



Exploring global surface temperature pattern scaling methodologies and assumptions from a CMIP5 model ensemble

Cary Lynch¹, Corinne Hartin¹, Ben Bond-Lamberty¹, and Ben Kravitz²

¹Pacific Northwest National Laboratory, Joint Global Change Research Institute, 5825 University Research Court, Suite 3500, College Park, MD 20740
 ²Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352

Correspondence to: Cary Lynch (cary.lynch@pnnl.gov)

Abstract. Pattern scaling is used to explore the uncertainty in future forcing scenarios. Of the possible techniques used for pattern scaling, the two most prominent are the delta and least squared regression methods. Both methods assume that local climate changes scale with globally averaged temperature increase, allowing for spatial patterns to be generated for multiple models for any future emission scenario. We explore this assumption by using different time periods and scenarios, and exam-

- 5 ine the differences and the statistical significance between patterns generated by each method. Regardless of epoch chosen, the relationship between globally averaged temperature increase and local temperature are similar. Temperature patterns generated by the linear regression method show a better fit to global mean temperature change than the delta method. Differences in patterns between methods and epochs are largest in high latitudes (60-90 degrees N/S). Error terms in the least squared regression method are higher in lower forcing scenarios, and global mean temperature sensitivity is higher. These patterns will be used to
- 10 examine feedbacks and uncertainty in the climate system.

1 Introduction

The Representative Concentration Pathways (RCPs) are a collection of scenarios of future climate change that span a range of potential changes in aerosols, greenhouse gases, and natural forcings consistent with various assumptions about societal choices, technological development, and socio-economic assumptions in the 21st century (*Moss et al.*, 2010; *Meinshausen*

15 *et al.*, 2011; *van Vuuren et al.*, 2011). It is far too computationally costly to explore a large number of emission scenarios in fully coupled general circulation models (GCMs). However, uncertainties in climate effects stemming from future forcing projections are difficult to explore with a limited number of future forcing scenarios. Uncertainty in future climate is particularly problematic for stakeholders who attempt to adapt to or mitigate future impacts.

In the absence of a large sample of model experiments to draw upon, scaled scenarios are used for reducing uncertainty

20 (*Dessai et al.*, 2005). Creating spatial patterns of change scaled from GCMs is called 'pattern scaling'. Pattern scaling was initially established to enable the creation of transient climate projections from the steady state response of a GCM to a doubling of the preindustrial CO_2 concentration (*Santer et al.*, 1990). It can is used to generate climate change scenarios under changes in anthropogenic forcings that have not been simulated by full GCMs, but instead simulated by using simple climate





models (SCMs) that emulate the more complex behavior of GCMs (Moss et al., 2010). In essence, pattern scaling combines the computational efficiency of a SCM with a GCMs ability to represent spatial patterns of change. In order to include uncertainty estimates or derive optimal policies, through use of an Integrated Assessment Model (IAM), highly efficient calculations of global change are generally required; such simulations often use pattern scaling to derive spatial distributions of change (Frieler et al., 2012; Collins et al., 2015).

5

The main assumption in pattern scaling is that the time-dependent, spatially varying response of a climate variable (e.g., local temperature) to change (e.g., in the CO_2 concentration) is separable into (1) a spatial pattern that is invariant and scaled with global mean temperature and (2) a timeseries of global mean temperature change. The spatial pattern of temperature change per degree of global average temperature increase has been shown to remain relatively constant with increasing global mean

temperature increases (Mitchell, 2003; Tebaldi and Arblaster, 2014; Leduc et al., 2016). For temperature, the generated pattern 10 explains a large portion of the variability of the externally forced change over time and across scenarios within a given model. Pattern scaling methodologies have evolved into two general types. The first and most common pattern scaling technique is the time-slice method or the "delta" method, which is simply local future change, normalized by global future change both

averaged over some chosen time period, hereafter referred to as an epoch (Tebaldi and Arblaster, 2014; Herger et al., 2015; 15 Osborn et al., 2015). The second is the linear regression method, which uses ordinary least squares regression coefficients to

fit local trends on a model gridbox scale to a global trend (Mitchell, 2003; Ruosteenoja et al., 2007; Lopez et al., 2013).

Assumptions

25

In the delta method, the underlying assumption is that responses to external forcing and internal variability are independent, implying that anthropogenic forcings do not modify the internal variability of the climate system (Mitchell, 2003; Herger et al.,

2015). This premise is known to be false, but in practice, estimation errors introduced through this assumption are small. In 20 this method there is also the assumption that the relationship of local change to global mean change is independent of the trend, which implies that length of epochs used should not alter the resulting pattern. Barnes and Barnes (2015) argue that the ideal epoch length is dependent on minimizing variable variance by selecting a epoch length with a high signal-to-noise ratio, which is largely dependent on length of time series and whether the trend in the time series is linear. They found that for temperature,

one-third the length of the time series is ideal, and for a 100 year time series the standard thirty year epoch length is sufficient. Choice of reference epoch does matter for assessing future change (Lopez et al., 2013; Hawkins and Sutton, 2015), but it is unclear whether choice of epoch will affect the resulting scaled pattern. Throughout the IPCC Fifth Assessment Report (AR5; Stocker et al., 2013) a 30-year reference epoch of 1986-2005 was used for discussion of projected anomalies; previous assessment reports used earlier epochs. In impacts studies, a later reference period is more suitable in that it is more representative

30 of the current climate, and hence what socio-economic systems are already somewhat adapted to. In adaptation/mitigation analyses, a pre-industrial control simulation epoch is often used as the baseline from which change is diagnosed, as this period is likely to provide the largest deviation from projected future climate.

