

Emulating mean patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments

Stacey E. Alexeeff¹ · Doug Nychka¹ · Stephan R. Sain¹ ·
Claudia Tebaldi²

Received: 15 February 2015 / Accepted: 8 September 2016
© Springer Science+Business Media Dordrecht 2016

Abstract There are many climate change scenarios that are of interest to explore by climate models, but computational power limits the total number of model runs. Pattern scaling is a useful approach to approximate mean changes in climate model projections, and we extend this methodology to build a climate model emulator that also accounts for variability of temperature projections at the seasonal scale. Using 30 runs from the NCAR/DOE CESM1 large initial condition ensemble for RCP8.5 from 2006 to 2080, we fit a pattern scaling model to grid-specific seasonal average temperature change. We then use this fitted model to emulate seasonal average temperature change for the RCP4.5 scenario based on its global average temperature trend. By using a linear mixed-effects model and carefully resampling the residuals from the RCP8.5 model, we emulate the variability of RCP4.5 and allow the variability to depend on global average temperature. Specifically, we emulate both the internal variability affecting the long-term trends across initial condition ensemble members, and the variability superimposed on the long-term trend within individual ensemble members. The 15 initial condition ensemble members available for RCP4.5 from the same climate model are then used to validate the emulator. We view this approach as a step forward in providing relevant climate information for avoided impacts studies, and more broadly for impact models, since we allow both forced changes and internal variability to play a role in determining future impact risks.

This article is part of a Special Issue on “Benefits of Reduced Anthropogenic Climate Change (BRACE)” edited by Brian O’Neill and Andrew Gettelman.

Electronic supplementary material The online version of this article (doi:10.1007/s10584-016-1809-8) contains supplementary material, which is available to authorized users.

✉ Stacey E. Alexeeff
salexeeff@ucar.edu

¹ Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, PO Box 3000, Boulder, CO, USA

² Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, PO Box 3000, Boulder, CO, USA

Keywords Initial condition ensemble · Internal variability · Pattern scaling · Emulator

Table of Contents

1. Introduction to the special issue (B. O'Neill, A. Gettelman)
2. Comparing societal impacts and mitigation between RCP8.5 and RCP4.5: A review (B. O'Neill, B. van Ruijven, et al.)

Methodological issues

3. On the dependency of climate variability and extremes with mean climate state (B. Sanderson et al.)
4. Patterns of mean temperature and variability across and within scenarios: using the RCP8.5 large ensemble to emulate RCP4.5 (S. Alexeeff et al.)
5. Detecting differences in extreme temperature and precipitation between RCP8.5 and RCP4.5 (C. Tebaldi et al.)
6. Precipitation extremes via pattern scaling from an initial condition ensemble of the CESM (M. Fix et al.)

Heat Extremes & Health

7. Avoided impacts of urban and rural heat waves over the U.S. using large climate model ensembles for RCP8.5 and RCP4.5 (K. Oleson et al.)
8. Population exposure to heat-related extremes: Demographic change vs climate change (B. Jones et al.)
9. Avoided extreme heat-related health impacts in U.S. cities (B. Anderson et al.)
10. Climate change influences on extreme heat vulnerability in Houston, Texas (Marsha et al.)
11. Climatic suitability for the dengue virus vector mosquito *Aedes aegyti*: Historical and future geographic patterns under RCP8.5 and RCP4.5 scenarios (Monaghan et al.)

Agriculture and land use

12. Implications of alternative land use patterns for the carbon cycle and regional climate outcomes in RCP8.5 and RCP4.5 (Lawrence et al.)
13. Simulated twenty-first century changes in large-scale crop water requirements and yields (Levis et al.)
14. Avoided economic impacts of climate change on agriculture (Ren et al.)
15. Avoided impacts of mean and extreme temperatures on crops (Tebaldi & Lobell)

Tropical Cyclones

16. Hurricane/tropical cyclone activity in 2070–2100 in RCP8.5 vs RCP4.5 (Bacmeister et al.)
17. Estimating avoided tropical cyclone impacts with an index of damage potential (Done et al.)
18. Tropical cyclone damage assessments in the twenty-first Century: Climate and development contributions (Gettelman et al.)

Sea Level Rise

19. Avoided Sea level rise in RCP8.5 vs RCP4.5 (Hu & Bates)

Drought and conflict

20. Avoiding drought risks and potential water management conflict under climate change (Towler et al)

Air Pollution

21. Climate impacts and trade-offs of air quality policies in RCP8.5 (Lamarque et al)

BRACE: background and rationale

Understanding the potential consequences of climate change for ecosystems and society is necessary for an informed response to the climate issue. A wide body of literature on the impacts of climate change has developed over the past 20 years, as assessed in successive reports of IPCC's Working Group 2. Yet much remains to be done. A particularly important task is improving our understanding of how impacts differ across alternative levels of future climate change. Better understanding the consequences of a world in which the radiative forcing driving climate change increases by small, medium, or large amounts, for example, can inform policy discussions of desired long-term climate outcomes.

Costs of both mitigation and adaptation vary with future climate outcomes as well. Mitigation costs have been better studied, indicating that reducing emissions enough to limit radiative forcing to about 4.5 W/m² above pre-industrial levels is relatively inexpensive, while costs to achieve forcing below that level increase rapidly (Clarke et al., 2014). What is less well understood is what the benefits of such mitigation would be, in terms of reduced climate change impacts and adaptation costs.

This special issue on the Benefits of Reduced Anthropogenic Climate change (BRACE) is aimed at helping to fill a gap in understanding how impacts vary across different future climate outcomes. This “avoided” or “differential” impacts framing is intended to derive the costs and benefits of various long-term climate outcomes. The BRACE project contributes to this literature by assessing the differences in impacts between two specific climate futures: those associated with Representative Concentration Pathways (RCPs) 4.5 and 8.5. The latter would lead to a likely global average temperature change of 2.6–4.8 °C relative to recent temperatures by the end of the century, the former to a likely range of 1.1–2.6 °C degrees of warming (Collins et al., 2013).

