Replies to comments by Reviewer 1 (J. Scinocca)

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: Dependence of results on value of τ in the classical method

In Scinocca and Kharin (2024), hereafter referred to as SK24, a systematic analysis was presented in Section 3 of how the properties of the nudging applied in the N_0 calibration simulations, most notably the value of the nudging timescale τ (and, more generally, the spatial filtering of the nudging tendencies), affect the amount of bias reduction achieved in the corresponding ERBC model simulations (i.e. C_0). The results of that analysis were summarized in Fig. 1 of SK24. For convenience, I reproduce that figure here. Panel a shows the global bias reduction of the prognostic fields corrected in the nudging runs (R-ADAPT in that study or N_0 here), while panel b shows the associated global bias reduction in the ERBC model simulations (R-ERBC in that study or C_0 here) as a function of tau and the spatial filtering applied to the nudging tendencies. As the present study does not employ spatial filtering, the relevant information corresponds to the information along a horizontal line at the top of this figure (i.e. grid-point nudging).

In the absence of spatial filtering, SK24 identified an optimal nudging timescale of approximately τ = 3d (for the N_0 runs) that yields the smallest biases in the corrected ERBC model simulations (C_0). As discussed in the final three paragraphs of Section 3 of SK24, this optimal value lies between weaker nudging, which limits the effectiveness of the bias correction, and stronger nudging, which introduces artifacts associated with the action of unbalanced motions. There, SK24 argued that, for the classical approach, a separation between balanced and unbalanced dynamics cannot be achieved in principle, and can only be approximated in practice through appropriate choices of the temporal (and spatial) properties of the relaxation applied in the nudging runs.

While the specific optimal value of tau will be model dependent, the analysis presented in SK24 suggests that there may be scope to further reduce biases in the C_0 correction runs of the present study through an adjustment of τ toward longer timescales. Given that the authors explicitly consider both the efficiency of the process (e.g. Section 4.5) and the pursuit of "more perfect" bias reductions, it would therefore seem important to identify the optimal value of tau for their configuration, denoted here as τ_{opt}. This could likely be accomplished with a relatively small number of additional experiments and would provide a useful baseline against which the performance of the iterative approach could be assessed.

Consideration of the role of τ_{opt} in this study raises several important questions:

a) Does the reduction of biases achieved using the iterative approach with τ = 1 d exceed the reduction obtained in a C_0 simulation using τ_{opt}?

b) If τ_{opt} were used instead of τ=1d within the iterative framework, would the approach still yield a meaningful additional improvement? The authors show in Fig. 1 that the effective strength of the nudging decreases with each successive calibration simulation N_i. Reducing τ from 1 d to the optimal value (approximately 3 d) similarly weakens the nudging. This raises the possibility that the iterative approach may, at least in part, represent a more computationally expensive means of approaching the same optimal bias reduction that could be achieved more directly through an appropriate choice of tau in the classical method.

Clarifying these issues would seem essential for assessing the practical utility of the iterative approach proposed in this study.

Reply : This is indeed an important point, and it is also raised in similar terms by Reviewers 2 and 3. We have indeed carried out simulations with varying nudging time constants and found that both in a classic (non-iterative) approach and in the case of iterations (point of the paper here), the selected time constant $\tau = 1$ d is the best choice in our model, for the range of time constant tested. We show this in detail recently submitted work (Champouillon et al., 2026) where we compare systematically the "classical" approach, a revised implementation of the CABCOR approach in LMDZ (see one of the following comments), the iterative approach described here, and a state-dependent variant of the "classical" approach (the article is not published on the EGU website yet, but can be downloaded here: https://cloud.univ-grenoble-alpes.fr/s/5SM9Hqq9xcfLwWX). The scope of the present work is to present the iterative procedure as such, and in the discussion we therefore mention here that $\tau = 1$ d is the optimum choice in our context (in the new subsection "Effect of the nudging time constant τ" under "Discussion"), and refer to Champouillon et al. (2026) for details. This new subsection (4.2) also compares correction tendencies of a non-iterated simulation with $\tau = 1$ d to those of a test simulation with twice iterated bias correction and $\tau = 3$ d, in response to the reviewer's question b) above. A new figure added in this new section of the revised version of the article shows that the iteration procedure is not, as the reviewer puts it, "a more computationally expensive means of approaching the same optimal bias reduction that could be achieved more directly through an appropriate choice of tau in the classical method". We then refer to the above-mentioned recently submitted article for a more detailed evaluation.