In the linear regression pattern scaling method, the underlying assumption is that local change scales proportionally with global mean temperature (GMT) change, and that the relationship is stationary over time. This assumption is not always true in





the climate system, especially considering different forcing scenarios and spatial heterogeneity of projected change. Transient forcing is likely to scale the local temperature sensitivity to the trend in global mean temperature. *Mitchell* (2003) found that the linear regression method reduces the influence of the non-linearities arising from faster (or slower) warming. For temperature-related variables the assumption of stationarity is valid, but the magnitudes of estimation errors vary between scenarios for

- 5 non-temperature variables (*Frieler et al.*, 2012). *Lopez et al.* (2013) found that when pattern scaling patterns of temperature extremes, the magnitude of the error in the pattern estimates was substantially large. In linear regression, only the error term (ϵ) is assumed to have a normal distribution (based on the central limit theorem), so it is highly likely that climate extremes would yield high error terms. This can be problematic when constructing confidence intervals but is not necessarily a limitation in the method itself, nor in the resulting patterns (*Lustenberger et al.*, 2014).
- 10 The differences between the two methods are clear, but the differences between the patterns generated by each method are not. In terms of computational efficiency, the delta method is the fastest, which is a reason why it is predominantly used (*Herger et al.*, 2015). However, in terms of skill in trend estimation and adaptability of additional predictors, the linear regression method is preferable (*Frieler et al.*, 2012; *Mitchell*, 2003; *Lustenberger et al.*, 2014; *Barnes and Barnes*, 2015).

In both methodologies, it is implied that the patterns generated under different scenarios are not significantly different. *Tebaldi and Arblaster* (2014) found that patterns from different scenarios were highly spatially correlated, and that choice of

15 Tebaldi and Arblaster (2014) found that patterns from different scenarios were highly spatially correlated, and that choice of scenario did not explain a significant proportion of variability in patterns when using the delta pattern scaling method. However, would the patterns generated through the regression method show significant difference between scenarios, and if so, where?

In this manuscript, we use a simplified approach for each method to assess the differences in pattern strength and skill between each method's generated pattern. We begin by examining how the choice of epoch changes the global and local re-

20 lationship especially considering potential long-term trends and variability. We then compare how the delta and regression patterns differ, with respect to spatial heterogenity and pattern strength. And finally, we explore the assumption that the relational pattern is consistent across forcing scenarios where the forcings and potential mitigation would be different.

2 Data/Methods

Climate Models

- 25 We employ three sets of experiments from the Coupled Model Intercomparison Project Phase 5 (CMIP5; *Taylor et al.*, 2011). The 'historical' experiment includes most known climate forcings from the 19th and 20th Centuries, and is used to construct historical climatologies for use in creating a pattern via the delta method. Model historical runs varied in length, so we used 1861 as the start of the historical period, and 2005 as the end. For future projections, we used climate model output for two of the four available RCPs. The bulk of our efforts employ the high-forcing RCP8.5 scenario where radiative forcing increases to
- 30 8.5 W/m² through the 21st Century (*Riahi et al.*, 2011). We also use the mid-forcing RCP4.5 scenario where radiative forcing increases to 4.5 W/m² through the 21st Century (*Thomson et al.*, 2011) achieved by limiting future emissions yielding a weaker response. For the future simulations, the start year was 2006, and the end year was 2100.





5

Twelve climate models are examined for the historical and the RCP4.5/RCP8.5 experiments (Table 1). Accurate simulations of global mean temperature are not necessarily an essential pre-requisite for predicting global temperature trends *Rupp et al.* (2013); *Eyring et al.* (2016). However, culling the list of all available models is sometimes necessary, and developing a performance based rubric 'score card' of models is helpful. Also, because most modeling centers released multiple models based on different physical parameterizations and/or interactions, the assumption of model independence is not valid (*Sanderson et al.*, 2015). For these reasons, we developed a small set of simple performance metrics based on the structure from *Rupp et al.* (2013). We then used the 'best' representative model from each modeling center, to limit the list of models used in this analysis to twelve. However, it should be noted that by calculating the climate change pattern, individual model bias is not likely to skew ensemble mean results, and as such excluding individual models based on observed bias is not necessary (*Mitchell*, 2003;

10 Lopez et al., 2013).

For this study, only seasonal and annual surface air temperature was analyzed, and in some cases, the influence of the landocean contrast was examined. For such cases, a land mask was applied using each model's native resolution, with 100% of the grid cell either land or ocean. Uncertainty (model to model variability) in GCM output results from individual model parameterizations, random climate variability, and input for model initialization (*Tebaldi and Knutti*, 2007). Therefore, this analysis

15 incorporates the multi-model ensemble mean and median as well as the uncertainty represented by the model spread. Because models varied in spatial resolution, when appropriate, the models were first regridded to T85 resolution prior to calculating ensemble mean or median.

