Emulating climate extreme indices

C Tebaldi, A Armbruster, H P Engler and R Link

1 Joint Global Change Research Institute, College Park, MD, United States of America
2 Georgetown University, Washington, DC, United States of America
E-mail: claudia.tebaldi@pnnl.gov

Keywords: extreme indices, emulation, scenarios, pattern scaling, time shift, internal variability, error metric

Abstract
We use simple pattern scaling and time-shift to emulate changes in a set of climate extreme indices under future scenarios, and we evaluate the emulators’ accuracy. We propose an error metric that separates systematic emulation errors from discrepancies between emulated and target values due to internal variability, taking advantage of the availability of climate model simulations in the form of initial condition ensembles. We compute the error metric at grid-point scale, and we show geographically resolved results, or aggregate them as global averages. We use a range of scenarios spanning global temperature increases by the end of the century of 1.5°C and 2.0°C compared to a pre-industrial baseline, and two higher trajectories, RCP4.5 and RCP8.5. With this suite of scenarios we can test the effects on the error of the size of the temperature gap between emulation origin and target scenarios.

We find that in the emulation of most indices the dominant source of discrepancy is internal variability. For at least one index, however, counting exceedances of a high temperature threshold, significant portions of the globally aggregated discrepancy and its regional pattern originate from the systematic emulation error. The metric also highlights a fundamental difference in the two methods related to the simulation of internal variability, which is significantly resized by simple pattern scaling. This aspect needs to be considered when using these methods in applications where preserving variability for uncertainty quantification is important.

We propose our metric as a diagnostic tool, facilitating the formulation of scientific hypotheses on the reasons for the error. In the meantime, we show that for many impact relevant indices these two well established emulation techniques perform accurately when measured against internal variability, establishing the fundamental condition for using them to represent climate drivers in impact modeling.

1. Introduction
Simple pattern scaling (pattern scaling from now on) (Santer et al 1990, Huntingford and Cox 2000, Tebaldi and Arblaster 2014, Herger et al 2015, Kravitz et al 2017) is widely used to emulate fields of future changes in climatic variables under arbitrary forcing scenarios. Output from general circulation models (GCMs) is used to train the emulator, using global mean temperature change as predictor variable. Since global mean temperature can be efficiently simulated by simple climate models, emulators can explore a wider range of scenarios than those available from GCM runs, a capability that proves valuable in characterizing impacts of climate change for a broad range of futures. A heightened attention to the role of additional sources of uncertainty besides scenarios, i.e. model structure and internal variability (Knutti et al 2010, Hawkins and Sutton 2009, Deser 2012) further increases interest in computationally efficient methods able to explore those uncertainty dimensions. Ideally these capabilities can be soft- or hard-coupled to impact models to robustly characterize risk, and even develop modeling frameworks where feedbacks between earth and human systems can be represented. At the same time, the role of extreme events and their future changes in causing significant impacts makes the development of emulators of extreme behavior a priority and with it the need for rigorous performance evaluation.

