



Iterative run-time bias corrections in an atmospheric GCM (LMDZ v6.3)

Gerhard Krinner¹, Aude Champouillon¹, Juliette Blanchet¹, and Frédérique Chérut²

¹Univ. Grenoble Alpes, CNRS, INRAE, IRD, IGE, 38000 Grenoble, France

²Laboratoire de Météorologie Dynamique-IPSL, Sorbonne Université/CNRS/École Normale Supérieure-PSL
Université/École Polytechnique-Institut Polytechnique de Paris, Paris, France

Correspondence: Gerhard Krinner (gerhard.krinner@cnrs.fr)

Abstract. Run-time bias corrections of atmospheric circulation models can be based on nudging (newtonian relaxation) to an atmospheric reanalysis. In this case, the time increments of selected state variables are modified by adding the nudging terms obtained with an uncorrected version of the model. This is a well-known method to improve the models' representation of large-scale circulation patterns. In this work, we propose and evaluate a variant of this method, consisting of iterative nudging: the corrected model is itself nudged towards the reanalysis, and the resulting nudging terms are added to the initial ones to calculate the new, iterated correction terms. This procedure can be iterated an arbitrary number of times. Evaluating the LMDZ atmospheric general circulation model (AGCM) for a varying number of iterations of nudging to the ERA5 reanalysis for the period 1981-2000, we show that the simulated large-scale circulation patterns over the period 2001-2020 are consistently improved when the bias correction procedure is iterated compared to the non-iterated original procedure. However, while there is a clear benefit of one or two iterations of the bias correction method, signs of over-correction appear after about three iterations.

1 Introduction

Despite constant and continuing progress over several decades, climate models (in a broad sense, ranging from Earth System Models to limited-area dynamical atmospheric circulation models) still exhibit biases in their representation of current climate and of large-scale indicators of climate change (Eyring et al., 2021; Arias et al., 2021). Nevertheless, in the face of rapid climate change and its increasingly detrimental impacts, climate models remain the primary source of information to quantify future climate change from global to regional scales (Lee et al., 2023). However, for most use cases, run-time bias corrections of climate models, or bias adjustments of their output, are necessary (Doblas-Reyes et al., 2021; Ranasinghe et al., 2021).

The type of run-time bias corrections of global atmospheric models that is used in the present work has been, to our knowledge, first described by Guldberg et al. (2005). In this approach, time increments to the model's state variables are corrected using terms derived from the corrected increments that are obtained using a nudged simulation (see Section 2). In a perfect model framework, this approach has been shown to substantially improve simulated future climate projections (Krinner et al., 2020). Together with the fact that even in the context of a strong climate change ($4\times\text{CO}_2$), large-scale climate model bias patterns have been shown to be stationary to a high degree (Krinner and Flanner, 2018), this justifies the use of this method,



for example, for projections of future Antarctic climate on a centennial time scale (Krinner et al., 2019; Beaumet et al., 2019). Compared to post-hoc bias adjustment of climate model output, a potential advantage of run-time bias corrections is that the improved large-scale circulation patterns can lead to a more physically consistent multivariate representation of climate change. Moreover, even if the output of a climate model using such a run-time bias correction method may require further statistical bias correction because necessarily some biases remain (e.g., Scinocca and Kharin, 2024), a better representation of large-scale circulation patterns may help avoid pitfalls of post-hoc statistical bias corrections linked to the misplacement of critical circulation features (e.g., Hall, 2014; Maraun et al., 2017).

Nudging-based run-time bias corrections only partly eliminate model biases. Root mean square error (RMSE) reductions of about 30-50% are typical for this method (Krinner et al., 2020; Scinocca and Kharin, 2024), depending on the model and the assessed variable. This is the main reason why in this paper, we examine the question whether an iterative application of this method can lead to a more perfect bias reduction. However, as with any model tuning or bias correction, care has to be taken to avoid over-correction of model biases. The aims of this paper are therefore a) to describe the iterative bias correction procedure implemented in the LMDZ6 atmospheric GCM (Hourdin et al., 2020); b) to evaluate the simulated present-day climate in the uncorrected model and for a varying number of iterations of the correction procedure; c) to determine limits of the procedure, in particular by evaluating possible signs of over-correction.

The next section describes the iterative run-time bias correction method and the simulations carried out to test it. We then present results for circulation-related variables on large spatial and varying temporal scales, and discuss the potential of the method and its limitations.

2 Methods

2.1 Iterative run-time bias correction

We start from the “classical” nudging-based run-time bias correction method described by Guldberg et al. (2005). Following the notation used by Scinocca and Kharin (2024), the original model increments for a state variable X (e.g. the meridional or zonal wind component) are noted $M(X, t)$:

$$\partial_t X = M(X, t). \quad (1)$$

When the model is nudged (Jeuken et al., 1996) to a (re)analysis, these time-varying model increments become:

$$\partial_t X = M(X, t) + \frac{X_a - X}{\tau} \equiv N_0(X, t), \quad (2)$$

where X_a denotes the reanalysis state variable, τ the nudging timescale (here, 1 day), and N_0 the nudged version of the model.

The climatological nudging increments $G_0 = \overline{(X_a - X)}/\tau$ (the overbar denoting a cyclostationary time average, taking into account seasonal and daily cycles) are then used in a free-running, bias-corrected simulation denoted C_0 :

$$\partial_t X = M(X, t) + G_0 \equiv C_0(X, t). \quad (3)$$



55 Until this stage, this is, as stated, the “classical” run-time bias correction method described by Guldberg et al. (2005).

We can now simply iterate this procedure for a first time, using the initial corrected model C_0 :

$$\partial_t X = C_0(X, t) + \frac{X_a - X}{\tau} \equiv N_1(X, t). \quad (4)$$

Here, N_1 stands for the first iterated nudging. The cyclostationary climatological nudging increments $G_1 = \overline{(X_a - X)}/\tau$ of this first iteration are then added to the initial correction terms G_0 : $G_{0+1} = G_0 + G_1$. The first iterated corrected model run C_1
 60 then uses these combined correction terms:

$$\partial_t X = M(X, t) + G_{0+1} \equiv C_1(X, t). \quad (5)$$

This iteratively corrected model can then again be nudged to the (re)analysis. In general terms, the n^{th} iteration of the bias-corrected model is then given by:

$$\partial_t X = M(X, t) + G_{0+1+\dots+n} \equiv C_n(X, t). \quad (6)$$

65 2.2 Simulations and analysis

We use the LMDZ6 AGCM (Hourdin et al., 2020) at low resolution (96 (longitude) $\times$ 95 (latitude) horizontal grid points and 79 vertical levels). The model is nudged towards the ERA5 reanalysis (Hersbach et al., 2020) for the period 1981-2000 (plus the year 1980 for model spinup). Only the meridional and zonal wind components above about $P = 0.85P_s$ (where P and P_s are the atmospheric pressure at a given level and at the surface, respectively) are nudged towards ERA5 with a time
 70 constant of $\tau = 1$ day in the nudged simulations N_i ($i = 0 \dots 3$). The free-running corrected simulations C_i ($i = 0 \dots 3$) and the free-running reference simulation denoted M are run for the evaluation period 2001-2020 (plus the year 2000 for model spinup), distinct from the nudging period. The free-running reference M and the corrected simulations C_i are run twice for 2001-2020 with varying initial conditions. The presented results for these simulations M and C_i are the average of these $2 \times 20 = 40$ model years. Table 1 provides a summary of the simulations described in this work.