Main comment #2: Out-of-sample validation.

An important aspect of the present study is the evaluation of the effectiveness of the ERBC model outside the period used to derive the bias correction. The authors have used this approach to identify the point at which bias reduction in the out-of-sample period ceases and iterations should be stopped – even if the correction continues to reduce biases when validated during the calibration period (e.g. ll 213-216).

If the whole system were stationary, then this out-of-sample validation approach would be appropriate. However, the period 1981-2020 contains some of the strongest historical climate-change forcings. So it is not strictly stationary. For example, in Krinner et al. (2020), idealized experiments were performed to evaluate the efficacy of the ERBC approach under time-evolving climate change forcings. There, it was found that the bias reduction achieved through ERBC evolved with time (e.g. Fig. 1 of that study). In principle, the out-of-sample validation period (2021-2020) has more of a climate-change signal than the calibration period (1981-2000). It is expected, therefore, that there will be some degradation of the impact of the ERBC during the validation period due to the change in external forcings. While the climate-change signal might be small during this period, so too is the degradation of the bias reduction induced by the ERBC after 2-3 iterations.

This is a central aspect of the study's method. It is not clear how this can be evaluated within the study's experimental setup. At minimum, the authors should to include a discussion on this point and place caveats on their conclusions about terminating the number of iterations.

Reply : This is an excellent point. Indeed the climate system has been undergoing rapid change, as we all know. Therefore, as the reviewer states, there can be some confusion between the effects of out-of-sample-testing and climate change. One could indeed devise experiments such as calculating the bias-correction terms for all pair years between 1981 and 2020 (i.e., 1982, 1984, 1986,...2020), and testing the effect of these correction for all odd years (i.e., 1981, 1983, 1985,...2019). Or one could calculate the bias-correction terms for the latter half of the entire period (2001-2020) and evaluate the effect of the bias corrections for the first half of the period (1981-2000), and compare with the current setup. However, the main motivation for our use of various ERBC approaches is to eventually use it in climate change simulations, and as the reviewer states, we do have arguments to use it for that kind of applications. In that sense, testing the effect of the bias corrections in a changing climate is not necessarily a drawback; it is, to some degree, a prerequisite for the intended use of the ERBC approach. However, potential interference of the climate change signal with "pure" out-of-sample effects cannot be excluded. As requested by the reviewer, 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, evaluating the effect of the bias corrections in a climate that is warmer than the the calibration period, but still known, is a relevant test for the intended use of the ERBC approach.

Main comment #3: The relationship of ERBC and Model tuning.

The final step in the development of all climate models is a tuning exercise in which all free model physics parameters are assigned values based on a process which generally minimizes model biases during the historical period. ERBC is typically applied to a finalized model (what is called the free model in this study). Different finalized models have different inherent annual-cycle climatological biases. In this sense, the model physics was tuned to operate optimally in the presence of such biases. When ERBC is applied to the model, say on u, v, and T, it reduces the biases in the finalized model by construction. In principle, this will push the behaviour of physical parameterizations out of optimal performance. In practice, the extent to which this is problematic depends on the magnitude of biases in the finalized model, their overlap with specific physical processes of interest, and the ability of the ERBC to significantly reduce those biases.

There is a large cold bias throughout the troposphere of free (finalized) LMDZ model (Fig. 5), which parameterizations such as convection have been tuned to compensate for. Significantly reducing this bias by applying ERBC to T would throw such compensation out of balance. It is not unexpected then, that the convection in runs of LMDZ with T correction would alter convective activity in a negative way as indicated by the authors (l 177). Other models with less of a temperature bias in their finalized models might not suffer the same degradation in convective behaviour and so benefit from T runtime bias correction.

This model tuning argument would seem to be the better explanation for why some models benefit from T runtime correction while others do not. The authors, however, attempt to argue that there is a conceptual reason for limiting the runtime bias corrections to the dynamics (winds) to avoid interaction with the model physics. Model physics formulations are often developed and validated offline against observed inputs of winds, temperature, and specific humidity. The performance of the parameterizations should not inherently suffer when biases in such inputs are reduced. They suffer because the values of their free parameters were set in the presence of model biases in the finalized model configuration. The authors should offer up this explanation in Section 4.1 and where appropriate throughout the paper.

