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 two methods of pattern scaling for annual mean precipitation (regression and epoch difference) and evaluate which method is "better" in particular circumstances by quantifying their robustness to interpolation/extrapolation in time, inter-model variations, and inter-scenario variations. Both the regression and epoch difference methods (the two most commonly used methods of pattern scaling) have good absolute performance in reconstructing the climate model output, measured as an area-weighted root mean square error. We decompose the precipitation response in the RCP8.5 scenario into a CO_2 portion and a non-CO_2 portion. Extrapolating RCP8.5 patterns to reconstruct precipitation change in the RCP2.6 scenario results in 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 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.

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. However, their complexity means that these models are often computationally expensive, 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 AOGCM uncertainties and how they affect projections of climate change (Qian et al., 2016).

Emulators of 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 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. 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, 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.

Through these investigations, we hope to better reveal in what circumstances methods of pattern scaling of precipitation perform 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 AOGCMs.

2 Pattern Scaling Methods

2.1 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.

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

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

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

All of the analysis conducted here uses simulations from 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
5   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
10   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
15   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
20 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
25 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
30 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:

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

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 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, 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 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 two methods of reconstruction. Both methods show qualitatively similar features. 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.

Figure 3 shows a more quantitative comparison between the different reconstruction methods and the actual model output. Overall error (RMS; equation 3) in the regression and epoch different methods are very small (0.04 and 0.03 mm day^{-1}, respectively; see Table 3), and no region in the reconstruction is statistically different from the actual model output.

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 were time-invariant, then the results presented in this section would be identical to those previously discussed.

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 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 patterns calculated by using the regression and epoch difference methods only show small changes between the two periods, virtually
5   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 both methods (Supplemental Figure 2) than the results shown in Figure 3. RMS error increases by an order of magnitude (not shown), although few areas show statistically significant differences from the actual model output over this time period. This is likely due to the noise introduced by building $P(x)$ on
10   the early years of the simulation when the climate change signal is weak.

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 increases by a
15   factor of two for both methods, which potentially indicates the presence of nonlinearity. As before, no difference is statistically significant.

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
20   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
25   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: both methods have similar magnitudes of error (except for small numbers of models). The RMS error values (Table 3) for Group 1 (13 models) are consistent with the RMS error ranges depicted in Figure 4, indicating that Group 1 is not an outlier.

The regression method shows a dependence of RMS error on the number of models, whereas with the exception of low
30   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. 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

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). However, when examining interpolation in time, error increases for both methods, indicating issues with robustness to timescale (Supplemental Section 1). Also, the pattern shows increased error in many places when 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.

4 Pattern Scaling for 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 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 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. 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

5 Figure 5 shows the RCP8.5 scaling pattern $P_{\text{RCP8.5}}(\mathbf{x})$ and the difference from the CO_2-only pattern. Patterns are nearly identical to those in Figure 1. 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

10 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 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,

15 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. Although the two portions of the reconstruction generally show similar features, the regional effects have quite different magnitudes in many

20 regions. In particular, the non-CO_2 response is weaker over the tropical Pacific than the CO_2 response 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 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

25 greenhouse gases (most prominently methane), it appears to have captured an aerosol signature. It would be a useful future area of investigation to 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

30 are driven by different forcing. Even though Figure 5 shows that the patterns $P_{\text{RCP8.5}}$ 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

$$\Delta \hat{B}_{\text{RCP8.5}} = \Delta \bar{T}_{\text{CO}_2} P_{\text{CO}_2} + \Delta \bar{T}_{\text{non}-\text{CO}_2} P_{\text{non}-\text{CO}_2} \tag{6}$$

We set P_{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. $P_{\text{non}-\text{CO}_2}$ is defined to be $4P_{\text{RCP8.5}} - 3P_{\text{CO}_2}$ (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 $\Delta \bar{T}_{\text{CO}_2}$, 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 $\log_2([\text{CO}_2])$ 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 ppm$_\text{v}$.) Then $\bar{T}_{\text{CO}_2} = \alpha \log_2([\text{CO}_2]) + \beta$. $\bar{T}_{\text{non}-\text{CO}_2}$ is calculated as the residual $\bar{T}_{\text{RCP8.5}} - \bar{T}_{\text{CO}_2}$. 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. Both CO$_2$ 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.

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 period 2076–2100 at the end of the RCP8.5 simulation (Figure 10), 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), 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 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. Both the epoch difference and regression methods show strong differences that are consistent with the patterns displayed in Figure 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 pattern build on RCP8.5 was not effective for explaining non-CO_2 behavior in RCP2.6, indicating that there are limits to the applicability of a "universal" non-CO_2 forcing. A future area of investigation could explore these limits: for example, would the non-CO_2 pattern built on RCP8.5 work on RCP6.0 or RCP4.5 instead of the extreme RCP2.6 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).

In particular, we found limited ability in reconstructing the RCP2.6 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 RCP6.0 or RCP4.5.

