Exploring precipitation pattern scaling methodologies and robustness among CMIP5 models
Kravitz et al., Geoscientific Model Development
Response to reviewers

Reviewer comments in plain text. **Responses in bold.**

General response

We thank the reviewers for their comments on our paper. Both reviewers were critical of the "physically-based" method, and we have carefully considered their points. We agree that it is important to evaluate this method more carefully, including additional checks of the accuracy of our implementation. Exploring this method would also require a discussion of its usefulness as a pattern scaling method and why we obtained the results that we did. Given the large increase in scope this would require, which would distract from our assessments of the performance of the other two methods, we have elected to remove mention of this method from the present manuscript. We will do a better job with it in a future study.

Reviewer #1

In my opinion, the paper constitutes an interesting contribution that should be published in Geoscientific Model Development after adequate revisions. An evaluation of the performance of different pattern scaling methods for climate variables other than temperature (here: precipitation) is of significant practical importance. The evident main conclusion of the work is that two of the methods work reasonably but the third one does not. This should be clarified in the text.

We thank the reviewer for his/her comments. As stated above, we have removed the third method, so our conclusions will change slightly. We have updated the text to accommodate this.

General comments

Regarding the two traditionally-used methods, here termed the regression and epoch methods, I mainly agree with the conclusions presented by the authors. These methods appear to be fit for scaling precipitation. However, the verbal assessments given for the "physically-based" method in the manuscript do not seem to be supported by the quantitative results presented in the figures and Table 3. Examples of statements that I find unjustified: "the physically-based method shows a greater degree of robustness (less relative root-mean-square variation than the other two methods) and could be a particularly advantageous method if outstanding biases could be reduced" (abstract, l. 7–9); "This indicates the potential for robustness of the physically-based method" (p. 5, l. 2); "The overall performance of the physically-based method is still worse in all cases, but these results suggest that if the overall bias in the physically-based method could be reduced or corrected, it holds great promise in being a useful pattern scaling method..." (p. 8, l. 26–28); "The physically-based method has substantially worse performance than the other two methods but shows some features of robustness that could be advantageous if overall biases in the method could be reduced" (p. 12, l. 16–18).

In all the examples studied, the performance of the "physically-based" method appears to be inferior to the other two methods (Table 3). In some cases (e.g., that depicted in Fig. 6), the gap between the performances is apparently somewhat smaller. Note, however, that in these experiments the magnitude of the projected change B is small, which makes the scaling error $\hat{B} - B$ small as well. Accordingly, in these cases the small RMS error produced by the "physically-based" method is likely to be a trivial consequence of the smallness of B. (See further discussion in "specific comments".)

Furthermore, on l. 10–11 of p. 3 it is stated that "There are many possibilities for physically-based approaches". Therefore, I suggest that the authors should use some other, more specified name for the version of the method examined in this paper. Note also that 'physically-based' inherently sounds very positive and thus a more neutral term should be preferred; particularly, taking into account the low performance of that method.

We agree with all of the points in the previous several paragraphs of the reviewer's assessment. Per the general response above, we have removed the physically-based method from this manuscript and all text associated with it.

The number of figures in the paper, 19, is excessive. In particular, there are plenty of figures (14) that visually very similar, consisting of a set of six global map panels. A high level of concentration is required for a reader in order to study this large manifold of illustrations. I find that it is mainly Figs. 4, 9, 10, 12, 14 and 16 that include key information. Conversely, Figs. 5–7, 11, 15 and 17–19 are not that essential and mainly relevant for readers of special interest. For the majority of readers, it would facilitate reading the article if these figures (or a significant portion of them) would be shifted into an electronic supplemental file that is available in conjunction with the article.

We agree with the reviewer that there were too many figures. After reviewing the paper, we have moved Figures 5-7 and 17-19 to supplemental material. We have also removed Figure 2.

Is the precipitation variable discussed in the paper an annual mean? That should be specified in the abstract, introduction, conclusions and, perhaps, in some of the figure captions as well.

Agreed. We have added mentions of this throughout the paper.

Compared to the other two methods, the performance of the "physically-based" method is very poor. The poorness is so striking that I recommend that the authors should check the correctness of their algorithms once again.

Per the general response above, we have removed the physically-based method from this manuscript.

Specific comments

Interpolation, extrapolation. For readers less familiar with the idea, please specify that you are dealing with interpolation (extrapolation) in time (p. 1, 3, 7, 8 and 12).

Thanks for pointing that out. We have added more specificity where appropriate.

Earth System Models (ESMs) vs. Atmosphere-Ocean General Circulation Models (AOGCMs). According to the definition applied in Chapter 9 of IPCC (2013), ESMs are those climate models that include an interactive carbon cycle. All models listed in Table 1 of your paper do not fulfill this criterion but belong to the category of AOGCMs. I recommend that you would use the same terminology as IPCC (2013). — This does not have any impact on the quantitative findings as you have used concentration-driven model runs alone (p. 5, l. 7).