Reply : This is a fair point. More generally, the parameters of the ERBC approach (most importantly the nudging time constant τ), the set of corrected variables, the temporal and spatial resolution, the place of the ERBC within the GCM's time stepping scheme, and possibly other choices, are all to some degree model-dependent. And it is clear that, if we hadn't encountered problems with temperature ERBC in LMDZ, we would probably be using these. We have tested model tuning strategies to eliminate the cold bias of our model. However, there is no simple way to eliminate the cold bias in the model without simultaneously inducing unacceptable errors in other fundamental climate metrics such as TOA radiative fluxes, and, as stated, tuning of the LMDZ model was done with a clear focus on global radiative fluxes, disregarding mid-tropospheric temperature. Moreover, (Hourdin et al., 2021) report that "jointly tune radiation and convection is probably something which cannot be handled" with the tools at disposal at the time of the model development.

We are happy to clearly state this, as this was a point we had intended to make anyway and apparently failed to state it clearly enough. We therefore add the following sentences in the discussion:

> It is very likely that the package of physical parameterizations in LMDZ causes this cold bias, and at the same time counteracts temperature ERBC because it was consistently developed and tuned to comply with a certain set of observed metrics, on particular TOA radiative fluxes (Hourdin et al., 2021). In this model, complying with these radiative metrics appears to be in contradiction with an absence of a pervasive cold bias in the middle atmosphere, as tests of tuning strategies to eliminate the cold bias did not yield satisfying results in terms of TOA radiative metrics. We have to leave this problem unsolved for future work.

To further clarify our argument, we added the following sentences to this discussion:

> It can therefore make sense to separate atmospheric circulation structures from other variables such as temperature and humidity in bias-correction approaches. The effect of misplaced circulation features (for example, a misplaced storm track) often cannot be corrected a posteriori (Maraun et al. 2017), while a posteriori bias corrections of "physical" variables related to surface climate, such as temperature and precipitation, are quite usual.

Main comment #4: Relevance to SK24

ll 280–281. "While tests of the 'direct compensation' method recently proposed by Scinocca and Kharin (2024) yielded unsatisfying results with the LMDZ AGCM, . . . ". This statement is problematic.

The CABCOR method itself is not model dependent, in the same sense that the classical nudging-based method is not model dependent. Irrespective of the model, they both reduce annual-cycle climatological biases by construction. In SK24, a systematic analysis demonstrated that, for a fixed climate model version, the CABCOR approach consistently produced significantly larger bias reductions in the ERBC model than the classical nudging approach, including in cases where the classical approach employed optimal relaxation parameters. This was shown to be a robust methodological result rather than a model-specific outcome.

Stating that the CABCOR approach yielded "unsatisfying results" relative to the classical approach for a single LMDZ model version, therefore, appears to be at odds with the primary conclusions of SK24. If the authors would like to refute the results of SK24 by asserting that CABCOR performs poorly relative to the classical method, this claim must be substantiated and explained in much greater detail. Alternatively, if the authors do not intend to refute the results of SK24, they should clearly clarify what is meant by "unsatisfying results," how this assessment was made, and why it is not inconsistent with the findings reported in that study.

Reply : The relevant sentence in the submitted version of the article was clumsy at the very least, and we apologize for possible misunderstandings this may have caused. After submitting the current version of the article, we realized that our initial implementation of the method was not exactly equivalent to the original method described by Scinocca and Kharin (2024). After correcting the implementation, the results obtained using the CABCOR method are substantially improved, and a strong sea-level pressure bias induced by the alternative initial implementation was strongly reduced. New analysis shows that in

LMDZ and without temperature ERBC (and this caveat is certainly important), the iterative method described here yields results that are broadly equivalent (but overall slightly better, at least for τ = 1d) than the corrected CABCOR method. These results are presented in more detail in a recently submitted article (Champouillon et al., 2026) that we have already mentioned. We therefore propose to mention the CABCOR method here, state that it yields broadly comparable results in terms of bias reduction, and refer to a very recently submitted article for a detailed comparison. This initially problematic last paragraph of the conclusion now reads as follows:

> Several other variants or further developments of run-time bias corrections have been proposed, such as an interesting method based on direct compensation of diagnosed model biases ("CABCOR": Scinocca and Kharin (2024)) or inference of bias-correction terms using machine learning (Watt-Meyer et al., 2021). We have implemented the CABCOR method in LMDZ and found it to yield very good results, in many respects equivalent to the iterative method presented here. We have also developed a method of state-dependent bias corrections based on the simulated instantaneous synoptic situation (as expressed in the regional 500 hPa geopotential field) as a variant of the "classical" nudging-based ERBC. These approaches are evaluated in the context of LMDZ and without temperature correction by Champouillon et al. (2026). In addition, we are currently implementing machine-learning-based inference of state-dependent ERBC in LMDZ, following Watt-Meyer et al. (2021).