5 Conclusions

We have explored 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. Both the regression and epoch difference methods perform well and approximately similarly.

Most of the errors that arise for 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 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 AOGCM may be required. However, given the difficulties 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 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 are available on GitHub at http://github.com/JGCRI/CMIP5_patterns. All code used to generate figures in this manuscript 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

5   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.

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.

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 were conducted using the models in both Group 1 and Group 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	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%)

Table 3. RMS error values calculated over the entire globe (Equation 3) for each of the figures. All units are in mm day^{-1} for differences (Figures 3, 6, 7, 10, 11, and 12) or mm day^{-1} K^{-1} for patterns (Figure 5).

	Regression	Epoch difference
Figure 3	0.04	0.03
Figure 5	0.02	0.02
Figure 6	0.10	0.10
Figure 7	0.09	0.09
Figure 10	0.10	0.07
Figure 11	0.10	0.09
Figure 12	0.21	0.20
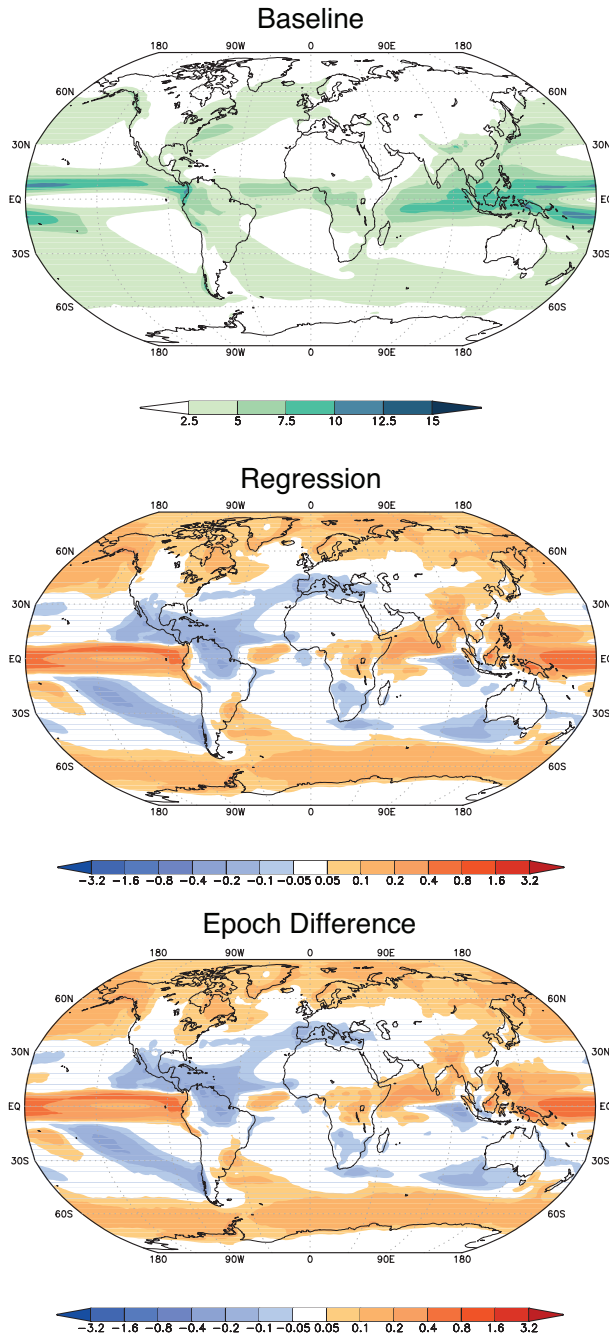
Figure 1. The components needed for pattern scaling of the precipitation response to CO_2 forcing, averaged over the 13 models in Group 1 (Table 1). Top shows the baseline precipitation pattern for the multi-model average: $B(x,0)$ in Equation 1 (mm day^{-1}; averaged over years 1–25 of the 1pctCO2 simulation). The other panels show the time-invariant pattern $P(\mathbf{x})$ in Equation 1 (mm day^{-1} K^{-1}) for the regression method (middle), and the epoch difference method (bottom).