2.1 Data Analysis

- Reanalysis output from the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) (*Kalnay et al.*, 1996) is used to validate the model ensemble annual and seasonal climatology because it spans the later half of the 20th Century and has continuous spatial coverage. It incorporates observations and numerical weather prediction model output to create a dataset that includes surface and atmospheric variables. The NCEP/NCAR reanalysis data for temperature captures observed surface temperature trends and variability reasonably well (*Simmons et al.*, 2004). It seems likely
 that biases in the reanalysis could contribute to errors in pattern scaling methodologies, and using different reanalyses could
- shed light on the relative contribution of this source of error in comparison to other potential sources; exploring this would be a large undertaking and is the subject of future work.

Scaled temperature patterns were created for each model of the ensemble and also for mean and median values of the ensemble. It is unclear how the choice of epoch changes the pattern between global and local temperature, considering how

30 different sectors place importance on certain reference periods. To obtain a delta pattern with a high signal-to-noise ratio, ideally one would use an ensemble of models, with a pre-industrial historical epoch forced by the highest emission scenario with the future epoch being as far in the future as possible. Each of these contributions could improve signal clarity and robustness. Here we calculate the delta pattern for each model, and then use the ensemble median and/or mean in our analysis.



5

20

The delta pattern (DP) is described as follows:

$$DP_{MS} = \frac{\Delta T L_{MS}}{\Delta T G_{MS}} \tag{1}$$

For each model (M) and future scenario (S), local, grid-box change (TL) is normalized by global mean change (TG), with respect to a 30 year reference climatology from the CMIP5 historical simulation. The global mean for the historical and future periods were weighted by the cosine of the latitude because the CMIP model data is on a gaussian grid.

The linear regression pattern scaling method is not dependent on historical climatology, and as such uses only the future forcing scenario. We use a least squares approach, which provides the best fit for calculating the regression pattern:

$$TL_{MS} = \alpha_{MS} + \beta_{MS} * TG_{MS} + \epsilon_{MS} \tag{2}$$

As previously, TG is the global mean surface temperature time series (unsmoothed), and TL is the gridded time series. β
10 is a two-dimensional field of regression slopes, and ε is the residual term (error) stemming from linearly fitting the dependent variable to the predictor. α is the *y*-intercept, which we take to be 0 by only computing change, not absolute temperature.

Solving for β (the regression coefficient) using ordinary least squares results in:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (TL_i - \overline{TL}) (TG_i - \overline{TG})}{\sum_{i=1}^{n} (TG_i - \overline{TG})^2}$$
(3)

Multiplying β by the number of time steps in the series results in the 21st Century trend. A time series for ϵ can then be 15 determined.

To examine the significance differences between patterns, we use a 2-tailed Student's *t*-test probability using the incomplete beta function ratio. This was done because the ensemble consists of only twelve models, which is a small portion of all available models, and because we assume the ensemble variance for each pattern is the same. In other words, models with a weak relationship between TG and TL in the delta method are also weak when using the regression method. The resulting probability indicates that for small values (≤ 0.05) there is a significant difference between patterns.

To assess the significance of LSR patterns, we calculated *p*-values of β , indicating whether the regression predictor is statistically significantly associated with changes in the response. High *p*-values indicate that the dependent variable and the independent variable are not significantly related, and that the linear fit is poor.

A Principal Components Analysis (also called an Empirical Orthogonal Function, or EOF, analysis) was also conducted, but only to test the assumption that the global warming signal explains the largest percentage of variance across scenarios,

25 but only to test the assumption that the global warming signal explains the largest percentage of variance across scenarios, which has been shown to be the case in previous CMIP experiments (*Vecchi and Soden*, 2007; *Dai*, 2016). We did not evaluate the EOFs (spatial patterns) themselves for further physical causation, or for construction of patterns as done by *Holden and Edwards* (2010).





3 Results/Discussion

3.1 Epoch Differences

Differences between historical and future epochs are analyzed by first comparing the multi-model ensemble against reanalysis data to show how well the multi-model ensemble data fits the observed period, as well as to show the projected trend for two

- 5 future scenarios (Figure 1). For the annual time series, the multi-model ensemble shows strong agreement (<0.3°C) with the reanalysis data and the standard deviation of the ensemble mean is small (+/-0.2°C). The ensemble significantly underestimates the observed epoch in December through February (DJF) as the reanalysis data is generally outside 1 sigma of the ensemble mean. The ensemble overestimates the June through August (JJA) period, but the reanalysis data is mostly within the one sigma of the ensemble mean. For all future scenarios, the higher forcing scenario shows an increase of about 5°C by the end of the</p>
- 10 21st Century with no sign of leveling off, whereas the lower forcing scenario shows an average of 2°C change and appears to level off by the end of the century (Figure 1).

