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.

Reviewer #2

The submitted manuscript compares several methods for the pattern scaling of precipitation across time periods and scenarios. They compare a regression based approach, an epoch difference and a 'physically' approach. I cannot recommend this paper for publication because of two significant errors in the methodology, combined with a manuscript which is too long, without a clear structure.

We have substantially shortened the paper and provided outlining and clearer desciptions as to our main findings.

Firstly, the 'physically-based' approach, which is based on the work of Lau (2013), is very likely incorrectly applied. In Figure 4, which is basically a test of whether the methods are able to reconstruct an in-sample pattern of precipitation using the same ensemble and time period as a test response pattern as was used to produce the pattern itself. In this case, the method produces errors an order of magnitude greater than the other approaches - which suggests that there is an error in application. If there is no error, this huge discrepancy requires an explanation.

However, even taking this into account, there is little logic that this approach is 'physically-based' at all. The precipitation rates are binned by different monthly rain rates, averaged over the ensemble and recombined into a single pattern. If a single pattern is being scaled - the ability to treat differently rain rates in different regimes has already been lost. The entire concept is not clearly defensible.

After careful consideration (see general response above), we have removed the physically-based method from this manuscript.

The separation of response patterns into CO2 and non-CO2 components could potentially be useful, but the implementation is flawed. The authors assume in Figure 14 that the non-CO2 response pattern is given by the difference between the RCP8.5 and 1pctCO2 patterns. This is not correct.

Assume there is a 'pure CO2' precipitation response which can be measured from the 1pctCO2 simulation:

$B_{CO2} = \Delta P_{1pctCO2}/\Delta T_{1pctCO2}$ If we assume things are linear, the precipitation response in RCP8.5 is this pure CO2 response, multiplied by the pure CO2 warming, plus a non-CO2 response:

$\Delta P_{RCP\ 85} = \Delta T_{RCP\ 85,CO2} B_{CO2} + \Delta T_{RCP\ 85,nonCO2} B_{nonCO2}$ so - by solving this, we get the B_{nonCO2} pattern and could reconstruct the ΔP_{RCP85} exactly.

We thank the reviewer for this comment. We agree that we were not as careful as we should have been in the previous iteration of this manuscript. We have added Supplemental Section 1, which goes through this derivation and arrives at a more accurate formulation for the non-CO2 pattern.

However, it's still not clear that CO2/nonCO2 is the correct way to break this problem down. The nonCO2 component is a broadly mix of aerosols, and other greenhouse gases (CH4, N2O etc). These two groups can have opposite effects on global mean temperature - potentially making $\Delta T_{RCP85,nonCO2}$ near zero and making the above equation ill-posed.

Furthermore, CH4 and aerosols have very different precipitation response fingerprints. RCP8.5 and RCP2.6 have very similar aerosol forcings, but very different CH4 trajectories, so the nonCO2 pattern appropriate for RCP8.5 would be very different than that for RCP2.6.

A far more logical decomposition would be between GHG and nonGHG forcing. The authors could solve this by treating the 1pctCO2 response as the GHG response pattern, and then in RCP8.5 calculating the effective CO2 concentration using the emission factors for each of the non CO2 gases, and then computing the $\Delta T_{RCP85,GHG}$ as before using effCO2 rather than CO2 itself.

We acknowledge the reviewer's excellent point. We have opted to keep the division into CO2 and non-CO2 because dividing into GHG and nonGHG components results in nonlinearities that violate the conditions of pattern scaling. The new Supplemental Section 1 provides more details as to why we made this choice. We have also added a new paragraph of text in Section 4 that describes the above issues that the reviewer raises.

The general formulation of the rest of the paper, and the treatment of the other two pattern scaling approaches, is broadly correct - but the presentation is often frustratingly vague. It is often not made clear what is in sample, and what is being tested. In Figure 8, are the same models being used to make the patterns and the test the errors? In Figure 11, is it 1pctCO2 or RCP85 being reconstructed?
The authors should correct the major errors above and restructure the paper to ensure concise and clear communication before resubmission.

We acknowledge both of the items the reviewer points out. We have clarified

our description of Figure 8, also in line with a comment from Reviewer #1. For Figure 11, we have clarified what we are doing in the text.

Exploring precipitation pattern scaling methodologies and robustness among CMIP5 models

Ben Kravitz[1], Cary Lynch[2], Corinne Hartin[2], and Ben Bond-Lamberty[2]

[1]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA.
[2]Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA.

Correspondence to: Ben Kravitz, P.O. Box 999, MSIN K9-30, Richland, WA 99352, USA. (ben.kravitz@pnnl.gov)

Abstract. Pattern scaling is a well established method for approximating modeled spatial distributions of changes in temperature by assuming a time-invariant pattern that scales with changes in global mean temperature. We compare ~~three~~ two methods of pattern scaling for annual mean precipitation (regression ~~, epoch difference, and a physically-based method~~and epoch difference) and evaluate which ~~methods are~~ method is "better" in particular circumstances by quantifying their robustness to interpolation/extrapolation in time, inter-model variations, and inter-scenario variations. ~~Although~~ Both the regression and epoch difference methods (the two most commonly used methods of pattern scaling) have ~~better~~ good absolute performance in reconstructing the climate model output~~by two orders of magnitude (~~, measured as an area-weighted root mean square error~~), 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~~. We decompose the precipitation response in the RCP8.5 scenario into a CO_2 portion and a non-CO_2 portion~~; these two patterns oppose each other in sign. Due to low signal-to-noise ratios, extrapolating~~ . Extrapolating RCP8.5 patterns to reconstruct precipitation change in the RCP2.6 scenario results in ~~double the error of reconstructing the RCP8.5 scenario for the regression and epoch difference methods~~large errors due to violations of pattern scaling assumptions when this CO_2/non-CO_2 forcing decomposition is applied. The methodologies discussed in this paper can help provide precipitation fields ~~for~~ to be utilized in other models (including integrated assessment models or impacts assessment models) for a wide variety of scenarios of future climate change.

1 Introduction

Quantifying uncertainties in projections of climate change is one of the cornerstone investigative areas in climate science. There are numerous sources of uncertainty, including parametric (which parameter values are the "right" ones), structural (which key processes are missing or poorly characterized), and scenario (how climate forcing agents will change in the future). One commonality among these sources is that uncertainties in each of them can be explored using climate models.

~~Earth System Models (ESMs~~Atmosphere-Ocean General Circulation Models (AOGCMs) are the gold standard of climate models used for projections of global change, as they incorporate many of the fundamentally climatically important processes, including atmosphere, land, ocean, and sea ice responses and feedbacks, as well as interactions between these different areas. ~~Their~~ However, their complexity means that these models are often computationally expensive, ~~however,~~ so any sensitivity studies or uncertainty quantification efforts using them are necessarily limited. No modern uncertainty quantification technique is capable of fully characterizing the space of ~~ESM~~ AOGCM uncertainties and how they affect projections of climate change (Qian et al., 2016).

Emulators of ~~ESMs~~ AOGCMs are often an effective compromise for exploring uncertainty by sacrificing precision for vastly improved computational efficiency. This allows other models, such as integrated assessment models or impacts assessment models, to include ~~a ESM-emulating~~ an AOGCM-emulating climate component and incorporate feedbacks between the climate and other sectors. There are many methods of building emulators (see MacMartin and Kravitz, 2016, for a discussion of different linear, time-invariant approaches), but one of the most commonly used methods is pattern scaling, described in more detail in Section ~~2.1~~2. This methodology involves computing a time-invariant pattern of change in a variable in response to change in global mean temperature, which vastly reduces the dimensionality of input needed to produce projections of climate change.

Pattern scaling has a fairly long history of research (e.g., Mitchell, 2003) and has been shown to be reasonably accurate for a variety of purposes. Lynch et al. (2017) provide a review of pattern scaling of temperature, as well as an in-depth exploration of two commonly used pattern scaling methods (regression and epoch difference methods, described later in Section 2.1). Both of these methods perform quite well in reproducing the actual model output for temperature. Conversely, comparatively little work has been done on pattern scaling for annual mean precipitation. Ruosteenoja et al. (2007) found that local precipitation changes are generally linear with global mean temperature change, with errors of 15–30% over 90 years of simulation. Holden and Edwards (2010) identified the importance of covariance between local temperature change and local precipitation change, and Frieler et al. (2012) furthered this discovery, concluding that no single fit (e.g., regression coefficients) will be applicable to all grid points. Herger et al. (2015) used a novel method of piecing together results associated with the desired global mean temperature change and found excellent agreement with model output (errors rarely exceed 0.3 mm day^{-1}). In a different style of emulation, Castruccio et al. (2014) trained a statistical model on pre-computed climate model simulations and found that it was capable of capturing nonlinearities in the response in ways that pattern scaling inherently cannot. Xu and Lin (2017) compared several different methods (akin to what we do in Section 4) to assess pattern scaling on temperature, precipitation, and potential

evapotranspiration in the CESM Large Ensemble (Kay et al., 2015). To the best of our knowledge, no previous study has compared different methods of pattern scaling of precipitation, particularly with a focus on robust model response.

Here we provide a systematic (although non-exhaustive) assessment of the robustness of pattern scaling of precipitation. Section 3 focuses on pattern scaling the response to temperature changes solely due to carbon dioxide increases, looking at interpolation in time, extrapolation in time, ~~extrapolation,~~ and inter-model robustness. Section 4 explores inter-scenario robustness, i.e., whether the patterns obtained for CO_2 are useful for pattern scaling other scenarios.

~~One common feature of all pattern scaling methods is that they are largely statistical approaches. While this is often suitable for obtaining scaling factors that accurately approximate simulations conducted with more complicated models (here, Earth System Models), there are potential issues that are necessarily inherent to statistical approaches, primarily dealing with nonlinearity in the response and extrapolation. If the climate response is perfectly linear, then any pattern scaling method will work equally well and will be highly accurate. However, if the climate response is nonlinear, as might be expected to some degree, then any linear approximation will have reduced fidelity, and error will increase as one extrapolates to time periods farther from the training data set. Conversely, physically-based approaches are less prone to issues that arise from extrapolation, provided that all of the relevant system dynamics are captured in the emulated pattern. There are many possibilities for physically-based approaches; here we evaluate one (described below) and compare it to the other two methods.~~

Through these investigations, we hope to better reveal ~~which~~ in what circumstances methods of pattern scaling of precipitation perform ~~better than others~~well. We will also provide some (limited) guidance as to which situations pattern scaling is likely to provide a computationally efficient, reasonably accurate result, versus which situations require actual simulation using ~~Earth System Models~~AOGCMs.

2 Pattern Scaling Methods

2.1 ~~Three~~ Two Methods of Pattern Scaling for Precipitation

Pattern scaling involves approximating a timeseries of the pattern of change in a field of interest $\Delta B(\mathbf{x},t)$ by $\Delta\hat{B}(\mathbf{x},t)$:

$$\Delta B(\mathbf{x},t) \approx \Delta\hat{B}(\mathbf{x},t) = P(\mathbf{x})\Delta\bar{T}(t) \tag{1}$$

where $P(\mathbf{x})$ describes a time-invariant spatial pattern (the spatial dimension is denoted by $\mathbf{x}$), and $\Delta\bar{T}(t)$ describes a time-varying (the time dimension is denoted by t) series of the change in global mean temperature, starting from a reference period $t=0$ (often the preindustrial era). This notation will be used repeatedly throughout the manuscript. There are two commonly used methodologies for ascertaining $P(\mathbf{x})$: regression and epoch differencing (Barnes and Barnes, 2015). In the regression method, $P(\mathbf{x})$ is obtained by regressing $\Delta B(\mathbf{x},t) = B(\mathbf{x},t) - B(\mathbf{x},0)$ against $\Delta\bar{T}(t)$ at each point in $\mathbf{x}$. In the epoch method,

$$P(\mathbf{x}) = \frac{B\left(\mathbf{x},[k,n+k]\right) - B\left(\mathbf{x},[0,n]\right)}{\bar{T}\left([k,n+k]\right) - \bar{T}\left([0,n]\right)} \tag{2}$$

where the intervals $[0,n]$ and $[k,n+k]$ indicate averaging over n-year time periods at the beginning and end of the simulation, respectively. All values calculated are over a multi-model mean; Ruosteenoja et al. (2007) showed that pattern scaling for

precipitation over a model mean outperforms results obtained from using single models. Frieler et al. (2012) argued that no single set of regression coefficients will be applicable to all grid points. We circumvent this issue by (for example) regressing $\Delta\bar{T}$ against ΔB at each grid point. By the results of Lynch et al. (2017), who showed excellent pattern scaling relationships for temperature, this approach automatically accounts for correlations between local temperature and local precipitation changes.