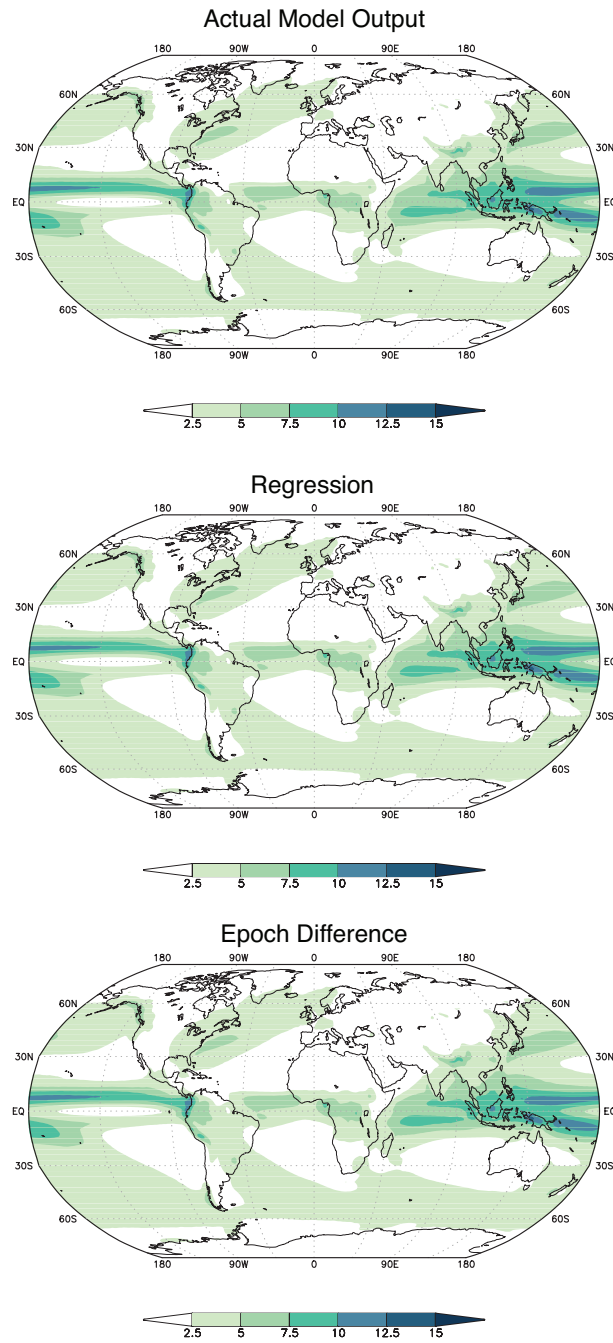
Figure 2. Comparison between the actual Group 1 multi-model average precipitation output (top) and the reconstructions produced by pattern scaling ($\hat{B}$ in Equation 1). All values are in mm day^{-1} and represent averages over years 116–140 of the 1pctCO2 simulation. Middle panel shows the regression method, and bottom panel shows the epoch difference method.
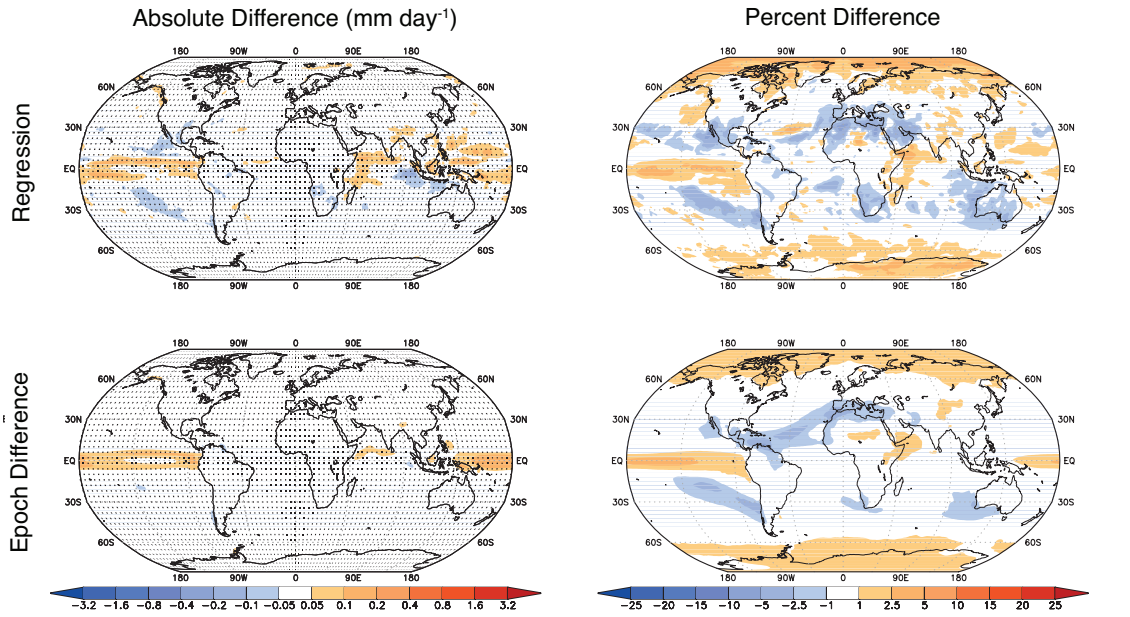
Figure 3. Differences between the reconstructions produced by pattern scaling ($\hat{B}$) and the actual model output for precipitation (B). Left column shows absolute values of $\hat{B} - B$ (mm day^{-1}), and right column shows percent change. Top row shows results for the regression method, and bottom 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).
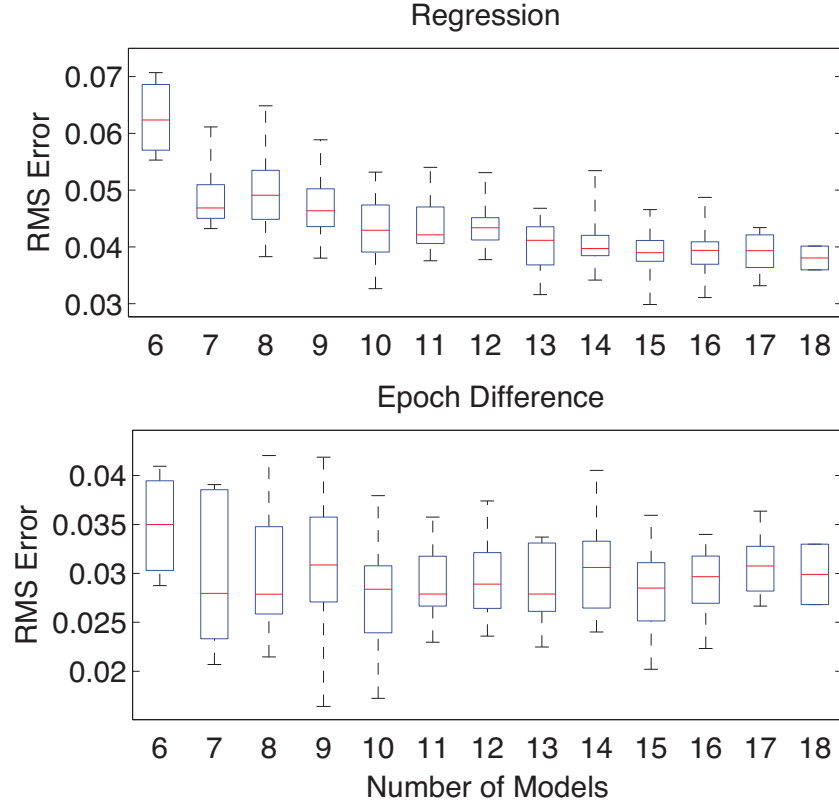