The BRACE project is not alone in this goal. Other recent or current studies focused on avoided impacts include the US EPA study on Climate change Impacts and Risk Analysis (CIRA), the UK AVOID (and now AVOID2) project, and less directly, ongoing EU projects such as IMPRESSIONS and HELIX. Each has a particular focus: CIRA is a US-only study with wide sectoral coverage; AVOID is a global study with a substantial mitigation and emissions pathway component, focusing in particular on the benefits of stringent mitigation

pathways; IMPRESSIONS and HELIX focus on Europe, with smaller case studies in other parts of the world.

The BRACE study complements these other activities and breaks new ground in several areas. It is global in scope, with a regional focus in a subset of papers on the US. It includes a substantial component on the differences in physical impacts (e.g., heat waves, tropical cyclones, drought, sea level rise) between the scenarios, drawing heavily on multi-member initial condition ensembles of the Community Earth System Model (CESM). These ensembles include a recently produced Large Ensemble (30 members) for RCP 8.5 (Kay et al., submitted) and a smaller (15 member) ensemble for RCP 4.5 newly developed as part of the BRACE project (Sanderson et al.). Societal impact studies are based on the new Shared Socioeconomic Pathways (SSPs; O'Neill et al., 2013); the study therefore contributes to a nascent literature based on the new scenario framework (van Vuuren et al., 2013) combining RCPs and SSPs. These ensembles are also used to make several methodological advances in accounting for variability even when ensemble simulations are not available. The study also includes the first application of new global scenarios of spatial population distribution, the first use of crop models within CESM in a global agricultural impact assessment, and new high resolution (25 km) global simulations of tropical cyclone activity.

BRACE authors cut across disciplines. Many are trained in economics or other social sciences and focus on societal impacts of climate change. Other authors are physical climate scientists who build and analyze fully coupled Earth System Models (ESMs), including CESM. Many of the papers have co-authors from both types of scientists.

Most of the BRACE analyses involve multiple papers approaching a specific type of impact – those related to tropical cyclones, agriculture, and heat extremes – from different perspectives, building on each other. For example, Bacmeister et al. present new high resolution simulations of future tropical cyclones; Done et al. present a new index of cyclone damage potential that is then calculated based on those simulations; and Gettelman et al. estimate actual economic damages from future cyclone activity based on a spatial model of the economic value of physical assets. Similarly, on the topic of heat extremes, Oleson et al. analyze the difference in heat wave occurrence in urban and rural areas in CESM between the two scenarios; Jones et al. combine these climate model outcomes with projected spatial distributions of the population to estimate future exposure to heat waves; Anderson et al. use these projections to estimate future heat-related mortality in US cities; and Marsha et al. focus on mortality consequences for an individual city (Houston) in which they can account for within-city spatial heterogeneity in urban form and socioeconomic conditions. For agriculture, Levis et al. use the CESM land surface model to assess the impacts on crop yields of the alternative RCPs, while Ren et al. use these yield effects to investigate their economic implications for global agriculture in the NCAR integrated assessment model. Tebaldi and Lobell address a shortcoming of the CESM yield modeling by assessing the potential direct impacts of extreme heat on crop production.

The large and medium ensembles also afford the opportunity to examine critical methodological issues in the treatment of variability and extremes in impact studies. Sanderson et al. present the RCP4.5 Medium Ensemble and examine the linearity of variability with differences in global average temperature, identifying where such linearity breaks down and suggesting physical reasons for nonlinear responses. Alexeeff et al. and Fix et al. propose and test methods of pattern scaling that go beyond the usual approaches to provide not just spatial patterns of temperature and precipitation but also estimates of their variability. This group of papers as a whole offers methods that can expand the types of avoided impacts studies that are possible to

carry out in the future by extending them to situations in which ensembles of simulations are not available for all scenarios to be assessed. Further, Tebaldi and Wehner, and Fix et al., employ the ensembles to identify when differences in extreme temperature and precipitation events become apparent (and statistically significant) between them.

Overall, the BRACE study combines modeling of the physical and human systems with methodological advances in the treatment of variability and extremes to advance understanding of climate change impacts in a number of sectors at the US and global level.

References

- Clarke, L., et al. (2014) Assessing transformation pathways. *Climate Change 2014: Mitigation of Climate Change*. Intergovernmental Panel on Climate Change, Report of Working Group III.
- Collins, M., et al. (2013) Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*. Intergovernmental Panel on Climate Change, Report of Working Group I.
- Kay, J. E., C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. Arblaster, S. Bates, G. Danabasoglu, J. Edwards, M. Holland, P. Kushner, J. -F. Lamarque, D. Lawrence, K. Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, and M. Vertenstein, 2014: 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. Met. Soc.*, submitted April 17, 2014.
- O'Neill, B.C., Kriegler, E., Riahi, K., Ebi, K., Hallegatte, S., Carter, T.R., Mathur, R., van Vuuren, D. (2013) A new scenario framework for Climate Change Research: The concept of Shared Socio-economic Pathways. Special Issue on "A new scenario framework for climate change research: background, process, and future directions," *Climatic Change*, DOI 10.1007/s10584-013-0905-2.
- van Vuuren, D., Kriegler, E., O'Neill, B.C., Ebi, K., Riahi, K., Carter, T., Edmonds, J., Hallegatte, S., Kram, T., Mathur, R., Winkler, H. (2013) A new scenario framework for Climate Change Research: Scenario matrix architecture. Special Issue on "A new scenario framework for climate change research: background, process, and future directions," *Climatic Change*, DOI 10.1007/s10584-013-0906-1.