75 Evaluation is carried out against 2001-2020 ERA5 circulation characteristics such as climatological mean biases for meridional (v) and zonal (u) wind at selected standard pressure levels (850, 500 and 200 hPa), mean sea-level pressure biases, inter-annual patterns of sea-level pressure variability using empirical orthogonal functions, blocking frequencies (Davini et al., 2012; Palmer et al., 2023) based on 500 hPa geopotential indices, band-pass (between 2.5 and 7d) sea-level pressure variability linked to storm track location and intensity (Chang, 2009), and frequencies of synoptic weather patterns using 5×4 self-organizing
 80 maps (Kohonen, 1982) of hemispheric extratropical (polewards of 40°N and S) daily-mean geopotential heights. Furthermore, the amplitude of the u and v correction terms is analyzed to evaluate convergence of the correction procedure.

Contrary to previous studies (Krinner et al., 2019, 2020), we refrained from bias-correcting atmospheric temperatures because we have realized that run-time temperature bias corrections in the LMDZ AGCM destructively interfere with the physical parameterizations in the tropics. In particular, the temperature corrections induce a large-scale tropospheric warming, leading
 85 to reduced convective instability and thus to reduced simulated convective activity, accumulation of water vapour in the atmo-



Simulation	Type	Period
M	Free-running, uncorrected	2001-2020 (2 runs)
N ₀	Nudged to ERA5	1981-2000
C ₀	Free-running, corrected using N ₀	2001-2020 (2 runs)
N ₁	Nudged to ERA5, corrected using N ₀	1981-2000
C ₁	Free-running, corrected using N ₀ +N ₁	2001-2020 (2 runs)
N ₂	Nudged to ERA5, corrected using N ₀ +N ₁	1981-2000
C ₂	Free-running, corrected using N ₀ +N ₁ +N ₂	2001-2020 (2 runs)
N ₃	Nudged to ERA5, corrected using N ₀ +N ₁ +N ₂	1981-2000
C ₃	Free-running, corrected using N ₀ +N ₁ +N ₂ +N ₃	2001-2020 (2 runs)

Table 1. Simulations described in this work.

spheric boundary layer, and a substantial surface energy flux imbalance linked to reduced surface evaporation and increased downwelling longwave radiation.

3 Results

3.1 Nudging and correction increments

90 The amplitude of the nudging increment decreases for higher iterations, as shown in Figure 1 for the zonal mean absolute zonal wind nudging increments. While the spatial patterns are broadly similar for all simulations N_i, the amplitude of the nudging increments decreases by approximately a factor of 2–3 from N₀ to N₃. Contrary to the nudging increments, which decrease, the amplitude of the bias correction increments applied in simulations C_i increases for higher iterations, as the correction increments for simulation C_i are the sum of the average nudging increments of N₀ to N_i (see Equation 6). The spatial structure
 95 of the bias correction increments also remains similar for the successive iterations C_i, but for the higher iterations it is not a simple spatially constant multiple of the initial correction increments for C₀, as shown by the ratio of the zonal mean absolute zonal wind correction increments between C₃ and C₀, displayed in Figure 2. Instead, this ratio has distinctive spatial structures varying in space and in time through the annual cycle (not shown), indicating that the iterative procedure is not identical to a simple uniform amplification of the initial correction.

100 3.2 Mean errors

Plots for wind bias reductions obtained with the standard bias correction method and its successive iterations are provided in Figure 3 (maps for *u* at 500 hPa) and in Figure 4 (zonal means for *u* and *v*). By construction, mean wind biases at various levels in the atmosphere are decreased in the corrected simulation C₀ compared to the uncorrected simulation M. Relative to M, the root mean square error (RMSE) of the zonal and meridional wind components *u* and *v* is decreased by ≈ 20% near the

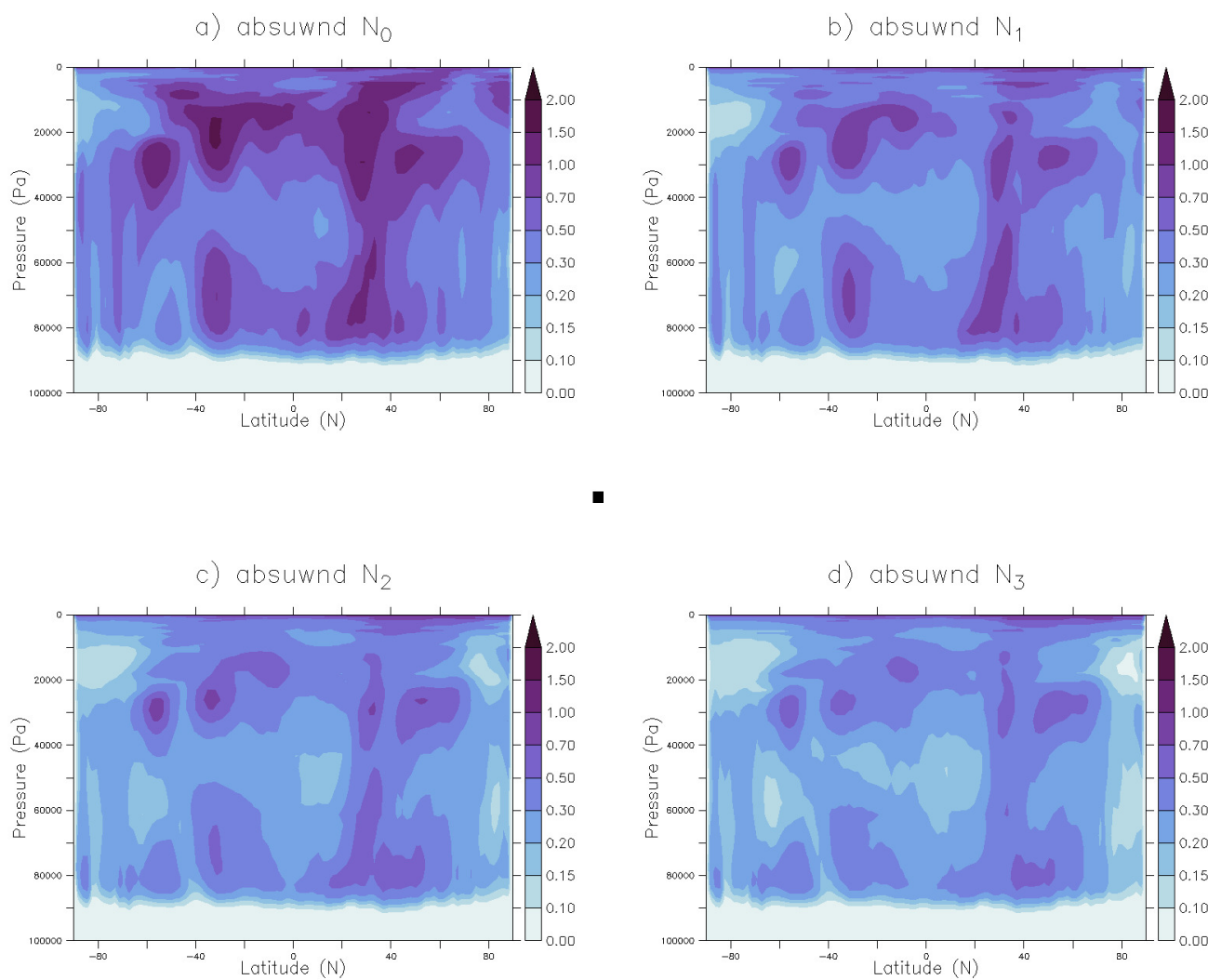


Figure 1. Zonal mean absolute zonal wind nudging increments (in m/s/day) calculated from a) N_0 (used in C_0); b) N_1 ; c) N_2 ; d) N_3 , average for January 1981 to 2000.