© 2020 The Author(s). Published by IOP Publishing Ltd
Several recent studies and assessments (Lustenberger et al 2014, Seneviratne et al 2016, Wartenburger et al 2017, Masson-Delmotte et al 2018) have shown that not only average temperature and precipitation, but also many indices of extremes appear to change in direct, often linear relation to global drivers (exp. global mean temperature) supporting the idea of building emulators for extremes driven by simple models like MAGICC (Meinshausen et al 2011) or Hector (Hartin et al 2015) (which through SCENGEND (Wigley 2008) and LEdgen (Link et al 2019) respectively already implement pattern scaling for average temperature and precipitation). Our aim in this paper is to consider pattern scaling and an alternative approach, time-shift (also known as time-sampling) (James et al 2017, King et al 2018, Herger et al 2015), and evaluate their performance in approximating a set of extreme indices. Both methods have been previously validated for average quantities (e.g. Tebaldi and Arblaster 2014, Kravitz et al 2017, Tebaldi and Knutti 2018) under increasing and stabilized scenarios, showing that the error introduced by the approximation is of comparable size to internal variability, and of much smaller size than the uncertainty introduced by models’ alternative structural choices (Tebaldi and Knutti 2018). Here we use as targets of the emulation ten indices among the Expert Team on Climate Change Detection Indices (ETCCDI) suite (Alexander 2016). We then test the accuracy of the approximations when the origin and target scenarios are relatively closer or farther apart within a range of trajectories that span radiative forcings from less than 2 W m$^{-2}$ to approximately 8.5 W m$^{-2}$ by 2100. We introduce a metric of emulator error that explicitly separates the portion that is unavoidable, coming from internal variability of the target quantity, from the portions that are artifacts of the emulation choice, i.e. the internal variability of the emulator output, and the proper emulator error. We further separate the two latter components, as only the emulator error needs minimizing. When this is done at geographically resolved scales, further analysis of the reasons for the emulator performances can be undertaken, zooming into specific regions of the globe and specific processes at play. Explicitly addressing internal variability is made possible by the use of a set of initial condition ensemble simulations, available for all future scenarios considered. The actual model output for the target scenario is taken to be the ground truth, against which our emulations are validated. A large source of variability that is not addressed in this study, however, is model uncertainty, the source of variability in model output that derives from each model’s structural choices, as we use ensembles of simulations from only one GCM. A multi-model ensemble, CMIP5 (Taylor et al 2012), was used in Tebaldi and Arblaster (2014), which found that pattern scaling produces more accurate emulations for temperature than for precipitation, but that the uncertainty in emulations of both is significantly smaller than the disagreement over the patterns among models. The same result was confirmed by Tebaldi and Knutti (2018) in the context of the emulation of highly mitigated scenarios. This study extends the analysis beyond average temperature and precipitation to ten climate extreme indices, using initial condition uncertainty as benchmark for accuracy, rather than inter-model variability. We note that inter-model variability for long-term changes is known to dominate internal variability (Hawkins and Sutton 2009), therefore a method found accurate when measured against the latter is also accurate against the former.

The paper is structured as follows. Section 2 describes the CESM ensembles and their characteristics, the ten ETCCDI indices and our error metric. Since the two emulation methods are well established and documented, we only briefly introduce them here, and we describe them in detail in the supplementary material (stacks.iop.org/ERL/15/074006/mmedia). Section 3 describes the results of the emulations: we give an overview of the performance over the suite of indices, and we then focus on a particular index for which the globally aggregated error metric and its regional patterns show a comparatively larger emulation error. We conclude in section 4 by summarizing the validity of these emulation methods when applied to indices of extremes, pointing at the promise of implementing these emulators into impact models. We also highlight the value of our metric for uncertainty quantification, in particular the possibility of identifying regions and processes where linear scaling and time-shift fail to perform.

2. Data and methods

2.1. Climate model output, scenarios, indices
NCAR-DOE CESM1-CAM5 (CESM) (Hurrell et al 2013) was used to simulate historical and future climate according to several pathways of future emissions, producing ensembles from perturbed initial conditions under each of the different scenarios. The so called large ensemble (LENS) was documented in Kay et al (2015). It simulates conditions under historical and RCP8.5 forcings, with initial conditions perturbed at 1920. A subsequent project, BRACE (O’Neill and Gettelman 2018), motivated the simulation of an alternative, lower-emission scenario, RCP4.5, for which the first 15 members of the historical simulations were used as starting future trajectories at 2005 (Sanderson et al 2018). After the Paris agreements spurred a special report by the IPCC (Masson-Delmotte et al 2018) it was decided to produce a further set of three ten-members ensembles that reached the 1.5 C and 2.0 C targets at the end of
the 21st century (Sanderson et al. 2017). For these low-warming scenarios the first ten members of the historical simulations of the LENS were used as starting points at 2005. From the first ten members of all 4 scenarios (1.5 C, 2.0 C, RCP4.5 and RCP8.5) we use daily output of maximum, minimum and mean temperature (TREFHTMX, TREFHTMN and TREFHT) and precipitation (PRECT) to compute ten extreme indices, which take the form of annual summaries of daily weather behavior.

Specifically, we target ten ETCCDI extreme indices (Alexander 2016). Five indices relate to the behavior of extreme precipitation: total annual precipitation when daily rainfall exceeds the 95th climatological percentile (r95ptot), maximum consecutive five day accumulated precipitation amount (rx5day), annual count of days when precipitation exceeds 10 mm (r10mm), average precipitation intensity (sdii) defined as the total precipitation amount during the year divided by the number of wet days (defined in turn as days with at least 1 mm of precipitation), and maximum number of consecutive dry days (i.e. with precipitation less than 1 mm) (edd). Five indices relate to the behavior of extreme temperatures: warm spell duration index (wspdi), defined as the longest stretch of days in the year with maximum temperature above the 90th climatological percentile; annual number of frost days (fd) when minimum temperature falls below freezing, growing season length (gsi) defined as the period between the last freeze in the spring and the first freeze in the fall, minimum temperature of the coldest day of the year (tnn), and maximum temperature of the hottest day of the year (ttx). The choice of these indices is somewhat arbitrary but aims at spanning a wide range of behaviors in temperature and precipitation, more or less extreme values, singular events or spells, cold and heat, dry and wet extremes.