When using the delta method to construct a pattern, the assumption is that regardless of epoch chosen, the trend is the same. This assumption is not true when comparing a higher forcing scenario with a steeper trend (Figure 1 and 2) to a lower forcing scenario, where the difference in thirty year trends can be as much as 2° C. The largest differences between future and past

- 15 climate are when the earliest epoch is chosen from the historical, and the latest epoch is chosen from the future, regardless of scenario (Figure 3), which is also when differences in trend are the largest. In the higher forcing scenario, there are larger differences between global future and historical epochs, up to 5.5°C change in DJF for the ensemble mean. This is reinforced in Table 2, where the 21st Century ensemble median trend is more similar to the late 21st, late 19th Century difference than other epochs, but the 21st Century trend has a larger spread. DJF has the largest 21st Century trend/difference. If using the late
- 20 21^{st} Century epoch, the difference between historical epochs result in an average of 0.5° C difference. This is also consistent when using the mid 21^{st} Century epoch.

There is a weaker (stronger) temperature sensitivity in the Northern Hemisphere (Southern Hemisphere) when using the late 19th Century as a epoch rather than the late 20th Century epoch (Figure 4), as observed warming over the Northern Hemisphere has been greater than the observed warming over the Southern Hemisphere (*Friedman et al.*, 2013). However the difference

- in patterns are minor, $\leq 0.2^{\circ}$ C between the two historical epochs. The difference in patterns for the two future epochs are larger in the Northern Hemisphere at higher latitudes, but there is little hemispheric symmetry in these differences particularly at mid and high latitudes where pattern differences are not equal in strength or direction. These future epoch differences in pattern strength in the high Northern latitudes exhibit strong seasonality, with weaker temperature sensitivity in the Northern Hemisphere cool season (DJFM) and summer (JJA), and stronger in the late spring (AM) and autumn (SON). The difference in
- 30 variance between the two reference epochs is not significant (probability ≤ 0.05 (S1 and S2)) but appears to show that where the pattern is weaker, the variance is less. This suggests that the strength of the relational pattern is somewhat dependent on variance, with larger differences in variance showing stronger temperature sensitivity.





3.2 Pattern Differences

The key idea in either pattern scaling method is that local temperature change scales with global climatic change. The different methods used to quantify global change are first examined by comparing future minus present change to the 21st Century projected linear trend (Figure 5). In the higher forcing scenario, the 21st Century trend is most similar to the difference between

- 5 the late 21^{st} Century and late 19^{th} Century. Using the later historical epoch results in ~ 1°C difference between projected trend and epoch change (S3). The opposite is true when using a lower forcing scenario. In this case, the later 20^{th} Century epoch difference is most similar to the 21^{st} Century linear trend. When using the lower forcing scenario, it is not necessarily implied that there is less future variability (*Knutti and Sedlacek*, 2013), despite the suggestion of less variability from the twelve model ensemble used in this study as shown in Figure 1.
- 10 Regardless of epoch chosen for the delta method, the patterns are similar (Figure 6),and with the exception of the high latitudes, the differences in pattern are small (< 0.2°C). In both plots, the regression method has a significantly stronger temperature sensitivity in the Northern Hemisphere high latitudes from December to July and is weaker in boreal autumn (SON). This stems from how each methodology captures the effect of Arctic amplification, where the warming trend in the Arctic is almost twice as large as the trend in the global average, but this question is not explored here. It is interesting to</p>
- 15 note that in the S. Hemisphere, the Antarctic ice sheet margin (60-80°S) region has a weaker relationship with global mean temperature when using the regression method than using the delta method.

There are few regions where the patterns differ significantly (Figure 7), and there are fewer significant differences between the regression method and the delta method using the late 20th Century epoch than the late 19th Century epoch. Significant differences between method patterns are shown in the Baltic/ N. European region for both epochs, but in the earlier epoch, significant differences are shown in the Northwest Pacific region.

The global mean trend/difference forces an overestimation of the spatial pattern, particularly over land masses and high latitudes in the delta method (Figure 8), possibly due to Arctic amplification. Also, as shown in Figure 6, the Baltic/N. European region pattern has a stronger temperature sensitivity in the delta method. Overall, it appears that the regression pattern method overestimates the relationship between global temperature and local temperature, but the degree to which it overestimates the

25 relationship is small (< 0.8° C). When using the delta pattern, the Antarctic region is both overestimated and underestimated by up to 0.15° C, which is generally larger than the error in the regression pattern estimates. Temperature patterns fit well to GMT change, but when constructing precipitation patterns the relationship to GMT is not as strong. Nevertheless precipitation patterns have been shown to scale linearly with GMT (*May*, 2011). These patterns are valid in the 'wet get wetter and dry get drier idea, but the resulting patterns are likely to have larger errors, for many regions.

30 3.3 Scenario Differences

20

It is assumed that the local temperature sensitivity to global mean temperature change, regardless of methodology, is consistent across scenarios despite less accuracy in stronger mitigation scenarios (May, 2011). Assuming the trend in global mean temperature is linear, the first EOF in global annual temperature, in both future scenarios, is the warming trend (S4). This explains





5

the majority of the variance in the time series of temperature change (Table 3), despite differing rates of change from each scenario (Figures 1, 2, 3, and 4). This is expected, as the lower forcing scenario would result in less warming, and GMT change would be smaller accounting for less variance explained.