Minor Points

Minor point #1: ll 52-53. For 20 year calibration runs, there will be significant sampling variability in G_0. Some sort of smoothing of the averaged nudging tendencies is appropriate. Was any smoothing used? If so, please specify. If not, this seems like an important source of sampling artifacts.

Reply : Yes, we applied a 20-day running mean average. We now specify this:

> To reduce the high-frequency weather noise generated by averaging over only 20 years for any given day of the year, we apply a 20-day running-mean average to the final correction terms, but preserve the mean daily cycle.

Minor point #2: l 70. How was τ=1d selected? It differs from the optimal value identified in SK24 by a factor of 3. (See Major point 1).

Reply : We refer to the discussion here and use the opportunity to add an information about the reduced nudging near the surface and near the top of the atmosphere that had been omitted in the previous version of the article:

> The influence and the choice of the nudging time constant is discussed in section 4.2. The nudging strength is reduced to 0 near the surface (using a hyperbolical tangent factor transitioning from close to 1 to close to 0 over a range of about 500 m around 1500 m above the surface, and similarly at the upper limit of the atmosphere (beyond 5 hPa).

Minor point #3: l 70. "The free-running corrected simulations". This seems to be mixing two model variants. M is the free running model and C_i is the corrected model, and N_i is the nudged model. Perhaps just define them as such and stick to those terms throughout.

Reply : We intended to use the word "free-running" as indicating that the simulations are not nudged, but we see that this can lead to confusion. We do not use this expression in the revised version. We also clarify this aspect in Table 1 where we add "not nudged" where relevant instead of "free-running".

Minor point #4: ll 72-73. Why are there 2 out-of-sample corrected model runs but only 1 in-sample corrected model run?

Reply : The in-sample corrected runs are nudged to ERA5, so they are very similar. We see no point in running this period twice with nudging.

Minor point #5: l 84. "In particular, the temperature corrections induce a large-scale tropospheric warming". This would be better state as, "the temperature corrections significantly reduce a large-scale tropospheric cold bias in the freely running model" (i.e. see Major Point 3.).

Reply : Implemented as suggested by the reviewer.

Minor point #6: ll 90-91. The addition of global integrals of the absolute wind nudging tendencies would be helpful here in gauging the amplitude reduction with iteration number

Reply : We added a paragraph on this issue in 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 potentially vanishing final nudging tendencies is still far away after 3 iterations. 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 G_0 is identical to the mean nudging tendencies of N_0), 0.83 m/s/day for G_{0+1}, 1.09 m/s/day for G_{0+1+2}, and 1.30 m/s/day for $G_{0+\ldots+3}$, 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 compensation occurs between different iterations, as already shown by Figure 2.

Minor point #7: l 94. It would be helpful to look at the contribution to the bias correction after each iteration - not just the total. This would help identify when the iteration converges.

Reply : As stated in revised Section 3.1 (see preceding comment), the bias correction terms do not converge yet, while there are signs of over-correction at iteration 3 and signs of convergence in the simulated climate (no more substantial bias reduction after iteration 2, see Figure 3). The numbers given in the additional paragraph in Section 3.1 (see response to preceding comment) do show the contributions of each iteration to the correction terms, as requested by the reviewer.

Minor point #8: ll 113-120. Because the wind corrections act through the model dynamics, its influence on temperature is expected to arise primarily through dynamical balance (e.g. thermal wind, geostrophic and hydrostatic adjustment), which constrains horizontal temperature gradients rather than the global-mean temperature at a given pressure level. As a result, wind-only bias correction can reasonably be expected to improve regional temperature anomalies about the level-wise global mean while leaving the global-mean temperature largely unaffected, since the latter is controlled by the column-integrated energy balance and model physics. It may be helpful to clarify the physical interpretation of the temperature response in this way to motivate the decomposition of Temperature biases into the level mean and regional anomalies.