2.2. Emulation methods

We use two different methods for approximating lower warming scenarios from higher ones. The first one, simple pattern scaling (Santer et al. 1996), depends upon the assumption that local climate can be modeled as a linear function of Global Surface Air Temperature (GSAT from now on), where the slope coefficient differs from grid-point to grid-point but is scenario independent. The field of such coefficients determines a normalized pattern that is used to provide an approximation of the actual field by multiplying by values of GSAT from the target scenario. We apply the method to each of the ten ensemble members available for any choice of origin-target scenario, for each of which we obtain one field of coefficients. Using each member separately let us explore the effects of internal variability.

We also test and compare a second emulation method, the time-shift approach (James et al. 2017). In order to approximate lower warming scenario A using higher warming scenario B, the method identifies a window of a number of years (here we use 10) when GSAT from origin scenario B is on average the same as GSAT from the target scenario A. Once those years from scenario B are isolated, the quantities of interest are extracted and used straightforwardly as surrogate for the same quantities under scenario A. We add details about both methods as supplementary material.

2.3. Error metric

Our model output consists of ensembles of 10 members obtained by perturbing initial conditions. We perform the emulation for each of these 10 separately, so that we end up comparing 10 emulated fields against 10 ground truth fields for each method. Note however that one should not expect a given ensemble member from the target scenario and the corresponding ensemble member from the origin scenario to be comparable. This is because the two future trajectories soon become independent of one another due to the short memory of the climate processes at play and the difference in forcings immediately after 2005. We now show how to distinguish internal variability (uncertainty that is present in the climate model output, and is thus treated as unavoidable) from systematic error, introduced in areas where pattern scaling or time-shift are not effective approximation methods. In addition, since the two emulation methods don’t propagate internal variability through the emulation in the same way, we can distinguish the variations introduced by the internal variability of the target data from that represented in the emulated quantities.

At any grid-point, consider the vectors \( y = (y_1, y_2, \ldots, y_K) \) of values from the target (truth) ensemble and \( \hat{y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K) \) from the emulation ensemble. In our case, \( K = L = 10 \). As usual we write \( \bar{y}, \bar{\hat{y}} \) for the ensemble averages of the \( y_i \) and \( \hat{y}_i \). The mean square difference between all possible pairs of emulated and true quantities is a measure for their closeness. It can be decomposed as follows:

\[
\frac{1}{KL} \sum_{i,j} (y_i - \hat{y}_j)^2 = ||\bar{y} - \bar{\hat{y}}||^2 + \langle (\bar{y} - \bar{\hat{y}})^2 \rangle + \langle (y - \hat{y})^2 \rangle. \tag{1}
\]

The angle brackets on the right denote averages that are computed over one of the two ensembles each, e.g. \( \langle (y - \hat{y})^2 \rangle = \frac{1}{K} \sum_i (y_i - \hat{y}_i)^2 \). A derivation of this decomposition is given in the supplementary material. As pointed out above, this decomposition holds at the grid-point level, but we can also think of it as a vector equation that holds at multiple locations. Only the first term in this decomposition represents a true emulation error. The second term is related to internal variability of the emulated quantity and
Figure 1. Both methods (circles for pattern scaling, diamonds for time-shift) and all indices (color coded according to legend), for all pairings of origin and target scenarios (as indicated at the top of the plot): The size of the emulation error scaled by the size of the internal variability of the target quantity (first component of our metric (2)) is shown as the y-axis dimension. The x-axis orders the scenario pairings according to the size of the GSAT gap between them at the end of the century. The metric is first computed at each grid-point of the land regions of the globe between 60S and 60 N, then aggregated with weights proportional to cosine of latitude.

the third term captures the internal variability of the true quantity. We therefore choose to define our error metric as the two-dimensional quantity

$$ Er(y, \hat{y}) = \left( \frac{|y - \hat{y}|}{\sqrt{(y - \bar{y})^2}}, \frac{\sqrt{(y - \hat{y})^2}}{\sqrt{(y - \bar{y})^2}} \right). \quad (2) $$

This means that we scale the two error components related to the emulation by the unavoidable variation of the target quantity.