The choice of future forcing scenario may significantly ($\geq 0.05^{\circ}$ C) alter the resulting pattern. At high latitudes, the localglobal temperature relationship is significantly different and therefore not predicted well when using a pattern scaling method

- (Figure 9). In the higher forcing scenario, small portions along the Antarctic margin are significant at the 95% level. In the lower forcing scenario, there are significant fit errors along the Antarctic continent and in the North Atlantic, as well as throughout the mid and high latitudes. While this shows that the local-global relationship in the lower forcing scenario has more errors in fit, it does not imply that the relational pattern is significantly different across scenarios.
- The difference in patterns from future forcing scenarios is small ($< 0.5^{\circ}$ C) except over the Barents Sea region (Figure 10 10). For the delta patterns the largest relational difference in patterns is in the Northern Hemisphere at high latitudes, but these differences are not significant likely due to high variability in these regions. The difference in patterns generated by the regression method under different forcing scenarios are generally larger with more extreme deviations at mid-high latitudes. This is due to the fact that scenarios with stronger mitigation practices have a smaller GMT trend and the resulting local
- 15 temperature sensitivity to GMT is stronger.

This is further examined by separating the land and ocean patterns (Figure 11). The difference in pattern between scenarios for the regression method when isolating the land/ocean pattern is comparatively large, especially over the Arctic and Antarctic regions. The ocean (land) GMT sensitivity of the lower forcing scenario is $\geq 0.5^{\circ}$ C than the higher forcing scenario over the Arctic (Antarctic) region. In the higher forcing scenario the local fit to GMT has less significant errors, but this is likely

because the GMT trend is stronger and it is shown that the local temperature scales well with GMT. The difference in the 20 pattern between scenarios for the delta method when isolating the land/ocean pattern is small except over the Arctic region, which shows strong seasonal differences ($\geq 0.5^{\circ}$ C) in boreal autumn (SON). In this way the delta method is more consistent across future forcing scenarios, and this should be taken into consideration when choosing methodology.

Conclusions 4

This study looks specifically at differences in spatial temperature sensitivity patterns using simplistic methods. The goal in 25 pattern scaling is to efficiently generate statistical emulators from GCMs for use in SCMs, which we have done using two of the basic pattern scaling methodologies.

The assumption that choice of epoch does not alter the resulting delta pattern holds true. The difference in patterns from using epochs a century apart is not statistically significant. The GMT sensitivity and the spatial pattern is very similar for the 30 epochs used, despite the fact that epoch trends were dissimilar.

The difference in patterns generated by each method are minor except at Northern Hemisphere high latitudes and along the Antarctic margin. With the assumption of a future linear trend in GMT, the regression pattern works well with a better global to local fit. This results in the regression method pattern being closer to the modeled trend than the delta method pattern. The





simplistic design of the regression method allows for additional predictors in the pattern equation, and confidence intervals to be easily calculated, and as such more ideal for future use in an integrated assessment model experiment. The delta method introduces further complexity in choice of historical epoch, and assumes that there is no observed trend in the historical simulation.

- 5 Choice of scenario can affect the resulting pattern, particularly when using the regression method. In this case, the GMT temperature sensitivity is stronger when using a lower forcing scenario because the GMT trend is proportionally smaller and changes in GMT have a stronger effect on local temperature, particularly under strong mitigation. Delta method patterns are more consistent across scenarios with less heterogeneity in local temporal and spatial GMT sensitivity. With the assumption that different future forcing scenarios should not change the resulting pattern, the delta pattern is more consistent across scenarios.
- 10 regardless of epoch chosen, despite differences in epoch trends being large.

5 Code and/or data availability

CMIP5 model data is publicly available via the Earth System Grid Federation website (ESGF, https://pcmdi9.llnl.gov/). Reanalysis data is publicly available online via the Earth System Research Laboratory website

(http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html). Code used to construct this analysis is available on
15 GitHub through the Joint Global Change Research Institution repository. Any additional data can be obtained from Cary Lynch (cary.lynch@pnnl.gov).

Acknowledgements. This research is based on work supported by the US Department of Energy, Office of Science, Integrated Assessment Research Program. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.





References

35

Barnes, E. A., and R. J. Barnes (2015), Estimating linear trends: Simple linear regression versus epoch differences, *Journal of Climate*, 28(24), 9969–9976, doi:10.1175/JCLI-D-15-0032.1.

Collins, W. D., A. P. Craig, J. E. Truesdale, A. V. D. Vittorio, A. D. Jones, B. Bond-Lamberty, K. V. Calvin, J. A. Edmonds, S. H. Kim, A. M.

