Authors' Response to Reviews of

Improving Thermodynamic Nudging in the E3SM Atmosphere Model Version 2 (EAMv2): Strategy and Hindcast Skills on Weather Systems

Shixuan Zhang and Ruby Leung (on behalf of all authors)

Geoscientific Model Development (GMD), `10.5194/egusphere-2025-4277`

RC: ***Reviewers' Comment***, AR: Authors' Response, ☐ Manuscript Text

We would like to thank the anonymous reviewer for their useful comments to help us improve the manuscript. Our responses to individual comments from the reviewer are included below.

1. Comments from Reviewer #2

The manuscript proposes an improved nudging strategy for EAMv2 based on vertically modulated tendencies, aiming to mitigate distortions of physical processes caused by improperly designed nudging of temperature and humidity. Overall, this study addresses a practical problem in constrained atmospheric modeling and demonstrates a certain degree of innovation. Despite these positive aspects, several important elements of nudging implementation are insufficiently documented. In addition, some key conclusions would benefit from additional sensitivity experiments to better demonstrate the robustness and general applicability of the proposed method. In summary, this work represents a contribution with practical value and some novelty. I recommend that the manuscript undergo major revision prior to publication.

RC: *Nudging is intended to steer the model state toward the observed state and typically involves weighting parameters in both space and time. However, the methodological description in Section 2.2 only addresses vertical spatial weighting. In the authors' experiments, each model grid point obtains an interpolated value at every integration time step. For observations with longer temporal intervals, is also derived through spatial and temporal interpolation in order to perform nudging? If so, could this introduce additional errors for highly nonlinear variables such as water vapor? The authors are encouraged to provide a more detailed explanation of the horizontal spatial weighting and temporal weighting schemes. In general, observations should exert a smooth and continuous influence before and after the observation time to avoid excessive shocks to the model. Furthermore, it is currently unclear whether a relaxation coefficient is included to control the rate at which the nudging tendency is introduced into the model state. If such a parameter exists, it should also be explicitly specified and described.*

AR: We thank the reviewer for the careful reading and for raising these important questions regarding the nudging implementation and its application. The implementation of nudging in E3SM—including the interpolation of observational fields, the choice of relaxation time scales, and the overall application strategy—has been extensively documented and evaluated in a series of our previous studies (Sun et al., 2019; Zhang et al., 2022). Those works examined the sensitivity of simulations to these implementations and parameter choices and identified configurations that provide reasonable constraint without introducing notable numerical instabilities or unrealistic physical behavior. For this reason, Section 2.2 provides only a concise description of the nudging setup specific to this study, while directing the readers to Sun et al. (2019) and Zhang et al. (2022) for a complete methodological discussion and justification. In the present paper, we focus specifically on extending nudging to temperature and, in particular, humidity fields in addition to the large-scale winds, and on evaluating the revised vertical weighting designed to improve their incorporation into the nudging simulations. These aspects were not fully addressed in the earlier studies and therefore form the central contribution of the current work, as highlighted in the Introduction (lines 41–49):

> Motivated by the emerging need for temperature and humidity nudging in model analysis and evaluation, this study extends the work of Sun et al. (2019) and Zhang et al. (2022) on atmospheric nudging within the Energy Exascale Earth System Model atmosphere component (EAM; Rasch et al., 2019; Xie et al., 2018; Leung et al., 2020; Golaz et al., 2019, 2022). We focus on developing strategies that incorporate temperature and humidity, in addition to large-scale wind fields, to improve the realism of atmospheric simulations nudged toward reanalysis products while avoiding significant drifts in other fields important for evaluation, such as precipitation and radiative fluxes. Our goal is to demonstrate that this enhanced nudging approach enables more reliable, constrained hindcast simulations that are especially valuable for high-impact weather analysis and for machine learning applications using low-resolution models such as EAM. These aspects of nudging strategy and application remain underexplored in previous EAM-based studies (e.g., Sun et al., 2019; Zhang et al., 2022).

While we agree that the reviewer's questions are valid, a full re-presentation and re-analysis of the previously documented interpolation and nudging choices would largely duplicate material already available in earlier publications (Sun et al., 2019; Zhang et al., 2022) and would divert the focus of the current paper. Nevertheless, we have reorganized Section 2.2 and Section 2.3 by adding additional clarification and discussion (see lines 74–204 in the revised manuscript) to address the reviewer's concerns. In addition, we summarize below the relevant information documented in Sun et al. (2019) and Zhang et al. (2022) to provide helpful context for the reviewer.