Importantly, the first component of the metric should be as small as possible to reduce systematic errors, while the second should be close to 1 to show that the emulation does not inflate or reduce the variability that is in the target ensemble. Both components of the measures become unitless and we can compare the performance of the emulator across quantities with different units, magnitudes, and variability. This allows one to rank indices with respect to the faithfulness of emulation. We will assess which combinations of origin, target, and emulation method produce the most accurate emulations. In particular it is usually assumed that the closer the origin and target scenarios are, the better the performance of the emulation will be, and we will test this assumption here.

3. Results

3.1. Globally aggregated metric

Figures 1 and 2 summarize our results aggregated at a global scale (after computing the metric at each grid-point), separately for each component of the error metric (2). Each of the symbols in the plot corresponds to an index/method combination, with circles referring to pattern scaling results, diamonds referring to time-shift results. We used warm colors (yellow to red) for temperature dependent indices and cold colors (green to purple-blue) for precipitation-dependent indices. For these plots, however, the identification of the indices is only aimed at assessing if the relative performance of the methods across the index set remains substantially the same, i.e. the same index
is found easy (accurately emulated) by both methods, or found challenging (poorly emulated) by both methods, and the ranking from easy to challenging remains the same across the different origin-target combinations. We are plotting all the combinations of origin-target scenarios we have tested as a function of the gap in GSAT between the two scenarios at the end of the century (x-axis), expecting that the smaller the gap the better the emulation performance.

First, we note that for the large majority of indices and emulation choices, and for both methods, the value of the emulation error (figure 1, along the y-axis) is less than one, i.e. is less than the size of internal variability of the true quantity. The only exception is the index \textit{wsdi}, which quantifies changes in the duration of warm spells. The index appears to be particularly challenging for pattern scaling suggesting that the assumption of linearity adopted by pattern scaling is the source of the error. We will further explore this behavior in the next subsection. As for all other indices, we can assess that indeed the ranking with respect to difficulty of emulation remains substantially constant across methods and origin/target scenario pairs. Precipitation indices are usually at the lower end of the range, likely thanks to the relatively larger size of the internal variability of precipitation-based quantities (the denominator in definition (2)). The traditional assessments that find emulating precipitation more challenging than emulating temperature may therefore be the result of failing to distinguish discrepancies due to real emulation error from those due to internal variability. Figure 1 also shows that between the two methods pattern scaling is always at least equivalent, in most cases better than time-shift (except for \textit{wsdi}). Also interestingly, there is no dependency between the size of the emulation errors and the size of the gap in GSAT between scenarios. Our study is framed around the comparison of the two emulation methods, and as time-shift cannot approximate higher scenarios, only the emulation of lower scenarios from higher one can be assessed for both. In the supplementary material we show a similar plot to figure 1 for simple pattern scaling, where we evaluate emulating higher scenarios by lower ones. The larger magnitude of the errors, increasing with the GSAT gap between the origin and target scenarios confirms that it is advisable to emulate down, rather than up. As pattern scaling aims at

---

**Figure 2.** As in figure 1, but showing the second component of our metric (2), i.e. the size of the internal variability of the emulation, scaled by the size of the internal variability of the target. As in figure 1, the metric is first computed at each grid-point of the land regions of the globe between 60S and 60 N, then aggregated with weights proportional to cosine of latitude.
estimating the forced signal of change it is to be expected that using a high scenario facilitates its extraction, making the exercise more robust.