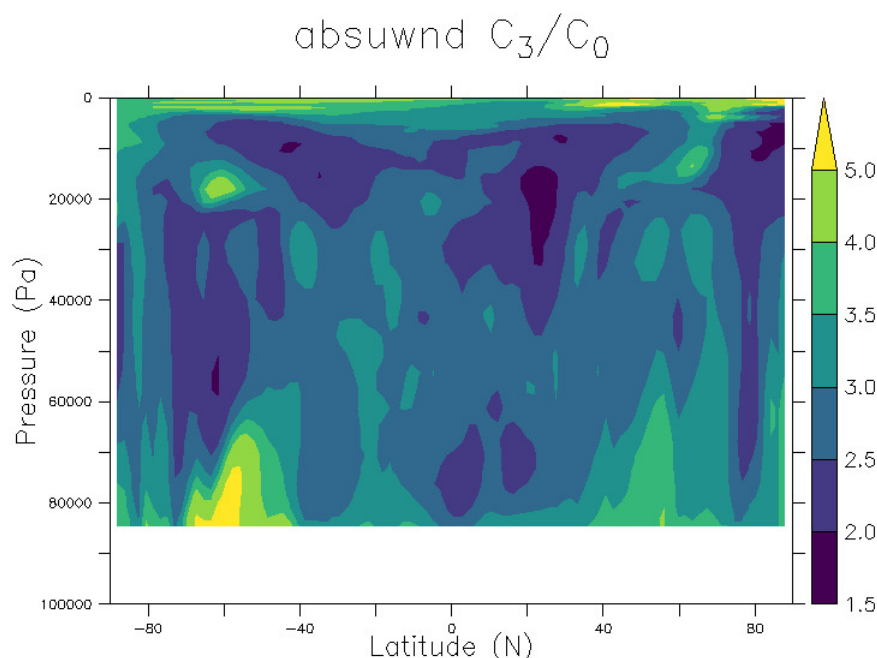


Figure 2. Ratio of the zonal mean absolute zonal wind bias correction increments (averages for 1981 to 2000) between C_3 and C_0 . Values between 100000 and 85000 Pa are not shown in the lower part of the figure. They are not relevant as, by construction, the absolute values of the increments vanish at these altitudes (see Figure 1).

105 atmospheric boundary layer (where no correction is applied) to $\approx 60\%$ (for u) at 200 hPa. When the correction procedure is iterated once (simulation C_1), a further substantial RMSE reduction is obtained. In most cases except close to the atmospheric boundary layer, no substantial further improvement is obtained in subsequent iterations of the correction procedure.

In the absence of run-time temperature corrections in this study, substantial large-scale air temperature biases remain in our wind-only corrected simulations C_i (see Figure 5). While the wind corrections lead to improvements in the zonal-mean temperature simulation in the upper atmosphere, above ≈ 300 hPa, the bias reduction is only modest below this level. In addition, further bias reduction between 200 and 300 hPa in polar regions during higher iterations (>2) of the bias correction procedure come at the expense of a degradation of the simulated temperature patterns at these latitudes at about 150 hPa. However, when for a given pressure level the global mean bias is accounted for (i.e. subtracted, yielding a centered temperature $\tilde{T} = T - \Delta T$, where ΔT is the global mean temperature bias at a given atmospheric level), the remaining regional bias patterns are reduced in the corrected simulations except very close to the surface, as can be seen in Figure 6 which displays the RMSE of