First, as described in Section 2.3, the ERA5 data are interpolated horizontally and vertically from their native grid and temporal frequency to the E3SM grid prior to the simulations and are then provided as external forcing fields. These interpolations are performed using well-tested tools developed for CESM2, as described in the revised manuscript (lines 187–194):

> The ERA5 atmospheric and land-surface reanalysis fields were originally produced on a $0.25^\circ \times 0.25^\circ$ horizontal grid with 37 pressure levels. These data were remapped to the EAMv2 model grid (the "pg2" horizontal grid with 72 hybrid sigma–pressure levels) and subsequently used as reference fields for the nudging simulations. The remapping procedures (both horizontal and vertical) followed the algorithms developed for nudging in the Community Earth System Model version 2 (CESM2) (CESM2, 2025). Furthermore, linear-function nudging was applied to the 3-hourly ERA5 reanalysis to derive nudging tendencies at each model time step (30 minutes). This configuration reflects a balance between the computational cost of online data processing and the demonstrated performance benefits reported in Zhang et al. (2022).

For the nudging experiments in this study, we use the linear-function nudging approach developed for E3SM, following Sun et al. (2019). This approach was chosen because it provides a continuous constraint on the model state, favors the adjustment of slow and large-scale motions, and yielded the best overall performance among the nudging strategies evaluated in that study. In the linear-function approach, two neighboring time slices of the constraining data are linearly interpolated to the model time step and then compared with the predicted model state to compute the nudging tendency applied during that step. We note that the E3SM atmospheric model uses a time step of 1800 s (30 minutes). In relation to the reviewer's question, the linear-function nudging scheme can operate on the constraining data (ERA5 in this study) regardless of its native temporal frequency, because the required values at each model time step are obtained through linear temporal interpolation performed online during the nudging process. This design allows the reanalysis constraint to be applied smoothly and continuously in time. Regarding temporal interpolation errors and the choice of constraining-data frequency, a thorough evaluation of the sensitivity of E3SM nudging performance to these factors has been documented in Zhang et al. (2022). As mentioned above, revisions in Section 2.2 and Section 2.3 have been made to provide additional clarification in response to the reviewer's comments, and we have also explicitly cited our two previous studies so that reviewers and readers can obtain further methodological details if needed.

Second, as described in Eq. (2) in Section 2.2, the nudging scheme in E3SM is implemented as a Newtonian relaxation,

$$\left(\frac{\partial X_m}{\partial t}\right)_{\mathrm{ndg}} = \frac{1}{\tau}\big(X_{\mathrm{ERA5}} - X_m\big),$$

where τ is an effective relaxation time scale, which can be treated as a coefficient controlling the rate at which the nudging tendency is introduced into the model state. In the original implementation, τ is globally uniform, and a value of $\tau = 6$ h was identified as an optimal choice for practical applications in E3SM based on our previous studies (Sun et al., 2019). This value has also been widely adopted because, historically, the reanalysis fields used to constrain the model were typically available at 6-hourly intervals (Kooperman et al., 2012; Subramanian and Zhang, 2014; Ma et al., 2014, 2015; Lin et al., 2016; Fast et al., 2016). We note, however, that more recent reanalyses such as ERA5 provide data at hourly frequency, which in principle opens the possibility of exploring shorter relaxation time scales or alternative nudging strategies. In our previous work (Sun et al., 2019; Zhang et al., 2022), the choice of τ, the constraining-data frequency, and the associated implications were discussed in detail, and we refer the reviewer to those studies for a more comprehensive treatment of this topic. In response to the reviewer's comment, we have also added clarification and additional context in the revised manuscript (lines 185–204):

> It is important to note that all nudged EAMv2 simulations listed in Table 1 differ only in their choice of nudged variables and vertical weighting scheme; all other aspects of the nudging configuration follow the recommendations from prior studies (Sun et al., 2019; Zhang et al., 2022). The ERA5 atmospheric and land-surface reanalysis fields were originally produced on a $0.25° \times 0.25°$ horizontal grid with 37 pressure levels. These data were remapped to the EAMv2 model grid (the "pg2" horizontal grid with 72 hybrid sigma–pressure levels) and subsequently used as reference fields for the nudging simulations. The remapping procedures (both horizontal and vertical) followed the algorithms developed for nudging in the Community Earth System Model version 2 (CESM2) (CESM2, 2025). Furthermore, linear-function nudging was applied to the 3-hourly ERA5 reanalysis to derive nudging tendencies at each model time step (30 minutes). This configuration reflects a balance between the computational cost of online data processing and the demonstrated performance benefits reported in Zhang et al. (2022). The nudging relaxation time scale τ in Eq. (3) was set to 6 hours for the upper-atmospheric layers and 1 hour for the land-surface layer. The choice of $\tau = 6$ hours for the upper atmosphere is motivated by both practical experience and demonstrated performance in previous EAM/E3SM nudging studies, where it provides an effective constraint while avoiding excessive distortion of the model's internal dynamics (Sun et al., 2019; Zhang et al., 2022). For the land-surface layer, the choice of $\tau = 1$ hour was empirically determined in this study. The shorter relaxation time is intended to provide a tighter constraint on the model state, given the large variability associated with the spatial inhomogeneity of land-surface processes and the fact that the atmospheric model in EAMv2 is coupled to the land model at a 1-hour frequency. We acknowledge that the performance of the nudging may exhibit some sensitivity to alternative choices of the parameters discussed here, as well as to the tunable parameters

P_0 and Z_b used in the proposed strategy in Eq. (4). However, we believe that these sensitivities do not alter our primary goal of demonstrating how the revised nudging strategy proposed in this study benefits E3SM nudging simulations, as discussed in the following sections.