Reply : Thank you for this comment. This was exactly the reseasoning behind the idea to show the regional-scale temperature bias patterns $\tilde{T}$. We are very happy to take the reviewer's suggestion on board to explicitly provide the motivation of this analysis. We now write, using the reviewer's words (Thank you! As non-native speakers, we would have a hard time formulating this thought in an equally precise and elegant way.):

> Because the wind corrections act through the model dynamics, its influence on temperature is expected to arise primarily through dynamical balance (e.g. thermal wind, geostrophic and hydrostatic adjustment), which constrains horizontal temperature gradients rather than the global-mean temperature at a given pressure level. As a result, wind-only bias correction can reasonably be expected to improve regional

> temperature anomalies about the level-wise global mean while leaving the global-mean temperature largely unaffected, since the latter is controlled by the column-integrated energy balance and model physics. Therefore, when for a given pressure level the global mean bias is accounted for...

Minor point #9: ll 140-141. The caption for Table 2 says that the results are reported for 20 years (2001-2020) not 40.

Reply : We clarify this in the text:

> The results reported here are obtained for the total 40 years of the two 20-year runs of each experiment C_i ...

Minor point #10: The C_0 entry for 30-90°N, MAM in table 2 should also be bold.

Reply : It is actually smaller than the C_1 entry – this is a rounding issue. But the difference is very small (1.189 compared to 1.191), so bolding the C_0 value makes sense. We note that we had stated the following sentence in the caption: "The corrected simulation with the highest relative r^2 is bolded (although differences between simulations are not necessarily significant)."

Minor point #11: ll 185-186. There are two papers this year in Climate Dynamics on this topic: Scinocca et al. (2025; https://doi.org/10.1007/s00382-025-07814-5) and Labonte et al. (2025; https://doi.org/10.1007/s00382-025-07911-5). These provide a detailed analysis of the utility of driving regional climate models with the output of runtime bias corrected global model data.

Reply : Yes, it's indeed appropriate to cite these two papers here. Done.

Minor point #12: ll 186-189 The discussion would benefit from explicitly distinguishing between the target application of the bias-corrected fields. If the corrected model is intended to be used as a standalone global LMDZ model, avoiding temperature bias correction is clearly justified given the strong impact on model physics documented (also see Major comment 3). However, if the purpose is to provide driving fields for regional downscaling, the relevant criterion is the quality of the lateral boundary conditions (u, v, T, q) rather than the internal physical behaviour of the global model. In that context, bias-correcting temperature could be appropriate even if it degrades the standalone AGCM, since an RCM uses different physics and is primarily sensitive to the climatological realism of the boundary fields. Explicitly articulating this distinction would help clarify the scope and implications of the conclusions.

Reply : This is true, but the aim of the paper is to present the iterative ERBC procedure as such, independent of possible applications with LMDZ or any other model. In that sense, the distinction is not central to the paper. We therefore write:

> One could argue that, if the aim of the bias-corrected simulations is to serve as boundary conditions for a regional climate model, then the relevant criterion will be the quality of the lateral boundary conditions (u, v, T, q) provided to the regional model rather than the internal physical behaviour of the global model, justifying temperature ERBC in LMDZ. In any case, the aim of the present paper is to present the iterative ERBC method as such. Therefore, we limited the ERBC here to the zonal and meridional wind components.

Minor point #13: ll 231-235. "However, an in-depth analysis of the reasons for regional bias persistence might be necessary in specific regional use cases." The persistent regional biases in the global model are not really relevant to the dynamical downscaling problem. Regional responses within RCM domains are not really sensitive to regional GCM behaviour in overlapping locations. RCM regional responses are determined by their own physics packages and properties of the large-scale boundary driving, not regional GCM performance per se. The experimental design of RCM experiments includes boundary placement far from the region of interest, which deliberately filters out such interior GCM pathologies.

Reply : We delete this sentence and refer to very recently submitted work where an analysis of bias reduction using difference ERBC methods over the European region is carried out:

A detailed analysis for the reasons of bias persistence in specific regions is beyond the scope of this paper, and these reasons can depend on various factors, possibly including resolution, and are probably model-dependent. Regional bias persistence is analysed in more detail in Champouillon et al. (2026).