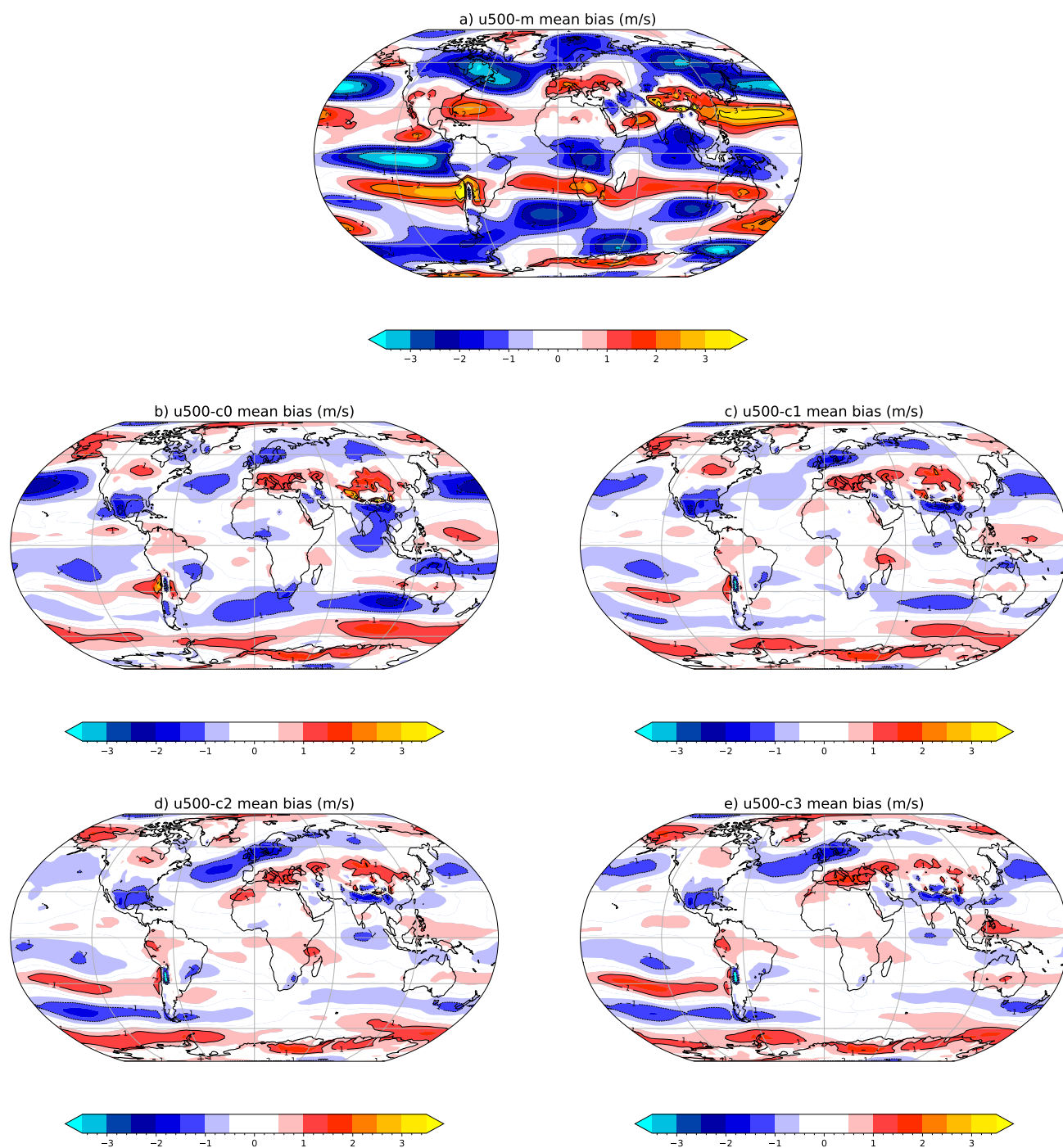


Figure 3. Annual mean 500 hPa zonal wind mean bias with respect to ERA5 (in ms^{-1}), for different simulations. a) uncorrected simulation M; b) corrected simulation C_0 ; c) first iterated corrected simulation C_1 ; d) second iterated corrected simulation C_2 ; e) third iterated corrected simulation C_3 .

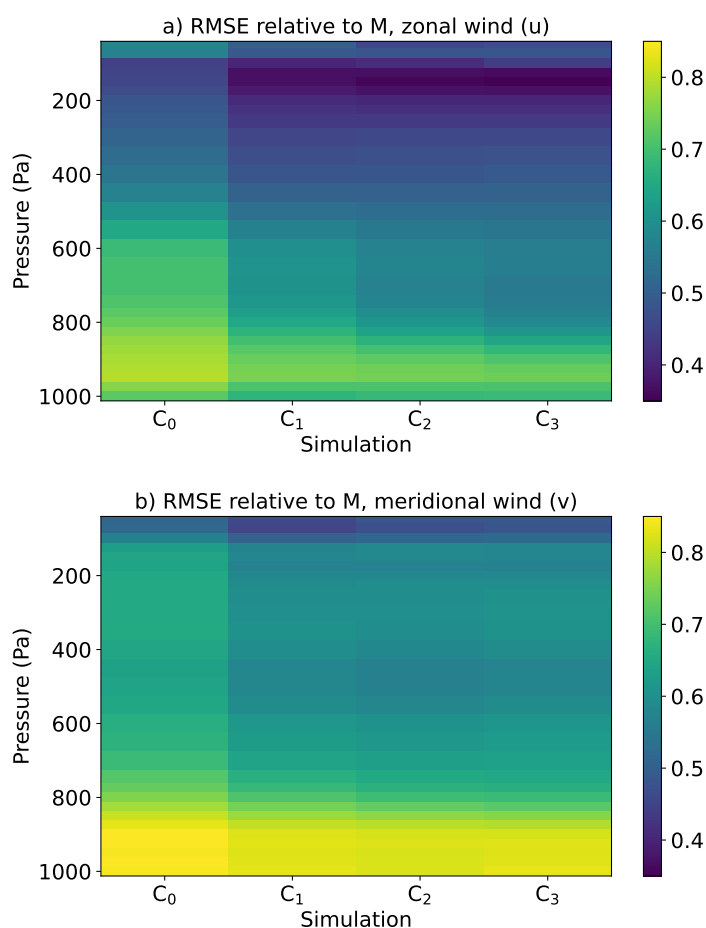


Figure 4. Annual mean of monthly mean RMSE (with respect to ERA5) for zonal (a) and meridional (b) wind, for the corrected simulations C_0 – C_3 , relative to the same quantity for the uncorrected simulation M. Values below 1 indicate that the RMSE in the corrected simulations is smaller than the RMSE of the uncorrected simulation.



this centered temperature $\tilde{T}$. In this case the iterative run-time bias correction procedure leads to further improvement in most parts of the middle and upper atmosphere (with the exception of a small degradation of the biases at about 850 hPa, which are already not well corrected in C_0 , and an already-mentioned degradation at about 150 hPa). This means that, while the global mean temperature biases – predominantly cold below about 150 hPa and very likely caused by deficiencies in the model physics – are not corrected, regional-scale temperature bias patterns are smoothed because of the corrected model dynamics.

Sea-level pressure biases are only weakly reduced in our simulations that only bias-correct wind patterns. In C_0 , the global RMSE of the annual mean sea-level pressure (with respect to ERA5) is reduced by 10% compared to the uncorrected simulation M, and iterations of the bias correction method only lead to insignificant overall improvements (the reduction compared to M is 13% for simulation C_3). The weak correction of sea-level pressure errors is in contrast to earlier results for the Southern Hemisphere obtained using run-time temperature corrections (Krinner et al., 2019).

3.3 Monthly to interannual circulation variability

As stated, building on cyclostationary correction terms, the run-time bias correction procedure used here is, by construction, designed to better represent climatological mean values of the corrected variables. However, it is certainly of equal importance to assess whether these run-time bias corrections also improve the way the model represents circulation variability on different timescales.

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 reference simulation 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 correlation coefficient between simulated (LMDZ) and “observed” (ERA5) dominant modes of monthly 500 hPa geopotential anomalies as identified by principal component analysis.

Only seasons for which the first principal component explains more than 40% of the variance in the ERA5 dataset are retained here in order to restrict the analysis to the most stable and physically meaningful patterns. Although the 20-year observational period might be a bit short and could be the reason for some remaining noise in the results reported in Table 2 – the tendencies for a given season and hemisphere from C_0 to C_3 are not always steady –, it seems that an iteration of the run-time bias correction leads in most cases to a better representation of the interannual circulation variability patterns. The results reported here are obtained for the stacked 40 years of the two runs of each experiment C_i , but the results for individual 20-year runs of these experiments are very similar, increasing the confidence in the robustness of the results. Furthermore, the squared correlation coefficients overall increase between C_1 and C_2 .

3.4 Short-term circulation variability

Diagnostics of short-term atmospheric circulation variability indicate an overall positive impact of the wind-based bias corrections for spatial means of bandpass-filtered extratropical winter sea-level pressure variability and average blocking frequencies (Table 3), but there is no clear improvement over multiple iterations of the correction procedure.