In addition to the clarifications and revisions described above, we have also added a statement in the Conclusions section to clarify the choices of the nudging parameters used in this study and their implications for our findings (lines 539–542):

> We also note that the performance of the nudging strategy based on Eq. (4) may exhibit some sensitivity to alternative choices of P_0, Z_b, and τ, as discussed in Sections 2.2 and 2.3. However, both exploratory tests and the physical arguments outlined above indicate that the main conclusions of this study are not dependent on the specific values adopted here.

References:

- Sun, J., Zhang, K., Wan, H., Ma, P.-L., Tang, Q., and Zhang, S. (2019): Impact of nudging strategy on the climate representativeness and hindcast skill of constrained EAMv1 simulations, *J. Adv. Model. Earth Syst.*, 11, 3911–3933, https://doi.org/10.1029/2019MS001831.

- Zhang, S., Zhang, K., Wan, H., and Sun, J. (2022): Further improvement and evaluation of nudging in the E3SM Atmosphere Model version 1 (EAMv1): simulations of the mean climate, weather events, and anthropogenic aerosol effects, *Geosci. Model Dev.*, 15, 6787–6816, https://doi.org/10.5194/gmd-15-6787-2022.

- Kooperman, G. J., Pritchard, M. S., Ghan, S. J., Wang, M., Somerville, R. C. J., and Russell, L. M. (2012): Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the Community Atmosphere Model 5, *J. Geophys. Res. Atmos.*, 117, D23204, https://doi.org/10.1029/2012JD018588.

- Lin, G., Wan, H., Zhang, K., Qian, Y., and Ghan, S. J. (2016): Can nudging be used to quantify model sensitivities in precipitation and cloud forcing?, *J. Adv. Model. Earth Syst.*, 8, 1073–1091, https://doi.org/10.1002/2016MS000659.

- Ma, H.-Y., Chuang, C. C., Klein, S. A., Lo, M.-H., Zhang, Y., Xie, S., Zheng, X., Ma, P.-L., Zhang, Y., and Phillips, T. J. (2015): An improved hindcast approach for evaluation and diagnosis of physical processes in global climate models, *J. Adv. Model. Earth Syst.*, 7, 1810–1827, https://doi.org/10.1002/2015MS000490.

- Ma, P.-L., Rasch, P. J., Fast, J. D., Easter, R. C., Gustafson Jr., W. I., Liu, X., Ghan, S. J., and Singh, B. (2014): Assessing the CAM5 physics suite in the WRF-Chem model: implementation, resolution sensitivity, and a first evaluation for a regional case study, *Geosci. Model Dev.*, 7, 755–778, https://doi.org/10.5194/gmd-7-755-2014.

- Fast, J. D., Berg, L. K., Zhang, K., Easter, R. C., Ferrare, R. A., Hair, J. W., Hostetler, C. A., Liu, Y., Ortega, I., Sedlacek III, A., Shilling, J. E., Shrivastava, M., Springston, S. R., Tomlinson, J. M., Volkamer, R., Wilson, J., Zaveri, R. A., and Zelenyuk, A. (2016): Model representations of aerosol layers transported from North America over the Atlantic Ocean during the Two-Column Aerosol Project, *J. Geophys. Res. Atmos.*, 121, 9814–9848, https://doi.org/10.1002/2016JD025248.

RC: *According to the authors' description, the proposed nudging configuration appears conceptually similar to the grid-nudging approach used in WRF, in which ERA5 is treated as pseudo-observations and interpolated to each model grid point to compute the local nudging tendency. If this is indeed the case, the physical meaning and practical role of the spatial influence weighting become difficult to interpret, since each grid point already possesses its own "true" value derived from ERA5. The authors should clarify how the spatial influence weighting operates within this framework and explain its necessity and impact when the nudging target field is already defined at every grid point.*

AR: We thank the reviewer for this helpful comment. The reviewer is correct that our nudging configuration in E3SM is conceptually similar to the grid-nudging approach used in WRF. However, the practical implementation of nudging in E3SM differs from that in WRF because of the distinct model infrastructures and scientific objectives. In our application, nudging in E3SM is designed primarily for reanalysis nudging, where the reanalysis fields are first interpolated horizontally to a grid that matches the E3SM model grid. In this way, reanalysis information is available at every model grid point, and the horizontal weight of nudged fields is simply set to $w_m = 1$ (i.e., it does not vary geographically). Based on previous tests and applications, horizontal interpolation errors have not been found to pose a significant problem for nudging in E3SM (Sun et al., 2019; Zhang et al., 2022), although sensitivity to certain numerical implementation choices has been documented and discussed in detail in Zhang et al. (2022). This is partly because, in most cases, reanalysis products such as ERA5 have horizontal resolutions that are higher, or at least comparable, to those of low-resolution climate models (e.g., ERA5 at 0.25° versus E3SM at roughly 1°), so the properly designed horizontal interpolation (e.g., the method from CESM2 (2025) used in our study) is generally quite accurate. In line with the reviewer's comment, however, we acknowledge that the treatment here can be contrasted with the observational-nudging framework in WRF, where real instrument observations are spatially sparse and cannot be mapped directly onto the model grid. In that context, spatial weighting schemes are essential to distribute observational influence appropriately across the domain.