- 5 Thomson, P. Patel, Y. Zhou, J. Mao, X. Shi, P. E. Thornton, L. P. Chini, and G. C. Hurtt (2015), The integrated earth system model version 1: formulation and functionality, *Geoscientific Model Development*, 8, 2203–2219.
 - Dai, A. (2016), Future warming patterns linked to today's climate variability, Scientific Reports, 6, 19,110 EP -.
 - Dessai, S., X. Lu, and M. Hulme (2005), Limited sensitivity analysis of regional climate change probabilities for the 21st Century, *Journal of Geophysical Research: Atmospheres*, *110*(D19), n/a–n/a, doi:10.1029/2005JD005919.
- 10 Diffenbaugh, N. S., J. S. Pal, R. J. Trapp, and F. Giorgi (2005), Fine-scale processes regulate the response of extreme events to global climate change, *Proceedings of the National Academy of Sciences of the United States of America*, *102*(44), 15,774–15,778.
 - Eyring, V., M. Righi, A. Lauer, M. Evaldsson, S. Wenzel, C. Jones, A. Anav, O. Andrews, I. Cionni, E. L. Davin, C. Deser, C. Ehbrecht,P. Friedlingstein, P. Gleckler, K. D. Gottschaldt, S. Hagemann, M. Juckes, S. Kindermann, J. Krasting, D. Kunert, R. Levine, A. Loew,J. Mäkelä, G. Martin, E. Mason, A. S. Phillips, S. Read, C. Rio, R. Roehrig, D. Senftleben, A. Sterl, L. H. van Ulft, J. Walton, S. Wang,
- 15 and K. D. Williams (2016), ESMValTool (v1.0) –a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP, *Geosci. Model Dev.*, 9(5), 1747–1802, doi:10.5194/gmd-9-1747-2016.
 - Friedman, A. R., Y.-T. Hwang, J. C. H. Chiang, and D. M. W. Frierson (2013), Interhemispheric temperature asymmetry over the Twentieth Century and in future projections, *Journal of Climate*, 26(15), 5419–5433, doi:10.1175/JCLI-D-12-00525.1.
- Frieler, K., M. Meinshausen, M. Mengel, N. Braun, and W. Hare (2012), A scaling approach to probabilistic assessment of regional climate
- change, *Journal of Climate*, 25(9), 3117–3144, doi:10.1175/JCLI-D-11-00199.1.
 Hawkins, E., and R. Sutton (2015), Connecting climate model projections of global temperature change with the real world, *Bulletin of the*
 - American Meteorological Society, doi:10.1175/BAMS-D-14-00154.1.
 - Herger, N., B. M. Sanderson, and R. Knutti (2015), Improved pattern scaling approaches for the use in climate impact studies, *Geophysical Research Letters*, 42(9), 3486–3494, doi:10.1002/2015GL063569.
- 25 Holden, P. B., and N. R. Edwards (2010), Dimensionally reduced emulation of an AOGCM for application to integrated assessment modelling, *Geophysical Research Letters*, 37(21), doi:10.1029/2010GL045137.
 - Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhur, M. Chelliah, W. Evisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jeene, and D. Joseph (1996), The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American Meteorological Society*, *77*(3), 437 471.
- 30 Knutti, R., and J. Sedlacek (2013), Robustness and uncertainties in the new CMIP5 climate model projections, *Nature Clim. Change*, *3*(4), 369–373.
 - Kunkel, K. E., T. R. Karl, H. Brooks, J. Kossin, J. H. Lawrimore, D. Arndt, L. Bosart, D. Changnon, S. L. Cutter, N. Doesken, K. Emanuel, P. Y. Groisman, R. W. Katz, T. Knutson, J. O'Brien, C. J. Paciorek, T. C. Peterson, K. Redmond, D. Robinson, J. Trapp, R. Vose, S. Weaver, M. Wehner, K. Wolter, and D. Wuebbles (2012), Monitoring and understanding trends in extreme storms: State of knowledge, *Bulletin of the American Meteorological Society*, *94*(4), 499–514, doi:10.1175/BAMS-D-11-00262.1.
 - Leduc, M., H. D. Matthews, and R. de Elia (2016), Regional estimates of the transient climate response to cumulative CO₂ emissions, *Nature Clim. Change, advance online publication*, –.