Figure 4. Root mean square (RMS) error (Equation 3, calculated on the difference expressed in mm day^{-1} between the reconstruction and actual model output $\hat{B} - B$) as a function of the number of models used to conduct the scaled precipitation reconstruction. From the full set of 26 models (Table 1), anywhere between 6 and 18 models (x-axis) were chosen randomly 20 different times. Each box in the plots represents those 20 sets of models. 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.
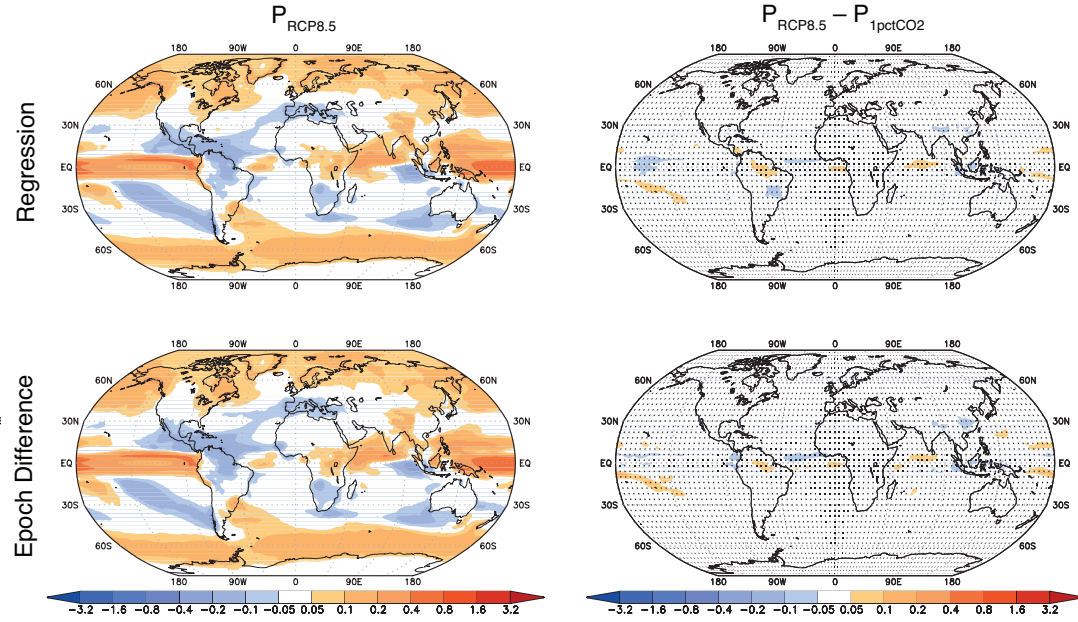
Figure 5. Absolute values (left) of and differences (right) in the precipitation scaling pattern $P(\mathbf{x})$ (Equation 1) when different scenarios are used to construct the pattern (RCP8.5 vs 1pctCO2). Left column shows values of $P_{\text{RCP8.5}}$, and right column shows values of $P_{\text{RCP8.5}} - P_{\text{1pctCO2}}$ (mm day^{-1} K^{-1}). Top row shows results for 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).
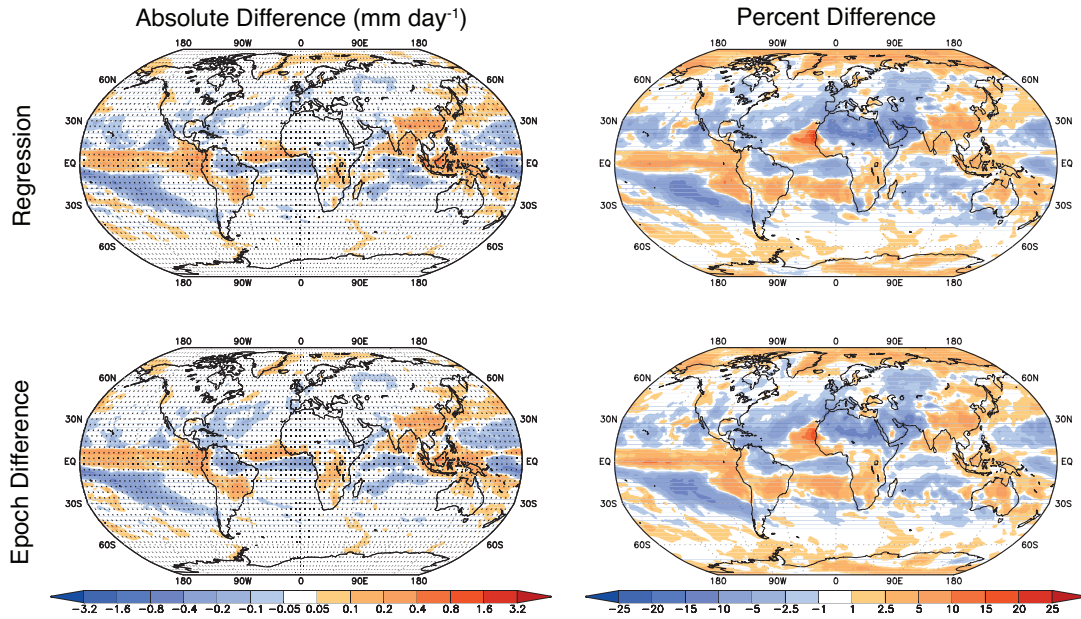
