Replies to comments by Reviewer 2

Gerhard Krinner, Aude Champouillon, Juliette Blanchet and Frédérique Chéruy

April 2, 2026

We sincerely thank all three reviewers for their thoughtful comments and suggestions which we have taken into account in the revised version of this article.

Major points

Main comment #1: The introduction in its current form does not sufficiently articulate the motivation for the proposed approach or clearly explain its potential impacts on model performance and predictive skill. For example, the study by Guldberg et al. (2005), which is cited in this manuscript, along with the related literature therein, provides an important foundation on empirical model correction and the reduction of long-term systematic errors. A more explicit discussion of this prior work in the introduction would help readers better understand the scientific context, value, and fundamental advances underlying the exploration presented in this study. In addition, I would emphasize that it is important to clearly distinguish the target of the bias-correction problem addressed in this study—namely, the correction of long-term systematic errors—from the weather-scale bias correction typically associated with classical nudging. Clarifying this problem setup in the introduction would help readers better understand the intended scope of the proposed method and avoid potential confusion regarding its relationship to standard nudging approaches. Also, as will be mentioned in comments below, some existing similar studies should be mentioned as an background for the introduction of this study.

Reply : We agree that a clearer motivation of this study, placing the objectives more explicitly within the framework of existing literature, is useful. We thank the reviewer for suggestion of additional work to cite (in the following point they made). In the revised version, we aimed at more clearly distinguishing this work, aiming at developing bias-correction methods for climate change applications, from another cluster of published literature aiming at bias corrections in a weather and seasonal prediction context. This leads to substantial changes in the introduction, where we now cite several of the studies mentioned by the reviewer, refer to studies of technical choices in nudging applications as suggested by several reviewers, and situate our work more clearly with respect to these studies.

Main comment #2: As described in Section 2.1, the nudging-based bias-correction framework in this study consists of two key steps: the construction of cyclostationary climatological nudging increments using a classical nudging procedure that constrains the model toward a reanalysis product, and the subsequent application of these increments, either directly or recursively, to free-running model simulations to reduce long-term systematic biases in the simulated climate. Although the primary focus of this study is on the second step, the methodology used to generate the nudging increments in the first step is essential in determining the behavior of the recursive nudging approach, including the number of iterations required to achieve optimal performance. In particular, the choice of the nudging relaxation timescale τ directly affects the strength of the diagnosed increments, raising the question of whether stronger nudging (e.g., τ = 6h instead of τ = 24h) would require fewer iterations to reach performance comparable to the two-iteration case shown here, while weaker nudging (e.g., τ = 48h) might require more iterations to achieve optimal performance. More generally, as discussed in many previous studies (e.g., Zhang et al., 2014; Sun et al., 2019; Zhang et al., 2023), a key challenge in nudging formulations such as Eq. (2) is determining an optimal choice of τ that is strong enough to reduce model biases while remaining weak enough to avoid undue interference with the model's intrinsic physics and dynamics.

The appropriate choice often depending on the scientific purpose of the nudged simulations. Similarly, for the recursive approach proposed in this study, if the number of iterations required for optimal performance is highly sensitive to the nudging configuration, this may pose a challenge for the generalization of the method to other applications and model and raise questing on the value of the method to the modeling community proposed by this study. From a machine-learning perspective, the iterative nudging framework can be viewed as a form of repeated residual correction, in which correction tendencies are successively accumulated to reduce systematic error. In this context, the sensitivity of performance to parameters such as the nudging timescale and iteration number is closely related to issues of stability, regularization, and generalization, and therefore merits further discussion.

Overall, I think that the current version of the manuscript would benefit from additional discussion addressing the points raised above, that an expanded discussion in this regard would help strengthen the value and impact of the proposed method in this study.