The vertical interpolation, on the other hand, can introduce problems, especially when the model top in E3SM differs from that in the reanalysis datasets. This issue is part of the motivation for our further exploration of the vertical modulation (or vertical weighting) scheme in this study on nudging in E3SM. Following the reviewer's comment, we add some clarifications in our revised manuscript to explicitly convey how weighting is applied in the default and revised nudging framework in our study when we discuss the method (lines 83–100):

> In the conventional EAMv2 configuration (Sun et al., 2019; Zhang et al., 2022), the nudging tendency is written as
>
> $$\dot{\boldsymbol{R}} = -\, w_m \, \frac{\boldsymbol{X_m} - \boldsymbol{X_r}}{\tau}, \qquad (2)$$
>
> where $\boldsymbol{X_m}$ is the model state variable (U, V, T, or Q), and $\boldsymbol{X_r}$ denotes the corresponding value from the reference dataset, often provided by ERA5 reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF, Hersbach et al., 2020). The scalar weight w_m controls the nudging strength. In the default design we have
>
> $$w_m = 1.0, \qquad (1)$$
>
> which is spatially and vertically uniform (i.e., identical at all model levels and grid points within the nudging domain) and has been shown to perform efficiently in previous applications (e.g., Sun et al., 2019; Zhang et al., 2022).
>
> In this study, we propose a more generalized formulation of Eq. (3) by allowing w_m to vary with height.

Specifically, we define w_m as a function of model-layer pressure and geopotential height,

$$w_m(P_m, Z_m) = \begin{cases} 0, & P_m \leq 1\,\mathrm{hPa}, \\ \dfrac{P_m}{P_0}, & 1\,\mathrm{hPa} < P_m \leq P_0, \\ \dfrac{1}{2}\left[1 + \tanh\left(\dfrac{Z_m - Z_b}{0.1\,Z_b}\right)\right], & P_m > P_0 \text{ and } Z_m \leq Z_b, \\ 1, & \text{otherwise.} \end{cases} \tag{4}$$

where P_m and Z_m denote, respectively, the pressure and geopotential height at the midpoints of the model layers. The parameter Z_b represents a user-defined height threshold below which model layers are treated as near-surface layers, while P_0 is a user-defined pressure threshold. Distinct from the default formulation in Eq. (3), the proposed Eq. (4) allows the nudging strength to (i) transition smoothly to zero in the upper atmosphere and (ii) weaken gradually within the PBL, while retaining full strength in the free troposphere. A schematic illustration of the weighting implied by Eqs. (3)–(4) at E3SM model levels is shown in Fig. C1. These modifications to the nudging formulation are motivated by two key considerations discussed below.

We hope that these revisions and additional explanations improve clarity, address the reviewer's concerns, and help readers better understand the nudging strategy used in this study.

RC: *For the third group of experiments, it is recommended to include additional nudging experiments using the original formulation in Eq. (2). I noticed in Fig. 3 that, particularly for temperature (c3, c4) and water vapor (d1–d4), the errors near the lower and upper boundaries of the atmosphere are noticeably larger than those in the mid-troposphere. I would like to know how the assimilation results would appear when the original vertical weighting scheme is applied, and to quantitatively assess the improvement achieved by the revised vertical weighting approach.*

AR: Thank you for this instructive suggestion. We would like to emphasize that the primary motivation of this study, as stated in the Introduction (lines 44–46), is to develop nudging strategies that incorporate temperature and humidity—together with large-scale wind fields—in order to **improve the realism of simulations nudged toward reanalysis while avoiding undesirable drifts in fields important for evaluation, such as precipitation and radiative fluxes**.

As discussed in Section 3, our short-term sensitivity experiments using both the original and revised formulations of the vertical weighting scheme show that the original formulation, when applied to humidity, leads to substantial degradation of the simulated hydrological cycle. In particular, humidity nudging under the original formulation produces a pronounced weakening of precipitation systems (Figure 1) and noticeably amplifies errors in cloud and radiative-flux fields. Because these impacts are large and systematic, additional long integrations (11 years) with the original formulation would not be scientifically informative for our objectives, which focus on physically realistic nudged configurations suitable for evaluation and downstream applications (see Section 4). Moreover, completing a full set of 11-year simulations would require substantial computational resources and time, which further underscores the need to design experiments carefully when there is no clear scientific benefit.