Figure 6. As in Figure 3 but where the reconstruction $\hat{B}$ is built on the pattern P for the RCP8.5 simulation (Group 1 average over years 227–251), and global mean temperature $\Delta\bar{T}$ is averaged over years 227–251 (2076–2100) of the RCP8.5 simulation. That is, $\hat{B} = P_{RCP8.5}(\mathbf{x})\Delta\bar{T}_{\mathrm{RCP8.5}}(227-251)$. Results shown are for the difference between the reconstruction and the actual model output of the RCP8.5 simulation $\hat{B} - B_{\mathrm{RCP8.5}}(\mathbf{x}, 227-251)$.
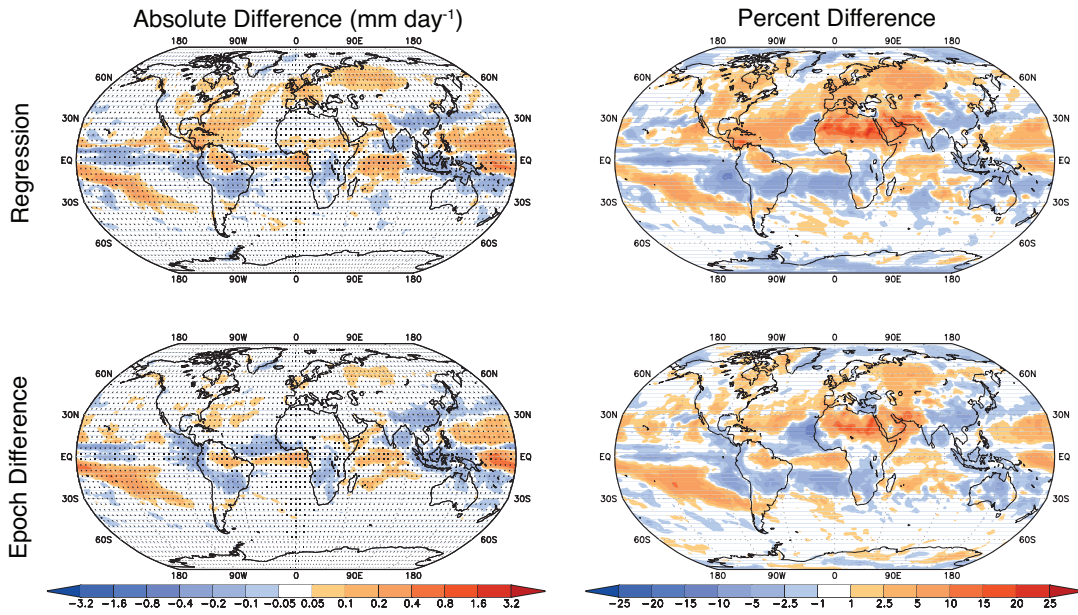
Figure 7. As in Figure 3 but where the reconstruction $\hat{B}$ is built on the pattern P for the RCP8.5 simulation (Group 1 average over years 227–251), and global mean temperature $\Delta\bar{T}$ is averaged over years 116–140 (1965–1990) of the historical/RCP8.5 simulation. That is, $\hat{B} = P_{RCP8.5}(\mathbf{x})\Delta\bar{T}_{\mathrm{RCP8.5}}(116-140)$. Results shown are for the difference between the reconstruction and the actual model output of the historical simulation $\hat{B} - B_{\mathrm{RCP8.5}}(\mathbf{x}, 116-140)$.