In addition to these two methods, which were explored by Lynch et al. (2017) for use in pattern scaling temperature, we introduce a third, physically-based method. The physically-based reconstruction is founded on the work of Lau et al. (2013), who discovered a robust hydrological cycle response to global warming amongst the models participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor et al., 2012). The idea of this method is based on the "rich-get-richer, poor-get-poorer" concept wherein areas that receive high amounts of precipitation will receive more as global temperature increases, and areas with low precipitation will receive even less (e.g., Trenberth, 2011). Instead of finding differences in annual averages, this method calculates differences in statistics of precipitation intensity and spatial distribution.

The pattern $P(x)$ for the physically-based method is constructed as follows:

1. For each model, monthly mean precipitation values over the first 25 years of the simulation being evaluated are binned into low (< 0.3 mm day^{-1}), medium ($0.9 - 2.4$ mm day^{-1}), and high precipitation (> 9 mm day^{-1}). The same is done for a later epoch of length 25 years. The difference between the two epochs is then calculated, yielding three maps of precipitation change for low, medium, and high precipitation.

2. The multi-model mean of the results of Step 1 is calculated for each of the three maps. Map values are set to 0 where fewer than 65% of the models agree on the sign of the response. Then all three maps are added together.

3. For each epoch, global mean temperature is averaged over the entire 25-year period, and then the results for the two epochs are subtracted to obtain an estimate of change in global mean temperature over that period. The pattern $P(x)$ is then defined as the map from Step 2 divided by this temperature change.

Lau et al. (2013) found that for 14 models participating in CMIP5, there is a robust hydrological cycle response both in terms of frequency of occurrence of precipitation events, as well as the spatial distribution of precipitation change. Preliminary tests using a different set of models (Table 1 below, results not shown) indicate that we can replicate the findings and spatial patterns of precipitation change as depicted by Lau et al. (2013) rather well.

2.2 Methodology

In the following sections, we quantify differences between the reconstruction $\hat{B}$ and the actual model output B via the root mean square (RMS) over the area-weighted difference $\hat{B} - B$, calculated as

$$\mathrm{RMS} = \frac{\sqrt{\sum_x \left[\left(\hat{B}(x) - B(x)\right) \cdot A(x)\right]^2}}{\sqrt{\sum_x \left[A(x)\right]^2}} \tag{3}$$

where $A(x)$ is the area of grid box x, and sums are calculated over all x.

~~Because Lau et al. (2013) only define their methodology for grid boxes between 60°S and 60°N, we compared RMS values restricted to that range with RMS values calculated over the entire globe. The results were quite similar in both cases (comparison not shown), so we only report RMS values calculated over the entire globe. This indicates that, on average, errors in the range of 60°S to 60°N are similar to errors at high latitudes, which is somewhat inconsistent with a conclusion of Tebaldi and Arblaster (2014) that scalings that include zonal mean temperature have better fidelity to the actual model output because they can account for the effects of polar amplification. This indicates the potential for robustness of the physically-based method.~~

All of the analysis conducted here uses simulations from ~~Earth System Models~~ AOGCMs contributed to CMIP5. The models used in the bulk of the analysis in this study (Table 1, Group 1) are identical to those used by Lynch et al. (2017) with two exceptions (due to model output availability):

1. The present study used NorESM1-ME instead of NorESM1-M. NorESM1-ME includes prognostic biogeochemical cycling and has the capability of being emissions-driven, but when using concentration-driven scenarios (as is the case here), the two versions of the model will produce nearly identical results (Bentsen et al., 2013).

2. The present study used CMCC-CM instead of CMCC-CMS. The difference between these two versions is that CMCC-CMS has a fully-resolved stratosphere, whereas CMCC-CM is the lower-top version of the model (Davini et al., 2014; Sanna et al., 2013). Cagnazzo et al. (2013) describe some of the differences between these two models. In general, the models agree on qualitative climate features, although as might be expected, CMCC-CMS better matches observations in situations where a fully resolved stratosphere is important for capturing the effects, including dynamical feedbacks of stratospheric circulation and ozone chemistry on surface climate. Although these effects are non-negligible, they are generally of lower order than the changes that occur over the course of the scenarios analyzed in this study (to be discussed presently), so we anticipate that differences between these two models will not substantially affect results for the model mean.

Throughout this study, we evaluate three scenarios. The 1pctCO2 scenario involves a 1% per year increase in the CO_2 concentration, beginning at its preindustrial value. This simulation is run for 140 years to an approximate quadrupling of the CO_2 concentration. The RCP8.5 and RCP2.6 scenarios (Representative Concentration Pathways, or RCPs; Moss et al., 2010; Meinshausen et al., 2011) describe the results of two socioeconomic narratives that produce particular concentration profiles of greenhouse gases, aerosols, and other climatically relevant forcing agents over the 21st century. The RCP8.5 scenario reflects a "no policy" narrative, in which total anthropogenic forcing reaches approximately 8.5 W m^{-2} in the year 2100. Conversely, the RCP2.6 scenario involves aggressive decarbonization, causing radiative forcing to peak at approximately 3 W m^{-2} around 2050 and decline to approximately 2.6 W m^{-2} at the end of the 21st century. Table 2 provides additional forcing details for the two RCP scenarios, as calculated by Hector (Hartin et al., 2015), a climate, carbon-cycle model that is used as the climate component of the Global Change Assessment Model (GCAM), a state-of-the-art Integrated Assessment Model. Both RCPs are appended to simulations of the historical period, for total simulation lengths of 251 years (1850–2100).

Throughout the remainder of the paper, subscripts on P, $\bar{T}$, $\hat{B}$, and B are used to denote the scenario (e.g., RCP8.5), the model group (e.g., Group 2), or the years over which the patterns are computed (e.g., $1-50$). If there is no subscript specified,

then the associated value corresponds to the Group 1 (see Table 1) multi-model mean of the 1pctCO2 simulation, averaged over years 116–140 of the simulation (the last 25 years of the 1pctCO2 simulation, approximately at quadruple the preindustrial CO_2 concentration).

Statistical significance was calculated using Welch's t-test, which is analogous to a Student's t-test, but where the variances s_1 and s_2 of the two samples x_1 and x_2, respectively, do not need to be equal. We use this statistic here because the ensemble for each method is small, and the ensemble pattern distribution is assumed to be normal. The test statistic is defined by

$$t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \tag{4}$$

where n_1 and n_2 are the number of models in each sample, respectively. Once the t statistic is calculated for each grid box, the value in any given grid box is determined to be statistically significant if the test value exceeds a threshold computed from the inverse of the Student's t cumulative probability distribution at the 97.5% confidence level (which is the 95% confidence level for a two-sample test). The number of degrees of freedom df used to generate that threshold is approximated by the Welch-Satterthwaite Equation:

$$df = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{s_1^2/n_1}{n_1-1} + \frac{s_2^2/n_2}{n_2-1}} \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1} + \frac{\left(s_2^2/n_2\right)^2}{n_2-1}} \tag{5}$$

In all figures, stippling is used to obscure values that are not statistically significant, i.e., the t-statistic failed to exceed the 95% confidence threshold.

3 Comparisons Between Pattern Scaling Methods for CO_2-Only Forcing

3.1 Pattern Scaling for CO_2 Concentration Changes

Figure 1 shows the baseline (preindustrial) annual mean precipitation pattern $B(x,0)$ and the scaling patterns $P(x)$ for ~~each of the three~~ both of the pattern scaling methods generated from the Group 1 (see Table 1) model average for the 1pctCO2 simulation. The regression and epoch difference methods have very similar scaling patterns, ~~and both are quite different from the comparatively more responsive physically-based method. All~~ no differences greater in magnitude than 0.05 mm day^{-1} K^{-1}, and no differences are statistically significant (not shown). Both patterns show similar broad features: an increase in tropical precipitation with global warming, particularly over the oceans; increases at high latitudes, again over the oceans; and decreases in the South Pacific, North Atlantic, and South Indian Oceans, as well as Central America and the Mediterranean basin.

Figure ?? ~~shows differences in $P(x)$ between each of the three methods. The epoch difference and regression methods have no differences greater in magnitude than 0.05 mm day^{-1} K^{-1}, and no differences are statistically significant. In comparison to those two methods, the physically-based method is more responsive in the tropics (except the equatorial Pacific, where it is less responsive) and generally less responsive in other regions where the patterns show a nonzero response. Most of these differences in patterns are statistically significant.~~

3.2 ~~Pattern Scaling for CO_2 Concentration Changes~~

~~Figure~~ 2 shows a comparison between the actual model output (Group 1 averaged over the mean of years 116–140 of the 1pctCO2 simulation) and the ~~three~~ two methods of reconstruction. ~~All of the~~ Both methods show qualitatively similar features. ~~The physically-based reconstruction has too much precipitation in the tropics and Northern Hemisphere subtropics, especially over the tropical oceans~~In general, they reproduce the actual model output well, with possible exceptions in the tropics. Tebaldi and Arblaster (2014) note that pattern scaling methodologies have difficulty in representing convection processes, so departures in these areas might be expected. ~~The regression and epoch difference methods are quite similar to each other, and both reproduce the qualitative features of the actual model output well.~~

Figure 3 shows a more quantitative comparison between the different reconstruction methods and the actual model output. ~~The physically-based method has larger error than the other two methods nearly everywhere.~~ Overall error (RMS; ~~Equation~~equation 3) ~~is approximately 50 times greater than the other two methods (Table ??). Error~~ in the regression and epoch different methods are very small (0.04 and 0.03 mm day^{-1}, respectively; see Table **??**), and no region in the reconstruction is statistically different from the actual model output. ~~Conversely, the physically-based reconstruction often exceeds 25% error.~~

3.2 Interpolation/Extrapolation

In this section, we examine robustness of the methods to interpolation or extrapolation in time. If the scaling pattern $P(\mathbf{x})$ truly ~~is~~ were time-invariant, then the results presented in this section ~~will~~ would be identical to those previously discussed.

~~The poor performance of the physically-based method extends to extrapolation. Figure ??~~ Supplemental Figure 1 shows the patterns $P(\mathbf{x})$ obtained by conditioning the reconstructions only on years 1–50 of the 1pctCO2 simulation. In the ~~physically-based and epoch difference methods~~epoch difference method, the second epoch is calculated over years 26–50 instead of years 116–140. In the regression method, the regression coefficients are calculated only using the first 50 years of simulation. The ~~pattern $P(x)$ is nearly zero everywhere for the physically-based method, indicating that the statistics of precipitation do not differ appreciably between the two periods. This results in large areas of statistically significant differences from the patterns calculated using the full 140 years. Conversely, the patterns calculated~~ patterns calculated by using the regression and epoch difference methods only show small changes between the two periods, virtually none of which is statistically significant.

Despite similarities, using patterns conditioned on the earlier period to reconstruct the precipitation in the later period (years 116–140) results in considerably poorer performance for ~~the regression and epoch difference methods (Figure ??~~both methods (Supplemental Figure 2) than the results shown in Figure 3. RMS error ~~for the regression and epoch difference methods~~ increases by an order of magnitude (~~Table ??~~not shown), although few areas show statistically significant differences from the actual model output over this time period. ~~RMS error for the physically-based method is reduced by a factor of three and is similar to the RMS error of the other two methods. The physically-based reconstruction still has areas of statistically significant error, but they are generally fewer and smaller in absolute magnitude.~~ This is likely due to the noise introduced by building $P(x)$ on the early years of the simulation when the climate change signal is weak.