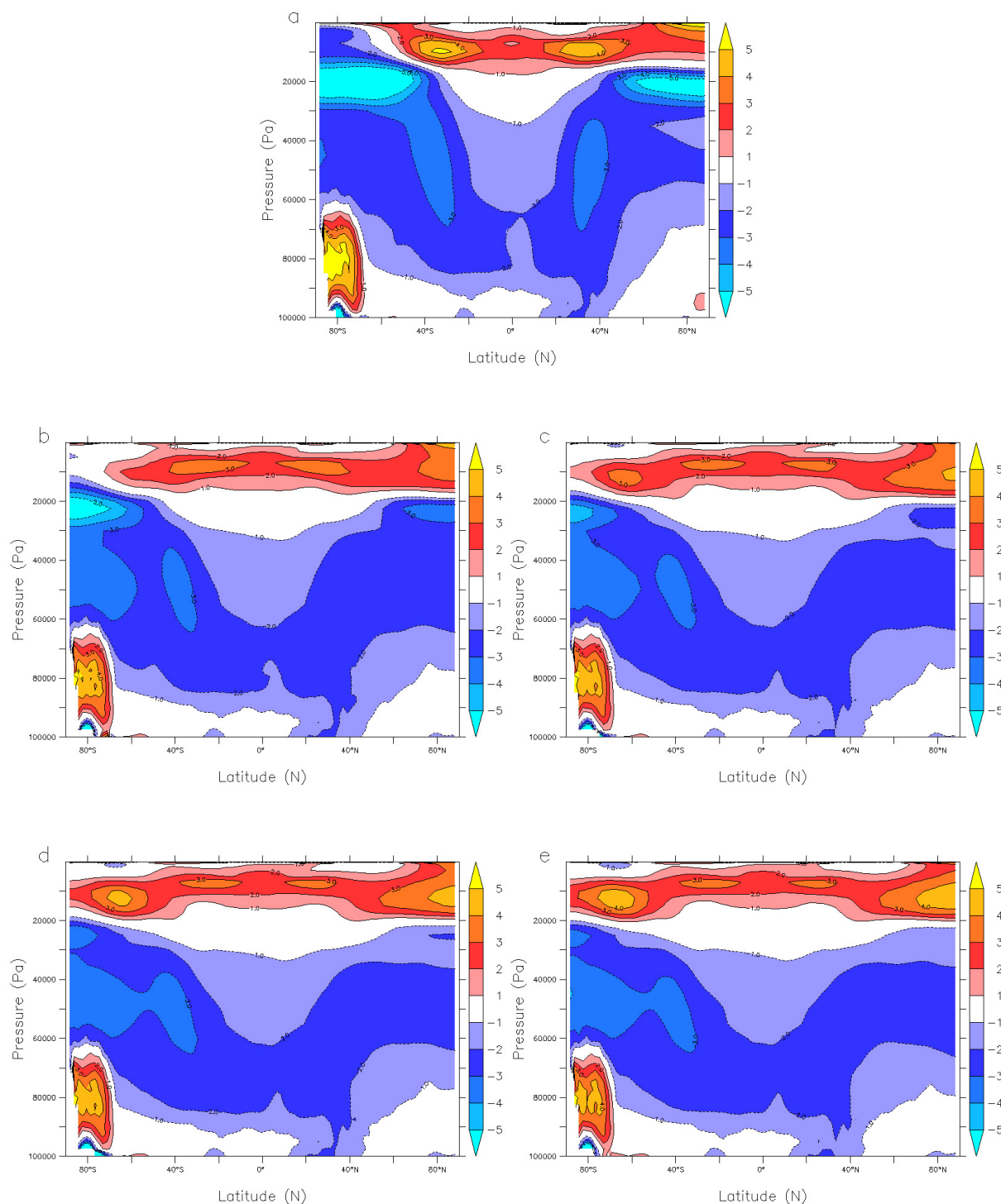


Figure 5. Zonal and annual mean temperature biases of the free-running simulations M (a) and the corrected simulations C₀-C₃ (b,c,d,e) against ERA5 (°C).

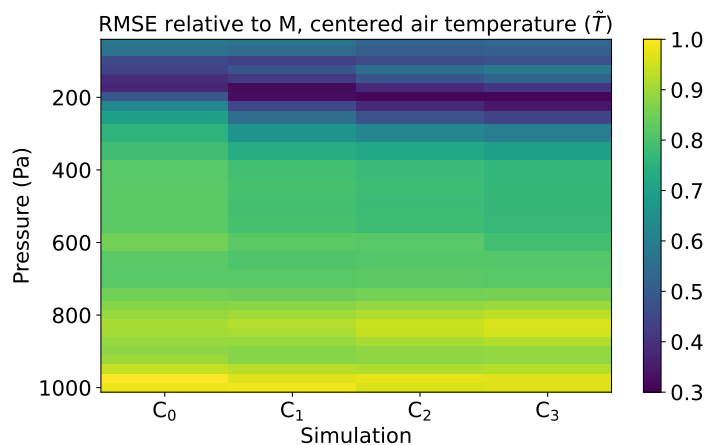


Figure 6. Annual mean of monthly mean RMSE (with respect to ERA5) of the centered air temperature $\tilde{T}$ (see text), for the corrected simulations $C_0 \dots C_3$, relative to the same quantity for the uncorrected simulation M, with global mean bias for the respective level subtracted. Values below 1 indicate that the RMSE in the corrected simulations is smaller than the RMSE of the uncorrected simulation.

Hemisphere and season	r_M^2	r^2/r_M^2			
		C_0	C_1	C_2	C_3
30-90°N, MAM	0.56	1.19	1.19	1.12	1.14
30-90°N, SON	0.47	1.06	1.15	1.21	1.24
30-90°S, DJF	0.76	1.03	0.96	1.03	1.01
30-90°S, JJA	0.54	1.13	1.24	1.24	1.25
30-90°S, SON	0.49	1.12	1.23	1.24	1.18

Table 2. Squared spatial correlation coefficient r^2 between corrected LMDZ runs C_i and ERA5 for the first EOF of monthly extratropical (30-90° latitude) variability of the 500 hPa geopotential height ($\phi_{500\text{hPa}}$) for selected seasons and hemispheres, 2001-2020, relative to r^2 of the uncorrected simulation M. Only cases for which the variance explained in the ERA5 reanalysis exceeds 40% are shown. The squared spatial correlation coefficient for the uncorrected reference simulation, r_M^2 , is given in a separate column for reference. The corrected simulation with the highest relative r^2 is bolded (although differences between simulations are not necessarily significant).



The winter short-term (2.5–8 day) mid-latitude (40–60°N and 40–60°S) sea-level pressure variability ($\sigma_{2.5-8d}$) shows no consistent improvement, and even slight degradation, for the iterated corrections $C_{1...3}$ compared to the simple non-iterated bias correction applied in C_0 . However, the short-term pressure variability is overall better depicted in the corrected than in the uncorrected simulations.

While the iterated procedure seems to improve the model score for the 30–75°N annual and spatial mean blocking frequency (Davini et al., 2012), with best results obtained for 2 and 3 iterations, the bias-corrected simulations show degraded performance compared to the free-running simulations 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 .

In the Southern Hemisphere, the mean RMSE of the frequencies of typical hemispheric-scale daily circulation patterns, as identified using 20 (5x4) self-organizing maps for each season (DJF and JJA) and denoted as f_S in Table 3, is slightly degraded in the corrected simulations compared to the uncorrected simulation (see f_S for 40–90°S in the table). Frequency errors are increased by about 15% on average compared to the uncorrected simulation. Iterations of the bias correction procedure ($C_{1...3}$) do not show any consistent change (neither improvement nor degradation) compared to the non-iterated bias correction applied in C_0 , even if C_2 and C_3 show the least degradation in JJA and DJF, respectively, compared to C_0 . In the Northern Hemisphere, the non-iterated bias correction C_0 procedure produces the best results of all simulations (including the uncorrected simulation M). In other words, while the bias correction method appears to improve the model performance in this particular respect, iterating the bias correction procedure does not lead to further improvement of the frequency of the simulated main synoptic weather patterns (see f_S for 40–90°S in the table).