1 Introduction

Pattern scaling is a popular method for approximating average regional changes in future climate model projections using global average temperature change. Typically, pattern scaling models assume that regional changes in climate over time follow a linear function of the global mean change in temperature (Tebaldi and Arblaster 2014). Many studies have demonstrated that decadal-average regional temperature changes can be closely approximated by linear pattern scaling (Mitchell 2003, Ruosteenoja et al., 2007, Cabré et al., 2010). In contrast, this paper focuses on pattern scaling of annual seasonal mean temperatures and considers variability around those means, developments which are needed so that pattern scaling can be better used for impact studies. This study is part of a larger project on the Benefits of Reducing Anthropogenic Climate change (BRACE; O'Neill and Gettelman, in prep.), which focuses on characterizing the difference in impacts driven by climate outcomes resulting from the forcing associated with Representative Concentration Pathways (RCPs) 8.5 and 4.5.

To study impacts of particular climate forcing scenarios, outputs from climate model simulations can be used as inputs to process-based or empirical models of natural or human systems. Many climate impacts, such as agriculture, heat waves, droughts and air pollution, require climate model outputs at a regional spatial scale and, at a minimum, at a seasonal or annual temporal scale rather than in the form of decadal means (Maracchi et al., 2005, Patz et al., 2005, Piao et al., 2010). At these spatio-temporal scales, the internal variability of the climate system plays a greater role and may represent a large fraction of total variability across models and scenarios (Hawkins and Sutton 2009). In addition, to compare the impacts of different climate scenarios, there is a need for climate model simulations under *multiple scenarios* as well as simulations of multiple ensemble members *within a given scenario*. Creating these model ensembles allows separation of the forced response from the noise of internal variability. The inability to generate a large number of high cost simulations is the driving reason for creating a statistical emulator, which seeks to mimic the behavior of climate model simulations through computationally inexpensive statistical methods. Thus, emulators of key climate variables under alternative climate scenarios can be a valuable tool by providing inputs to impact models at a low computational cost. This statistical emulation would facilitate the quantitative assessment of avoided impacts, i.e. impacts that could be avoided if climate change mitigation could reduce the rate of the climate change forcing effects.

To tailor the pattern scaling methodology to the needs of impact studies, this paper proposes and illustrates a pattern scaling approach for building a statistical emulator using the NCAR Community Earth System Model (CESM). The CESM Large Ensemble (CESM-LE), a 30-member initial condition ensemble of CESM simulations branched at 1930 (and therefore regarded as independent realizations of internal variability by the time they reach the baseline period considered here, of 1976–2005) for RCP8.5 (Kay et al., 2014 (to appear)) is used to build a pattern scaling emulator of CESM. We then use the emulator to generate approximate realizations of regional temperatures under the RCP 4.5 forcing scenario. Finally, the similarly constructed CESM Medium Ensemble (CESM-ME), a 15-member initial condition CESM ensemble for RCP4.5 (Sanderson et al., in prep) is used to validate the emulator. Note that the next phase of the Coupled Model Intercomparison Project (CMIP6) will include the request for a sizable initial condition ensemble to be run by the participating models for one of the scenarios planned to inform future projections (O'Neill, personal comm.). Thus, we see our method as timely in providing a way to extend the information about internal variability that will be available under a single scenario to others for which only a single run is planned.

2 Pattern Scaling Emulator of RCP 4.5 Ensemble using RCP 8.5 Ensemble

2.1 Pattern scaling model and its assumptions using RCP 8.5

We first describe our pattern scaling model and the assumptions for the mean and variance of the statistical model. Our main innovation is that we propose adding the variance by resampling the residuals to allow variability to change over time by grid point in response to the external forcing. We assume a linear relationship between the grid-specific temperatures and the global average temperature, typical of pattern scaling models. Using a linear mixed-effects model, random slopes and intercepts for each ensemble member model the variation among

ensemble members around the mean pattern common to the ensemble. Specifically, for each grid point i and season s , we propose the model

$$T_{isk} - T_{is*0} = (a_{isk} + a_{is}) + (b_{isk} + \beta_{is}) \cdot (g_t - g_0) + \varepsilon_{isk}$$

where t denotes time in years, k denotes ensemble member and 0 indicates the baseline period of 1976–2005, and T denotes surface temperature. Then, T_{is*0} is the average surface temperature over all K ensemble members at baseline, g is global average temperature over all ensemble members. The parameter β is the fixed-effect slope, α is the fixed-effect intercept, the random-effects a , b are assumed to have a multivariate normal distribution with mean zero, $\text{Var}[a_{isk}] = \tau_{ais}^2$ and $\text{Var}[b_{isk}] = \tau_{bis}^2$ with a Matern covariance that accounts for the spatial correlation among grid points. We assume the residuals are $\varepsilon_{isk} \sim N(0, \sigma_{is}^2)$. We fit this model separately to each grid point on land for the seasons of boreal summer (JJA) and winter (DJF), where each model includes all $K = 30$ ensemble members of RCP 8.5 large ensemble for years $t = 2006, \dots, 2080$. Because of the large number of data points, it is convenient to use the well-established two-stage random effects model formulation. The first stage models the random effects for each ensemble member k at each grid point i and the second stage models the fixed effects representing the average responses across all ensemble members (Fitzmaurice et al., 2011). A nonstationary covariance function is used to model the spatial covariance of the random effects and is fit as a third step in this modeling framework.