- Lopez, A., E. B. Suckling, and L. A. Smith (2013), Robustness of pattern scaled climate change scenarios for adaptation decision support, *Climatic Change*, *122*(4), 555–566, doi:10.1007/s10584-013-1022-y.
- Lustenberger, A., R. Knutti, and E. M. Fischer (2014), The potential of pattern scaling for projecting temperature-related extreme indices, *International Journal of Climatology*, *34*(1), 18–26, doi:10.1002/joc.3659.
- 5 May, W. (2011), Assessing the strength of regional changes in near-surface climate associated with a global warming of 2°C, *Climatic Change*, *110*(3), 619–644, doi:10.1007/s10584-011-0076-y.
 - Meinshausen, M., S. J. Smith, K. Calvin, J. S. Daniel, M. L. T. Kainuma, J.-F. Lamarque, K. Matsumoto, S. A. Montzka, S. C. B. Raper, K. Riahi, A. Thomson, G. J. M. Velders, and D. P. P. Vuuren (2011), The rcp greenhouse gas concentrations and their extensions from 1765 to 2300, *Climatic Change*, 109(1), 213–241, doi:10.1007/s10584-011-0156-z.
- 10 Mitchell, T. D. (2003), Pattern scaling: An examination of the accuracy of the technique for describing future climates, *Climatic Change*, 60(3), 217–242, doi:10.1023/A:1026035305597.
 - Moss, R., J. Edmonds, K. Hibbard, M. Manning, S. Rose, D. van Vuuren, T. Carter, S. Emori, M. Kainuma, T. Kram, G. Meehl, J. Mitchell, N. Nakicenovic, K. Riahi, S. Smith, R. Stouffer, A. Thomson, J. Weyant, and T. Wilbanks (2010), The next generation of scenarios for climate change research and assessment, *Nature*, 463(7282), 747–756.
- 15 Osborn, T. J., C. J. Wallace, I. C. Harris, and T. M. Melvin (2015), Pattern scaling using ClimGen: monthly-resolution future climate scenarios including changes in the variability of precipitation, *Climatic Change*, pp. 1–17, doi:10.1007/s10584-015-1509-9.
 - Riahi, K., S. Rao, V. Krey, C. Cho, V. Chirkov, G. Fischer, G. Kindermann, N. Nakicenovic, and P. Rafaj (2011), RCP 8.5—a scenario of comparatively high greenhouse gas emissions, *Climatic Change*, 109(1), 33–57, doi:10.1007/s10584-011-0149-y.
- Ruosteenoja, K., H. Tuomenvirta, and K. Jylhä (2007), GCM-based regional temperature and precipitation change estimates for Europe
 under four SRES scenarios applying a super-ensemble pattern-scaling method, *Climatic Change*, 81(1), 193–208, doi:10.1007/s10584-006-9222-3.
 - Rupp, D. E., J. T. Abatzoglou, K. C. Hegewisch, and P. W. Mote (2013), Evaluation of CMIP5 20th Century climate simulations for the Pacific Northwest USA, *Journal of Geophysical Research: Atmospheres*, 118(19), 10,884–10,906, doi:10.1002/jgrd.50843.
- Sanderson, B. M., R. Knutti, and P. Caldwell (2015), Addressing interdependency in a multimodel ensemble by interpolation of model
 properties, *Journal of Climate*, 28(13), 5150–5170, doi:10.1175/JCLI-D-14-00361.1.
 - Santer, B., T. Wigley, M. Schlesinger, and J. Mitchell (1990), Developing climate scenarios from equilibrium GCM results, *Tech. rep.*, Hamburg, Germany.
 - Simmons, A. J., P. D. Jones, V. da Costa Bechtold, A. C. M. Beljaars, P. W. Kållberg, S. Saarinen, S. M. Uppala, P. Viterbo, and N. Wedi (2004), Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature,
- Journal of Geophysical Research: Atmospheres, 109(D24), n/a–n/a, doi:10.1029/2004JD005306.
 Stocker, T., D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley (Eds.) (2013), IPCC,
 - 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, doi:10.1017/CBO9781107415324.
- 35 Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2011), An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, doi:10.1175/BAMS-D-11-00094.1.
 - Tebaldi, C., and J. M. Arblaster (2014), Pattern scaling: Its strengths and limitations, and an update on the latest model simulations, *Climatic Change*, *122*(3), 459–471, doi:10.1007/s10584-013-1032-9.





Models	Spatial Resolution	Organization		
ACCESS1-0	1.2 x 1.8	Commonwealth Scientific and Industrial Research Organization, and		
		Bureau of Meteorology, Australia		
CanESM2	2.8 x 2.8	Canadian Centre for Climate Modeling and Analysis, Canada		
CCSM4	1 x 1	NCAR, University Corporation for Atmospheric Research, U.S.A		
CMCC-CMS	1.8 x 1.8	Centro Euro-Mediterraneo per Cambiamenti Climatici, Bologna, Italy		
CNRM-CM5	1.4 x 1.4	Centre National de Recherches Meteorologiques/Center Europeen de		
		Recherche et Formation Avancees en Calcul Scientifique, France		
CSIRO-Mk3.6	1.8 x 1.8	Commonwealth Scientific and Industrial Research Organization, Australia		
GFDL-CM3	2 x 2	NOAA, Geophysical Fluid Dynamic Laboratory, U.S.A		
HadGEM-ES	1.2 x 1.8	Meteorological Office Hadley Centre, UK		
INMCM4	1.5 x 2	Institute for Numerical Mathematics, Russian Academy of Sciences, Russia		
IPSL-CM5A-MR	1.25 x 2.5	Laboratoire de Meteorologique Dynamique, Institut Pierre-Simon Laplace, France		
MIROC-ESM	3 x 3	Atmosphere and Ocean Research Institute, National Institute for Environmental		
		Studies, and Japan Agency for Marine-Earth Science and Technology, Japan		
MPI-ESM-MR	1.8 x 1.8	Max Planck Institute for Meteorology, Germany		
NorESM1-M	2 x 2	Norwegian Climate Centre, Norway		

Table 1. List of the CMIP5 models and their respective spatial resolution and organization used in this analysis.

- Tebaldi, C., and R. Knutti (2007), The use of the multi-model ensemble in probabilistic climate projections, Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 365(1857), 2053–2075.
- Thomson, A. M., K. V. Calvin, S. J. Smith, G. P. Kyle, A. Volke, P. Patel, S. Delgado-Arias, B. Bond-Lamberty, M. A. Wise, L. E. Clarke, and J. A. Edmonds (2011), RCP4.5: a pathway for stabilization of radiative forcing by 2100, Climatic Change, 109(1), 77-94, doi:10.1007/s10584-011-0151-4.
- van Vuuren, D., J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. Hurtt, T. Kram, V. Krey, J. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S. Smith, and S. Rose (2011), The representative concentration pathways: an overview, Climatic Change, 109(1-2), 5-31, doi:10.1007/s10584-011-0148-z.

Vecchi, G. A., and B. J. Soden (2007), Global warming and the weakening of the tropical circulation, Journal of Climate, 20(17), 4316–4340, doi:10.1175/JCLI4258.1.