Figure 2 repeats the lay-out of figure 1, but shows the second component of the error metric, which should be close to 1. Here the two methods show a consistently different performance, with pattern scaling systematically underestimating, time-shift closely replicating the internal variability of the target. The former result can be explained by the definition of pattern scaling which, by construction, multiplies the normalized pattern by a constant that is less than one (the ratio of GSAT for the lower target scenario and the higher origin scenario) which dampens the variability of the patterns. Time-shift does not apply any rescaling, and, if anything its results suggest a slight overestimation of internal variability, which can be due to the use of scenarios with larger emissions to emulate lower emission targets. Again, in the supplementary material we show a similar plot for the results of applying pattern scaling to the emulation of higher scenarios from lower ones, and again we can assess the larger discrepancies between emulated and true internal variability, in this case by construction showing higher variability, as a factor greater than one is applied.
Figure 3 shows the comparison of the two methods for the emulation of the 1.5°C and 2.0°C target scenarios using RCP4.5 as the origin (results from other origin-target combinations are available as supplementary material) highlighting the relative contribution of the two terms of the error metric (blue/lighter for the internal variability term, purple/darker for the emulation error term, adding up to the total error as the height of each bar). A dashed line in each plot marks the unit line. Time-shift can be seen replicating a measure of internal variability closer to one (from above) than pattern scaling, whose blue bars are relatively farther from the unit line (from below). The index wsdi stands out with the larger portion due to emulation error, which remains for all other indices smaller than one. Note that more traditional metrics measuring emulation performance through root mean square errors (RMSEs) without distinguishing the two components would have used the total height of the bars in figure 3 without distinguishing the two components of each bar. In that case the message would have been substantially, but we argue deceptively, simpler with one of the methods, pattern scaling, clearly a winner showing lower RMSEs for all but one index. Having decomposed the error metric into a component that is the true emulation error and a component due to internal variability of the emulation results, we instead can interpret the results more subtly, connecting the differences in performance to a difference in the way the two methods reproduce internal variability.

Concluding this portion of the assessment we underline that the characteristics of smaller or larger internal variability in the emulation results are not necessarily a shortcoming or a benefit, depending on the application. However, since most applications of emulating extreme indices are likely used in impact models we would expect the representation of internal variability to be an important component of any risk assessment.

3.2. Geographically resolved metric

We now present a disaggregated view of the same results, examining geographical patterns of error. The global measures gave us a means to assess the overall emulation performance for the indices, and enabled a quick comparison among them. A spatial analysis can highlight geographic features specific to the various sources of error and variability, can give insight into the causes of inaccuracies, and can help identify those physical processes responsible for the emulation error. We use two indices showing very different outcomes for the two components of the error metric, rx5day, which measures the wettest 5 consecutive days of the year, and wsdi, already identified as the index with a large emulation error, which counts the longest spell in the year of warm days (i.e. days above
the estimated 90th percentile of the climatological distribution of local temperature). We illustrate the analysis using results from the emulation of the 1.5 C scenario (the target) based on RCP4.5 (the origin). In figure 4 for rx5day and 5 for wsdi we show 4 panels, 2 for each method (pattern scaling on the left, time-shift on the right), with the first metric component (the emulation error) on the top row, the second (the internal variability) on the bottom row. As is indicated by the corresponding bars in figure 3 for rx5day (second bar from the left in both top panels), figure 4 shows that both approaches produce a similarly small emulation error (significantly less than one when globally aggregated). Consistently with their behavior over other indices, pattern scaling estimates too little internal variability, while time-shift reproduces it accurately. Geographically, patterns are consistently and smoothly below 1 for the emulation error, and noisily below or around 1 for the internal variability component. No feature stands out that raises questions about regional effects worth investigating, at least at this fairly broad, bird’s eye perspective. In contrast, the same analysis for wsdi in figure 5 shows that both emulation methods find challenges in similar regions, particularly pattern scaling, with values of the emulation error well above 1 in large coherent regions of the subtropics. In lesser measure the same regions also appear in the error patterns of time-shift. Internal variability is slightly underestimated overall by pattern scaling, but significantly overestimated in many regions by time-shift. This is definitely an index that poses challenges to the emulation to a significantly higher degree than the rest of the ETCCDI suite. We already expected this from the size of the top portion of the bars for the index in figure 3, top row (last bar to the right in both panels) but now we can identify where the sources of the error are, geographically, and attempt formulating hypotheses for the causes of the errors.

The pattern of emulation errors in the top two panels of figure 5 identifies specific regions of the subtropics as the sources of the larger discrepancies between emulated and true index values. A map showing Net Primary Productivity of vegetation from the land component of CESM would identify many of the same areas as locations where broad-leaf forest grows (see figure 1 in the supplementary material) and one could be justified in speculating about the differential role of CO$_2$ fertilization effects on vegetation in the two scenarios (origin and target), possible surface-atmosphere interactions when more or less moisture is available from vegetation transpiration and therefore possible effects on precipitation and temperature in these regions. This is the type of hypothesis that could be used as a starting point for scientific investigation, further experiments, closer analysis of time series and spell behavior. That would be beyond the scope of this work, however, and we mention it only as an example of the type of analysis enabled by the close inspection of the regional behavior of the error. In the supplementary material we elaborate on the source of the error further.
Maps similar to those in figures 4 and 5 for the remaining 8 indices are available in the supplementary material, from which we can confirm that no emulation error presents itself as starkly as for the index wsdi. Specific applications that focus on limited regional areas may still find the decomposition eye-opening, and stimulating more in-depth error analysis.

