AGU100 ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL079176

Key Points:
- Model-based estimates of changes in probability of temperature extremes at 1.5 degrees Celsius global warming are sensitive to the experimental setup
- Changes in odds of annual warm extremes in tropics more than 5 times larger in prescribed SST than fully coupled setup of same GCM
- Experimental design needs to be taken into account when interpreting projected changes in probability of extremes

Supporting Information:
- Supporting Information S1

Correspondence to:
E. M. Fischer,
erich.fischer@env.ethz.ch

Citation:
Fischer, E. M., Beyerle, U., Schleussner, C. F., King, A. D., & Knutti, R. (2018). Biased estimates of changes in climate extremes from prescribed SST simulations. *Geophysical Research Letters*, *45*, 8500–8509. https://doi.org/10.1029/2018GL079176

Received 12 JUN 2018
Accepted 31 JUL 2018
Accepted article online 13 AUG 2018
Published online 19 AUG 2018

Biased Estimates of Changes in Climate Extremes From Prescribed SST Simulations

E. M. Fischer[1], U. Beyerle[1], C. F. Schleussner[2,3,4], A. D. King[5], and R. Knutti[1]

[1]ETH Zurich, Institute for Atmospheric and Climate Science, Zurich, Switzerland, [2]Climate Analytics, Berlin, Germany, [3]Potsdam Institute for Climate Impact Research, Potsdam, Germany, [4]IRI THESys, Humboldt University, Berlin, Germany, [5]ARC Centre of Excellence for Climate Extremes, School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia

Abstract Large climate model ensembles are widely used to quantify changes in climate extremes. Here we demonstrate that model-based estimates of changes in the probability of temperature extremes at 1.5 °C global warming regionally differ if quantified using prescribed sea surface temperatures (SSTs) instead of using a fully coupled climate model. Based on the identical climate model used in two experimental setups, we demonstrate that particularly over the tropics and Australia estimates of the changes in the odds of annual temperature extremes can be up to more than a factor of 5 to 10 larger using prescribed SSTs rather than a fully coupled model configuration. The two experimental designs imply a different perspective on framing projections. If experiments conditional on prescribed observed SSTs are interpreted as unconditional real-world projections, they project changes in extremes that are systematically biased high and overconfident. Our results illustrate the importance of carefully considering experimental design when interpreting projections of extremes.

Plain Language Summary There is great interest in understanding the likelihoods and associated risks of potential future climate extremes, especially at the Paris Agreement global warming targets of 1.5 and 2 °C warming above preindustrial conditions. In this study, we assess the implications of the model setup for the quantification of changes in the odds of temperature extremes between different global warming levels. Our analysis illustrates the strong sensitivity in the outcomes of such analyses related to the use of different model experiments. We demonstrate that despite using the exact same global climate model the projected changes in the probability of extreme annual temperature anomalies for a climate consistent with a 1.5 °C warming target are in some cases much larger if sea surface temperatures are prescribed over a decade rather than if the model is run in a fully coupled configuration. If prescribed sea surface temperature experiments are interpreted as a projection for the real world at the end of the 21st century independent of ocean variability, they regionally lead to estimates of changes in extremes that are systematically biased high and overconfident. Our results illustrate the importance of carefully considering experimental design when interpreting projected changes in extremes.

1. Introduction

Changes in climate extremes at distinct levels of warming like the targets defined in the Paris Agreement are challenging to assess as the occurrence of extremes is rare by definition. Some studies used transient Coupled Model Intercomparison Project Phase 5 (CMIP5) multimodel experiments to quantify the occurrence of extremes at levels of 1.5 and 2 °C global warming (King et al., 2017; Schleussner et al., 2016). In addition, targeted model experiments have been designed as large initial condition ensembles, of both atmosphere-only (Mitchell et al., 2017) and coupled (Sanderson et al., 2017) simulations, sampling internal variability for different levels of warming. Thereby the signal-to-noise ratio can be maximized, which implies that the forced response of extremes can be determined despite large internal variability even for comparatively small levels of warming (Fischer et al., 2014). Running large ensembles further allows to better sample the tails of the distribution and to explore changes even in high return level extremes (Otto et al., 2012; Pall et al., 2011). However, due to finite computational resources there is a trade-off between producing large initial condition ensembles and (a) running at high resolution, which potentially allows to better resolve the atmospheric dynamics driving extremes (e.g., Schiemann et al., 2017) and to better represent precipitation extremes (Kopparla et al., 2013; Wehner et al., 2010), or (b) running at higher model complexity such as using more complex land surface models, or (c) running models over a global rather than a regional climate model domain, or (d) using a fully coupled atmosphere-ocean-sea-ice model instead of prescribing sea surface temperatures (SSTs) in a so-called Atmospheric Model Intercomparison Project (AMIP) configuration.