Overall, taking these diagnostics together, the wind bias correction does lead to a modestly improved representation of short-term atmospheric circulation variability, but iterating the bias correction procedure does not necessarily lead to clear further improvement. We note that in the cases where the model performance is degraded by applying the bias correction once (that is, C_0 shows degraded performance relative to M), iterating the bias correction can lead to improved (less degraded) results.

4 Discussion

4.1 Choice of bias-corrected variables

Focusing on polar regions (Krinner et al., 2019), previous studies with empirical run-time corrections of wind and temperature in the LMDZ GCM have yielded promising results. However, at least with the most recent version of the LMDZ AGCM, we have found strong detrimental effects of temperature corrections in the lower latitudes. As shown in Figure 5, LMDZ has a wide-spread cold bias of about 2°C in the middle troposphere. Empirical run-time corrections of this bias physically correspond to an additional radiative heating. As mentioned before (section 2), we suspect that this leads to a weakening of the temperature lapse rate and reduced convective activity. As a result, humidity accumulates in the lower troposphere, which reduces the upwards surface latent heat flux on large scales, in particular over the oceans, and the (approximate) global surface energy balance is strongly perturbed - in an ocean-atmosphere coupled model, this would lead to a very strong oceanic warming exceeding by far the observed oceanic heat uptake (Von Schuckmann et al., 2023).



Variable	C ₀	C ₁	C ₂	C ₃
$\sigma_{2.5-8d}(p_{msl})$ (DJF 40-60°N)	0.94	0.97	1.00	0.94
$\sigma_{2.5-8d}(p_{msl})$ (JJA 40-60°S)	0.96	0.96	1.06	1.03
f_B (30-75°N)	0.83	0.86	0.69	0.75
f_B (30-75°S)	1.18	1.09	1.06	1.08
f_S (DJF 40-90°S)	1.14	1.24	1.09	1.07
f_S (JJA 40-90°S)	1.14	1.14	1.04	1.15
f_S (DJF 40-90°N)	0.76	0.91	0.79	0.86
f_S (JJA 40-90°N)	0.88	1.01	1.03	0.95

Table 3. RMSE of selected model results, calculated with respect to ERA5 (2001-2020) and normalized with respect to the uncorrected free-running reference simulation M, for high-frequency circulation-related quantities. $\sigma_{2.5-8d}(p_{msl})$: bandpass-filtered (between 2.5 and 8 d) sea-level pressure standard deviation between 40 and 60° latitude in both hemispheres, limited to grid points below 1000 m surface height; f_B : annual mean frequency of blocking situations (based on daily mean 500 hPa geopotential height) between 30 and 75°N and S; f_S : individual frequencies of the 20 most representative hemispheric extratropical large-scale synoptic situations based on daily 500 hPa geopotential maps. The simulation with the lowest RMSE of the corrected simulations is bolded (note however that values > 1 indicate that the model performance is degraded with respect to the uncorrected reference simulation M). Differences between simulations are not necessarily significant.

It is noteworthy that this destructive effect of empirical temperature correction is highly model-dependent. For example, no such effect has been noted in different version of the Canadian atmospheric model CanAM (Krinner et al., 2020; Scinocca and Kharin, 2024).

This motivated the choice to only bias-correct the wind in this study. As a consequence, as shown above, the widespread cold bias in the middle atmosphere is not corrected. Of course this raises questions about the usefulness of such bias-corrected simulations for driving limited-area models for simulations of the present climate and projections. The uncorrected large-scale cold bias of LMDZ would certainly also lead to a cold bias in the nested limited-area model which would, in most cases, not be able to correct for the bias through its own, possibly less biased, radiative-convective process representations, given the time scales of atmospheric circulation involved.

However, the choice to only correct the horizontal wind components, and not other variables such as temperature or atmospheric humidity, can make sense if one thinks of the partly conceptual separation between model dynamics and physics. The horizontal wind components, in particular above the atmospheric boundary layer, are quite directly determined by the model dynamics. Atmospheric temperature and humidity, in contrast, and in particular on large spatial scales, can be seen as primarily determined by the model physics. One can argue that the pervasive cold bias throughout the almost entire middle atmosphere, partly “hidden” by stronger local temperature biases caused e.g. by the latitudinal misplacement of jet streams, should be preferably corrected by improved model physics or specific tuning. In contrast, localized biases caused by the misplacement of circulation features are often resolution-dependent and as such more obvious aims for empirical run-time bias corrections



limited to the horizontal wind components. In short, it is conceptually consistent to limit the empirical run-time bias correction to errors primarily linked to the model dynamics, and to avoid strong interference with the model physics which primarily determines the model response to external forcings, notably in climate change experiments.

4.2 Over-correction: Out-of-sample vs. in-sample evaluation

An obvious risk of iterative bias correction procedures is over-correction of model biases, leading to degraded or physically less meaningful results when the model is applied outside of the calibration period. To detect possible signs of over-correction, the difference of the global mean RMSE reduction in corrected simulations between the calibration period (1981–2000, "in-sample") and the normal evaluation period (2001–2020, "out of sample") is displayed in Figure 7 for mean annual zonal and meridional wind speed at different levels in the atmosphere. The idea is that if, as the number of iterations increases, the performance during the out-of-sample period (2001–2020) is systematically degrading compared to the performance during the in-sample period (1981–2000), then this means that the real added value of additional iterations is decreasing. As Figure 7 suggests, this seems to be the case at least after three iterations. In that case, while for several variables the in-sample score still continues to increase, the out-of-sample score starts to decrease. This is particularly the case for variables which have low overall bias reduction scores, for example the 700 hPa meridional wind. The difference between the evolution of the in-sample and the out-of-sample scores is weaker for variables which have strong bias correction scores, such as the 200 hPa zonal wind.

It therefore seems that after about three iterations of the bias-correction method, the performance during the out-of-sample period is degrading compared to the performance during the in-sample period, in particular for variables for which still some potential for correction is left. We interpret this as a sign of over-correction, suggesting that two iterations of the bias correction method should be the maximum number in practical applications.

4.3 Persistence and attenuation of bias patterns

We have shown that globally the mean biases are progressively reduced during the iteration of the bias correction, at least until the point when over-correction starts to occur. However, it is clear that the overall structure of the bias patterns is robust, as can be seen very clearly for example in Figure 3. This is not surprising, as climate model bias patterns are quite persistent even under strong climate change (Krinner and Flanner, 2018; Krinner et al., 2020).

What is noteworthy nevertheless is that the strength of attenuation of mean biases is to some degree spatially variable. For example, while Figure 3 shows a steady decrease of the annual mean 500 hPa zonal wind mean bias in the Southern Indian Ocean with increasing number of iterations of the bias correction, the corresponding bias pattern over Western Europe (positive over the Mediterranean and negative over North-western Europe) quickly stabilises, and is even amplified in simulations C_2 and C_3 relative to C_0 and C_1 . Similarly, while biases continue to decrease over the North Pacific, they already start to re-increase in simulation C_2 over the South Pacific. The reasons for regional persistence of biases are not clear, and they probably are region-specific. It is for example possible that these regionally varying behaviours are linked to region-specific factors such as topography, or that they are caused by teleconnection patterns in interaction with error compensation (in the sense that in two regions linked by teleconnection mechanisms, well correcting an error in one region might unmask errors in the other region).