4. Discussion and conclusions

Two emulation methods, simple pattern scaling and time-shift, already well established for the emulation of average quantities, are here tested for a set of indices of temperature and precipitation extremes. After introducing a new error metric able to distinguish true emulation error from discrepancies introduced by internal variability we find that for almost all these indices both methods perform accurately. This is a promising finding as we aim at representing more, and more impact relevant, climate drivers in our impact models, expanding on recent efforts identifying a broad range of hazards and their indices Mora et al (2018), Forzieri et al (2018). Having tested emulation performance on a broad range of indices we are confident that for most temperature and precipitation indices the implementation of their emulation will be successful, but hazards come from different weather and climate variables too, like winds, extreme sea levels, snow and ice storms etc for which the emulation is sure to present new challenges.

Our work is focused on the characteristics of the error metric as much as on the actual results of the emulation exercise: we show that there is value in distinguishing the true emulation error from the component of the discrepancy due to internal variability generated by the emulator. A traditional view of emulation error would seek a method that delivers an unbiased estimate (a small emulation error term in our metric) and is affected by small variance (a small internal variability term in our metric). The first requirement is unquestionable and in our application sees pattern scaling taking a comparative advantage over time-shift for all but one index. Contrary to the second requirement, for many uses of the emulated values an ability to recreate internal variability besides average behavior should in fact be desirable. If we embrace this point of view, the comparative advantage is reversed between the methods: simple pattern scaling shows too small variability while time-shift reproduces in most cases the variability of the target. Interestingly the closer analysis of time-shift internal variability suggests that the output under the higher scenarios appears in most cases not to show significantly different variability compared to the lower target scenarios at the time when global average temperature is similar. That also suggests that many of these quantities do not show significantly larger trends in the higher origin scenarios’ windows than in the lower target scenarios, likely because the variability of these extreme quantities mutes any trend that may be present in a ten year window. Along with these effects of internal variability we noticed that emulation of precipitation indices appears to benefit from the explicit accounting of its role: we are used to think of emulation of precipitation as relatively more challenging than that of temperature until we recognize that the larger variability should provide larger ‘tolerance’ for the emulation errors. We were also able to make a systematic comparison of the performance of the emulation methods as a function of the distance between origin and target scenarios, and found that there is no definite gain in choosing the closest scenario available to the target, especially within a small range of the gap size (on the order of a fraction of a degree) where the differential in performance is not robust.

Acknowledgments

This work was funded by the U.S. Department of Energy, Office of Science, as part of research in the MultiSector Dynamics, Earth and Environmental System Modeling Program. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. The views and opinions expressed in this paper are those of the authors alone. The initial analyses originate from the summer project of Alex Armbruster’s undergraduate internship at the National Center for Atmospheric Research. That portion, together with Claudia Tebaldi’s time at NCAR was supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy’s Office of Biological & Environmental Research (BER) via National Science Foundation IA 1947282.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID iDs

C Tebaldi  🌐https://orcid.org/0000-0001-9233-8903
R Link  🌐https://orcid.org/0000-0002-7071-248X

