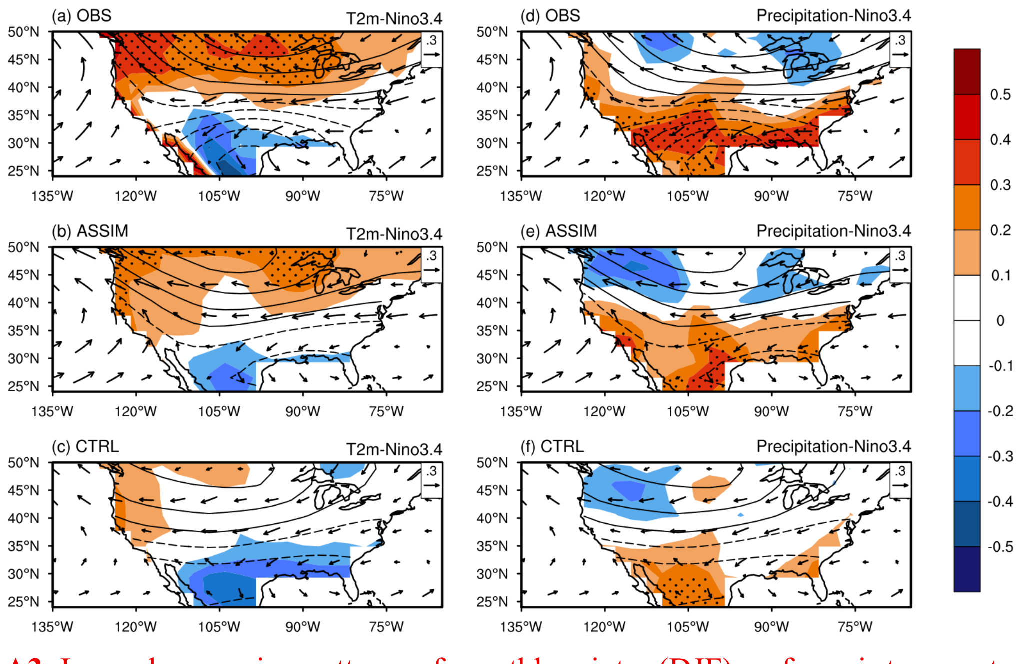


Figure A3. Lagged regression patterns of monthly winter (DJF) surface air temperature (left; shaded), precipitation (right; shaded), 500hPa geopotential height (contours), and 850hPa winds (vectors) onto the preceding month's (NDJ) standardized Niño 3.4 index for (a, d) the observation, (b, e) ASSIM, and (c, f) CTRL. Specifically, monthly atmospheric variables from December to February (DJF) are regressed onto the standardized monthly Niño 3.4 index from November to January (NDJ), corresponding to a one-month lead time. The Niño 3.4 index is calculated separately from each corresponding dataset. Dotted areas denote statistical significance at the 95% confidence level.

L438-442: To account for the potential delayed impact of ENSO, a one-month lagged regression analysis is also presented in Fig. A3. The resulting lagged patterns are highly consistent with the simultaneous winter regression results (Fig. 14), reinforcing the linkage between ENSO variability and U.S. winter climate anomalies.

Comment#15:

Overall, this manuscript contains interesting evaluation results, but I do worry it lacks model-development novelty required for GMD. It also provide insufficient dynamical insight. Some methodology should be clarified. I recommend major revision, with a strong suggestion that the authors either reframe the manuscript explicitly as a model-development and diagnostic paper, or consider submission to J. Clim. or other journals where the scientific results would be more appropriate.

Response:

Thank you for your recognition of the value of the evaluation results. Please refer to our detailed response to the General Comment above, where we clarify the manuscript type, its intended contribution, and the revisions made to better align with the GMD scope.

References:

Douville, H., Voldoire, A., and Geoffroy, O.: The recent global warming hiatus: What is the role of Pacific variability? Geophysical Research Letters, 42, 880–888, https://doi.org/10.1002/2014GL062775, 2015.

Chikamoto, Y., Timmermann, A., Widlansky, M. J., Zhang, S., and Balmaseda, M. A.: A drift-free decadal climate prediction system for the Community Earth System Model, Journal of Climate, 32, 5967–5995, https://doi.org/10.1175/JCLI-D-18-0788.1, 2019.

Golaz, J. C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W. Y., Bradley, A. M., Tang, Q., Maltrud, M. E., Forsyth, R. M., Zhang, C. Z., Zhou, T., Zhang, K., Zender, C. S., Wu, M. X., Wang, H. L., Turner, A. K., Singh, B., Richter, J. H., Qin, Y., Petersen, M. R., Mametjanov, A., Ma, P., Larson, V. E., Krishna, J., Keen, N. D., Jeffery, N., Hunke, E. C., Hannah, W. M., Guba, O., Griffin, B. M., Feng, Y., Engwirda, D., Vittorio, A. V., Cheng, D., Conlon, L. M., Chen, C., Brunke, M. A., Bisht, G., Benedict, J. J., Asay-Davis, X. S., Zhang, Y. Y., Zhang, M., Zeng, X. B., Xie, S. C., Wolfram, P. J., Vo, T., Veneziani, M., Tesfa, T. K., Sreepathi, S., Salinger, A. G., Jack Reeves Eyre, J. E., Prather, M. J., Mahajan, S., Li, Q., Jones, P. W., Jacob, R. L., Huebler, G. W., Huang, X. L., Hillman, B. R., Harrop, B. E., Foucar, J. G., Fang, Y. L., Comeau, D. S., Caldwell, P. M., Bartoletti, T., Balaguru, K., Taylor, M. A., McCoy, R. B., Leung, L. R., and Bader, D. C.: The DOE E3SM Model version 2: Overview of the physical model and initial model evaluation, Journal of Advances in Modeling Earth Systems, 14, e2022MS003156, https://doi.org/10.1029/2022MS003156, 2022.

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.

Morioka, Y., Doi, T., Storto, A., Masina, S., and Behera, S. K.: Role of subsurface ocean in decadal climate predictability over the South Atlantic, Scientific Reports, 8, 8523, https://doi.org/10.1038/s41598-018-26899-z, 2018.

Kosaka, Y. and Xie, S. P.: Recent global-warming hiatus tied to equatorial Pacific surface cooling, Nature, 501, 403–407, https://doi.org/10.1038/nature12534, 2013.

Kosaka, Y. and Xie, S. P.: The tropical Pacific as a key pacemaker of the variable rates of global warming, Nature Geoscience, 9(9), 669–673, https://doi.org/10.1038/ngeo2770, 2016.