References:

- Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman, G. J., Ma, P.-L., Rasch, P. J., Neubauer, D. and Lohmann, U.: Technical Note: On the use of nudging for aerosol–climate model intercomparison studies, Atmos. Chem. Phys., 14, 8631–8645, https://doi.org/10.5194/acp-14- 8631-2014, 2014.
- 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. Journal of Advances in Modeling Earth Systems, 11, 3911–3933. https://doi.org/10.1029/2019MS001831
- Zhang, S., Zhang, K., Wan, H., and Sun, J.: 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, 2022

Reply : This comment is similar to points made by the other reviewers. We agree that we needed to explicitly mention this aspect in the revised version. As stated also in the replies to the other reviewers, we have carried out simulations with different nudging constants and will present results in more detail in ongoing work that aims at comparing different ERBC methods. Again, the main objective of the present article is to present the iterative method as such, for which the chosen time constant is the most appropriate in our setting. Therefore, as stated before, we think that it is appropriate here to mention the fact that various time constants have been tested, that τ =1 d is overall the most convincing choice in our case, and refer to very recently submitted work for a more detailed discussion (Champouillon et al., 2026) (the article is not published on the EGU website yet, but can be downloaded here: https://cloud.univ-grenoble-alpes.fr/s/5SM9Hqq9xcfLwWX). This is mainly done in the new section 4.2 "Effect of the nudging time constant τ" in the revised version, which contains a new figure intended to illustrate the reviewer's point that iterating the ERBC procedure increases the nudging strength, but is not identical to a single iteration with an appropriately chosen nudging timescale. The papers referred to here by the reviewer are now cited in the introduction (see previous point).

Main comment #3: Closely related to the above comments, the approach discussed in this study appears to be conceptually equivalent to the tendency bias correction (TBC) framework described in Chang et al. (2019). It would be helpful for the authors to explicitly acknowledge this connection and clarify similarities and differences between the two approaches, particularly in terms of methodology, assumptions, and intended applications. The primary distinction between TBC and the approach proposed in this manuscript appears to lie in the method used to estimate the climatological tendency-bias correction terms. In TBC, the correction tendencies defined as climatological 6-hourly mean differences between the model and observations (reanalysis), such definition by itself is general for any model systems and in absent of the dependence on the empirical nudging relaxation time scale τ. To me, the approach presented in this study, [text seems to be missing?]

References:

- Chang, Y., S. D. Schubert, R. D. Koster, A. M. Molod, and H. Wang, 2019: Tendency Bias Correction in Coupled and Uncoupled Global Climate Models with a Focus on Impacts over North America. J. Climate, 32, 639–661, https://doi.org/10.1175/JCLI-D-18-0598.1.

Another factor that may influence the performance of the correction is the length of the correction simulations. By definition, the nudging-based tendency bias correction targets slowly evolving, long-term systematic errors, and therefore may be sensitive to the length of the period used to estimate the correction terms. For example, a 40-year continuous correction simulation may outperform a 20-year simulation, as a longer integration provides a more robust estimate of the climatological tendency biases and allows the cumulative effects of the correction to be more fully realized. A brief discussion of this sensitivity would help clarify the robustness of the proposed correction approach. Some analysis and discussions on this regard can be useful to enhance the value of the study. For example, the authors could compare the mean bias reductions for the first 10 years and the last 10 years of the free-running correction period (2001–2020) to assess whether the correction effectiveness show differences for two 10-year periods. Finally, the climatological mean bias-correction terms are estimated over the period 1981–2000, but are applied to correct the free-running model simulations during 2001–2020. However, the climatological states in both ERA5 and the model may differ between these two periods. How might such differences affect the effectiveness of the iterative bias-correction approach and the determination of the optimal number of iterations?

Reply : Concerning the first part of this comment, it is true that the approach used by Chang et al. (2019), based on the data assimilation procedure developed by Bloom et al. (1996), is similar but not identical to the nudging technique underlying the ERBC approach we use. The main difference, as the reviewer rightly states, consists in the nudging step that is slightly different in the work used by Chang et al. However, we think that a lengthy discussion of these differences would distract the reader from the main point of the paper. We therefore mention the work referred to by the reviewer in the revised introduction without going into specific details of the data assimilation method:

> A number of studies have used this approach, or slight variations thereof (e.g., Bloom et al., 1996), to reduce systematic model errors in a weather or seasonal prediction context (e.g., Kharin and Scinocca, 2012; Chang et al., 2019).