As a consequence, iterating the bias correction method might allow regional climate modelers to dispose of much improved atmospheric boundary conditions for the RCM in one specific region, but not in another. A detailed analysis for the reasons of bias persistence in specific regions is beyond the scope of this paper, and these reasons are probably model-dependent, and possibly resolution-dependent. However, an in-depth analysis of the reasons for regional bias persistence might be necessary in specific regional use cases.

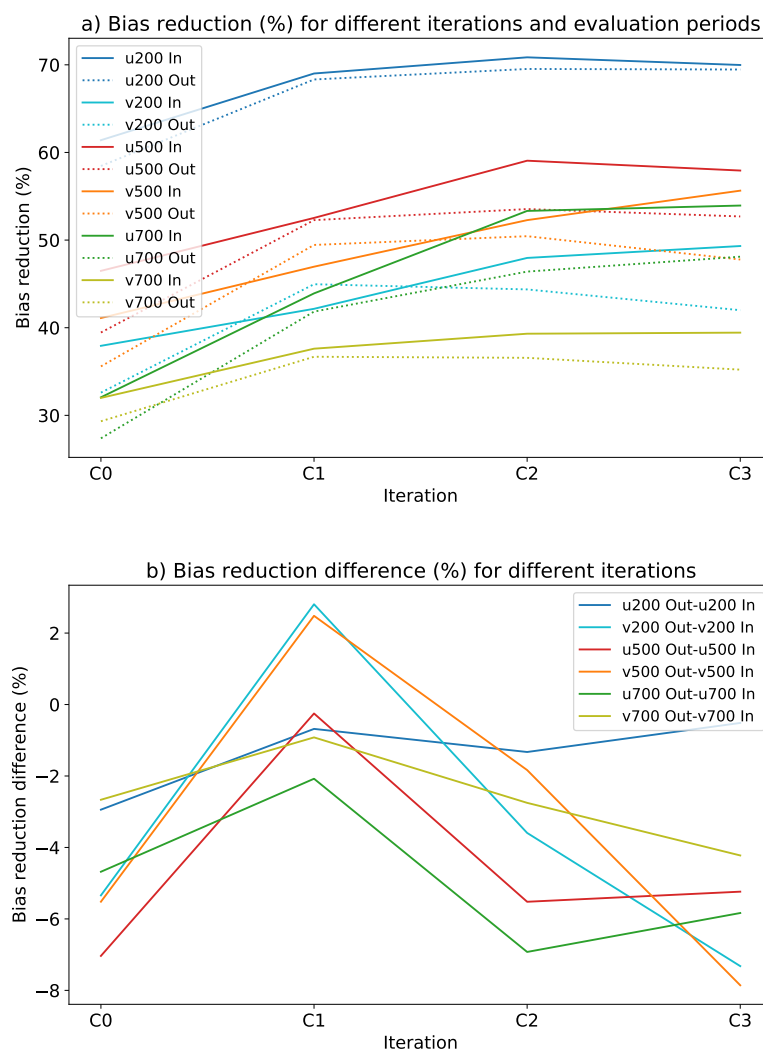


Figure 7. Comparison of bias reduction compared to the uncorrected simulation M (in %) between different evaluation periods. a) Bias reduction for in-sample and out-of-sample periods, for each variable; b) Difference between in-sample and out-of-sample periods, for a given variable. As the model is nudged during the years 1981–2000, the period 1981–2020 corresponds to in-sample evaluation (noted “In”), while 2001–2020 corresponds to out-of-sample evaluation (noted “Out”).



4.4 Simulation periods: Length of the calibration and evaluation periods

We made the choice in this study to calibrate the bias-correction terms over the 1981–2000 period and to evaluate the effect of the bias correction for the out-of-sample period 2001–2020. This means that both periods only cover 20 years, albeit with two model realizations for the evaluation runs. Although 20 years is today often considered to be a reasonably long evaluation period – see for example the IPCC AR6 WGI model evaluation chapter (Eyring et al., 2021) –, one might wonder whether a longer calibration period would lead substantially different results. However, there is a definite risk of over-correction when too long a calibration period is used, if only because this limits the possibility for out-of-sample testing. Of course this does not mean that a longer calibration period of 30 years would not be preferable when applying this method, as long as the conclusions for the experiment setup that have been drawn here with clear out-of-sample evaluation are kept in mind.

4.5 Simple multiplication of correction increments

As we have shown here (see for example Figure 3), the geographical patterns of the mean model biases are somewhat independent of the number of iterations, and the mean absolute bias correction increments increase with the number of iterations (see Figure 2). This might suggest that as a first simple approach one could try to simply use a multiple of the bias correction terms obtained during the first nudging (that is, those applied in C_0) as a surrogate for the more costly iteration procedure. We have carried out a simple test in which the correction terms of C_0 are amplified by 50%; otherwise this simulation (referred to as $C_{0,\times 1.5}$ in the following) is identical to C_0 . The analysis of $C_{0,\times 1.5}$ shows that the bias correction is less efficient than in C_0 . For example, with respect to ERA5, the annual mean global root mean square error of the 500 hPa zonal and meridional wind components is 7.5% (u_{500}) and 15.6% (v_{500}) higher in $C_{0,\times 1.5}$ than in C_0 . However we note that $C_{0,\times 1.5}$ still shows a substantial improvement compared to the uncorrected model run M.

It is fairly easy to understand why the model performance degrades between C_0 and $C_{0,\times 1.5}$. As one can see in Figure 2, although the amplitude of the correction terms does increase essentially everywhere for higher iterations of the correction procedure, the ratio of the correction terms between different phases of the iteration procedure is not uniform – it clearly has a spatial structure. A simple overall increase of the initial correction terms, although modest in our test, does not reproduce this spatial structure.

4.6 Model resolution

These simulations have been carried out at a rather low horizontal resolution of 3.75° (longitude) $\times$ 1.875° (latitude), lower than the standard IPSL-CM6A-LR setup used in CMIP6 (Bonnet et al., 2021) with its $2.5^\circ \times 1.25^\circ$. However, the vertical resolution, with 79 levels, corresponds to the standard CMIP6 setup for this model in order to reduce the need for re-tuning of the physical parametrisations. The reason for this choice is of course computational efficiency. Many model biases have similar structures at varying resolution, even if their amplitude can vary. The point here is to evaluate the benefit of iterated run-time bias corrections, rather than an optimum model setup with as weak as possible initial model biases. We are therefore rather certain that main our results and conclusions remain valid at a higher horizontal resolution.



5 Conclusions

This work has shown that iterating the “classical” empirical run-time bias correction method described by Guldberg et al. (2005) once or twice (as in our simulations C_1 and C_2 , respectively) allows one to substantially reduce the mean bias of the bias-corrected variables compared to simulations in which the bias correction procedure is not iterated. However, the representation of atmospheric circulation variability on interannual and shorter timescales is not always improved in simulations with iterated bias corrections. Signs of over-correction appear about three iterations of the bias correction procedure.

All in all, the method described is easy to implement, adds substantial value and does not represent a very high additional cost compared to the non-iterated procedure.