A point well taken. We have replaced all mentions of ESM with AOGCM in the manuscript.

P. 4, l. 8–17: The idea of the "physically-based" method should be explained in more detail. The present formulation is not adequate to make the idea understandable without consulting the reference.

Per the general response, we have removed mentions of the physically based method.

There is an error in Eq. (5): in the denominator, replace $s^2{}_1$ by $s^4{}_1$ and s^2 by $s^4{}_2$. Check whether this is an typing error only or whether you have used the wrong df in the calculations.

This was just a typo in the manuscript. Thanks for pointing that out.

P. 7, l. 18: the poorer performance of these two methods may be due to the large contribution of noise in the pattern of P that is determined from the early years of the simulation when the true climate change signal is weak.

Good point. We have added a sentence to this effect.

P. 7, l. 20–23: The error for the "physically-based" method is not similar but nearly double that produced by the other methods (Table 3). More importantly: the smallness of the error for the"physically-based" method may have been caused by the fact that P = 0 over the majority of the domain. Then, in these areas $\hat{B} = 0$ as well and, since B is small, the difference $B - \hat{B} = 0$ is likewise small. Thus, the smallness of the RMS error is not any indication of the good performance of the "physically-based" method. See also general comment 1 and the text that you have written on p. 8, l. 19–21 and p. 11, l. 12.

Per the general response, we have removed mentions of the physically based method.

P. 7, l. 27–28: "error is reduced by a factor of two for the physically-based method": is this a trivial consequence of the smallness of B? "and increases by a factor of two for the regression and epoch difference methods": this may be an indication of true non-linearity. Also, p. 8, l. 22–26 need revision.

Per the general response, we have removed mentions of the physically based method. We have revised what's left of the lines on page 8 to improve clarity.

Section 3.3 and Fig. 8: When you present the results for a certain number of models, have you used in each experiment the same sub-ensemble models in calculating P and B? Or are the models chosen randomly for that comparison? Please clarify.

The models are chosen randomly. We have clarified this in the text.

P. 9, l. 17–19: I did not understand the idea. How the dominance of the CO_2 response helps to apply the non-CO_2 pattern for the other scenarios? Please clarify. Note also that the non-CO_2 response includes both a warming (other GHGs) and cooling component (aerosol forcing) that may have different ratios in the various RCP forcing scenarios.

We have removed that part of the sentence that perplexed the reviewer. As to the other point, that is well taken. We have added an additional paragraph that discusses many of these issues.

P. 10, l. 28–32: Note that warming does not follow radiative forcing immediately but, due to the thermal inertia of oceans etc., with a lag. Has this been taken into account? If not, a caveat should be included in the text.

We have not accounted for lags like this. We have added a caveat to the text.

P. 34, l. 9–11: In my opinion, there is a contradiction between the text and the Figure captions 18–19. In the captions, it is stated that P is extracted from one group of models and ΔT and B from the other group. Thus, the experiments would be "antisymmetric" and accordingly, one would expect that the errors would be of a similar order of magni- tude. Differences in P between the groups 1 and 2 should affect Figs. 18 and 19 by about a similar magnitude. According to the text, figures and Table 3, however, this is not so. Please check and clarify.

We apologize for the confusion. We had a typo in the text, so the descriptions of the two figures appeared to be antisymmetric, but they weren't. We have fixed this so the paper better says what we actually did.

The discussion presented in the Appendix might be transferred into electronic supplementary material.

Agreed. We have moved this text and the associated figures to supplemental material.

Minor comments

P. 1, l. 12–13: for other models -> to be utilized in other models ?

Agreed and changed in the manuscript.

P. 2, l. 19: a long history of research -> a fairly long history of research (the method has been used for a few decades, not millenia).

Agreed and changed in the manuscript.

P. 2, l. 26–27: "no single fit (e.g., regression coefficients) will be applicable to all grid points" (and a similar statement on p. 3, l. 27–28). This is a trivial consequence of the fact that the modelled precipitation change is not geographically uniform. If you want to say something more, please clarify.

We agree with the reviewer's statement. We are simply reviewing what previous studies have shown.

P. 3, l. 6–7: "If the climate response is perfectly linear, then any pattern scaling method will work equally well and will be highly accurate." I would prefer a more conditional formulation, e.g.: "If the climate response were perfectly linear, then any pattern scaling method would work equally well and would be highly accurate."

Agreed and changed in the manuscript.

P. 3., l. 9: Conversely -> In principle; the findings of the present work do not favour the "physically-based" method.

We have removed mentions of the physically-based method in this manuscript.