~~Figure ??~~ Supplemental Figure 3 shows results for interpolation in time, where the patterns are conditioned on the full 1pctCO2 simulation (years 116–140), but the reconstruction predicts the average temperature in years 58-82 (halfway through the 1pctCO2 simulation). More specifically, $\hat{B} = P_{116-140}(\mathbf{x})\Delta\bar{T}(58-82)$. In general, the patterns for interpolation show similar qualitative features to those of reconstructing the later time period of years 116–140 (Figure 3). However, error ~~is reduced by a factor of two for the physically-based method and~~ increases by a factor of two for ~~the regression and epoch difference methods~~both methods, which potentially indicates the presence of nonlinearity. As before, no difference is statistically significant~~for the regression or epoch difference methods. The physically-based method has some areas with statistically significant differences, but they are fewer and smaller in extent.~~ .

3.3 Inter-Model Robustness

In this section, we explore the role of the number of models in improving robustness of the prediction, as well as inter-model robustness of pattern scaling by comparing reconstructions with actual model output where the scaling pattern $P(\mathbf{x})$ is conditioned on an entirely different set of models. More specifically, we examine two questions: (1) How does the prediction fidelity vary with the number of models used in the average? (2) If one conditions the pattern scaling on the average of Group 1, can one predict the response of Group 2 (or vice versa)?

Figure 4 shows the RMS error in the reconstruction (1pctCO2 simulation, averaged over years 116–140) as a function of the number of models used in the comparison. This figure was created by randomly sampling the space of all 26 models listed in Table 1 and then building P, $\bar{T}$, and $\hat{B}$ for the models in that set; each box/set of whiskers indicates 20 different sets of random samples. Results ascertained from this figure parallel those discussed in previous sections: ~~the regression and epoch difference~~ both methods have similar magnitudes of error (except for small numbers of models)~~, and both have error that is approximately two orders of magnitude lower than the physically-based reconstruction. The values in~~ . The RMS error values (Table **??**~~indicate that~~) for Group 1 (13 models) are consistent with the RMS error ranges depicted in Figure 4, indicating that Group 1 is not an outlier.

~~Both the physically-based reconstruction and regression methods show~~ The regression method shows a dependence of RMS error on the number of models, whereas with the exception of low model numbers (<10), there is much lower dependence for the epoch method. However, except for low model numbers, none of the boxes/whiskers is substantially different from any of the others, leading us to conclude that each of the methods is largely robust to changes in the number of models used to carry out pattern scaling. ~~Appendix 1~~ Supplemental Section 2 and the associated figures provide additional comparisons between the patterns generated for Groups 1 and 2.

3.4 Discussion of Pattern Scaling the Precipitation Response to CO_2

~~The~~ Both the regression and epoch difference methods show great promise in their usefulness as precipitation pattern scaling methods. Both are able to reconstruct the changes in precipitation due to CO_2 increases with errors of less than 5% in every region of the globe (Figure 3). ~~Conversely, the physically-based method has comparatively poor performance, with error regularly exceeding 25%. This method often scales too strongly with global mean temperature change. The test of extrapolation~~

shows that for the time periods analyzed here, using $P(x) = 0$ would result in better absolute performance than using the physically-based pattern $P(x)$ depicted in Figure **??**.

However, the physically-based method shows robustness in several ways that the regression and epoch difference methods do not. One example is with interpolation However, where error drops substantially for the physically-based method but increases for the other two. Another is related to inter-model robustness: the pattern of error is relatively unchanged for the physically-based method, regardless of the group of models that is used (Appendix when examining interpolation in time, error increases for both methods, indicating issues with robustness to timescale (Supplemental Section 1), whereas . Also, the pattern shows increased error in many places for the other methods. 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, particularly in situations (like the ones explored in Sections 3.2 and 3.3)where statistically-based or other non-physical methods might be expected to perform less wellwhen different models are used (Supplemental Section 2), indicating issues with inter-model robustness.

Like the temperature pattern scaling results of Lynch et al. (2017), we find that the regression and epoch difference methods have similar performance. In the present work, we find that the epoch difference method slightly outperforms the regression method, but the differences are relatively minor. Given the slight advantages in computational expense and reduced data input requirements, we profess a slight preference for using the epoch difference method to generate scaling patterns for the precipitation response to CO_2-induced global warming. In the next section, we explore a more broad application of pattern scaling by including non-CO_2 forcings.

4 Pattern Scaling for Non-CO_2 Additional Forcings

In this section, we compare the patterns and reconstructions between scenarios, primarily related to the RCP8.5 and 1pctCO2 simulations. We do this first as a test of robustness: does any one of the three methods one method perform "better" for CO_2-only simulations versus RCP8.5? If the fidelity of the reconstruction to the actual model output is similar for the two scenarios, then subtracting the reconstructions conditioned on RCP8.5 and 1pctCO2 could reveal a scaling pattern for non-CO_2 forcing. We note that this is one of the few ways of ascertaining the non-CO_2 response pattern without running separate simulations both with and without CO_2 forcing—without a scaling method to normalize for similar climate conditions, there is no way of obtaining meaningful results from directly subtracting a 1pctCO2 simulation from an RCP8.5 simulation. (The approach discussed here is analogous to the methodology of Herger et al. (2015), but where they attempted to ascertain similarities between patterns for a given change in global mean temperature, we are interested in the differences.)

We note several caveats with this approach. One is that, based on the results of Herger et al. (2015), the reconstructions of RCP8.5 and 1pctCO2 are likely to be similar have some similarities for a given temperature change because the dominant forcing in RCP8.5 is CO_2 (see Table 2). As such, ascertaining the non-CO_2 signal could be limited by low signal-to-noise ratios. A second caveat, one more germane to pattern scaling, is to ascertain whether the non-CO_2 pattern obtained from RCP8.5 can be used to reconstruct the non-CO_2 precipitation change for a different scenario. There is no *a priori* reason to expect that

this will work, as different scenarios have different combinations of forcings~~, but as long as the CO_2 portion dominates the response, such endeavors may still be useful~~. In Section 4.3, we investigate this problem using an extreme case, where we ascertain the scaling patterns from an RCP8.5 simulation and use them to attempt to reconstruct the RCP2.6 simulation.

We acknowledge that the non-CO_2 component is a combination of both non-CO_2 greenhouse gases and aerosols, which have opposite effects on global mean temperature. These two categories of forcing have different local responses as well. An alternative approach would be to split the RCP8.5 response into a CO_2 component, a non-CO_2 greenhouse gas component, and a non-greenhouse gas component. Supplemental Section 3 discusses the necessary calculations for both of these approaches. The CO_2/non-CO_2 approach proved to be quite amenable to pattern scaling. On the contrary, the CO_2/other greenhouse gas/non-greenhouse gas approach is not, due to distinct nonlinearities in the derived temperature responses for these particular forcing categories. As such, we have chosen to proceed with a CO_2/non-CO_2 division for the purpose of pattern scaling.

4.1 Inter-Scenario Differences

Figure 5 shows the RCP8.5 scaling pattern $P_{\mathrm{RCP8.5}}(\mathbf{x})$ and the difference from the CO_2-only pattern. Patterns are nearly identical to those in Figure 1. ~~The physically-based pattern shows some statistically significant changes, particularly in the tropics and at high latitudes. The~~ Both the regression and epoch difference methods show no differences exceeding 0.1 mm day^{-1} K^{-1} in magnitude and no statistically significant differences of any magnitude. This figure reinforces the findings of Herger et al. (2015) that patterns generated from commonly used scaling methods (regression and epoch difference) do not differ appreciably between scenarios, so pattern scaling can be accomplished by using periods in different scenarios with the same global mean temperature change.

Figures 6 and 7 show this in practice, where the reconstruction of the historical/RCP8.5 simulation $\hat{B}$ is built on the RCP8.5 pattern, multiplied by $\Delta\bar{T}$ averaged over years 227–251 (2076–2100) and 116–140 (1965–1990), respectively. The ~~physically-based method qualitatively and quantitatively matches the results in Figure 3. For the regression and epoch difference methods, the~~ reconstructed precipitation response in Figure 6 is generally too strong in the tropics and too weak in the midlatitudes (which is the same pattern in Figure 3), but Figure 7 shows the opposite pattern. None of these differences is statistically significant, and the RMS error is approximately the same in both figures (0.09–0.10 mm day^{-1} K^{-1}; 2–3 times greater than the error in Figure 3), but they suggest that there is a distinct non-CO_2 pattern that, while small, is still important in explaining precipitation differences in periods with large temperature change.

Figure 8 provides descriptions of the actual precipitation effects of both CO_2 and non-CO_2 forcing. ~~(In the following, we omit discussion of the results from the physically-based method due to the aforementioned strong biases in the results.) Large features of~~ Although the two portions of the reconstruction generally show similar features, the regional effects have quite different magnitudes in many regions. In particular, the non-CO_2 response ~~are the opposite of~~ is weaker over the tropical Pacific than the CO_2 response ~~, indicating an offsetting. In the non-CO_2 forcing case, equatorial precipitation is weakened, possibly indicating a Southward shift in the intertropical convergence zone. Precipitation~~ and is stronger over much of the Northern Hemisphere. One distinct difference between the two patterns is that precipitation is reduced over East Asia and ~~the maritime continent is also weaker, while precipitation in Amazonia, the South Pacific, the North Atlantic , and the Northern~~

~~Hemisphere subtropical Pacific is stronger~~India in the non-CO_2 response but increases in the CO_2 response. This is likely a result of global dimming from heavy aerosol emissions. Another source of differences, potentially attributable to dust, is the Saharan outflow over the Atlantic ocean and extending into the Amazon. This gives us confidence that although the non-CO_2 response is likely dominated by non-CO_2 greenhouse gases (most prominently methane), it appears to have captured an aerosol signature. It would be a useful future area of investigation to ~~ascertain whether these patterns of precipitation change arise in climate models forced with appropriate non-CO_2 forcings , and if so, the mechanistic reasons why these responses are different from CO_2 forcing. Such investigations would also aid in understanding the robustness of these signals, i.e., what portion of the reported response in Figure 8 is signal versus noise.~~ conduct pattern scaling studies on single-forcing simulations (e.g., Marvel et al., 2016) to reveal more robust signals and determine which forcings are amenable to pattern scaling, with a particular eye on inter-model variations in the responses to identical forcings. The results in Figure 8 also reinforce the conclusions of Frieler et al. (2012), who argue that the scaling patterns from one scenario are not in general translatable to scaling patterns for another scenario if the two scenarios are driven by different forcing. Even though Figure 5 shows that the patterns $P_{\mathrm{RCP8.5}}$ and P_{1pctCO2} are nearly identical, even small differences can affect reconstructions of precipitation change for large values of $\Delta\bar{T}$.

4.2 Non-CO_2 Forcing Pattern

Here we calculate a non-CO_2 pattern for use in pattern scaling. We begin by assuming that the effects of CO_2 forcing and non-CO_2 forcing are separable, that is, that there are no nonlinear interactions between the two forcings that would produce a non-additive response. Although this assumption is not strictly true, it is approximately true to a sufficient degree that such calculations are useful (MacMartin et al., 2015; MacMartin and Kravitz, 2016). Following the notation in Equation 1, separability means that

$$\Delta\hat{B}_{\mathrm{RCP8.5}} = \Delta\bar{T}_{\mathrm{CO_2}} P_{\mathrm{CO_2}} + \Delta\bar{T}_{\mathrm{non-CO_2}} P_{\mathrm{non-CO_2}} \tag{6}$$

We set $P_{\mathrm{CO_2}}$ equal to P_{1pctCO2} (from Section 3), because if pattern scaling holds, the time-invariant pattern of CO_2 forcing should be identical, regardless of the scenario from which it is derived. $P_{\mathrm{non-CO_2}}$ is ~~assumed to be $P_{\mathrm{RCP8.5}} - P_{\mathrm{CO_2}}$. This is the inherent assumption of the~~ defined to be $4P_{\mathrm{RCP8.5}} - 3P_{\mathrm{CO_2}}$ (see Supplemental Section 3 for the derivation of this expression). Embedded in this expression are inherent assumptions about the validity of a linear pattern scaling approach. If the approach fails, it is because either this pattern does not represent actual non-CO_2 forcing or because the pattern is too difficult to accurately estimate, perhaps due to internal variability. To calculate $\Delta\bar{T}_{\mathrm{CO_2}}$, we assume that global mean temperature scales linearly with radiative forcing (e.g., Gregory et al., 2004), and radiative forcing is known to scale logarithmically with the CO_2 concentration (Myhre et al., 1998). Performing linear regression of $\log_2([\mathrm{CO_2}])$ against global mean temperature change in the 1pctCO2 simulation yields a slope of $\alpha = 2.40$, an intercept of $\beta = -19.89$, and an R^2 value of 0.99. (Brackets $[\cdot]$ indicate the CO_2 concentration in $\mathrm{ppm_v}$.) Then $\bar{T}_{\mathrm{CO_2}} = \alpha \log_2([\mathrm{CO_2}]) + \beta$. $\bar{T}_{\mathrm{non-CO_2}}$ is calculated as the residual $\bar{T}_{\mathrm{RCP8.5}} - \bar{T}_{\mathrm{CO_2}}$. We note that this formulation does not explicitly account for lags in the climate response to radiative forcing, such as ocean thermal inertia.