Several other variants or further developments of run-time bias corrections have been proposed and implemented, such as an interesting method based on direct compensation of diagnosed model biases (Scinocca and Kharin, 2024) or inference of bias-correction terms using machine learning (Watt-Meyer et al., 2021). In addition, we are currently testing state-dependent bias corrections based on the simulated instantaneous synoptic situation (as expressed in the regional 500 hPa geopotential field). While tests of the “direct compensation” method recently proposed by Scinocca and Kharin (2024) yielded unsatisfying results with the LMDZ AGCM, one could imagine iterations of diagnosed state-dependent or machine-learned bias corrections, or of those directly derived from the diagnosed mean model biases, as a further development along the lines developed in this work.

Code and data availability. Configuration files and model code changes, scripts and model output data used to plot the graphics are available on Zenodo under doi:10.5281/zenodo.16363996 (Krinner, 2025a). The modipsl code infrastructure required to compile the codes, including the full model code (in particular LMDZ v6.3 but also ORCHIDEE and XIOS), is available on Zenodo under doi:10.5281/zenodo.17381281 (Krinner, 2025b).

Author contributions. GK designed the study, carried out the simulations, contributed to the analysis, and wrote the first draft of this article. AC contributed to the analysis and writing. JB and FC contributed to writing and overall discussions about the method.

Competing interests. The authors declare no competing interests.

Acknowledgements. This study has received funding from Agence Nationale de la Recherche – France 2030 as part of the PEPR TRACCS programme under the grants ANR-22-EXTR-0005, ANR-22-EXTR-0008, ANR-22-EXTR-00010, and ANR-22-EXTR-0011. It was pro-



vided with HPC computing and storage resources by GENCI at TGCC under the grants 2024-A0180116219 and 2024-AD010101523R on
295 the Joliot Curie supercomputer (SKL and ROME partitions).



References

- Arias, P. A., Bellouin, N., Coppola, E., Jones, R. G., Krinner, G., Marotzke, J., Naik, V., Palmer, M. D., Plattner, G.-K., Rogelj, J., Rojas, M., Sillmann, J., Storelmo, T., Thorne, P. W., Trewin, B., et al.: Technical Summary, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., et al., pp. 33–144, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/9781009157896.002>, 2021.
- Beaumont, J., Déqué, M., Krinner, G., Agosta, C., and Alias, A.: Effect of prescribed sea surface conditions on the modern and future Antarctic surface climate simulated by the ARPEGE atmosphere general circulation model, *The Cryosphere*, 13, 3023–3043, 2019.
- Bonnet, R., Boucher, O., Deshayes, J., Gastineau, G., Hourdin, F., Mignot, J., Servonnat, J., and Swingedouw, D.: Presentation and evaluation of the IPSL-CM6A-LR ensemble of extended historical simulations, *Journal of Advances in Modeling Earth Systems*, 13, e2021MS002 565, 2021.
- Chang, E. K.: Are band-pass variance statistics useful measures of storm track activity? Re-examining storm track variability associated with the NAO using multiple storm track measures, *Climate dynamics*, 33, 277–296, 2009.
- 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.
- Doblas-Reyes, F., Sörensson, A., Almazroui, M., Dosio, A., Gutowski, W., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W.-T., Lamptey, B., Maraun, D., Stephenson, T., Takayabu, I., Terray, L., Turner, A., and Zuo, Z.: Linking Global to Regional Climate Change, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., et al., book section 10, pp. 1363–1512, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/9781009157896.012>, 2021.
- Eyring, V., Gillett, N., Achuta Rao, K., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y.: Human Influence on the Climate System, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., et al., book section 3, pp. 423–551, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/9781009157896.005>, 2021.
- Guldborg, A., Kaas, E., Déqué, M., Yang, S., and Thorsen, S. V.: Reduction of systematic errors by empirical model correction: impact on seasonal prediction skill, *Tellus A*, 57, 575–588, <https://doi.org/10.1111/j.1600-0870.2005.00120.x>, 2005.
- Hall, A.: Projecting regional change, *Science*, 346, 1461–1462, 2014.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, 2020.
- Hourdin, F., Rio, C., Jam, A., Traore, A.-K., and Musat, I.: Convective boundary layer control of the sea surface temperature in the tropics, *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001 988, 2020.
- Jeuken, A., Siegmund, P., Heijboer, L., Feichter, J., and Bengtsson, L.: On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation, *Journal of Geophysical Research: Atmospheres*, 101, 16 939–16 950, 1996.
- Kohonen, T.: Self-organized formation of topologically correct feature maps, *Biol. Cybern.*, 43, 59–69, 1982.
- Krinner, G.: Krinner et al., Iterative run-time bias corrections in an atmospheric GCM (LMDZ v6.3) - LMDZ configuration files, output data and analysis scripts, Zenodo Dataset, doi=10.5281/zenodo.16363996, 2025a.



- Krinner, G.: Krinner et al., Iterative run-time bias corrections in an atmospheric GCM (LMDZ v6.3) - model infrastructure and code, Zenodo Dataset, doi=10.5281/zenodo.17381281, 2025b.
- 335 Krinner, G. and Flanner, M. G.: Striking stationarity of large-scale climate model bias patterns under strong climate change, *Proceedings of the National Academy of Sciences*, 115, 9462–9466, 2018.
- 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.
- Krinner, G., Kharin, V., Roehrig, R., Scinocca, J., and Codron, F.: Historically-based run-time bias corrections substantially improve model
 340 projections of 100 years of future climate change, *Communications Earth & Environment*, 1, 29, 2020.
- Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., Trisos, C., et al.: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland, <https://doi.org/10.59327/IPCC/AR6-9789291691647>, 2023.
- Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M., Hall, A., et al.:
 345 Towards process-informed bias correction of climate change simulations, *Nature Climate Change*, 7, 764–773, 2017.
- Palmer, T. E., McSweeney, C. F., Booth, B. B., Priestley, M. D., Davini, P., Brunner, L., Borchert, L., and Menary, M. B.: Performance-based sub-selection of CMIP6 models for impact assessments in Europe, *Earth System Dynamics*, 14, 457–483, 2023.
- Ranasinghe, R., Ruane, A., Vautard, R., Arnell, N., Coppola, E., Cruz, F., Dessai, S., Islam, A., Rahimi, M., Ruiz Carrascal, D., Sillmann, J., Sylla, M., Tebaldi, C., Wang, W., and Zaaboul, R.: Climate Change Information for Regional Impact and for Risk Assessment, in: *Climate
 350 Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., et al., book section 12, pp. 1767–1925, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/9781009157896.014>, 2021.
- Scinocca, J. F. and Kharin, V. V.: Climatological adaptive bias correction of climate models, *Journal of Advances in Modeling Earth Systems*, 16, e2024MS004563, 2024.
- 355 Von Schuckmann, K., Minière, A., Gues, F., Cuesta-Valero, F. J., Kirchengast, G., Adusumilli, S., Straneo, F., Ablain, M., Allan, R. P., Barker, P. M., et al.: Heat stored in the Earth system 1960–2020: where does the energy go?, *Earth System Science Data*, 15, 1675–1709, 2023.
- 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, e2021GL092555,
 360 <https://doi.org/https://doi.org/10.1029/2021GL092555>, e2021GL092555 2021GL092555, 2021.