To address the reviewer's concerns, we provide clear justification of our choices in our revised manuscript (lines 263–267):

> Finally, because the short-term experiments reveal persistent and systematic degradations when the default formulation in Eq. (3) is applied to humidity, and because similar degradations have also been documented in longer simulations by Sun et al. (2019) and Zhang et al. (2022), we restrict the subsequent long-term (11-year) integrations to the revised nudging configuration. In the next section, we evaluate how this configuration performs in multi-year integrations and whether it produces realistic long-term climate statistics in E3SM.

RC: *Table 1: The variables for the second group of experiments are incorrectly labeled.*

AR: Thank you for noting this mistake. The variable labels in Table 1 have now been corrected in the revised manuscript, and we have double-checked the table for consistency with the text.

RC: *Line 165: ERA5 data are interpolated to a 30-minute interval, while the temporal coefficients for the upper atmosphere and surface are set to 6 hours and 1 hour, respectively. What is the rationale for choosing these specific settings?*

AR: Thank you for the good question. In order to address the reviewer's questions, we have made revisions in our paper with extra clarifications and citations for the retionale for the nudging setups in our study (see lines 185–204 in revised manuscript):

> It is important to note that all nudged EAMv2 simulations listed in Table 1 differ only in their choice of nudged variables and vertical weighting scheme; all other aspects of the nudging configuration follow the recommendations from prior studies (Sun et al., 2019; Zhang et al., 2022). The ERA5 atmospheric and land-surface reanalysis fields were originally produced on a $0.25° \times 0.25°$ horizontal grid with 37 pressure levels. These data were remapped to the EAMv2 model grid (the "pg2" horizontal grid with 72 hybrid sigma–pressure levels) and subsequently used as reference fields for the nudging simulations. The remapping procedures (both horizontal and vertical) followed the algorithms developed for nudging in the Community Earth System Model version 2 (CESM2) (CESM2, 2025). Furthermore, linear-function nudging was applied to the 3-hourly ERA5 reanalysis to derive nudging tendencies at each model time step (30 minutes). This configuration reflects a balance between the computational cost of online data processing and the demonstrated performance benefits reported in Zhang et al. (2022). The nudging relaxation time scale τ in Eq. (3) was set to 6 hours for the upper-atmospheric layers and 1 hour for the land-surface layer. The choice of $\tau = 6$ hours for the upper atmosphere is motivated by both practical experience and demonstrated performance in previous EAM/E3SM nudging studies, where it provides an effective constraint while avoiding excessive distortion of the model's internal dynamics (Sun et al., 2019; Zhang et al., 2022). For the land-surface layer, the choice of $\tau = 1$ hour was empirically determined in this study. The shorter relaxation time is intended to provide a tighter constraint on the model state, given the large variability associated with the spatial inhomogeneity of land-surface processes and the fact that the atmospheric model in EAMv2 is coupled to the land model at a 1-hour frequency. We acknowledge that the performance of the nudging may exhibit some sensitivity to alternative choices of the parameters discussed here, as well as to the tunable parameters P_0 and Z_b used in the proposed strategy in Eq. (4). However, we believe that these sensitivities do not alter our primary goal of demonstrating how the revised nudging strategy proposed in this study benefits E3SM nudging simulations, as discussed in the following sections.

RC: *Line 185: Could the authors further explain why the precipitation magnitude becomes smaller after assimilating Q?*

AR: Thank you for this comment. To address the reviewer's question, we have added additional explanation and discussion in the revised manuscript regarding the impact of specific-humidity (Q) nudging, drawing on both the results of this study and insights from our previous work (Sun et al., 2019; Zhang et al., 2022), particularly in relation to the default nudging configuration (see lines 220–229 in the revised manuscript):

> In contrast, substituting temperature with humidity (DNDG-UVQ; panel e) results in a noticeable suppression of precipitation intensity, despite improved spatial correlation. This degradation becomes more pronounced when both temperature and humidity are nudged alongside winds (DNDG-UVTQ; panel f), revealing a key trade-off in the traditional nudging formulation: nudging Q improves the location and timing of precipitation (higher PCC) but substantially worsens its magnitude relative to observations (higher RMSE). As shown in previous studies (Sun et al., 2019), in addition to the realistic large-scale circulation provided by wind (U,V) nudging, further nudging of humidity (Q) improves the large-scale moisture distribution and leads to a more realistic spatial organization of precipitation systems, thereby increasing PCC. At the same time, those studies also demonstrated that improved pattern correlation does not necessarily guarantee a more realistic precipitation intensity, likely because humidity nudging can introduce unrealistic feedbacks between the large-scale moisture field and other model components (e.g., convection and microphysics) through changes to the condensational heating and vertical motion.

We hope these revisions help address the reviewer's concern.