Figure 9 shows all of the aforementioned $\bar{T}$ values, plotted as a function of the CO_2 concentration. ~~The temperature contribution of the non-CO_2 part increases with the~~ Both CO_2 ~~concentration, which monotonically increases~~ and non-CO_2 monotonically increase with time in the RCP8.5 simulation. This is consistent with the design of the RCP8.5 scenario, in which non-CO_2 radiative forcing increases over the period 2000–2100 (Table 2), largely due to a doubling of the methane concentration over this period. This change in forcing corresponds to a non-CO_2 induced temperature change (green line in Figure 9) from 0.31 to 1.36 K.

~~We next explore the ability of this decomposition to reconstruct~~ This decomposition of the scaling patterns into CO_2 and non-CO_2 components performs rather well for reconstructing the actual model output~~for the~~. For the period 2076–2100 at the end of the RCP8.5 ~~scenario~~ simulation (Figure 10)~~and a historical period in the late 20th century~~, the epoch difference method has fewer errors than the regression method, especially in the tropics and Northern Hemisphere midlatitudes. No error in the regression method exceeds 0.4 mm day^{-1} in magnitude, and no error in the epoch difference method exceeds 0.2 mm day^{-1} in magnitude. For the period 1965–1990 (Figure 11)~~. Qualitatively, the results for the physically-based method are similar to those of Figure 6 (RMS error is approximately 44% larger) , whereas the differences between the reconstruction and the actual model output ($\hat{B} - B$) have the opposite sign as compared to Figure 6 for the other two methods (RMS error is slightly higher but of a similar magnitude). Conversely, Figure 11 shows greater similarity to Figure 6for the epoch difference and regression methods (RMS error is nearly identical); while the physically-based method shows similar qualitative features in both Figures 6 and 11 , the magnitude of error is reduced~~, both methods perform nearly identically with slight (<0.2 mm day^{-1}) biases in the tropics, East Asia, Europe, and the South Pacific. The errors in Figure 10 are generally of the opposite sign of those in Figure 6, and some of the features in Figure 11 are of the opposite sign of those in Figure ~~11 by nearly a factor of 10. The epoch difference and regression methods show a hemispheric contrast in error in the historical period, but in the later, future period, the errors have more similar patterns to those in previously described figures~~7. This indicates that some features of non-CO_2 response offset features of the CO_2 response. Particular areas of these differences include the previously identified regions associated with strong aerosol forcing.

4.3 Scaling to Predict Other Scenarios

The final stage of inter-model exploration is to see how well the CO_2 and non-CO_2 patterns generated from one scenario can be used on another scenario. Here we choose the extreme case of predicting the pattern of precipitation change in RCP2.6, based on the patterns calculated from RCP8.5. In this scenario, the CO_2 concentration peaks and then drops slightly (Table 2). The non-CO_2 forcing comprises 29% of the total forcing in RCP8.5 in 2100 and 32% of the total forcing in RCP2.6 in 2100, according to simulations using Hector (Hartin et al., 2015).

Figure 12 shows the effectiveness of this reconstruction process. ~~The physically-based method shows a similar pattern of error, but with reduced magnitude. The~~ Both the epoch difference and regression methods show ~~too much CO_2 response and not enough non-CO_2 response as indicated by~~ strong differences that are consistent with the patterns displayed in Figure~~15. This can potentially be explained by the forcing values given in Table 2. Although the~~ 8. Supplemental Section 4 provides some additional derivations that explore the sources of these biases. There are two main conclusions from this section. First,

using the non-CO_2 ~~forcing is approximately the same percentage of the total forcing in both the~~ pattern build on RCP8.5 ~~and RCP2.6 simulations, the absolute forcing values are such that only 0.37 W m^{-2} of~~ was not effective for explaining non-CO_2 ~~forcing is exerted~~ behavior in RCP2.6~~in the year 2100. This indicates that global mean temperature change in RCP2.6 due to~~ , indicating that there are limits to the applicability of a "universal" non-CO_2 forcing~~($\Delta\bar{T}_{\mathrm{RCP2.6,non-CO_2}}$) is small, and as such, the values depicted in Figure 12 are almost entirely due to CO_2 forcing. Although more formal study is needed, this indicates that~~ . A future area of investigation could explore these limits: for example, would the non-CO_2 ~~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 .~~

~~One notable feature in Figure 12 is a difference at high latitudes due to polar amplification that is more pronounced in~~ pattern built on RCP8.5 ~~than~~ work on RCP6.0 or RCP4.5 instead of the extreme RCP2.6 ~~. This is an example of a nonlinearity that cannot be well captured by linear pattern scaling; future work involving scaling by zonal mean temperature (as was suggested by Tebaldi and Arblaster, 2014) could show promise in improving these sorts of inter-scenario comparisons~~case? The second conclusions is that the ability to perform this CO_2/non-CO_2 decomposition is itself limited. Supplemental Section 4 goes through detailed calculations showing that if one assumes that such a decomposition is possible, contradictions and inconsistencies arise. Determining why this decomposition failed for RCP2.6 would require a more thorough investigation, possibly including single forcing simulations, that is beyond the scope of this study. Such research could lead to an understanding of which scenarios would be more amenable to separable forcing treatments than others.

4.4 Discussion of Pattern Scaling for Non-CO_2 Forcings

In general, the pattern scaling results depicted in Section 4 are consistent with previous studies. Herger et al. (2015) found that the patterns between scenarios are rather similar, which Figure 5 confirms. However, the results for pattern scaling may be scenario-dependent (Figure 8) if global mean temperature change ($\Delta\bar{T}$) is sufficiently large, which confirms the conclusions of Frieler et al. (2012).

~~The patterns of change indicate some degree of nonlinearity in the response, particularly in that Figures 6 and 7 show opposite signs of error. Because the relationships between CO_2 concentration and precipitation are quite linear (Section 3), we conclude that the nonlinearities are due to the non-CO_2 response. Without conducting pattern scaling analyses on single forcing runs, we are unable to ascertain the exact sources of nonlinearity. However, this does explain in part why the RCP8.5 pattern was unable to reproduce some of the features of the~~ In particular, we found limited ability in reconstructing the RCP2.6 ~~response, in that they have substantially different magnitudes of non-CO_2 forcing, likely due to different combinations of forcing agents. We note that the differences between~~ model behavior from the RCP8.5 run, indicating limits in building scalings using one scenario and applying them to a different scenario. We deliberately chose an extreme case to understand whether such universal applications exist. The ability to do this might be improved for "closer" scenarios, such as RCP8.5 and ~~RCP2.6 are as extreme as was available, and the decomposition into CO_2 and non-CO_2 components may be more effective for scenarios that are more similar, particularly in their non-CO_2 components~~RCP6.0 or RCP4.5.

5 Conclusions

We have explored ~~three different~~two different, commonly used methods of pattern scaling for annual mean precipitation, with a focus on robustness to interpolation/extrapolation in time, inter-model variations, and inter-scenario differences. ~~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. The~~ Both the regression and epoch difference methods ~~have excellent, approximately similar performance~~perform well and approximately similarly.

~~One of the features that emerged was that the CO_2 and non-CO_2 components offset each other. Without a detailed assessment of the different single forcing agents that comprise the non-CO_2 patterns, we are unable to provide a mechanistic understanding of the causes of these features, but we highlight this as a promising area of future research.~~

Most of the errors that arise for ~~any of the methods~~ either method are either in areas dominated by convection (predominantly over the tropical oceans) or at high latitudes. Both of these areas are large sources of nonlinear responses to global mean temperature change, so pattern scaling might not be expected to perform well in these areas. The ~~physically-based method does include a bin for high precipitation, so despite providing too much precipitation in these regions, it is the best equipped to deal with these sources of nonlinearity. Moreover, the~~ approach of Tebaldi and Arblaster (2014) of using zonal mean temperature as a scaling parameter may prove useful in accounting for errors at high latitudes.

In terms of the usefulness of pattern scaling of precipitation, because the regression and epoch difference methods perform well over most land regions, they are likely quite suitable for a variety of applications, including societal models (like Integrated Assessment Models or impacts models) that mostly deal with land areas. If one's application requires good performance over tropical oceans, then pattern scaling may no longer be appropriate, and instead output from the full ~~Earth System Model~~ AOGCM may be required. However, given the difficulties ~~many Earth System Models~~ that many climate models have with proper representations of convective processes and the resulting precipitation biases those difficulties cause (e.g., Song and Zhang, 2009), there may be doubts as to how well ~~Earth System Models~~ AOGCMs represent precipitation in these areas in the first place.

The results presented in Section 4 indicated that while some scenarios are amenable to broad separations of pattern scaling forcings, some others are not. Much more systematic work needs to be done in this area to determine the usefulness of pattern scaling for different forcings. Single forcing experiments would be particularly useful, as they can allow a determination as to which forcings work best for pattern scaling, as well as whether there are any nonlinear effects that result from applying multiple simultaneous forcings. Another potential approach would be to use the "hybrid-pattern" method described by Xu and Lin (2017), in which a simple energy balance model is used to build separate forcings, circumventing the need for expensive single-forcing AOGCM simulations.

The results presented here have applications that extend beyond providing libraries of scaling patterns for Integrated Assessment Models (Lynch et al., 2017). Another more speculative application involves efficacy of climate forcings. Kravitz et al. (2015) developed a method of comparing forcing agents via analyses of their rapid adjustments (fast responses), that is, their responses in the absence of global mean temperature change. If our method of decomposing the response into CO_2 and non-

CO_2 components could be extended to single forcings, then one could isolate the feedback responses (slow responses), which are the portions of the responses that depend on global mean temperature change. Thus, there is potential to provide a more quantitative intercomparison of the different effects of climate forcing agents.

6 Code and/or Data Availability

All computations were performed using MATLAB 2012b, developed by MathWorks. The libraries of patterns ~~for each method will be included in an upcoming official code release of the Global Change Assessment Model (GCAM).~~are available on GitHub at http://github.com/JGCRI/CMIP5_patterns. All code used to generate figures in this manuscript ~~will be publicly released through a code repository~~is available upon request.

Author contributions. All authors collectively designed the study. BK was the lead on conducting analysis and writing the paper, with input from the other authors at all stages.

Acknowledgements. The authors thank two anonymous reviewers for their help in improving the manuscript. This research is based on work supported by the US Department of Energy, Office of Science, Integrated Assessment Research Program. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under contract DE-AC05-76RL01830.

References

Barnes, E. A. and Barnes, R. J.: Estimating linear trends: Simple linear regression versus epoch differences, J. Climate, 28, 9969–9976, doi:10.1175/JCLI-D-15-0032.1, 2015.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate, Geosci. Model Dev., 6, 687–720, doi:10.5194/gmd-6-687-2013, 2013.