Minor point #14: l 248. "as a first simple approach one could try to simply use" - redundant use of simple/simply.

Reply : Thank you. We deleted the second use of this word.

Minor point #15: Section 4.6. Are these runs really that expensive that you could not have used the standard IPSL-CM6A-LR version of the model? The total number of AMIP simulated years is 360 (i.e. 18 20y simulations for this whole study, based on Table 1). Had you used the standard version, would you have been more successful in the use of bias correction on T, as in Krinner et al. (2020)?

Reply : The problems with temperature correction are very similar in higher-resolution runs with this version (256×256 points), and are visible also in older runs. We could have used the standard version of the model (144x144 grid points instead of 96x96), but again, the main aim of the paper is to present the iterative method as such, so the added value of simulations with standard resolution would have been weak.

Minor point #16: ll 276-278. As discussed in the introduction portion of this review, the type of runtime bias correction investigated in this study is state-independent cyclostationary corrections to the model equations that directly target the correction of seasonal-cycle climatological biases. The state-dependent machine-learning variants of runtime bias corrections (e.g. Watt-Meyer et al. 2021) directly target model formulation, which have indirect impact (both positive and negative) on climatological biases of the free model. This is an important distinction to make as discussed in the introduction to SK24.

Reply : Yes, but both methods directly modify the prognostic model equations by adding a corrective term (which can be cyclostationary or based on a machine learning inference of nudging terms).

Minor point #17: Also, The new approach to deriving the ERBC in SK24 has been described as the "direct compensation" method in this study. It would be clearer if this study simply used the same terminology as SK24 – i.e. climatological adaptive bias correction method (or CABCOR).

Reply : We now mention the name "CABCOR" explicitly here:

> Several other variants or further developments of run-time bias corrections have been proposed, such as an interesting method based on direct compensation of diagnosed model biases ("CABCOR": Scinocca and Kharin (2024)) or inference of bias-correction terms using machine learning (Watt-Meyer et al., 2021).

References

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

Hourdin, F., Williamson, D., Rio, C., Couvreux, F., Roehrig, R., Villefranque, N., Musat, I., Fairhead, L., Diallo, F. B., and Volodina, V.: Process-based climate model development harnessing machine learning: II. Model calibration from single column to global, Journal of Advances in Modeling Earth Systems, 13, e2020MS002 225, 2021.

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.

Scinocca, J. F. and Kharin, V. V.: Climatological adaptive bias correction of climate models, Journal of Advances in Modeling Earth Systems, 16, e2024MS004 563, 2024.

Watt-Meyer, O., Brenowitz, N. D., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., Perkins, W. A., and Bretherton, C. S.: Correcting Weather and Climate Models by Machine Learning Nudged Historical Simulations, Geophysical Research Letters, 48, e2021GL092 555, https://doi.org/https://doi.org/10.1029/2021GL092555, e2021GL092555 2021GL092555, 2021.

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.

Replies to comments by Reviewer 3

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.

Overall comments

Main comment #1: My main difficulty with the manuscript is that most of the metrics presented and parts of the methodologies are loosely defined textually, leaving significant room for uncertainty in interpretation.

Reply : In the subsequent minor comments, the reviewer points to specific issues where clarification of definitions and methodologies is required. We will follow the requests made there.

Main comment #2: Can you mention why you chose a nudging time scale of one day? This choice does seem reasonable. Still, one might expect qualitatively but not quantitatively similar results for a different choice of time scale, as dycores tend to fight nudging tendencies to different degrees depending on this time scale (e.g. Kruse et al. 2022). If you chose, say, a nudging time scale of half a day, do you think your results on variability (short-term in particular), would still hold?

Reply : This point has also been raised in similar terms by the two other reviewers. We are currently finalizing work that compares the different ERBC approaches ("classical", CAB-COR, iterative, and an implementation of conditional ERBC) that also comprises a detailed evaluation of the effects of nudging time constants in the "classical" and iterative approach with the LMDZ model. In addition, the reviewers have indicated references to relevant published studies that, together with our experience, provide a rather complete picture of the question, including the finding that optimal nudging timescales can be strongly model- and application-dependent. In response to these comments, we have substantially expanded section 3.1 and added a new section 4.2 in the discussion, specifically addressing the question of nudging timescales. Concerning the question about shorter nudging timescales than one day, we have not tested such values here, but an older paper (Krinner et al., 2019) uses a time constant of 6 hours, and in that study, no particularly detrimental effect on short-term variability was seen. However, we also note that a recent study by Scinocca and Kharin (2024) showed strong degradation of ERBC results for small time constants under certain conditions, which lead us to refrain from simulations with small nudging time constants τ. We mention this now in section 4.2:

> However, in a study of the effect of "classical" ERBCs on the simulation of the Antarctic climate (Krinner et al., 2019), a nudging timescale of $\tau = 6h$ was used, substantially less than that used here (τ = 1 d), and that in that study, improvements of the simulated climate were shown across all timescales, including high-frequency variability. We note in this respect that the iterative procedure used here can be seen as a method that leads to a stronger effective nudging, as shown by the fact that the amplitude of the combined correction terms $G_{0+...+i}$ for higher numbers i of iterations increases (see Figure 2). Additional systematic tests of the effects of smaller time constants τ ¡ 1 d in LMDZ are planned for future studies.

Main comment #3: Just a comment, not a request: It'd be very interesting to repeat this analysis using two different models, or versions of models, that have different mean states (i.e. are "biased" relative to each other). Much stronger statistics could be achieved. Conclusions on the influence of such a method on short-term, smaller-scale variability might be much stronger.

Reply : We agree. We mention this need in the conclusions of the article, as a consequence of model-dependency of such results, as an outlook:

> More generally, because ERBC implementation choices and thus the corresponding results are at least partly model-dependent, it would be interesting to see the effect of iterative ERBC in other models, possibly in the framework of a coordinated multi-model intercomparison of various ERBC methods.

Minor comments

Minor comment #1: Line 52: It'd be very helpful to the readers if you define "cyclostationary time average". Is G_0 a function of x,y,z,t then?

Reply : Yes. We indicate this clearly now after the equation that first defines G:

> ... where G_0 is a function of space (latitude, longitude and vertical level) and time (day of the year).

Minor comment #2: Line 92-94: There's no cancellation between G_0, G_1, ...? Not too surprising, this but cannot be inferred from Fig. 1 since the absolute value is presented. If you were to further iterate, would you expect the sum of G_i to converge?

Reply : The reviewer is right. We analyze this in more detail in section 3.1 in the revised version:

> 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 potentially vanishing final nudging tendencies is still far away after 3 iterations. 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 G_0 is identical to the mean nudging tendencies of N_0), 0.83 m/s/day for G_{0+1}, 1.09 m/s/day for G_{0+1+2}, and 1.30 m/s/day for $G_{0+...+3}$, 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 compensation occurs between different iterations, as already shown by Figure 2.

Minor comment #3: Fig. 2: are you plotting C_3/C_0 (i.e. total model tendencies?) or G_3/G_0? I assume the latter, but this is not completely clear from the text or figure caption. An inline equation would be very helpful.

Reply : The reviewer is right to state that we have not been precise enough here. We modified the figure label to clearly indicate that the figure shows the ratio between the sum of G_i (i=0,...,3), which is used in C_3, and G_0 (used in C_0). We also specify this in the caption of the figure now:

> 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). ...)

Minor comment #4: Fig. 3: Are these biases in the corrected, but free-running runs (i.e. 2001-2020)?

Reply : Yes, these are the biases of the corrected, "free-running" simulations. We refrain from using "free-running" in the revised version because one reviewer found this expression confusing.

Minor comment #5: Fig. 4: What is plotted is not stated in the figure caption nor the text! Is this RMSE 2001-2020 time averaged u and v between the bias-corrected, free-running runs and ERA5? That is, the RMSE of a spatial bias?

Reply : We clarify this in the text:

> Figure 3 displays the 1981–2000 annual mean bias of the zonal wind component at 500 hPa for the different simulations. Figure 4 displays, in each of its two panels, the ratio of the 1981–2000 root mean square error (RMSE) in each of the corrected simulations C_i ($i = 0, \ldots, 3$) and the RMSE of the uncorrected simulation M : RMSE(C_i)/RMSE(M), for the annual mean of the zonal (a) and merdidional (b) wind speed, where the RMSE is calculate with respect to ERA5.

Minor comment #6: Line 117: Do you mean middle and upper troposphere and not middle-atmosphere (i.e. stratosphere/mesosphere)?