In the second part of the comment, the reviewer makes a valuable point that is also made by reviewer 1. This refers to the discussion section on over-tuning and out-of-sample testing. Indeed, potential interference of the climate change signal with "pure" out-of-sample effects cannot be excluded. As requested by both reviewers, we have addressed this point in the discussion (in section 4.3, "Over-correction: Out-of-sample vs. in-sample evaluation") by adding a paragraph on this issue :

> It is possible, however, that there is some interference between possible over-correction and the fact that the study period, 1981–2020, is a period of strong climatic change. Although it has been shown before that nudging-based ERBC remains valid under strong climate change (Krinner et al., 2020), part of the performance over the 2001–2020 period might therefore be influenced by the climate change signal. One could, as an additional test, use 2001–2020 as the ERBC calibration period and 1981-2000 for validation, or use pair years between 1981 and 2020 for the ERBC calibration and odd years of the same period for out-of-sample testing. However, the main motivation for our use of various ERBC approaches is to eventually use it in climate change simulations. In that sense, testing the effect of the bias corrections in a warmer climate is not necessarily a drawback – it is, to some degree, a prerequisite for the intended use of the ERBC approach.

We are a bit less sanguine about the reviewer's suggestion to evaluate the biases over 10-year periods, because that is a quite short period for meaningful evaluations, given natural climate variability. But we think we have taken the reviewer's point on board by referring to the possibility to use the last 20 years of the 1981–2020 period for ERBC calibration and the first 20 years for testing.

Main comment #4: The design of the iterative run-time bias correction is somewhat confusing. Based on Section 2.1 and Table 1, my understanding is the following:

- a. The N_x groups always employ Eq. (4), which corresponds to classical nudging using ERA5 during 1981–2001, with an additional climatological mean bias-correction term defined as the average of the nudging tendencies over 1981–2001 from the previous iteration (N_{x-1}).
- b. The C_x groups always employ Eq. (5), in which a climatological mean bias-correction term (independent of the model equation and estimated from the N_x simulations) is applied directly to the model equations.

Is this interpretation correct? In addition, it is unclear whether the Nx and Cx experiments are performed as independent integrations (each initialized from the same initial condition and differing only in the applied correction term) or whether subsequent iterations (e.g., N2) are continued from the end state of the previous iteration (N1). This distinction is important because it affects how the iterative procedure should be interpreted. I suggest that the authors provide a clearer and more explicit explanation of the experimental design and the iterative bias-correction framework to help reader better understand.

Reply : Concerning the first part of this comment, the reviewer's interpretation is correct. We have tried to clarify this in particular by adapting the description of the runs in Table 1. Concerning the second part of the comment, where the reviewer requests clarifications on the initial conditions used in the simulations, we have clarified this by adding specific information:

> The corrected simulations C_i ($i = 0 \dots 3$) and the uncorrected control simulations denoted M are run for the evaluation period 2001-2020 (plus the year 2000 for model spinup), distinct from the nudging period. These simulations are run twice, starting in 2000 with varying initial conditions. For these simulations we use atmospheric states obtained from other uncorrected LMDZ simulations for the years 1995 and 2000, respectively, and discard the first year of the integration as spinup, as just mentioned.

Because these corrected simulations are only driven by prescribed SST and atmospheric composition, their initial state is without much influence because chaotic unforced atmospheric variability rapidly becomes dominant and simulations can be considered as independent realizations after the first year of integration which is discarded as spinup.