Cagnazzo, C., Manzini, E., Fogli, P. G., Vichi, M., and Davini, P.: Role of stratospheric dynamics in the ozone-carbon connection in the Southern Hemisphere, Clim. Dynam., 41, 3039–3054, doi:10.1007/s00382-013-1745-5, 2013.

Castruccio, S., McInerney, D. J., Stein, M. L., Crouch, F. L., Jacob, R. L., and Moyer, E. J.: Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs, J. Climate, 27, 1829–1844, doi:10.1175/JCLI-D-13-00099.1, 2014.

Davini, P., Cagnazzo, C., Fogi, P. G., Manzini, E., Gualdi, S., and Navarra, A.: European blocking and Atlantic jet stream variability in the NCEP/NCAR reanalysis and the CMCC-CMS climate model, Clim. Dynam., 43, 71–85, doi:10.1007/s00382-013-1873-y, 2014.

Frieler, K., Meinshausen, M., Mengel, M., Braun, N., and Hare, W.: A Scaling Approach to Probabilistic Assessment of Regional Climate Change, J. Climate, 25, 3117–3144, doi:10.1175/JCLI-D-11-00199.1, 2012.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, Geophys. Res. Lett., 31, L03 205, doi:10.1029/2003GL018747, 2004.

Hartin, C. A., Patel, P., Schwarber, A., Link, R. P., and Bond-Lamberty, B. P.: A simple object-oriented and open-source model for scientific and policy analyses of the global climate system – Hector v1.0, Geosci. Model Dev., 8, 939–955, doi:10.5194/gmd-8-939-2015, 2015.

Herger, N., Sanderson, B. M., and Knutti, R.: Improved pattern scaling approaches for the use in climate impact studies, Geophys. Res. Lett., 42, 3486–3494, doi:10.1002/2015GL063569, 2015.

Holden, P. B. and Edwards, N. R.: Dimensionally reduced emulation of an AOGCM for application to integrated assessment modelling, Geophys. Res. Lett., 37, L21 707, doi:10.1029/2010GL045137, 2010.

Kay, J. E. et al.: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability, Bull. Amer. Meteor. Soc., 96, 1333–1349, doi:10.1175/BAMS-D-13-00255.1, 2015.

Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, Geophys. Res. Lett., 40, 1–6, doi:10.1002/grl.50256, 2013.

Kravitz, B., MacMartin, D. G., Rasch, P. J., and Jarvis, A. J.: A new method of comparing forcing agents in climate models, J. Climate, 28, 8203–8218, doi:10.1175/JCLI-D-14-00663.1, 2015.

Lau, W. K.-M., Wu, H.-T., and Kim, K.-M.: A canonical response of precipitation characteristics to global warming from CMIP5 models, Geophys. Res. Lett., 40, 3163–3169, doi:10.1002/grl.50420, 2013.

Lynch, C., Hartin, C., Bond-Lamberty, B., and Kravitz, B.: An open-access CMIP5 pattern library for temperature and precipitation: Description and methodology, Earth System Science Data, p. submitted, doi:10.5194/essd-2016-68, 2017.

MacMartin, D. G. and Kravitz, B.: Multi-model dynamic climate emulator for solar geoengineering, Atmos. Chem. Phys. Discuss., p. in review, doi:10.5194/acp-2016-535, 2016.

MacMartin, D. G., Kravitz, B., and Rasch, P. J.: On solar geoengineering and climate uncertainty, Geophys. Res. Lett., 42, 7156–7161, doi:10.1002/2015GL065391, 2015.

Marvel, K., Schmidt, G. A., Miller, R. L., and Nazarenko, L. S.: Implications for climate sensitivity from the response to individual forcings, Nature Climate Change, 6, 389–389, doi:10.1038/nclimate2888, 2016.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change, 109, 213–241, doi:10.1007/s10584-011-0156-z, 2011.

Mitchell, T. D.: Pattern Scaling: An Examination of the Accuracy of the Technique for Describing Future Climates, Climatic Change, 60, 217–242, doi:10.1023/A:1026035305597, 2003.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J.: The next generation of scenarios for climate change research and assessment, Nature, 463, 747–756, doi:10.1038/nature08823, 2010.

Myhre, G., Highwood, E. J., Shine, K. P., and Stordal, F.: New estimates of radiative forcing due to well mixed greenhouse gases, Geophys. Res. Lett., 25, 2715–2718, doi:10.1029/98GL01908, 1998.

Qian, Y., Jackson, C., Giorgi, F., Booth, B., Duan, Q., Forest, C., Higdon, D., Hou, Z. J., and Huerta, G.: Uncertainty Quantification in Climate Modeling and Projection, Bull. Amer. Meteor. Soc., 97, 821–824, doi:10.1175/BAMS-D-15-00297.1, 2016.

Ruosteenoja, K., Tuomenvirta, H., and Jylhä, K.: GCM-based regional temperature and precipitation change estimates for Europe under four SRES scenarios applying a super-ensemble pattern-scaling method, Climatic Change, 81, 193–208, doi:10.1007/s10584-006-9222-3, 2007.

Sanna, A., Lionello, P., and Gualdi, S.: Coupled atmosphere ocean climate model simulations in the Atmospheric Mediterranean region: effect of a high-resolution marine model on cyclones and precipitation, Nat. Hazards Earth Syst. Sci., 13, 1567–1577, doi:10.5194/nhess-13-1567-2013, 2013.

Song, X. and Zhang, G. J.: Convection Parameterization, Tropical Pacific Double ITCZ, and Upper-Ocean Biases in the NCAR CCSM3. Part I: Climatology and Atmospheric Feedback, J. Climate, 22, 4299–4315, doi:10.1175/2009JCLI2642.1, 2009.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bull. Amer. Meteor. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Tebaldi, C. and Arblaster, J. M.: Pattern scaling: Its strengths and limitations, and an update on the latest model simulations, Climatic Change, 122, 459–471, doi:10.1007/s10584-013-1032-9, 2014.

Trenberth, K. E.: Changes in precipitation with climate change, Clim. Res., 47, 123–138, doi:10.3354/cr00953, 2011.

Xu, Y. and Lin, L.: Pattern scaling based projections for precipitation and potential evapotranspiration: Sensitivity to composition of GHGs and aerosol forcing, Climatic Change, 140, 635–647, doi:10.1007/s01584-016-1879-7, 2017.

Table 1. Models used in the present analysis. Most of the analysis was conducted using the models in Group 1. Additional investigations described in Section 3.3 ~~(inter-model robustness)~~ were conducted using the models in both Group 1 and Group ~~2.~~ 2 to determine inter-model robustness of the different pattern scaling methods discussed in this study. All model names listed are the standard names used in contributions to the Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor et al., 2012). Knutti et al. (2013) provide an excellent description of these models and their provenance.

Group 1	Group 2
ACCESS1.0	ACCESS1.3
CanESM2	BCC-CSM1.1
CCSM4	BCC-CSM1.1M
CMCC-CM	BNU-ESM
CNRM-CM5	CESM1-BGC
CSIRO-Mk3.6	CESM1-CAM5
GFDL-CM3	FGOALS-g2
HadGEM2-ES	GFDL-ESM2M
INMCM4	IPSL-CM5A-LR
IPSL-CM5A-MR	IPSL-CM5B-LR
MIROC-ESM	MIROC5
MPI-ESM-MR	MPI-ESM-LR
NorESM1-ME	MRI-CGCM3

Table 2. Radiative forcing values (W m^{-2}) for RCP8.5 and RCP2.6 in 2000, 2050, and 2100. CO_2 forcing and total forcing were calculated using the simple climate model Hector (Hartin et al., 2015). Non-CO_2 forcing is calculated as the difference between total and CO_2 forcing. Percentages in parentheses indicate the percentage of the total forcing.

	2000	2050	~~2010~~ 2100
CO_2 forcing (RCP8.5)	1.226	3.289	6.167
Total forcing (RCP8.5)	1.991	5.049	8.686
non-CO_2 forcing (RCP8.5)	0.765 (38%)	1.760 (35%)	2.519 (29%)
CO_2 forcing (RCP2.6)	1.267	2.174	1.765
Total forcing (RCP2.6)	2.066	3.195	2.601
non-CO_2 forcing (RCP2.6)	0.799 (39%)	1.021 (32%)	0.836 (32%)

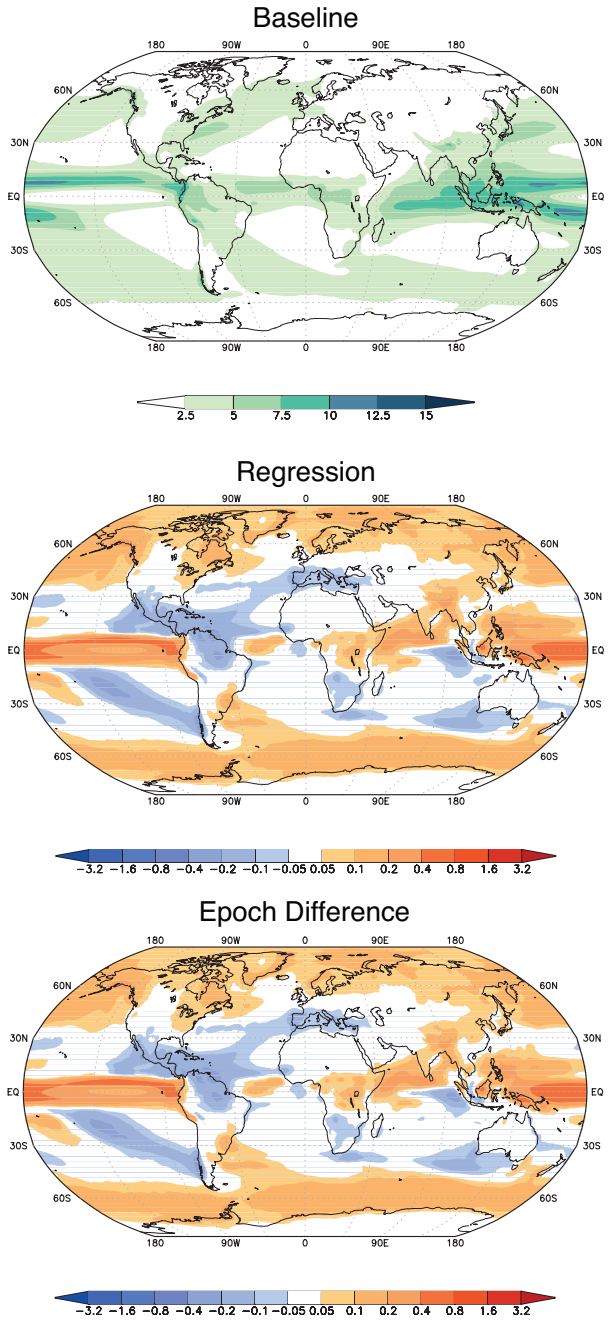

Figure 1. The components needed for pattern scaling of the precipitation response to CO_2 forcing, averaged over the 13 models in Group 1 (Table 1). ~~Top-left~~ Top shows the baseline precipitation pattern for the multi-model average: $B(x, 0)$ in Equation 1 (mm day^{-1}; averaged over years 1–25 of the 1pctCO2 simulation). ~~All~~ The other panels show the time-invariant pattern $P(\mathbf{x})$ in Equation 1 (mm day^{-1} K^{-1}) for the ~~physically-based method (top-right), the~~ regression method (~~bottom-left~~middle), and the epoch difference method (~~bottom-right~~bottom).