Reply : This was not clear indeed. We now write:

> In this case the iterative run-time bias correction procedure leads to further improvement in most parts of atmosphere above about 900 hPa [...]

Minor comment #7: Table 2: Again, some quantitative definition of these metrics would be helpful. I think I understand your metrics, but ideally there wouldn't be uncertainty in interpretation.

Reply : We revised the text to be more explicit:

> The dominant patterns of interannual variability of monthly circulation structures are indeed more realistically depicted in the corrected simulation C_0 than in the uncorrected control simulations M, and even more so in simulations with iterated run-time bias corrections (C_{1-3}). This can be seen in Table 2 which displays the seasonal squared spatial Pearson correlation coefficient r^2 between simulated (LMDZ) and "observed" (ERA5) dominant modes of monthly 500 hPa geopotential anomalies, as identified by principal component analysis for the different corrected model runs C_i, relative to the corresponding squared spatial correlation coefficient for the uncorrected control simulations M: $R = r^2_{C_i}/r^2_M$.

And we also made the table caption more explicit on the definitions:

> Squared spatial correlation coefficient r^2 between the corrected LMDZ runs C_i (i=0,...3) and ERA5, for the first EOF of monthly extratropical (30-90° latitude) variability of the 500 hPa geopotential height ($\phi_{500\mathrm{hPa}}$) for selected seasons and hemispheres, 2001-2020, relative to the corresponding squared correlation coefficient r^2_M of the uncorrected simulation M: $R = r^2_{C_i}/r^2_M$.

We also inverted the order of the table columns (and adapted the caption) such as to have the main results in the first columns, while the last column serves as a reference. We hope that this makes the table much easier to understand.

Minor comment #8: Table 3: Again, it's very difficult to be sure I understand all of the individual variability metrics. Could you please define these metrics (not ratios) quantitatively?

Reply : We defined the metric more properly in the text. For the sea-level pressure variability, we write:

> The winter short-term (2.5–8 day) mid-latitude (40-60°N and 40-60°S) sea-level pressure variability, diagnosed from the temporal standard deviation of sea-level pressure from its seasonal mean ($\sigma_{2.5-8\mathrm{d}}$, using a bandpass filtering in the frequency domain, and limited to grid points with surface height below 1000 m) ...

The definition of the blocking frequency score was indeed missing. We now write:

> Blocking anticyclones are a significant atmospheric phenomenon, although there is not one standard accepted definition (Lupo, 2021). Here we follow the widely used definition by Davini et al. (2012): Blocking events are diagnosed at a given point in time and space based on the reversal of the meridional gradient of geopotential height measured at 500 hPa, extending over at least 15° of continuous longitude, and persisting within a 5° latitude × 10° longitude box centered on a given grid point for at least 5 days. We then calculate the annual mean frequency of these events for each grid point, and calculate the spatial RMSE E of these frequencies compared to ERA5. We then calculate, for each corrected simulation, the ratio between E_{C_i} for that simulation and E_M of the uncorrected simulation, yielding a score $f_B = E_{C_i}/E_M$.
>
> While the iterated procedure seems to improve the model score for the 30-75°N annual and spatial mean blocking frequency, (i.e., f_B (30-75°N) < 1), with best results obtained for 2 and 3 iterations, the bias-corrected simulations show degraded performance compared to the uncorrected control simulations (i.e., f_B (30-75°S) > 1) in the Southern Hemisphere (30-75°S). However, in the Southern Hemisphere, the iterated bias correction procedure leads to less degraded performance compared to a single bias correction in C_0.

References

Davini, P., Cagnazzo, C., Gualdi, S., and Navarra, A.: Bidimensional diagnostics, variability, and trends of Northern Hemisphere blocking, Journal of Climate, 25, 6496–6509, 2012.

Krinner, G., Beaumet, J., Favier, V., Deque, M., and Brutel-Vuilmet, C.: Empirical run-time bias correction for Antarctic regional climate projections with a stretched-grid AGCM, Journal of Advances in Modeling Earth Systems, 11, 64–82, 2019.

Lupo, A. R.: Atmospheric blocking events: A review, Annals of the New York Academy of sciences, 1504, 5–24, 2021.

Scinocca, J. F. and Kharin, V. V.: Climatological adaptive bias correction of climate models, Journal of Advances in Modeling Earth Systems, 16, e2024MS004 563, 2024.