5





Table 2. Annual and seasonal mid (2041-2070) and late (2071-2100) 21st Century difference and 21st Century trend (2006-2100) in degrees Celsius for ensemble minimum, median, and maximum using the rcp8.5 future scenario.

		M21C - L19C Difference (°C)	M21C - L20C Difference (°C)	L21C - L19C Difference (°C)	L21C - L20C Difference (°C)	21C Trend (°C)
Ensemble Minimum	ANN	2.05	1.55	3.28	2.78	3.24
	DJF	2.18	1.71	3.65	3.17	3.71
	JJA	1.85	1.36	2.91	2.43	2.86
Ensemble Median	ANN	3.07	2.57	4.74	4.34	4.87
	DJF	3.22	2.75	5.07	4.62	5.25
	JJA	2.75	2.34	4.35	3.97	4.52
Ensemble Maximum	ANN	3.56	3.12	5.44	4.98	5.51
	DJF	3.74	3.46	5.77	5.47	6.04
	JJA	3.44	2.90	5.21	4.71	5.27

Table 3. Percent variance explained by empirical orthogonal functions of model ensemble mean of globally averaged surface air temperature for historical (1861-2005) and future (2006-2100) scenarios.

	Simulation	EOF1	EOF2	EOF3
Annual	Historical	18.39	11.35	7.60
	RCP 8.5	84.54	2.11	1.47
	RCP 4.5	56.33	5.70	4.29
DJF	Historical	21.60	14.136	11.04
	RCP 8.5	71.53	4.81	4.43
	RCP 4.5	39.42	13.30	7.79
JJA	Historical	19.04	10.37	8.0
	RCP 8.5	57.54	8.84	3.99
	RCP 4.5	21.24	16.03	8.60







Figure 1. Ensemble median global air temperature (degree Celsius). NCEP/NCAR reanalysis (1948-2010, purple line) compared to ensemble median historical simulation (1860-2005, black line), rcp4.5 (2006-2100, blue line), and rcp8.5 (2006-2100, red line) for annual (top plot), boreal winter (middle plot), and boreal summer (bottom plot). Shading indicates 1 sigma about the ensemble mean.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-170, 2016 Manuscript under review for journal Geosci. Model Dev. Published: 29 July 2016



© Author(s) 2016. CC-BY 3.0 License.





Figure 2. Ensemble mean 30 year trend difference in global air temperature in degrees Celsius between historical and future rcp8.5 (left column) and rcp4.5 (right column) CMIP5 scenarios for mean annual (top row), DJF (middle row), and JJA (bottom row).





Figure 3. Ensemble mean global air temperature in degrees Celsius difference between historical and future rcp8.5 (left column) and rcp4.5 (right column) CMIP5 scenarios for mean annual (top row), DJF (middle row), and JJA (bottom row).







Figure 4. Ensemble mean global air temperature delta method pattern difference based on differences in historical (top) and future (bottom) reference periods in degrees Celsius. Top plot shows late 21^{st} (2071-2100) and late 19^{th} (1861-1890) Century pattern and late 21^{st} and late 20^{th} (1971-2000) Century pattern difference. Bottom plot shows mid 21^{st} (2041-2070) and late 19^{th} (1861-1890) Century pattern and late 21^{st} (2071-2100) and late 19^{th} (1861-1890) Century pattern and late 21^{st} (2071-2100) and late 19^{th} (1861-1890) Century pattern and late 21^{st} (2071-2100) and late 19^{th} (1861-1890) Century pattern and late 21^{st} (2071-2100) and late 19^{th} (1861-1890) Century pattern difference.







Figure 5. Annual model ensemble global air temperature (degree Celsius) for late 21^{st} (2071-2100) vs. 19^{th} (1861-1890) Century difference and 21^{st} Century (2006-2100) trend. Future scenarios of rcp8.5 (brown) and rcp4.5 (green) are used.







Figure 6. Ensemble mean global air temperature regression method and delta method pattern (2071-2100 vs. 1861-1890 (left) and 2071-2100 vs. 1971-2100 (right)) difference for annual (top), DJF (middle) and JJA (bottom)in degrees Celsius.







Figure 7. Significance of difference for ensemble mean global air temperature regression method and delta method pattern (2071-2100 vs. 1861-1890 (left) and 2071-2100 vs. 1971-2100 (right)) for annual (top), DJF (middle) and JJA (bottom) using a Student's t-test probability statistic.







Figure 8. Difference in degrees Celsius between annual modeled trend for the 21st Century using forcing scenario rcp8.5 and delta method (2071-2100 vs. 1861-1890, top) and regression method (bottom).







Figure 9. Statistical significance (p-value) of regression patterns of rcp8.5 (left) and rcp4.5 (right) scenarios for annual mean (top), DJF (middle), and JJA (bottom).







Figure 10. Ensemble average difference annual delta (top, late 21st minus late 19th Century) and regression (bottom) patterns between rcp8.5 and rcp4.5 scenarios in degrees Celsius.







Figure 11. Ensemble annual average difference between rcp8.5 and rcp4.5 scenarios regression (left) and delta (right, late 21st minus late 19th Century) patterns (in degrees Celsius) for land only (top) and ocean only (bottom).