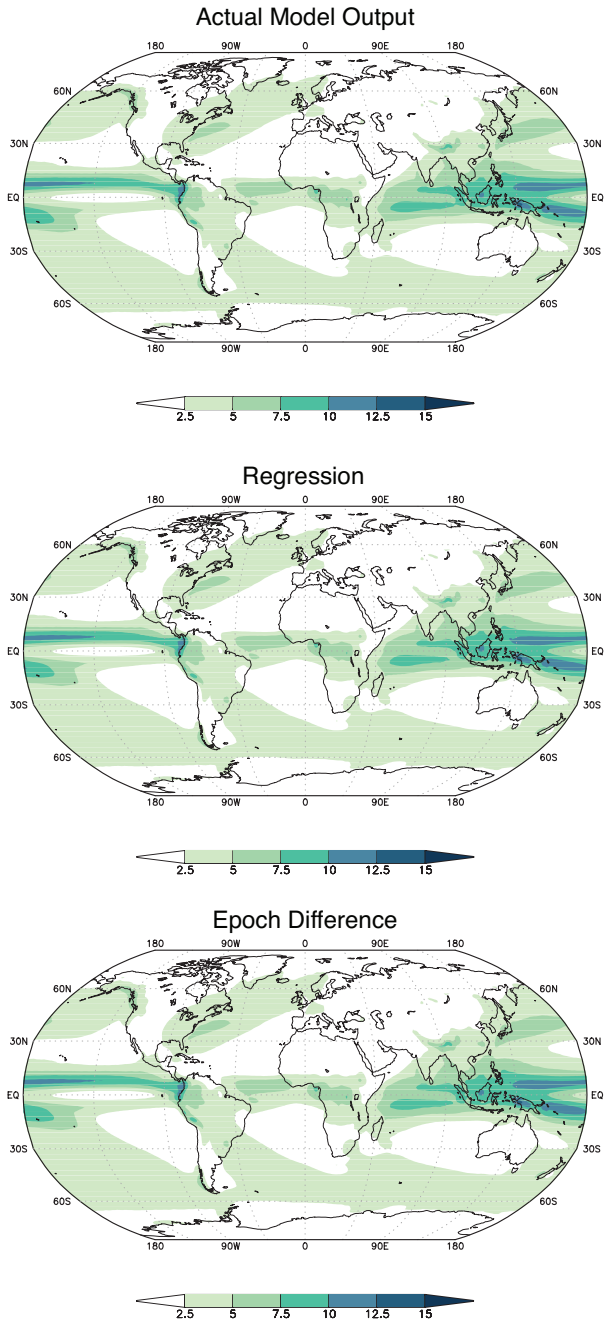

Figure 2. ~~Differences in~~ Comparison between the actual Group 1 multi-model average precipitation ~~scaling~~ output (top) and the reconstructions produced by pattern ~~$P(\mathbf{x})$~~ scaling ($\hat{B}$ in Equation 1). ~~Top panel shows $P_{\mathrm{phys}} - P_{\mathrm{epoch}}$, middle panel shows $P_{\mathrm{phys}} - P_{\mathrm{regr}}$.~~ All values are in mm day^{-1} and ~~bottom~~ represent averages over years 116–140 of the 1pctCO2 simulation. Middle panel shows ~~$P_{\mathrm{epoch}} - P_{\mathrm{regr}}$,~~ ~~where physdenotes~~ the ~~physically-based method, regrdenotes the~~ regression method, and ~~epochdenotes~~ bottom panel shows the epoch difference method. ~~All units are mm day^{-1} K^{-1} and were calculated for the multi-model average of Group 1 (Table 1). Stippling indicates a lack of statistical significance in the pattern of differences (Section 2.2).~~

~~Comparison between the actual Group 1 multi-model average precipitation output (top-left) and the reconstructions produced by pattern scaling ($\hat{T}$ in Equation 1). All values are in mm day^{-1} and represent averages over years 116–140 of the 1pctCO2 simulation. Top-right shows the physically-based method, bottom-left shows the regression method, and bottom-right shows the epoch difference method.~~

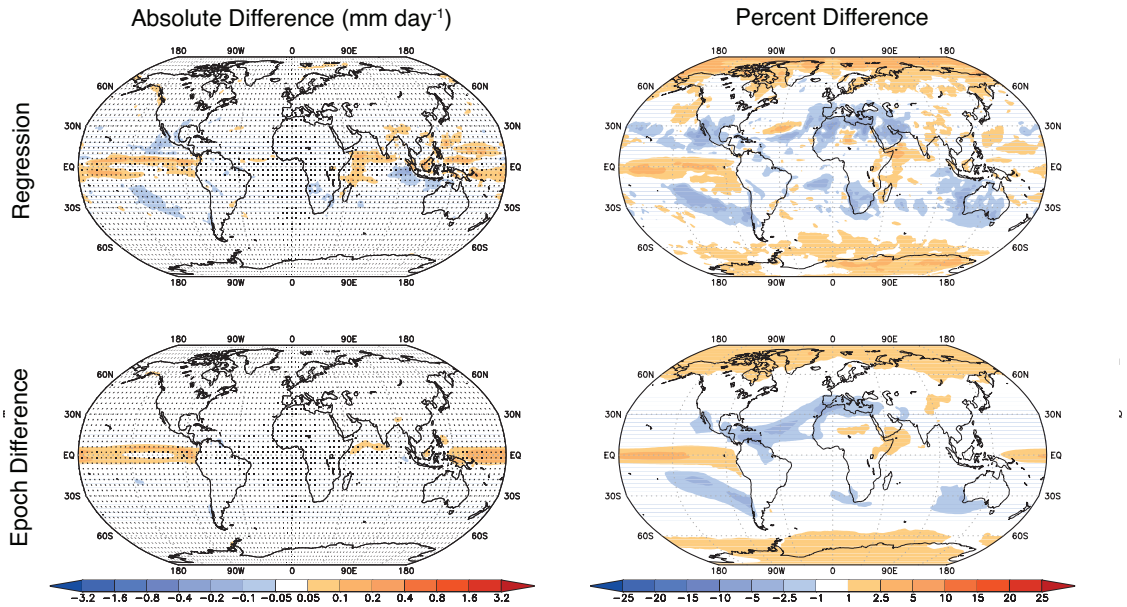

Figure 3. Differences between the reconstructions produced by pattern scaling ($\hat{B}$) and the actual model output for precipitation (B). Left column shows absolute values of $\hat{B} - B$ (mm day^{-1}), and right column shows percent change. Top row shows results for the ~~physically-based method, middle row shows the~~ regression method, and bottom row shows the epoch difference method. All values are calculated for a Group 1 multi-model average for the 1pctCO2 simulation over the years 116-140. Stippling indicates a lack of statistical significance in the pattern of differences (Section 2.2).

~~Differences in the precipitation scaling pattern $P(\mathbf{x})$ (Equation 1) when different time periods are used to construct the pattern (years 1–50 versus years 116–140 of the 1pctCO2 simulation). Left column shows values of P_{1-50}, and right column shows values of $P_{1-50} - P_{116-140}$ (mm day^{-1} K^{-1}). Values in subscripts denote that the associated quantities are calculated from an average over those years. Top row shows results for the physically-based method, middle row shows the regression method, and bottom row shows the epoch difference method. All values are calculated for a Group 1 multi-model average for the 1pctCO2 simulation. Stippling indicates a lack of statistical significance in the pattern of differences (Section 2.2).~~

~~As in Figure 3 but where the reconstruction $\hat{B}$ is built on the pattern P for years 1–50 (Group 1 average of the 1pctCO2 simulation), and global mean temperature $\Delta\bar{T}$ is averaged over years 116–140. That is, $\hat{B} = P_{1-50}(\mathbf{x})\Delta\bar{T}(116-140)$. Results shown are for the difference between the reconstruction and the actual model output $\hat{B} - B(\mathbf{x}, 116-140)$.~~

~~As in Figure 3 but where the reconstruction $\hat{B}$ is built on the pattern P for years 116–140 (Group 1 average of the 1pctCO2 simulation), and global mean temperature $\Delta\bar{T}$ is averaged over years 58–82. That is, $\hat{B} = P_{116-140}(\mathbf{x})\Delta\bar{T}(58-82)$. Results shown are for the difference between the reconstruction and the actual model output $\hat{B} - B(\mathbf{x}, 58-82)$.~~

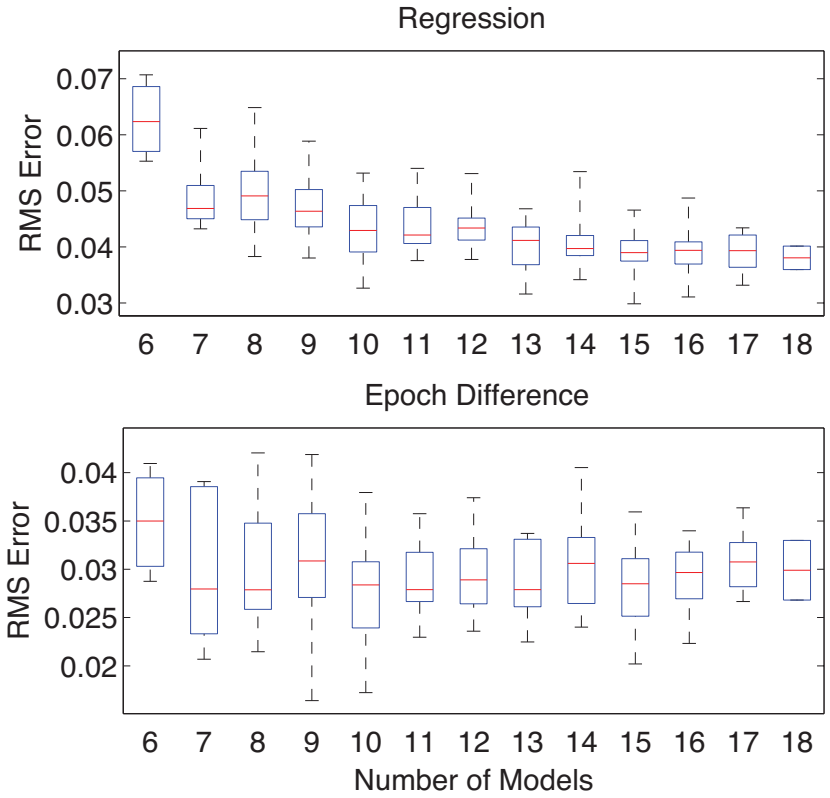

Figure 4. Root mean square (RMS) error (Equation 3, calculated on the difference expressed in mm day^{-1} between the reconstruction and actual model output $\hat{B} - B$) as a function of the number of models used to conduct the scaled precipitation reconstruction. ~~Models were chosen randomly from a~~ From the full set of 26 models (Table 1), anywhere between 6 and 18 models (x-axis) were chosen randomly 20 different times. ~~All values are calculated over an average of years 116–140.~~ Each box in the plots represents those 20 sets of models~~: red~~. Red lines indicate median values, blue boxes indicate the 25th and 75th percentiles, and whiskers indicate the full range of model response among the 20 sets of models. All values are calculated over an average of years 116–140.

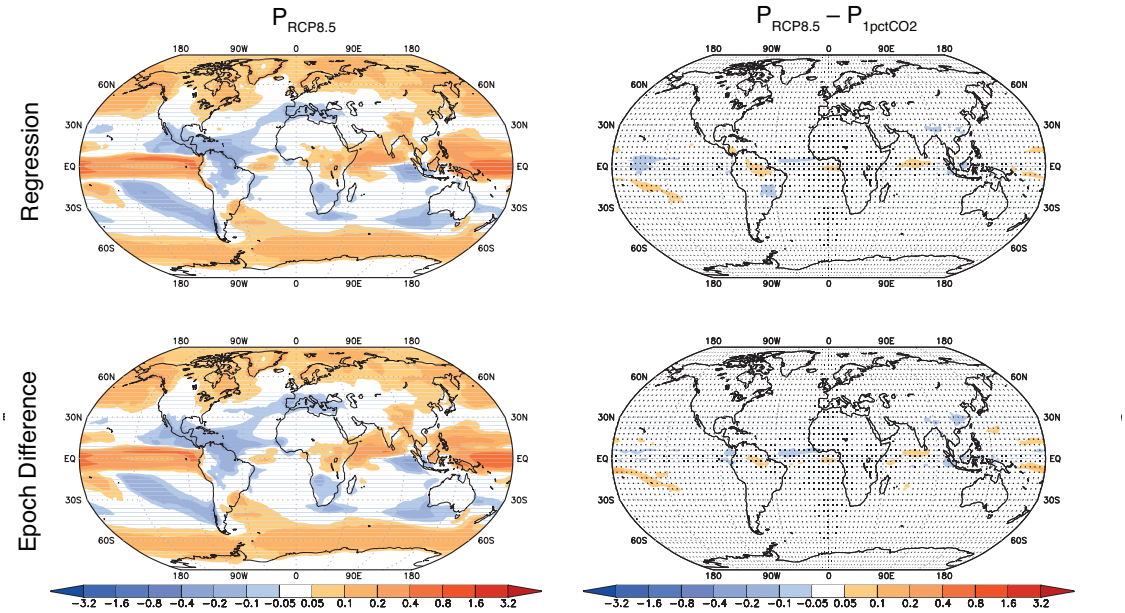

Figure 5. ~~Differences~~ Absolute values (left) of and differences (right) in the precipitation scaling pattern $P(\mathbf{x})$ (Equation 1) when different scenarios are used to construct the pattern (RCP8.5 vs 1pctCO2). Left column shows values of $P_{\mathrm{RCP8.5}}$, and right column shows values of $P_{\mathrm{RCP8.5}} - P_{\mathrm{1pctCO2}}$ (mm day^{-1} K^{-1}). Top row shows results for the ~~physically-based method, middle row shows the~~ regression method, and bottom row shows the epoch difference method. All values are calculated for a Group 1 multi-model average for the 1pctCO2 simulation. Stippling indicates a lack of statistical significance in the pattern of differences (Section 2.2).