The average slopes β of the fitted model for every grid point on land are shown in Fig. 1. We can see from their positive values how the regional temperatures at all the grid points on land are rising for every degree of increase in global average temperature. Since inclusion of the fixed-effect intercept term α is not standard in pattern scaling models, we first tested for statistical significance of this term at each grid point. Based on a two-sided t-test at level 0.05, we found that the fixed intercepts were statistically significant in 7851 grid points (60.5 %) for DJF, and in 9678 grid points (74.6 %) for JJA. The random slopes had an average standard deviation of 0.11 °C. Relative to the residual variance, the random intercepts and slopes for the individual ensemble members had small variances; thus, the within-ensemble-member variance was greater than the between-ensemble-member variance across grid points. Based on a likelihood ratio test for inclusion of random effect parameters at level 0.05, we found that the random effects were statistically significant in 1510 grid points (11.6 %) for DJF and in 1871 grid points (14.4 %) for JJA. These numbers are conservative estimates of the number of true

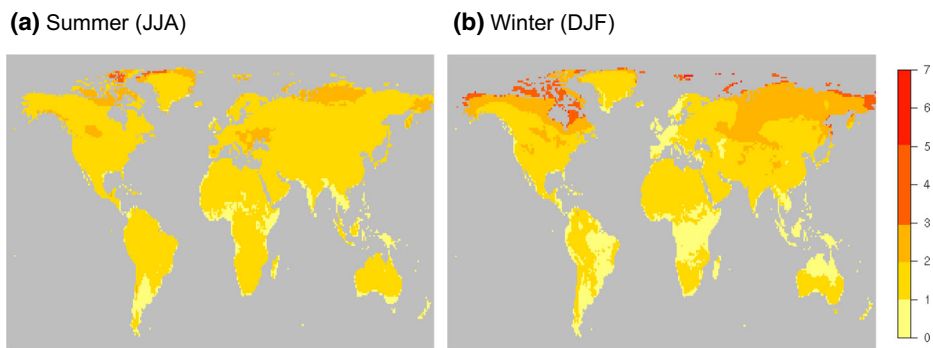


Fig. 1 The average slopes of the fitted pattern scaling model using the RCP 8.5 CESM-LE, showing the average rate of change in surface temperature at each grid point per degree increase in average global temperature

positives since this test for variance parameters in random effects models is known to be underpowered (Fitzmaurice et al., 2011). To combat this lack of power, using a less stringent cutoff level of 0.10 to determine statistical significance is often recommended, but we used a cutoff of level 0.05 for statistical significance for all tests for consistency in the manuscript.

2.2 Identifying key relationships in the pattern scaling model residuals

A key feature of our emulator is use of the residuals of the fitted pattern scaling model to the CESM-LE. We seek to emulate the internal variability of an ensemble of temperature simulations. To ensure that our emulator preserved meaningful temporal and spatial relationships in the variability of the CESM ensemble members, we first had to determine which temporal and spatial relationships were important. We evaluated three aspects of the detrended CESM-LE ensemble members: (i) multi-decadal trends in variance, (ii) short-term temporal auto-correlation, and (iii) spatial correlation within each year.

First, we examined the multi-decadal variability trends at each grid point from 2006 to 2080. We conducted an F-test at level 0.05 to compare the variance in local temperature during 2006 to 2030 to the variance during 2056 to 2080 at each grid point on land across all thirty CESM-LE ensemble members. We found that in 48 % of grid points there is no statistically significant change in variability of the local temperature as the global temperatures increase. In 15 % of grid points the local temperatures exhibit more variability as the global temperatures increase and in the remaining 37 % of grid points the variability of temperatures show a decreasing trend. Figure 2 illustrates examples of these changes in variability for the CESM-LE members at a few selected grid points.

We next calculated the seasonal autocorrelation of the detrended model residuals at each grid point on land up to a lag of 15 years. We found that after removing the pattern scaling trend the temporal autocorrelations in the residuals were negligible for each lag: on average 0.03 for lag 1, -0.10 for lag 2, -0.05 for lag 3, and similarly for lags 4–15. Thus, it is reasonable to consider detrended model residuals to be temporally exchangeable, i.e. independent, within a short-term period for our emulator. Finally, we examined the spatial correlation of the residuals by diagnostic plots of the residuals for selected ensemble members in the CESM-LE RCP8.5, shown in Supplementary Fig. 1. Spatial patterns are clearly visible, with different patterns for each ensemble member. Based on this evaluation, we decided to preserve all spatial correlation in the residuals and preserve any overall long-term variance trends.

2.3 Emulation of RCP 4.5 ensemble

Our emulator assumes that we can simulate the response of global average temperature to the forcing in RCP 4.5 (through a simple, computationally inexpensive climate model). In our specific application we can use the actual global average temperature ensemble mean time series from CESM-ME RCP4.5. We indicate global average temperature by g_t , for each year t from 2006 to 2080. We then use our fitted pattern scaling model estimates to generate members of the Emulated 4.5 ensemble. The algorithm for generating an ensemble follows the steps:

1. Compute the ensemble-average global mean temperature for each year of RCP4.5 and difference from the initial baseline period, $g_t - g_0$.
2. Add the estimated fixed effects, β_{is} , representing the overall average pattern across all ensemble members.

3. Generate the emulated random slopes, \tilde{b}_{isk} , using the estimated variance and spatial correlations.
4. Generate the spatially correlated residuals, $\tilde{\varepsilon}_{iskt}$, year by year such that the residuals are drawn from years of the CESM-LE with a global average mean temperature within a 1 °C window of the RCP4.5 global average mean temperature from step 1 above.
5. Combine the results of steps 1 to 4 to compute the emulated ensemble member

$$\tilde{T}_{iskt} = a_{is} + \left(\tilde{b}_{isk} + \beta_{is} \right) \cdot (g_t - g_0) + \tilde{\varepsilon}_{iskt}$$

In this way, we can generate hundreds of Emulated 4.5 ensemble members using this procedure. Note that for the emulation we set the intercept at its mean value from the fitted random effects model and do not simulate additional variability around the intercepts due to the complexity of the spatial cross correlations between the random intercepts and slopes. For further details about Step 3 and 4, see the description in Supplementary Materials.

