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 $\tau = 1$d) 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/s0038 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.