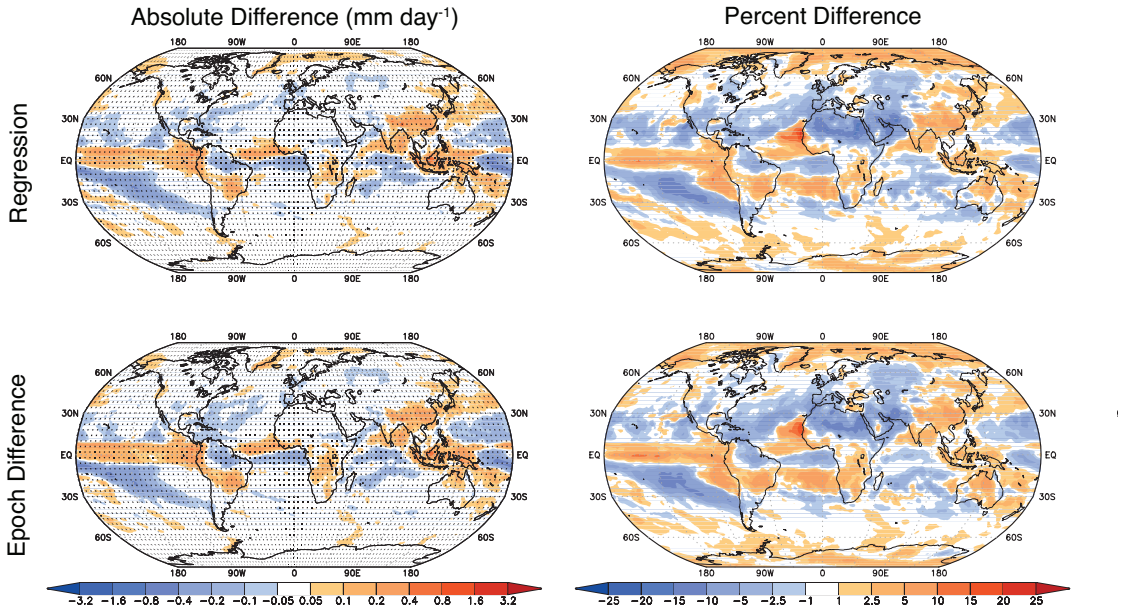

Figure 6. As in Figure 3 but where the reconstruction $\hat{B}$ is built on the pattern P for the RCP8.5 simulation (Group 1 average over years 227–251), and global mean temperature $\Delta\bar{T}$ is averaged over years 227–251 (2076–2100) of the RCP8.5 simulation. That is, $\hat{B} = P_{RCP8.5}(\mathbf{x})\Delta\bar{T}_{\mathrm{RCP8.5}}(227-251)$. Results shown are for the difference between the reconstruction and the actual model output of the RCP8.5 simulation $\hat{B} - B_{\mathrm{RCP8.5}}(\mathbf{x}, 227-251)$.

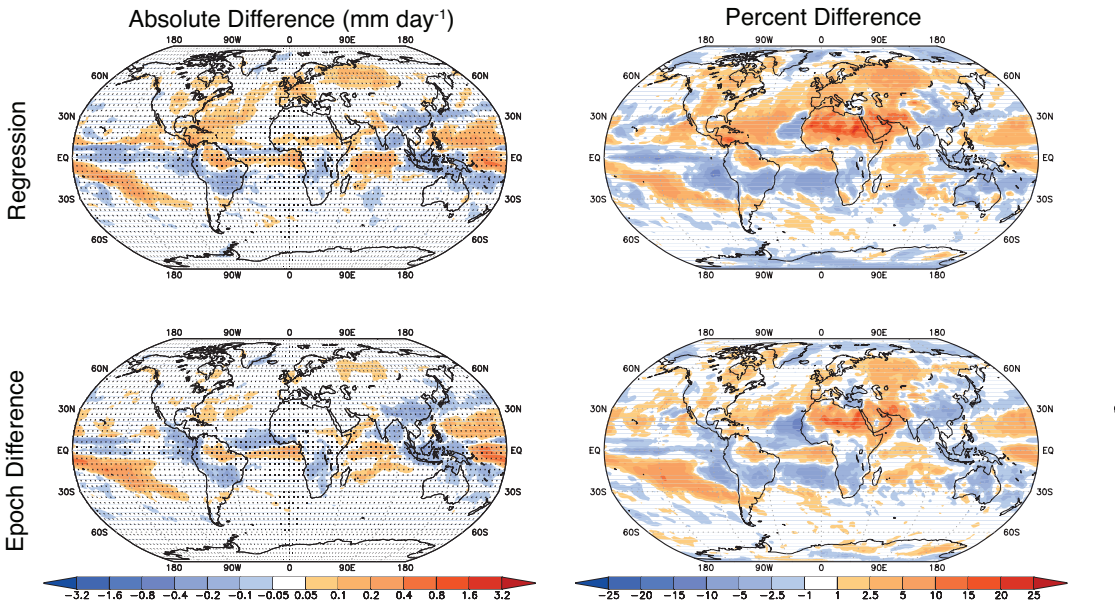

Figure 7. As in Figure 3 but where the reconstruction $\hat{B}$ is built on the pattern P for the RCP8.5 simulation (Group 1 average over years 227–251), and global mean temperature $\Delta\bar{T}$ is averaged over years 116–140 (1965–1990) of the historical/RCP8.5 simulation. That is, $\hat{B} = P_{RCP8.5}(\mathbf{x})\Delta\bar{T}_{\mathrm{RCP8.5}}(116-140)$. Results shown are for the difference between the reconstruction and the actual model output of the historical simulation $\hat{B} - B_{\mathrm{RCP8.5}}(\mathbf{x}, 116-140)$.

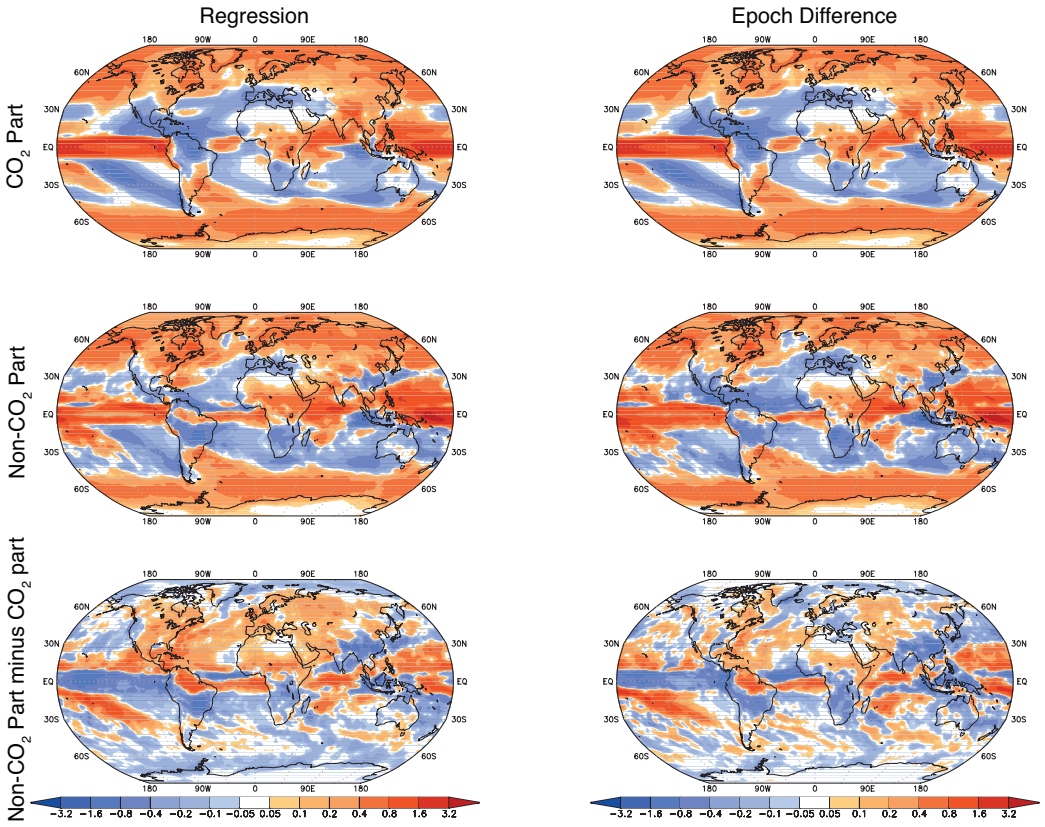

Figure 8. The CO_2 (~~left~~top) and non-CO_2 (~~right~~middle) responses over ~~the period~~ years 227–251 (2076–2100) of the RCP8.5 simulation, as well as the difference between the two (bottom). CO_2 response is calculated as $\Delta\hat{B} = P_{1pctCO2}\bar{T}_{RCP8.5}(227-251)$, and non-$CO_2$ response is calculated as ~~$\Delta\hat{B} = (P_{RCP8.5} - P_{1pctCO2})\bar{T}_{RCP8.5}(227-251)$~~ $\Delta\hat{B} = P_{\text{non-CO}_2}\bar{T}_{RCP8.5}(227-251)$ (see Equation 1 and the discussion surrounding Equation 6 for further details). ~~Top row~~ Left column shows results for the ~~physically-based method, middle row shows the~~ regression method, and ~~bottom row~~ right column shows the epoch difference method.

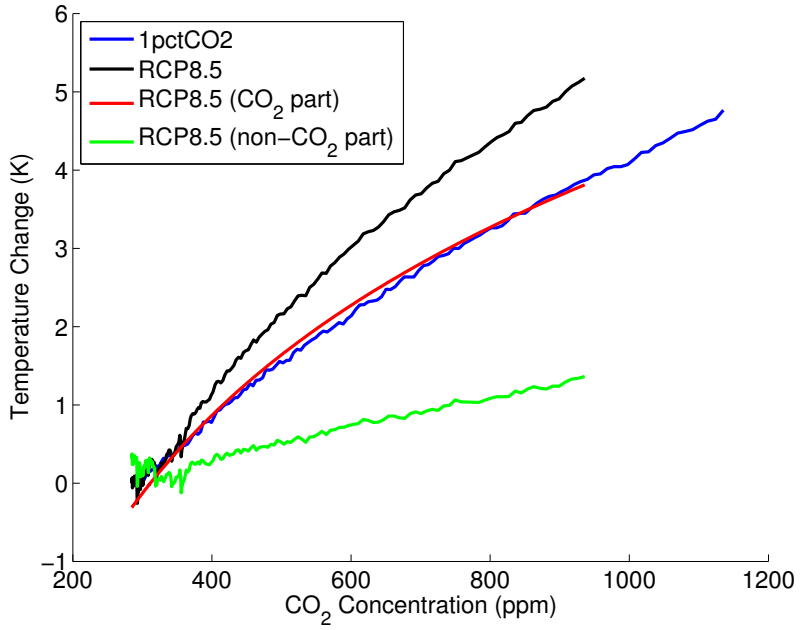

Figure 9. Decomposition of global mean temperature change (as a function of the CO_2 concentration) into its components, as described in Section 4.2.