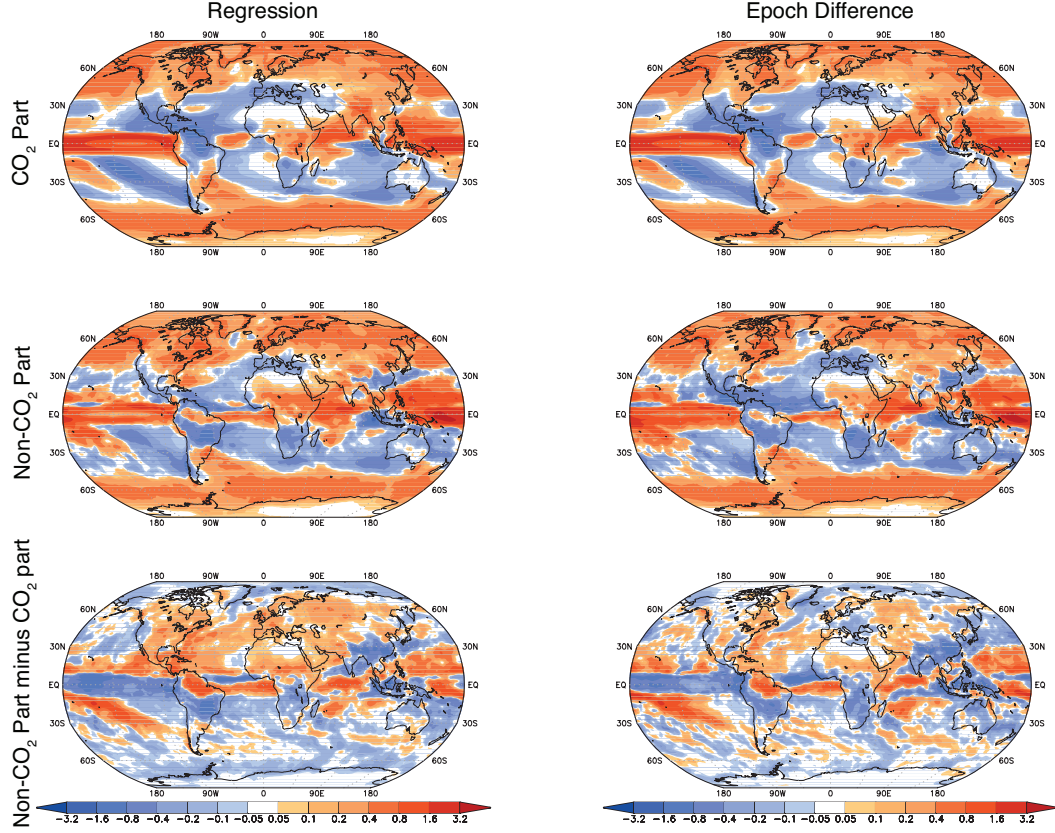
Figure 8. The CO_2 (top) and non-CO_2 (middle) responses over years 227–251 (2076–2100) of the RCP8.5 simulation, as well as the difference between the two (bottom). CO_2 response is calculated as $\Delta\hat{B} = P_{1pctCO2}\bar{T}_{RCP8.5}(227-251)$, and non-$CO_2$ response is calculated as $\Delta\hat{B} = P_{\text{non}-CO_2}\bar{T}_{RCP8.5}(227-251)$ (see Equation 1 and the discussion surrounding Equation 6 for further details). Left column shows results for the regression method, and right column shows the epoch difference method.
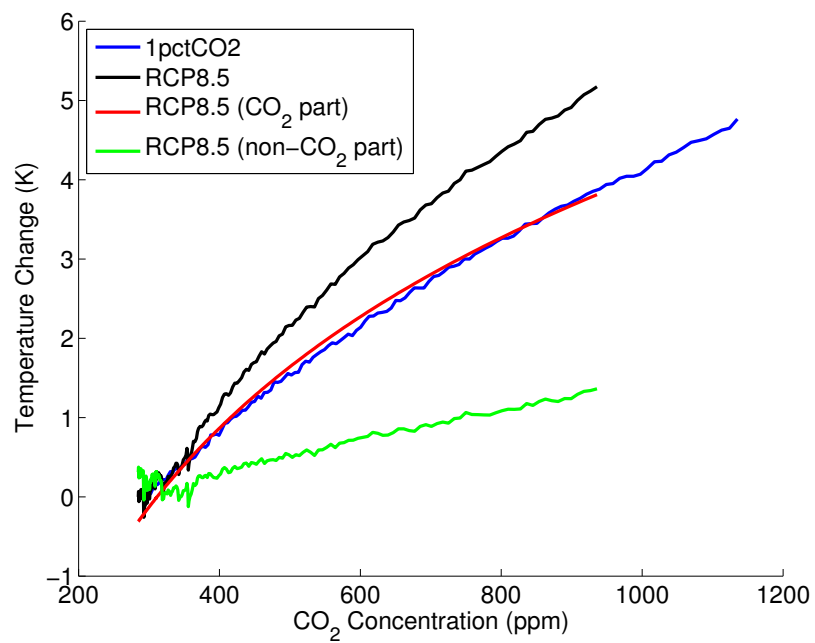
Figure 9. Decomposition of global mean temperature change (as a function of the CO₂ concentration) into its components, as described in Section 4.2.
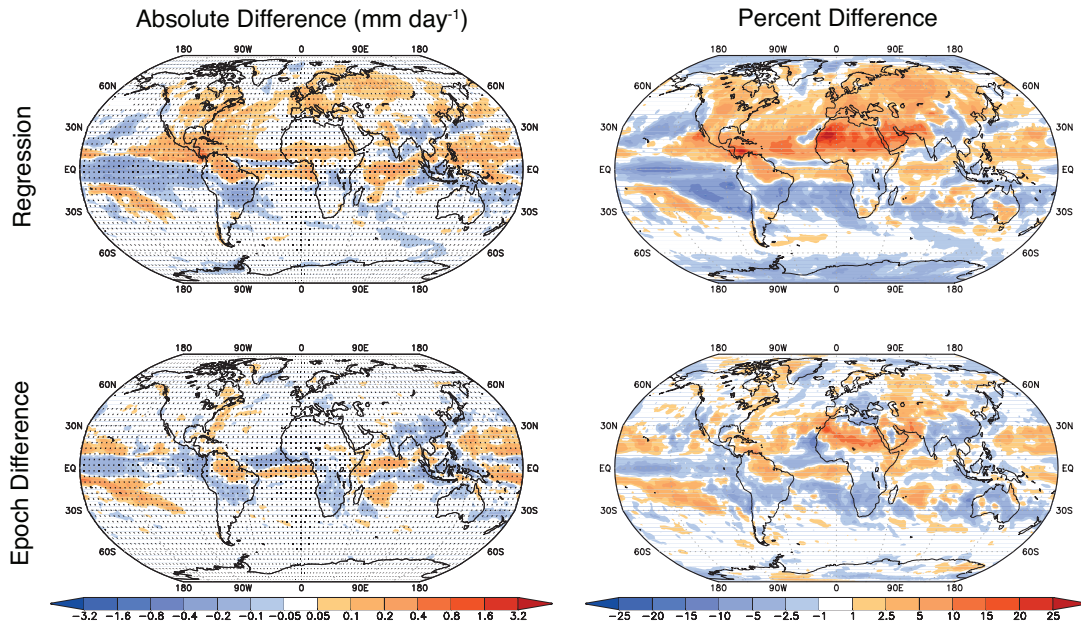