3 Results and validation of emulated ensemble for RCP 4.5

We consider validation of the mean and of the variance separately. To validate the emulation of the mean surface temperatures, we considered 20-year average seasonal surfaces. Figure 3 shows

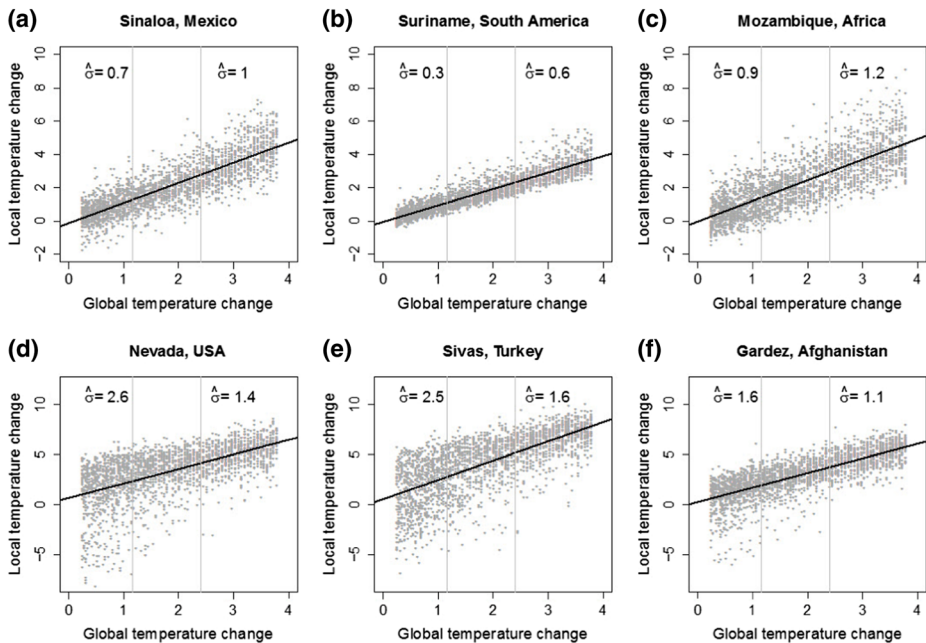


Fig. 2 An exploration of the long-term trends in variability of seasonal temperatures in the CESM-LE climate experiment. Change in local surface temperatures versus average global temperature change (degrees C) from the baseline period of 1976–2005 for CESM-LE members in selected grid points, and standard deviations of the local seasonal temperatures during 2006–2030 and during 2056–2080, illustrating examples of an increasing variance trend (a–c) and a decreasing variance trend (d–f)

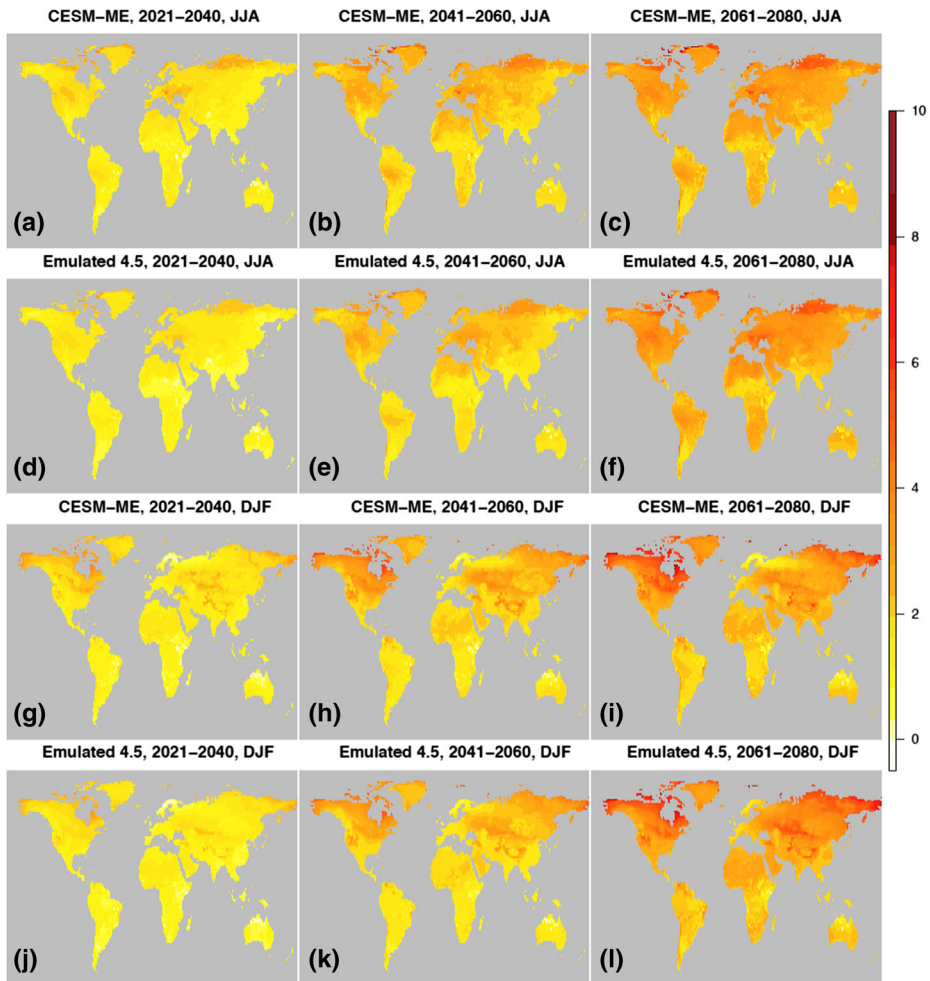


Fig. 3 The 20-year average increase in surface temperatures (degrees C) relative to the baseline period of 1976–2005 for the CESM-ME (a–c, g–i) and the Emulated 4.5 ensemble (d–f, j–l), for each 20-year period from 2020 to 2080, during summer (a–f) and winter (g–l)

the average increase in surface temperatures for the CESM-ME and the Emulated 4.5 ensemble, for summer (JJA), winter (DJF), and for each 20-year period from 2020 to 2080. The overall patterns of average temperature change are well captured by the emulated mean surfaces. To quantify predictive performance, we computed the absolute mean difference and the root mean squared error between the CESM-ME and the Emulated 4.5 ensemble for the 20 year average surfaces. Overall, we found that the root mean squared error (RMSE) of the multi-decadal mean grid point temperatures was 0.22 °C for summer and 0.34 °C for winter. For the predictive performance at each grid point on land, the absolute temperature difference was less than 0.2 °C for 66.1 % of the grid points and less than 1 °C for 98.6 % of the grid points. Supplementary Fig. 3 shows the absolute differences versus the percent of grid points in which the difference in mean temperatures falls within each range.