RC: *Figure 2b: Compared with DNDG-UV, DNDG-UVT shows reduced RMSE in LWCF, CLDTOT, and PRECT. However, why does DNDG-UVTQ exhibit a noticeably larger RMSE in these variables compared to DNDG-UVQ? Could the authors provide some discussion for this contrasting behavior?*

AR: Thank you for these valuable questions. We have revised the discussion in Section 3 related to Figure 2b, and added further explanation for the contrasting behavior of DNDG-UVQ and DNDG-UVTQ in lines 240–251:

> However, simulations that include humidity nudging using the default formulation (DNDG-UVQ and DNDG-UVTQ) show substantial increases in RMSE for precipitation, total cloud fraction (CLDTOT), and cloud-related radiative fluxes (FLNS, LWCF, and SWCF). Since consistent reductions are observed in large-scale model states such as PSL, TREFHT, and Z500, these degradations are likely attributable to the complex interactions of nudging with subgrid model physics. As reported in previous studies of nudging in the E3SM model series using configurations similar to the DNDG simulations here, including specific humidity in the nudging configuration (e.g., DNDG-UVQ) can directly influence parameterized convection by modifying moisture convergence (Sun et al., 2019), whereas temperature nudging (e.g., DNDG-UVT) has been shown to substantially suppress the model's adjustment to aerosol forcing, even in experiments nudged toward the model's own climatology (Zhang et al., 2022). When both temperature and humidity are nudged simultaneously (e.g., DNDG-UVTQ), these effects are easily triggered together, leading to an even stronger degradation of model performance in terms of simulated precipitation and aerosol life cycle, as reported by Ma et al. (2015).

RC: *Line 205: The relatively large errors in cloud, precipitation, and radiation variables may be partly attributed to the low-pass filtering nature of nudging, which can smooth out local temperature and moisture gradients. This filtering effect may be particularly detrimental to diagnostic variables with strong nonlinear characteristics, such as cloud, precipitation, and radiative fluxes, leading to amplified errors in their representation.*

AR: Thank you for pointing out this important aspect. We agree that the low-pass-filtering nature of nudging can influence the results, as discussed around line 205. Following the reviewer's comment, we have added additional discussion in the revised manuscript (lines 255–267):

> We also note that the low-pass-filtering nature of nudging, as implemented in Eq. (2), can smooth local temperature and moisture gradients, which in turn affects diagnostic variables with strongly nonlinear characteristics—such as clouds, precipitation, and radiative fluxes—leading to amplified errors in their representation. Such effects operate in both DNDG-UVTQ and RNDG-UVTQ in the free troposphere, but Eq. (4), relative to Eq. (3), is likely better at preserving strong thermodynamic gradients in the boundary layer owing to the omission of nudging there. With the current experiments, we cannot quantify the impact that near-surface temperature or moisture gradients have on diagnostic fields such as clouds or precipitation, and we leave such analyses to future research. Regardless, the comparison between RNDG-UVTQ and DNDG-UVTQ still demonstrates that the proposed nudging strategy significantly improves hindcast performance in EAMv2 by addressing key limitations of the default approach in handling humidity nudging. Finally, because the short-term experiments reveal persistent and systematic degradations when the default formulation in Eq. (3) is applied to humidity—and because similar degradations have also been documented in longer simulations by Sun et al. (2019) and Zhang et al. (2022)—we restrict the subsequent long-term (11-year) integrations to the revised nudging configuration. In the next section, we evaluate how this configuration performs in multi-year integrations and whether it produces realistic long-term climate statistics in E3SM.

RC: *Line 210: I remain cautious about the authors' conclusion that overly constraining thermodynamic fields necessarily damages the model's physical consistency. Although ERA5 is designed to be as physically consistent as possible, it does not guarantee strict consistency across all scales and variables. ERA5 is still a product of data assimilation under model constraints, and different variables are not assimilated simultaneously, which can lead to residual imbalances and physical inconsistencies. Therefore, conclusions based solely on comparison with ERA5 may not be fully convincing.*

AR: Thank you for this thoughtful comment. We would like to point out that the results in Figure 2 were derived by comparing E3SM predictions with multiple observational and reanalysis products, including ERA5 (large-scale dynamical and thermodynamic fields), CERES_EBAF (cloud and radiative fluxes), and GPCP (precipitation), as detailed in Table B1. In addition, we agree that the reviewer's concern regarding the physical consistency of ERA5 is reasonable and valid. Our original statement, however, was intended to describe behavior within the E3SM/EAM system itself. In this context, the phrase "model's physical consistency" referred specifically to the internal physical balance of E3SM/EAM, rather than to the physical consistency of ERA5. This interpretation is informed by previous work on nudging strategies in E3SM (Zhang et al., 2022), which showed that nudging thermodynamic fields such as temperature can introduce inconsistencies within the coupled cloud–radiation–dynamics system, even when the nudging target is derived from the model's own climatology. These inconsistencies arise because thermodynamic fields are strongly coupled to parameterized physics, and constraining them toward a climate state (e.g., reanalysis) that is not fully compatible with E3SM's internally generated balances — which themselves may be biased due to model errors and imperfections in physics parameterizations — can lead to unexpected model responses and unrealistic statistics in fields such as precipitation, clouds, and radiative fluxes. These related issues also motivate our discussion in Section 4.2 (lines 358–376 in the revised manuscript) regarding the behavior of the long-term nudging simulations and their potential connection to microphysics parameterizations. However, for this specific comment, we recognize that our wording may not have been sufficiently clear, which likely contributed to the reviewer's concern. To clarify, we have revised the text (lines 240–254 in the revised manuscript) as follows:

> However, simulations that include humidity nudging using the default formulation (DNDG-UVQ and DNDG-UVTQ) show substantial increases in RMSE for precipitation, total cloud fraction (CLDTOT),

and cloud-related radiative fluxes (FLNS, LWCF, and SWCF). Since consistent reductions are observed in large-scale model states such as PSL, TREFHT, and Z500, these degradations are likely attributable to the complex interactions of nudging with subgrid model physics. As reported in previous studies of nudging in the E3SM model series using configurations similar to the DNDG simulations here, including specific humidity in the nudging configuration (e.g., DNDG-UVQ) can directly influence parameterized convection by modifying moisture convergence (Sun et al., 2019), whereas temperature nudging (e.g., DNDG-UVT) has been shown to substantially suppress the model's adjustment to aerosol forcing, even in experiments nudged toward the model's own climatology (Zhang et al., 2022). When both temperature and humidity are nudged simultaneously (e.g., DNDG-UVTQ), these effects are easily triggered together, leading to an even stronger degradation of model performance in terms of simulated precipitation and aerosol life cycle, as reported by Ma et al. (2015). In contrast, the revised formulation in Eq. (4), as implemented in RNDG-UVTQ, effectively alleviates these issues, although it still shows some degradation in RMSEs for cloud and radiation fields relative to CTRL. As discussed further in Section 4.2, these residual errors may stem from the absence of nudging over land, compensating errors in the model system, as well as interactions between nudging and physics parameterizations.

RC: *A more reliable approach would be to conduct OSSE experiments, in which pseudo-observations are generated from EAMv2 forecasts, followed by nudging hindcast experiments and systematic evaluation. This would provide a more robust basis for assessing the physical consistency of the proposed nudging strategy.*

AR: Thank you for the valuable comment. If we understand correctly, the OSSE experiments suggested by the reviewer refer to Observing System Simulation Experiments, which are widely used in the data-assimilation community to assess the value of new or existing observing systems (e.g., satellites or radiosondes). In the context of nudging in E3SM, a related strategy has already been implemented in our previous work through a "nudging-to-climatology" configuration (e.g., Sun et al., 2019; Zhang et al., 2022), in which pseudo-observations are effectively produced by E3SM itself and subsequently used to constrain a nudged simulation toward its own climatology in order to evaluate the nudging strategy for physical consistency.

We agree with the reviewer that "nudging-to-climatology" experiments, implemented in a manner similar to OSSEs, represent a useful approach, as also demonstrated in our earlier studies. For example, in Zhang et al. (2022), we conducted a systematic comparison between the "nudging-to-climatology" and "nudging-to-reanalysis" configurations in the E3SM/EAM model family, and discussed how different nudging choices and strategies influence the physical consistency between nudged and free-running simulations. Such diagnostic investigations in the context of improving thermodynamic nudging, however, are beyond the scope of the present work. Nevertheless, we appreciate the reviewer's suggestion and have added a statement in the conclusion to explicitly highlight this point as an important future direction (lines 520–527 in the revised manuscript):

Moreover, we note that nudging temperature and humidity toward reanalysis can influence clouds, radiation, and precipitation by modifying how the model physics responds to the large-scale state (see discussion in Section 4.2). Although a full investigation of these interactions is beyond the scope of this study, they likely arise in part because many parameterizations were tuned within the biased mean state of the free-running control simulation, whose performance can be suboptimal when the system is subsequently operated under a nudging configuration. Future work using carefully designed experiments, including "nudging-to-climatology" frameworks as demonstrated by Zhang et al. (2022), will be important for developing a process-level understanding of these sensitivities and for clarifying how temperature and humidity nudging affects the model's internal physical consistency, thereby informing parameter tuning in future E3SM model development.