Figure 10. As in Figure 3 but where $\hat{B} = P_{\text{non}-\text{CO}_2}\bar{T}_{RCP8.5,nonCO_2}(227-251) + P_{\text{1pctCO2}}\bar{T}_{RCP8.5,CO_2}(227-251)$, and results are shown for $\hat{B} - B_{\text{RCP8.5}}(227-251)$. (See Equation 1 and the discussion surrounding Equation 6 for details.)
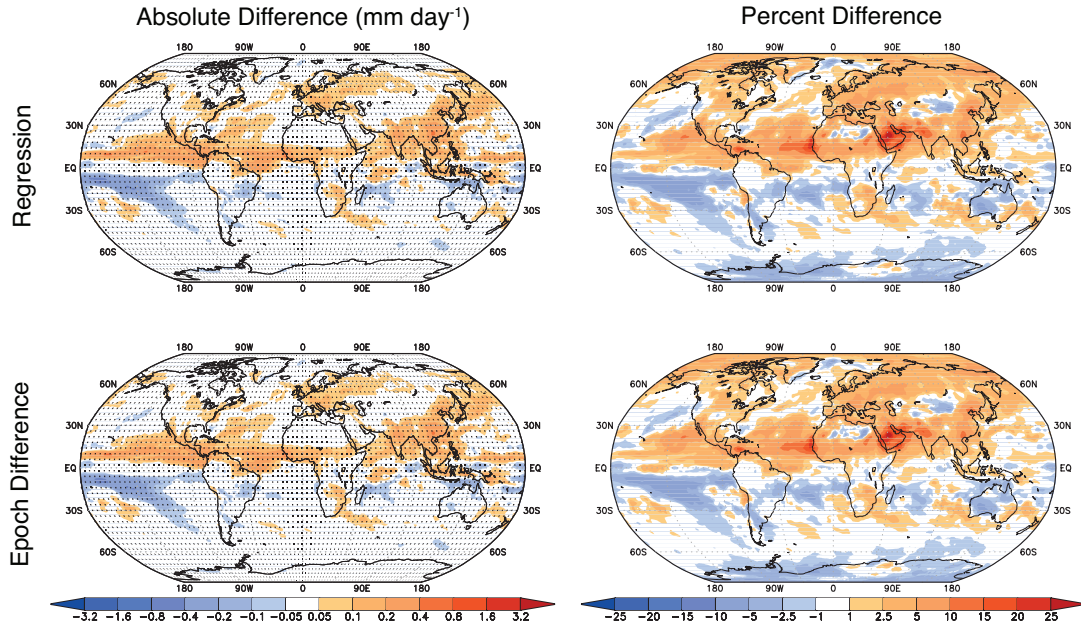
Figure 11. As in Figure 3 but where $\hat{B} = P_{\mathrm{non-CO_2}}\bar{T}_{RCP8.5,nonCO_2}(116-140) + P_{\mathrm{1pctCO2}}\bar{T}_{RCP8.5,CO_2}(116-140)$, and results are shown for $\hat{B} - B_{\mathrm{RCP8.5}}(116-140)$. (See Equation 1 and the discussion surrounding Equation 6 for details.)
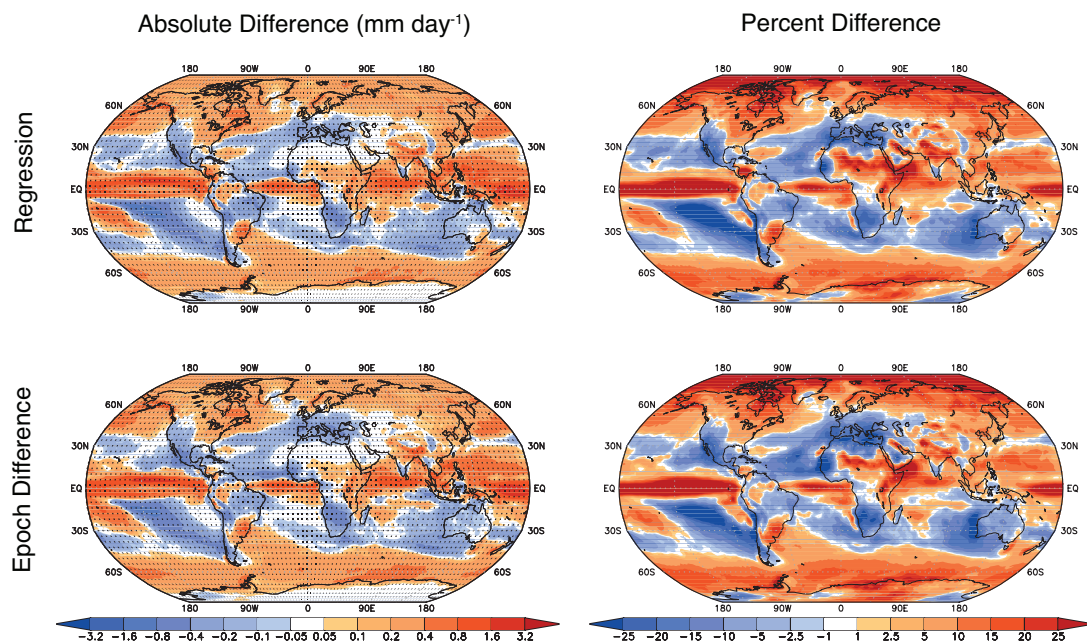
Figure 12. As in Figure 3 but where $\hat{B} = P_{\mathrm{non-CO_2}}\bar{T}_{RCP2.6,nonCO_2}(227-251) + P_{\mathrm{1pctCO2}}\bar{T}_{RCP2.6,CO_2}(227-251)$, and results are shown for $\hat{B} - B_{\mathrm{RCP2.6}}(227-251)$. (See Equation 1 and the discussion surrounding Equation 6 for details.)