We also compared the mean emulation performance of our random effects pattern scaling model to the usual simple pattern scaling model (with no fixed intercept and no random intercepts or slopes) and to an intermediate model that included only fixed and random slopes (with no fixed or random intercept). We emulated the 20-year mean surface temperature patterns from 2020 to 2080 under RCP 4.5 for each alternative model, for summer (JJA) and winter (DJF). We quantified the predictive performance of this model using the 20-year surface temperature patterns from CESM-ME as our validation. Overall, we found that the predictive performance of the alternative pattern scaling models was not quite as good as that of the full random effects pattern scaling model. Specifically, the usual simple pattern scaling model had a RMSE of 0.24 °C for summer and 0.39 °C for winter and the intermediate model had a RMSE of 0.30 °C for summer and 0.42 °C for winter. The percentage of grid points with an absolute temperature difference of <0.2 °C was 62.6 % for the simple pattern scaling model and 49.8 % for the intermediate model. Supplementary Fig. 3 shows the absolute differences versus the percent of grid points for the full random effects pattern scaling model in comparison to the simple pattern scaling model. We note that since the random slopes and residuals are constrained to have mean zero, we would expect the 20-year mean surface temperature patterns estimated by these two models to be similar. However, when examining residual plots of the intermediate model fits, the exclusion of the intercept terms brought additional heteroscedasticity to the residuals that impacted the centering of the residuals around zero over time (Supplementary Fig. 4).

To validate the variability assumed in the emulator across space, we used a principal components method. For a given year, we compute the principal components of the Emulated 4.5 Ensemble members and compute the variance of the Emulated 4.5 Ensemble explained by the first ten principal components. We then use the same basis functions of these principal components to compute the proportions of variance explained in the CESM-ME. If the two ensembles have the same patterns of spatial variation, then the same principal components will explain the same amount of the total variation. Further details can be found in the Electronic Supplemental Material. Because the years are assumed to be *independent* while principal components are useful to explain *correlations*, we examined each year separately. Supplementary Fig. 2 shows the variance explained by the first ten principal components in the emulated RCP 4.5 ensemble compared to the CESM-ME, and compared to random noise, for the selected years 2030, 2060, 2080. The cumulative variance explained by this basis is similar across the Emulated 4.5 Ensemble and the model output, demonstrating the similarity of the spatial variance and spatial correlations in the two datasets.

To validate the temporal trends in variance in the Emulated 4.5 Ensemble surface versus the CESM model output ensemble, we considered the properties of the individual time series for each grid point. Our emulator assumed that there was no dependence between the variability of temperatures of different years in the short term. We found that the first through fifteenth order auto-correlations were small in both the CESM-ME RCP4.5 output and the emulated RCP 4.5 ensemble. Because our emulator assumed that the regional variance may change over time in response to the forced climate change signal, we computed the variance over each twenty year period from 2021 to 2080 for the CESM-ME RCP4.5 output and the Emulated 4.5 Ensemble, and conducted an F-test at level 0.05 to test for differences in the variance. We found statistically significant differences in the variances in only 348 grid points (2.7 %) during 2021–2040, 585 grid points (4.5 %) during 2041–2060, and 891 grid points (6.9 %) during 2061–2080; thus, the variances were approximated well by the emulator.

4 Discussion and conclusions

This paper demonstrates that we can construct a pattern scaling emulator to approximate an initial condition ensemble for regional temperature under a different forcing scenario. The main features of building the statistical emulator are to fit a pattern scaling model with random slopes and intercepts to an existing ensemble of climate model runs, and to resample the model residuals in a way that preserves key spatial and temporal features. By sampling the residuals within a window of the corresponding global temperature change, we allow the variability to change based on the forcing component.

We found that the mean temperature changes at regional scales from 2006 to 2080 in the CESM-ME RCP4.5 were well represented by our linear pattern scaling model for the mean. In addition, we extended the traditional pattern scaling methodology to account for variability, so that many runs of an initial condition ensemble can be approximated rather than just the mean change. A number of climate model projection studies have found that regional variability in temperatures may change between the present and future (2050–2100) time periods (Schar et al., 2004, Salinger 2005, Beniston et al., 2007). This changing variability over time was also evident at some grid points in the CESM-LE RCP8.5. Our proposed pattern scaling methodology accounts for this phenomenon by assuming that the variability changes as a function of the forcing signal, and allows the variability of the emulator to also change relative to global temperature.

Our pattern scaling methodology made a number of assumptions and choices regarding the spatial and temporal properties of CESM. Each assumption was evaluated and tested for statistical significance. We conducted tests for statistical significance at every grid point on land at level 0.05, and thus the principles of multiple testing apply. When conducting a large number of independent statistical tests at level 0.05, this means that 5 % of the tests of effects that are truly null are expected to be statistically significant by random chance alone, i.e. false positives. In our setting 5 % of all grid points tested would be 648 grid points. However, it is important to note that in climate model settings, the spatial correlation of the temperature patterns may cause the tests to not be completely independent, so the actual false positive rate may be slightly lower or higher than 5 %.