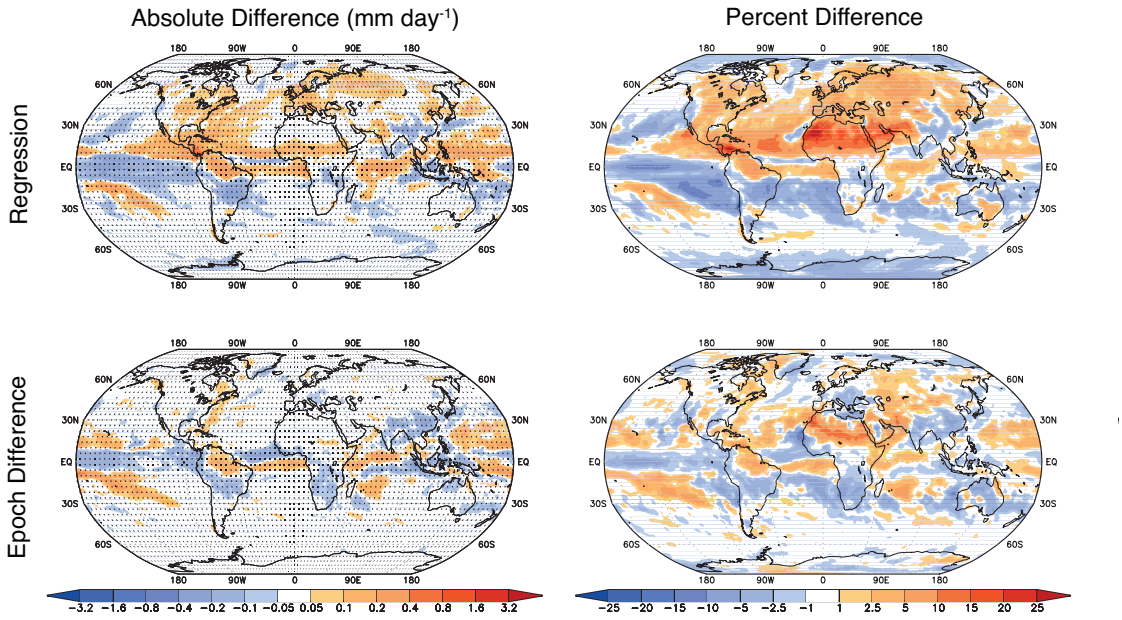

Figure 10. As in Figure 3 but where ~~$\hat{B} = (P_{\mathrm{RCP8.5}} - P_{\mathrm{1pctCO2}})\bar{T}_{RCP8.5,nonCO_2}(227-251) + P_{\mathrm{1pctCO2}}\bar{T}_{RCP8.5,CO_2}(227-251)$~~ $\hat{B} = P_{\mathrm{non-CO_2}}\bar{T}_{RC}$ and results are shown for $\hat{B} - B_{\mathrm{RCP8.5}}(227-251)$. (See Equation 1 and the discussion surrounding Equation 6 for details.)

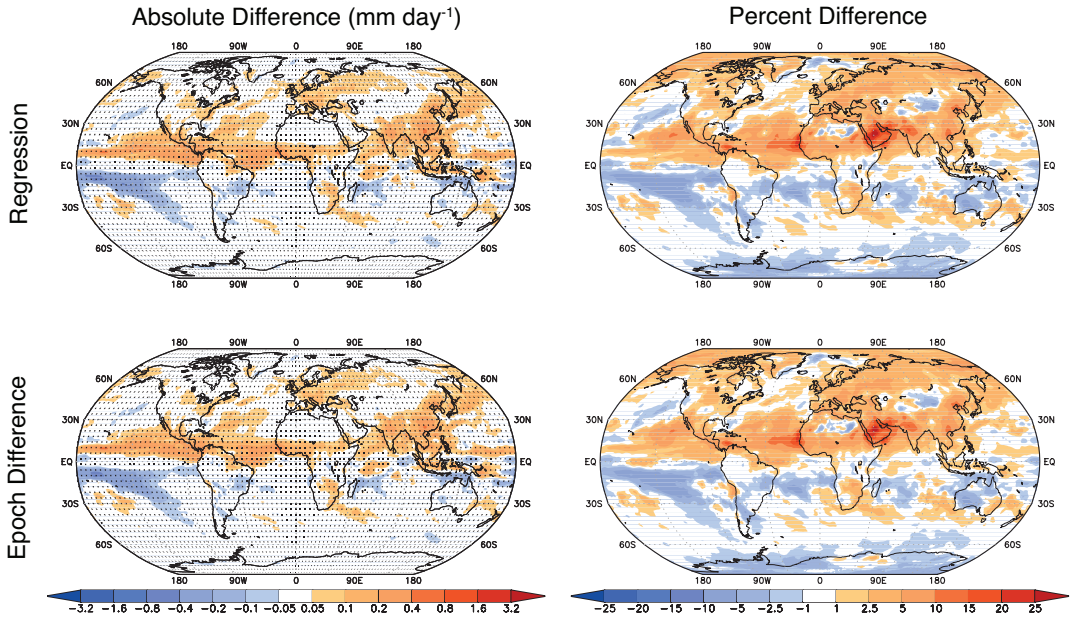

Figure 11. As in Figure 3 but where ~~$\hat{B} = (P_{\mathrm{RCP8.5}} - P_{\mathrm{1pctCO2}})\bar{T}_{RCP8.5,nonCO_2}(116-140) + P_{\mathrm{1pctCO2}}\bar{T}_{RCP8.5,CO_2}(116-140)$~~ $\hat{B} = P_{\mathrm{non-CO_2}}\bar{T}_{RC}$ and results are shown for $\hat{B} - B_{\mathrm{RCP8.5}}(116-140)$. (See Equation 1 and the discussion surrounding Equation 6 for details.)

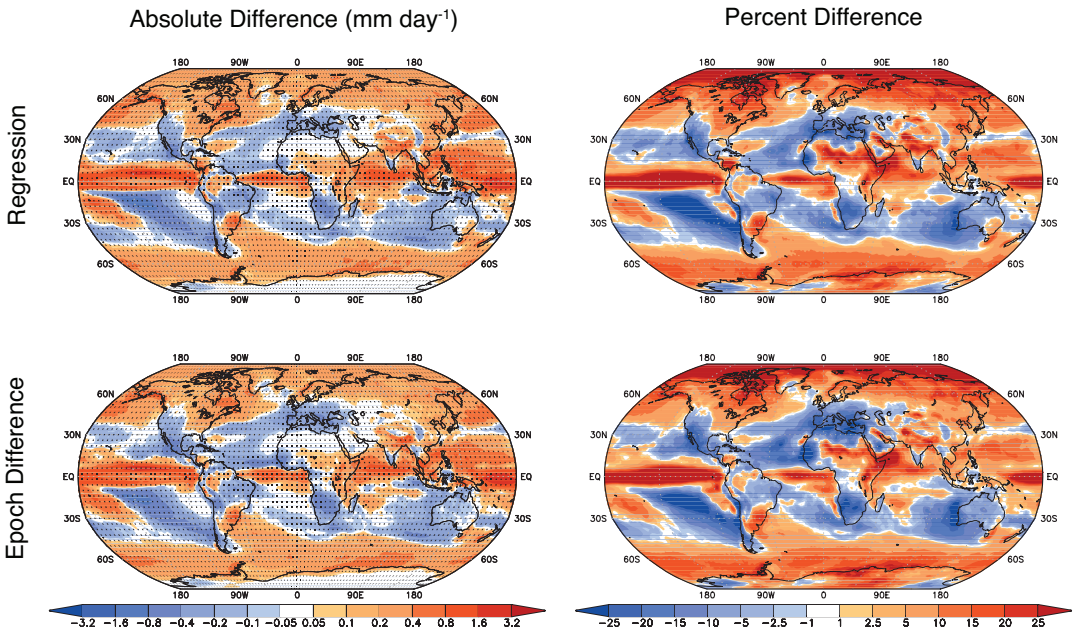

Figure 12. As in Figure 3 but where $\hat{B} = (P_{\mathrm{RCP8.5}} - P_{\mathrm{1pctCO2}})\bar{T}_{RCP2.6,nonCO_2}(227-251) + P_{\mathrm{1pctCO2}}\bar{T}_{RCP2.6,CO_2}(227-251)$ $\hat{B} = P_{\mathrm{non-CO_2}}\bar{T}_{RC}$ and results are shown for $\hat{B} - B_{\mathrm{RCP2.6}}(227-251)$. (See Equation 1 and the discussion surrounding Equation 6 for details.)

Comparison of Pattern Scaling Between Two Groups of Models

Figure **??** further supports the findings in Section 3.3 by showing that the patterns $P(\mathbf{x})$ are not statistically different for Groups 1 and 2 except for isolated areas. The results for the physically-based method indicate that the findings of Lau et al. (2013) are generally reproduced here, in that the pattern is largely robust across different groups of models.

Figures **??** and **??** show differences in the reconstructions, averaged over years 116–140. More specifically, Figure **??** shows differences $P_{\mathrm{Group\ 2}}\Delta\bar{T}_{\mathrm{Group\ 1}} - \Delta B_{\mathrm{Group\ 1}}$ and Figure **??** shows differences $P_{\mathrm{Group\ 1}}\Delta\bar{T}_{\mathrm{Group\ 2}} - \Delta B_{\mathrm{Group\ 2}}$.

The results in Figures **??** and **??** for the physically-based method are both qualitatively and quantitatively similar to those in Figure 3, and global RMS values are similar. Conversely, results for the regression and epoch difference methods, while similar to each other, have qualitatively more error than the results in Figure 3. Global RMS values are 2–3 times higher in Figure **??** than in Figure 3, but errors in Figure **??** are comparable to Figure 3. This might be expected, as on average, $\Delta\bar{T}_{\mathrm{Group\ 1}} \approx \Delta\bar{T}_{\mathrm{Group\ 2}}$, so differences in Figure **??** would be small, whereas differences in Figure **??** are driven by differences in the patterns $P_{\mathrm{Group\ 1}}$ and $P_{\mathrm{Group\ 2}}$ (Figure **??**). As discussed in Section 3.3, the physically-based method shows some statistically significant regions of error in both Figures **??** and **??**, whereas practically no region is statistically significant for the regression and epoch difference methods.

Differences in time-invariant patterns $P(\mathbf{x})$ among the two groups of models (Table 1), calculated for the 1pctCO2 simulation. Left column shows the multi-model average for Group 2, and right column shows the differences in multi-model averages among the two groups. All values shown have units mm day^{-1} K^{-1}. Stippling indicates a lack of statistical significance in the pattern of differences (Section 2.2).

As in Figure 3 but where the reconstruction $\hat{B}$ is built on the pattern P for Group 2 (average of years 116–140 of the 1pctCO2 simulation), and global mean temperature $\Delta\bar{T}$ is averaged over years 116–140 of Group 1. That is, $\hat{B} = P_{\mathrm{Group2}}(\mathbf{x})\Delta\bar{T}_{\mathrm{Group1}}(116-140)$. Results shown are for the difference between the reconstruction and the actual model output $\hat{B} - B_{\mathrm{Group1}}(\mathbf{x}, 116-140)$.

As in Figure 3 but where the reconstruction $\hat{B}$ is built on the pattern P for Group 1 (average of years 116–140 of the 1pctCO2 simulation), and global mean temperature $\Delta\bar{T}$ is averaged over years 116–140 of Group 2. That is, $\hat{B} = P_{\mathrm{Group1}}(\mathbf{x})\Delta\bar{T}_{\mathrm{Group2}}(116-140)$. Results shown are for the difference between the reconstruction and the actual model output $\hat{B} - B_{\mathrm{Group2}}(\mathbf{x}, 116-140)$.