Main comment #5: (This is actually the second half of the reviewer's long main comment #4.) Moreover, it is hard for me to understand what the purpose of the N1, N2 and N3 experiments and the resulting nudging tendency. The N_0 was the classical nudging towards reanalysis (i.e. ERA5), which was used to estimate the climatological bias correction tendency terms C_0 to be used and applied to the free running simulations. Here, C_0 can be interpreted as the mean tendency correction required to offset the model's systematic drift relative to reanalysis. However, for simulations starting from N_1, C_0 is added in addition to the ERA5 nudging terms. In this configuration, it becomes unclear how the subsequently diagnosed correction terms (e.g., C_1 derived from N_1) should be interpreted. As noted above, C_0 is diagnosed from the nudging tendency term, $-\frac{1}{\tau}(X - X_{\mathrm{ERA5}})$. The explicit dependence on the relaxation time scale τ implies that C_0 may vary with different choices of τ, and therefore may not represent an accurate or intrinsic estimate of the climatological mean model tendency bias. This may leave room for the subsequent iterations (N_1, N_2, etc.) to act toward a progressively improved estimate of the climatological bias-correction tendency through the iterative procedure. However, such convergence is not guaranteed, given the nonlinear and online interactions between the nudging terms and the model's physical and dynamical processes. Overall, the manuscript would benefit from a clearer and more systematic explanation of the theoretical basis and practical implementation of the iterative bias-correction framework, including its physical interpretation, potential deficiencies or limitations, and the key considerations underlying its implementation.

Reply : This comment is somewhat similar to points made by the other reviewers. First, we would like to clarify what might be a slight misconception: The initial correction term G_0 is very far from offsetting the model's systematic drift - and that is certainly one reason why large biases remain in the model when the "classical" ERBC method is used. We clarify this at the end of the "Methods" section:

> Because the nudging timescale τ is of the order of days and the model timestep of the order of minutes, the initial correction term G_0 derived from equation 1 only partially offsets the model's systematic drift relative to the reanalysis. Similarly, the subsequent correction terms G_{0+1} and more generally $G_{0+1+...+n}$ only partially offset this systematic drift. The question of potential convergence is addressed, among other results, in the remainder of this paper.

Following the reviewer's request, we then analyze the nudging and correction terms, and in particular the question of potential convergence to also comply with a request by reviewer 1, at the end of section 3.1:

> Do these successive correction terms converge towards some "final" correction term? Figure 1d shows that the nudging increments in the third iterated nudging step N_3 do not vanish, although they are substantially weaker than in N_0 (Figure 1a). The global mean of the absolute zonal wind nudging tendencies (in January, to be consistent with Figure 1) is 0.50 m/s/day for N_0, 0.35 m/s/day for N_1, 0.29 m/s/day for N_2, and 0.26 m/s/day for N_3. This means that the intensity of the remaining nudging tendencies decreases at higher iterations, but convergence towards a potentially vanishing final term is still far away. The combined correction terms arising from the sum of these absolute zonal wind nudging tendencies have global mean values of 0.50 m/s/day for G_0 (because they are identical to the mean nudging tendencies of N_O), 0.83 m/s/day for G_1, 1.09 m/s/day for G_1, and 1.30 m/s/day for G_3, and are thus somewhat lower than the corresponding sums of the global mean of the absolute zonal wind nudging tendencies (which would be 0.5, 0.85, 1.14 and 1.40 m/s/day, respectively), indicating that some local-scale drift compensation occurs between different iterations, as already shown by Figure 2.

Furthermore, the last part of the reviewer's comment, concerning the effect of the nudging time constant and more discussion of the presented approach, is very similar to a request by reviewer 1. As already stated in the reply to comments by reviewer 1, we have added a new subsection (4.2 "Effect of the nudging time constant τ") in the discussion section.

Other specific comments

Minor point #1: Line 9: The phrase "However, while . . . " sounds somewhat awkward. Please consider revising the sentence.

Reply : Agreed. We now simply write:

> However, signs of over-correction appear after about three iterations.

Minor point #2: Line 35: Consider replacing "more perfect" with "more effective."