Our choice of a one-degree-width window for sampling the residuals was based on empirical evidence and statistical and geo-physical considerations. Empirically, we observed changes in variance over longer periods (as shown in Supplementary Fig. 1). Geo-physically, we know that some changes in variance of temperatures are expected eventually in response to major changes in global forcing, while also knowing the climate system itself is relatively stable and not highly sensitive to very small changes in global forcing. A smaller window may improve the statistical accuracy of the emulator. Empirically, we observed that a slight reduction in the width of the sampling window corresponded to a slight decrease in the number of statistically significant F-tests for differences in the variance between the Emulated 4.5 Ensemble and the CESM-ME. However, we observed a satisfactory level of significant tests relative to the expected false positive rate using a one degree-width sampling window. In addition, the need to reduce the false positive rate must be balanced with the need to create different members of the Emulated 4.5 Ensemble, where a smaller sampling window reduces the number of residuals to choose from. We suggest that the choice of window may need to be tailored to specific climate variables.

Previous studies have found that linear pattern scaling models provide a good approximation to average regional temperature changes (Mitchell 2003, Ruosteenoja et al., 2007, Cabré et al., 2010), especially when the main drivers of change are increasing greenhouse gas

concentrations. We found better predictive performance of our pattern scaling emulator model with random effects than a simple pattern scaling model. Note that the marginal model of our pattern scaling emulator represents the ensemble average and reduces to the fixed effect coefficients because random intercepts and slopes have mean zero. The fixed effect coefficients are estimated by the generalized least squares estimator which incorporates the correlations induced by the random effects structure. The ordinary least squares estimator in a simple pattern scaling model assumes complete independence within and between ensemble members, which could lead to a biased estimate of the pattern scaling slopes if within-member correlations are present (Fitzmaurice et al., 2011). Thus, the difference between the estimates from two models depends on the degree of correlations within ensemble members.

An advantage of our emulator is using the CESM-LE output directly, to preserve spatial and temporal relationships in the emulator residuals. However, this strategy relies on the availability of a climate model experiment with a large initial condition ensemble. This may limit the wider applicability of the method depending on the design of future ensemble experiments.

We view our pattern scaling emulator approach as a step forward in providing relevant climate information for avoided impacts studies, but there are also some considerations and limitation to be overcome in extending this methodology to emulate scenarios other than RCP4.5. The emulation of strongly mitigated pathways (e.g. CMIP5 RCP2.6) or other trajectories such as overshoots with significant ramp-up and decline may pose challenges to our method, which we have not investigated. In particular, those scenarios with significant overshoot in forcing and therefore global mean temperature change could be more challenging because it may not satisfy the assumption of a time-invariant pattern if different parts of the climate system respond to the forcing change at different rates. We recommend additional investigation into the patterns of change in these types of scenarios to develop and validate an appropriate emulation technique.

In addition to the above mentioned mitigation cases, our pattern scaling approach may not hold for scenarios that include *regionally varying* forcing, such as emissions of aerosol precursors or land-use changes because these effects could result in different spatio-temporal patterns across the different scenarios of interest. This type of forcing signal could affect the mean spatial patterns in ways that cannot be accounted for by a simple linear regression of a single pattern on global average temperature changes, dominated by the effects of greenhouse gases. Approaches involving more than one scaling variable could be used to address this limitation. In this respect, it may be interesting to note that RCP4.5 assumes different land use change and different aerosol forcing than RCP8.5, yet those differences did not result in meaningfully different patterns of temperature changes or variability for the mean temperatures we examined in this study. However, other climate variables may be more sensitive. Here we also add as a note of caution that our exercise used the actual global average temperature time series under RCP4.5 from the CESM-ME experiment. More general implementations of this methodology would use a simple model to emulate also global average temperature under RCP4.5, introducing an additional degree of uncertainty in the emulation of the actual CESM results.

An overview of the CESM characteristics and performance metrics is beyond the scope of this paper, but Hurrell et al. (2013) and Kay et al. (2014) contain evaluations of the model output under a wide range of perspectives for both mean and variability patterns. In addition, results from applying many validation metrics evaluating the specific experiments that we use in this study have been made publically available online (NCAR CGD 2016). We emphasize that the eventual use of this type of emulator model for impacts work relies directly on the

validity of the underlying climate model in reproducing all relevant spatial and temporal trends in the climate variables. Thus, additional validation of the underlying climate model is a critical precursor to any impacts work, and the specifics of the validation should be focused to depend on the particular features most relevant to the impacts that the work addresses. Another concern is that we detected no short-term temporal auto-correlations when there are known decadal variability events in climate systems. Ideally, we would model these phenomena directly in the mean function, but if these signals introduce inconsistently timed non-linearities in the trend of temperatures that vary by ensemble member, they may not be detectable in the ensemble mean pattern. In addition, the statistical test examines the average correlation over a fixed time interval, so decadal patterns that have slight variability in time intervals (such as El Nino) may be hard to detect. For other applications of pattern scaling beyond seasonal average temperatures, temporal auto-correlations may be more pronounced. In the case where temporal correlations over a few years are important to incorporate in the emulator, an alternative strategy would be to sample several years of residuals simultaneously, thus preserving most of this correlation.

The emulation of some climate variables may not satisfy the assumption of linearity in local mean changes as global temperature increases. For example, precipitation mean changes in the future have been found to deviate from linearity in several studies (Wilby 1997, Watterson 2008, Cabré et al., 2010). Thus, some assumptions may need to be modified to work well for other climate variables. Another limitation is that our current emulator strategy is univariate, whereas impact modeling often requires more than just temperature. For example, one important priority for adaptation and mitigation planning is comparison of the agricultural yields of major crops such as maize, wheat, soy, and rice across global crop models and forcing scenarios (Rosenzweig et al., 2014), which would require extending this method to jointly emulate additional variables such as humidity, CO₂ concentration and precipitation, among others (Izaurrealde et al., 2006, Bondeau et al., 2007, Deryng et al., 2011).