RC: *Figure 3: The errors at different vertical levels should be analyzed in conjunction with the prescribed vertical weighting scheme, as discussed in Major Comment 3.*

AR: Thank you for this comment. We understand the reviewer to be requesting a more explicit discussion of the implications of the vertical weighting scheme proposed in this study for the results shown in Figure 3. In the revised manuscript, we have added an additional schematic (Figure C1) to illustrate the behavior of the weighting scheme and have expanded the discussion of Figure 3 (lines 291–301):

> One notable exception is the tropical lower troposphere (1000–800 hPa), particularly for temperature and humidity, where the nudging exerts relatively weak constraints compared to the upper troposphere. This behavior is partly explained by how the vertical weighting in Eq. (4) modulates the nudging within lower-tropospheric regions. As shown in Figure C1, the direct nudging tendency term is effectively zeroed out near the surface by Eq. (4). However, once the nudging tendencies are applied, the model dynamics and physical parameterizations respond immediately, such that corrections introduced in one constrained variable at one level or grid point influence other variables, as well as other layers and neighboring locations, through coupled dynamical and physical processes. Through this pathway, indirect constraints from variables and layers above the lower troposphere still act to reduce biases throughout most of the lower troposphere, as seen by comparing NDG-UV (Fig. 3a2–d2) with CLIM (Fig. 3a1–d1). A similar mechanism operates in NDG-UVT (Fig. 3a3–d3) and NDG-UVTQ (Fig. 3a4–d4); however, in these configurations the interactions between the nudging tendencies—particularly for temperature and humidity—and the model physics also influence the resulting bias structures. These effects are discussed further in the next subsection.

We believe these revisions, together with the added explanation, clarify the potential interactions between the vertical weighting scheme and the nudged simulations and their implications, and we hope that this addresses the reviewer's concern.

RC: *Section 4.2: add description of the parameter settings used for nudging the surface variables*

AR: Thank you for this comment. The parameter settings for the NDG-UVTQ_SRF1 experiment, including those related to nudging surface variables, are already described in Section 2.3 (now lines 173–184 in the revised manuscript). We believe Section 2.3 is the most appropriate place for this information, and repeating the same details in Section 4.2 would be redundant. In response to the reviewer's suggestion, however, we have added further clarification in Section 2.3 to present the parameter settings more explicitly. We hope that this explanation satisfactorily addresses the reviewer's concern.

RC: *Figure 9: Full-variable nudging (red line) appears to suppress the peak of the seasonal cycle of the GPI, please provide an explanation*

AR: Thank you for this comment. The comparison of Figure 9 is focused on how well the nudged simulations reproduce the seasonal cycle of GPI relative to the observational reanalysis (ERA5). We believe that the authors's questions here arise from features shown in panels (b) (c) for eastern and weather Pacific basins and panel (g) for overall Northern Hemisphere, where the evolution of GPI in NDG-UVTQ_SRF1 closely track those in ERA5, while the magnitude of seasonal peak of GPI are lower than ERA5. Following this understanding, we extended our discussion related to Figure 9 with additional explanations in our revised manuscript (lines 438–460):

Nevertheless, constraining both dynamical and thermodynamical fields is essential for improving the model's simulation of the seasonal and annual cycle of the GPI (Fig.9). Among all experiments, the NDG-UVTQ_SRF1 simulation (red lines, Fig. 9), which includes additional constraints on near-surface humidity, generally provides the best reproduction of the observed variability and seasonal peaks of GPI across most tropical-cyclone basins, with improved agreement relative to ERA5 (black lines). An exception occurs in the Eastern and Western Pacific, where the significant overestimation of GPI in the free-running control (CLIM) is reduced when large-scale wind and temperature nudging is applied, but the seasonal GPI peak becomes weaker when humidity nudging is further included in NDG-UVTQ_SRF1 (panels b and c). The underlying reason is partly informed by the results in Figure 8 and Eq. (A1): the GPI bias arises from compensation among the component fields that enter the GPI formulation. In CLIM, the positive biases in RH and vertical wind shear (Vshear) are larger in magnitude than the negative bias in Vpot in these regions, resulting in only partial cancellation. Because the RH bias is particularly large, it dominates the total GPI error and ultimately leads to the substantial overestimation seen in CLIM. For the nudged simulations without humidity (i.e., NDT-UV and NDT-UVT), the compensation occurs primarily between Vpot and RH, and the magnitudes of their errors are of similar size in these regions, leading to a smaller net GPI bias. However, when humidity nudging is included in NDG-UVTQ_SRF1, the RH bias is substantially reduced, breaking this compensation, and the remaining negative bias in Vpot results in a weaker seasonal GPI peak in the Eastern and Western Pacific. We note that such understanding would not have been possible without a complete comparison across the series of nudged simulations conducted using the strategy proposed in this study. This indicates that accurately constraining both dynamical and thermodynamical conditions in numerical models is valuable for representing the large-scale environment and for gaining diagnostic insight relevant to tropical cyclones. Given the strong statistical relationship between GPI and TC genesis density (r values in Fig. 9), the result here also highlight the importance of realistically simulating the large-scale temperature and humidity fields that support TC development. In particular, the humidity-nudging strategy proposed in this study provides a practical approach to improving the thermodynamic environment relevant to tropical-cyclone formation, thereby enabling more reliable TC downscaling and associated studies from low-resolution GCMs such as EAMv2.

We hope that these additional clarifications helps address the reviewer's concern.