P. 3, l. 30: "this approach automatically accounts for correlations between local temperature and local precipitation changes". How?

This sentence ended up being more confusing than illuminating, so we have removed it.

P. 4, l. 30: This may be caused (i) by the rather small area of the polar regions and (ii) by the fact that both B and $\hat{B}$ are relatively small there.

Per the general response above, we have removed this paragraph.

P. 7, 7–8: "If the scaling pattern P (x) truly is time-invariant, then the results presented in this section will be identical to those previously discussed." -> "If the scaling pattern P (x) truly were time-invariant, then the results presented in this section would be identical to those previously discussed." (They are not identical.)

Changed. Thanks for the phrasing.

P. 7, l. 16: none -> virtually none ?

Agreed and changed in the manuscript.

P. 8, l. 9: I did not understand "The values in Table 3 indicate that Group 1 (13 models) is not an outlier." Please clarify.

We have clarified this sentence.

P. 8, l. 27: I do not agree with "holds great promise".

Agreed. Keeping with the general response above, we have removed this paragraph.

Title of section 4 might be modified: you discuss the total forcing and its partition into the CO_2 and non-CO_2 components.

Changed to "Pattern Scaling for Additional Forcings".

P. 9, l. 17: "There is no a priori reason to expect this will work". Do you mean "There is no a priori reason to expect that this will work"?

Yes, changed.

P. 9, l. 30: Giving the actual years (e.g., 2076–2100 for model years 227–251?) would be informative (in figure captions as well).

Agreed. Changed in the text and the relevant figure captions.

P. 10, l. 11–12: One possible explanation is aerosol forcing.

Agreed. We have added a mention of aerosol forcing.

P. 10, 27–28: "If the approach fails, it is because this pattern does not represent actual non-CO_2 forcing." Noise due to unforced internal variability in the climate system may also have an influence.

Good point. We have added this.

P. 10, l. 30: $\log_2([CO_2])$. In what units $[CO_2]$ is expressed? This determines the values of the coefficients.

Added units of ppmv.

P. 11, l. 1–2: "The temperature contribution of the non-CO2 part increases with the CO2 concentration" was not entirely clear for me.

We have revised this sentence to be less confusing.

P. 11, l. 22–23: "The epoch difference and regression methods show too much CO2 response and not enough non-CO2 response as indicated by the patterns displayed in Figure 15." Would you please explain in more detail how one can see this?

We have substantially revised this section. The comment now references removed text.

P. 11, l. 27: "values depicted in Figure 16 are almost entirely due to CO2 forcing". According to Table 2, the ratio of non-CO_2 to CO_2 forcing does not differ substantially between these two RCP scenarios.

We have substantially revised this section. The comment now references removed text.

P. 11, l. 28–29: "this indicates that the non-CO2 forcing in RCP2.6 is insufficiently large to overcome issues with low signal-to-noise ratios in reconstructing patterns of precipitation change using this sort of decomposition." This was difficult to understand. Please explain in more detail.

We have substantially revised this section. The comment now references removed text.

P. 11, l. 30: “polar amplification of the precipitation response”. In general, polar amplification refers to the temperature response.

Added. Thanks!

P. 12, l. 18: the methods work relatively well, but i regard “excellent” as a too emphatic word.

Agreed. We have rephrased this sentence.

P.12,l.25–26:“it is the best equipped to deal with these sources of nonlinearity.” Perhaps in theory, but the present findings do not support this statement.

Agreed. We have removed the physically based method, so this sentence has been removed as well.

P. 12, l. 32 – p. 13, l. 1: “However, given the difficulties many Earth System Models have with proper representations of convective processes and the resulting precipitation biases those difficulties cause..” -> “However, given the difficulties that many Earth System Models (-> climate models (?)) have with proper representations of convective processes and the resulting precipitation biases that those difficulties cause..” (would be much more easy to understand for a non-native reader).

Changed. Thanks for the suggested phrasing!

Caption of Fig. 3: should there be $\hat{B}$ rather than $\hat{T}$?

Yes, fixed.

Caption of Fig. 5: “Differences in the precipitation scaling pattern...” Actually, the left column panels do not depict differences but the absolute distributions of P_{1-50}. Caption text needs revision. The same error occurs in the captions of Figs. 9 and 17.

Agreed. Thanks for catching that.

Fig. 8: Should the title of the top panel be “physically-based” rather than “reconstruction”?

Yes, thanks for catching that. Although we have removed this panel anyway, as we no longer include the physically-based reconstruction.

Fig. 11: this period, years 1965–1989 (if I have calculated correctly), actually does not yet belong to the RCP but to the historical period of the CMIP5 runs.

This is correct. We have clarified what we meant.