Overall, our pattern scaling emulator for seasonal temperature represents a step forward in providing relevant climate information for avoided impacts studies, and more broadly for impact models, since the Emulated 4.5 ensemble for temperature includes both forced and internal variability. Our emulator approach would allow impact studies to be carried out for many alternative scenarios seeing an increase in greenhouse gas forcing, all of them including internal variability. One may be interested in the impacts corresponding to each ensemble member, to look at how impacts might play out in conditions similar to the real world, where forced and internal variability always interplay. By looking at the ensemble of impact outcomes, one could quantify the relative role in the uncertainty in impacts due to either component. By emulating the variability of regional temperatures to approximate a large initial condition ensemble of temperature changes, we can obtain a more complete characterization of the impacts and the differences in impacts across scenarios. This methodology presents some promising features for avoided impact studies, and should be extended to more general assumptions, including non-linear mean trends and multivariate models with correlation among climate variables.

Acknowledgments The authors of this paper acknowledge the large contribution provided by the modeling groups to create the CESM model and to generate the initial condition ensembles used in this analysis. This work was supported in part by grants from the National Science Foundation (NSF). NCAR is managed by the University Corporation for Atmospheric Research under the sponsorship of the NSF. S. A. was supported by NSF Mathematical Sciences Postdoctoral Research Fellowship Award Number 1304321. C. T. was supported by the Regional and Global Climate Modeling Program (RGCM) of the U.S. Department of Energy's, Office of Science (BER), Cooperative Agreement DE-FC02-97ER62402.

References

- Beniston M, Stephenson D, Christensen O, Ferro CT, Frei C, Goyette S, Halsnaes K, Holt T, Jylhä K, Koffi B, Palutikof J, Schöll R, Semmler T, Woith K (2007) Future extreme events in European climate: an exploration of regional climate model projections. *Clim Chang* 81(1):71–95
- Bondeau A, Smith PC, Zaehle S, Schaphoff S, Lucht W, Cramer W, Gerten D, Lotze-Campen H, Müller C, Reichstein M, Smith B (2007) Modelling the role of agriculture for the twentieth century global terrestrial carbon balance. *Glob Chang Biol* 13(3):679–706
- Cabr  M, Solman SA, Nu ez MN (2010) Creating regional climate change scenarios over southern South America for the 2020's and 2050's using the pattern scaling technique: validity and limitations. *Clim Chang* 98(3–4):449–469
- Deryng D, Sacks WJ, Barford CC, Ramankutty N (2011) Simulating the effects of climate and agricultural management practices on global crop yield. *Glob Biogeochem Cycles* 25(2):GB2006
- Fitzmaurice GM, Laird NM, Ware JH (2011) Applied longitudinal analysis, 2nd Edition, Wiley
- Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteorol Soc* 90(8):1095–1107
- Hurrell JW, Marika M, Holland PR, Gent S, Ghan JE, Kay PJ, Kushner J-F, Lamarque (2013) The community earth system model: a framework for collaborative research. *Bulletin of the American Meteorological Society* 94(9):1339–1360
- Izaurrealde RC, Williams JR, McGill WB, Rosenberg NJ, Jakas MCQ (2006) Simulating soil C dynamics with EPIC: model description and testing against long-term data. *Ecol Model* 192(3–4):362–384
- Kay JE, Deser C, Phillips A, Mai A, Hannay C, Strand G, Arblaster, JM, Bates SC, Danabasoglu G, Edwards J, Holland M, Kushner P, Lamarque JWF, Lawrence D, Lindsay K, Middleton A, Munoz E, Neale R, Oleson K, Polvani L, Vertenstein M (2014 (to appear)) The community earth system model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*
- Maracchi G, Sirotenko O, Bindi M (2005) Impacts of present and future climate variability on agriculture and forestry in the temperate regions: Europe. *Clim Chang* 70(1–2):117–135
- Mitchell T (2003) Pattern scaling: an examination of the accuracy of the technique for describing future climates. *Clim Chang* 60(3):217–242
- NCAR CGD (2016) Climate Variability Diagnostics Package. Retrieved Tue Feb 2, 2016, from http://webext.cgd.ucar.edu/Multi-Case/CVDP_ex/CESM-CAM5-BGC-LE_1920-2014/
- Patz JA, Campbell-Lendrum D, Holloway T, Foley JA (2005) Impact of regional climate change on human health. *Nature* 438(7066):310–317
- Piao S, Ciais P, Huang Y, Shen Z, Peng S, Li J, Zhou L, Liu H, Ma Y, Ding Y, Friedlingstein P, Liu C, Tan K, Yu Y, Zhang T, Fang J (2010) The impacts of climate change on water resources and agriculture in China. *Nature* 467(7311):43–51
- Rosenzweig C, Elliott J, Deryng D, Ruane AC, M ller C, Arneth A, Boote KJ, Folberth C, Glotter M, Khabarov N, Neumann K, Piontek F, Pugh TAM, Schmid E, Stehfest E, Yang H, Jones JW (2014) Assessing agricultural risks of climate change in the twenty-first century in a global gridded crop model intercomparison. *Proc Natl Acad Sci* 111(9):3268–3273
- Ruosteenoja K, Tuomenvirta H, Jylh  K (2007) GCM-based regional temperature and precipitation change estimates for Europe under four SRES scenarios applying a super-ensemble pattern-scaling method. *Clim Chang* 81(1):193–208
- Salinger MJ (2005) Climate variability and change: past, present and future – an overview. *Clim Chang* 70(1–2): 9–29
- Schar C, Vidale PL, Luthi D, Frei C, Haberli C, Liniger MA, Appenzeller C (2004) The role of increasing temperature variability in European summer heatwaves. *Nature* 427(6972):332–336
- Tebaldi C, Arblaster JM (2014) Pattern scaling: its strengths and limitations, and an update on the latest model simulations. *Clim Chang* 122(3):459–471
- Watterson IG (2008) Calculation of probability density functions for temperature and precipitation change under global warming. *J Geophys Res Atmos* 113(D12):D12106
- Wilby RL (1997) Non-stationarity in daily precipitation series: implications for gcm down-scaling using atmospheric circulation indices. *Int J Climatol* 17(4):439–454