References

Alexander L V 2016 Global observed long-term changes in temperature and precipitation extremes: a review of progress and limitations in IPCC assessments and beyond Weather. Clim. Extremes. 11 4–16
Deser C, Phillips A, Bourdette V and Teng H 2012 Uncertainty in climate change projections: the role of internal variability Clim. Dyn. 38 527
Forzieri G, Bianchi A, Batista e Silva F, Marin Herrera M A, Leblois A, Lavalle C, Aerts J and Feyen L 2018 Escalating impacts of climate extremes on critical infrastructures in europe Glob. Environ. Change 48 97–107
Hartin C A, Patel P, Schwarber A, Link R P and Bond-Lamberty B P 2015 A simple object-oriented and open-source model for scientific and policy analyses of the global climate system vector 1.0 Geosci. Model. Dev. 8 939–55
Hawkins E and Sutton R 2009 The potential to narrow uncertainty in regional climate predictions Bull. Am. Meteorol. Soc. 90 1095–108
Herger N, Sanderson B M and Knutti R 2015 Improved pattern scaling approaches for the use in climate impact studies Geophys. Res. Lett. 42 3486–94
Huntingford C and Cox P M 2000 An analogue model to derive additional climate change scenarios from existing GCM simulations Clim. Dyn. 16 575–86
Hurrell J W, Holland M M and Gent P R 2013 The community earth system model: A framework for collaborative research Bull. Am. Meteorol. Soc. 94 1359–60
James R, Washington R, Schleussner C-F, Rogelj J and Conway D 2019 Characterizing half-a-degree-a-year: a review of methods for identifying regional climate responses to global warming targets Wiley Interdiscip. Rev. Clim. Change 8 e457
Kay J E, Deser C, Phillips A, Mai A, Hannay C, Strand G and Vertenstein M 2015 The community earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability Bull. Am. Meteorol. Soc. 96 1333–49
King A D, Knutti R, Uhe P, Mitchell D M, Lewis S C, Arblaster J M and Freychet N 2018 On the linearity of local and regional temperature changes from 1.5 °C to 2 °C of global warming J. Clim. 31 7495–7514
Knutti R, Furrer R, Tebaldi C, Cermak J and Meethal G A 2010 Challenges in combining projections from multiple climate models J. Clim. 23 2739–58
Krivitz B, Lynch C, Hartin C and Bond-Lamberty B 2017 Exploring precipitation pattern scaling methodologies and robustness among cmip5 models Geosci. Model. Dev. 10 1889–902
Link R, Snyder A, Lynch C, Hartin C, Krivitz B and Bond-Lamberty B 2019 Flгодen v1. 0: an emulator with internal variability and space-time correlation for earth system models Geosci. Model. Dev. 12 1477–89
Lustenberger A, Knutti R and Fischer E M 2014 Sensitivity of European extreme daily temperature return values to projected changes in mean and variance J. Geophys. Res.: Atmos. 119 3032–44
Masson-Delmotte V, Zhai P, Pörtner H, Roberts D, Skea J, Shukla P and et al 2018 IPCC (2018). Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty (Geneva: World Meteorological Organization)
Meinshausen M, Raper S C B and Wigley T M L 2011 Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 part 1: Model description and calibration Atmos. Chem. Phys. 11 1417–56
Mora C et al 2018 Broad threat to humanity from cumulative climate hazards intensified by greenhouse gas emissions Nat. Clim. Change 8 1062–71
O’Neill B and Gettelman A 2018 An introduction to the special issue on the benefits of reduced anthropogenic climate change (brace) Clim. Change 146 277–85
Sanderson B M et al 2017 Community climate simulations to assess avoided impacts in 1.5 and 2°C futures Earth Syst. Dynam. 8 827–47
Sanderson B, Oleson K W, Strand W G, Lehner F and O’Neill B 2018 A new ensemble of gcm simulations to assess avoided impacts in a climate mitigation scenario Clim. Change 146 303–18
Santer B D, Wigley T M L, Schlesinger M E, and Mitchell J 1990 Developing climate scenarios from equilibrium gcm results Report of the Max Plank Institut fur Meteorologie 47 29
Seneviratne S I, Donat M G, Pitman A J, Knutti R and Wilby R L 2016 Allowable CO2 emissions based on regional and impact-related climate targets Nature 529 477
Taylor K E, Stouffer R J and Meehl G A 2012 An overview of CMIP5 and the experiment design Bull. Am. Meteorol. Soc. 93 485–98
Tebaldi C and Arblaster J M 2014 Pattern scaling, its strengths and limitations and an update on the latest model simulations Clim. Change 122 439
Tebaldi C and Knutti R 2018 Evaluating the accuracy of climate change pattern emulation for low warming targets Environ. Res. Lett. 13 055006
Wartenburger R, Hirschi M, Donat M G, Greve P, Pitman A J and Seneviratne S I 2017 Changes in regional climate extremes as a function of global mean temperature: an interactive plotting framework Geosci. Model. Dev. 10 3609–34
Wigley T M L 2008 MAGICC/SCENGEN 5.3: User manual (version 2) (Boulder, CO: NCAR) 80 (http://www.ipcc.ch/site/assets/uploads/sites/2/2019/06/SR15_Full_Report_Low_Res.pdf)