Reply : Done.

Minor point #3: Line 104: "root-mean-square error (RMSE)" is defined multiple times throughout the manuscript. In addition, the terms "root-mean-square error," "RMSE," and "root-mean-square error (RMSE)" are used inconsistently. Please consider using a consistent format throughout the manuscript.

Reply : We now define the acronym at the first occurrence of "root mean square error" and subsequently only use RMSE.

Minor point #4: Section 3.1: Could you elaborate on how the results in Figure 1 should be interpreted in terms of the nudging procedure? In addition, you state that "Instead, this ratio has distinctive spatial structures that vary in space and time throughout the annual cycle (not shown), indicating that the iterative procedure is not identical to a simple uniform amplification of the initial correction." What is the underlying reason for this behavior? Finally, I am confused by Figure 2, why the ratio of C3/C0 are all above 1.5, given that the magnitude in Figure 1d are all smaller than those in Figure 1a?

Reply : In fact, the atmosphere being a highly nonlineear system, we would have been rather surprised if the iterative procedure would have turned out to be identical to a simple uniform amplification of the initial correction. We write:

> We interpret this as being linked to the non-linear and non-local nature of the atmospheric system, where the bias corrections implemented during a correction step can increase tendency errors elsewhere, which are then corrected during the next iteration, and where spatially unequal bias reduction at one iteration can also lead to modified tendency errors, and thus subsequent bias reduction, in the following iteration.

Concerning Figure 2, we apologize for a lack of precision in the first version of the article. We clarify the caption of Figure 2:

> Ratio of the zonal mean absolute zonal wind bias correction terms (averages for January 1981 to 2000) between those used in C_3 (bias correction terms $G_{0+...+3}$) and those used in C_0 (bias correction terms G_0). ...

This more clearly expressed that the Figure displays the ratio of the cumulative correction terms.

Minor point #5: Section 3: The authors present their discussion using only metrics focused on the U and V components. It would be valuable to include analyses of other fields, such as temperature, humidity, precipitation, and sea level pressure etc., which are not directly nudged but may still be significantly affected by the U and V nudging. This will be also valuable supports to your discussions in Section 4.

Reply : Although we see the reviewer's point, we note that Figures 5 and 6 do display temperature errors, and their reduction. Given that, as we state, the main objective of the work here is to reduce circulation errors, we think it is preferable to keep the paper focused on its main points. In-depth analysis of additional variables would be beyond the scope of the paper.

Minor point #6: Section 4: Some of the discussion in Section 4.4 could be made more meaningful by linking it more directly to the results of this study (e.g. section 4.1 and section 4.2), rather than relying on speculative or high-level arguments. I also suggest that the authors focus the discussion more closely on parameters and methodological choices related to the proposed approach, instead of emphasizing broader aspects such as model resolution (e.g. section 4.3). To me, broader considerations could be better briefly summarized in a short paragraph in the conclusion.

Reply : The two parts of the discussion that are least directly linked to the presented results were section 4.4 and 4.6, as the reviewer rightly states. But while it is true that these subsections are not directly linked to presented results, they address a point that is treated in the discussions section of almost every scientific paper, that is, the limitations of the study. We would therefore prefer to keep these points here, but to group them in one subsection for clarity. We suggest to combine subsections 4.4 and 4.6 with the subsection 4.1 ("Choice of bias-corrected variables") in a new subsection 4.1 "Limitations linked to the design of the model simulations", combining the old sections 4.1, 4.4 and 4.6 as new subsubsections 4.1.1-3.

References

Champouillon, A., Krinner, G., and Blanchet, J.: Intercomparison of run-time bias correction methods in LMDZ_v6.3, Geoscientific Model Development, submitted, 2026.

Krinner, G., Kharin, V., Roehrig, R., Scinocca, J., and Codron, F.: Historically-based run-time bias corrections substantially improve model projections of 100 years of future climate change, Communications Earth & Environment, 1, 29, 2020.