Particularly, in extreme event attribution, a common approach is to prescribe observed SSTs for a year or a decade and to compare it to a counterfactual climate, that is, an estimate of the *world that would have been* without human influence, constructed by removing the anthropogenic SST response (Massey et al., 2012; Pall et al., 2011; Schaller et al., 2016). Risser et al. (2017) argue that these experiments make the assumptions that (a) the anthropogenic influence on extreme weather is independent of the ocean state; (b) the nature of ocean variability, for instance, the frequency and spatial structure of El Niño events, is unaffected by anthropogenic emissions, beyond any prescribed mean response; and (c) short time scale atmosphere-ocean interactions are unimportant. In framing extreme event attribution analyses there is an important difference whether the attribution of an extreme event is tied to a specific state of the climate system such as the incidence of a strong El Niño anomaly. The attribution question would then be how human influence has changed the odds of a given extreme event conditional on the occurrence of this strong El Niño. AMIP experiments are the ideal setup to address this attribution question. However, if the attribution question is framed differently, for example, how anthropogenic influence altered the overall probability of a certain extreme event class for a given level of forcing or warming unconditional on the observed realization of ocean variability (Christidis et al., 2015; Fischer & Knutti, 2015; King et al., 2015a; Lewis & Karoly, 2013; Stott et al., 2004), fully coupled single or multimodel experiments are typically used. Thus, depending on the framing of the event attribution question, a certain type of model experiment is suited (Christidis et al., 2018; Stott et al., 2016). Particularly, for the tropics but also for the extratropics the experimental setup has been demonstrated to be crucially important for event attribution (Risser et al., 2017).

Likewise, AMIP and coupled experiments represent two alternative approaches for future projections of extremes, analogous to the two different attribution framings. One projects changes for a given level of warming conditional on prescribed SSTs (basically assuming that the same sequence of SST anomalies is repeated in the target period), and the other an unconditional projection for a given forcing consistent with the same level of warming. Prescribed SST experiments (in this case adding rather than subtracting a forced warming pattern) are attractive, since they allow to directly control distinct levels of warming. Such a targeted model experiment, called *Half a degree Additional warming, Projections, Prognosis and Impacts* (HAPPI) experiment was designed to assess and discriminate the 1.5 and 2 °C warming levels (Mitchell et al., 2017).

The direct comparison of prescribed SST 10-year time slice experiments and fully coupled projections is not straightforward, and many studies fail to discuss the implications of the choice of different model experimental designs for the framing of projections and the interpretation of the results. A series of studies have shown that overall the effect of ocean coupling is small for high atmospheric internal variability—sometimes referred to as weather noise (Chen et al., 2013). Regionally, however, there are differences as there is constructive or destructive interference between the SST variability and weather noise (Barsugli & Battisti, 1998; Chen et al., 2013). Deser et al. (2012) compared results from a fully coupled experiment with an experiment using a repeating climatological seasonal cycle and concluded that the range of seasonal trends is remarkably similar in the two setups, and significantly different only in the tropics. There is evidence particularly for the tropics that exceptional ocean states can affect even the remote occurrence of extreme temperature and rainfall anomalies (Arblaster & Alexander, 2012; Gershunov & Barnett, 1998; Meehl et al., 2007) as well as tropical cyclone activity (e.g., Camargo et al., 2007).

Here we analyze and discuss some of the consequences of the different model setups for assessing changes in extremes in a 1.5 °C world. To this end, we use the exact same global atmosphere and land surface model in two configurations.

2. Model Experiments

We analyze historical and RCP2.6 simulations from a 21-member initial condition ensemble covering the period 1950–2100. Fischer et al. (2013) performed with the Community Earth System Model (CESM) version 1.0.4 including the Community Atmosphere Model version 4 (CAM4) and fully coupled ocean, sea ice, and land surface components (Hurrell et al., 2013).

The prescribed SST experiment (AMIP) uses the exact same atmosphere and land surface model (see Table S1 in the supporting information for comparison of the two model setups). The experiments follow the HAPPI TIER1 protocol (Mitchell et al., 2017) and use prescribed observed SST and sea ice conditions for the period 2006–2015. The climate that is 1.5 °C warmer than preindustrial is represented by adding the CMIP5 multimodel mean SST response for an RCP2.6 scenario and using RCP2.6 forcing for the end of the 21st century. The AMIP experiment includes 500 members for two 10-year periods representing present-day and 1.5 °C conditions.

In the fully coupled model experiment the period 1995–2004 is used as present-day period to get the identical additional warming by the end of the 21st century as in the AMIP experiments. By using a slightly different reference period for the coupled CESM experiment the multimember global mean temperature difference with respect to the 1.5 °C warmer climate is almost identical. We do not expect that the slightly different present-day period affects any of our conclusions because in a coupled model the realization of variability for a given year is not expected to match the corresponding observed year. Due to sampling uncertainty of 21 members, we expect small differences between different choices of reference periods. However, they are not systematic and for the research question addressed here it is primarily important that the differences between the present-day and the 1.5 °C climate states are as similar as possible.

3. Results

The multimember mean global temperature difference between the present-day and 1.5 °C warming target in both the coupled (CESM-CAM4) and AMIP (CAM4-HAPPI) experiments is 0.64 and 0.65 °C, respectively (Figure 1a). Likewise, the spatial warming pattern is very similar in the two experiments (pattern correlation $r = 0.94$ and 0.98 land only). As expected, the SST response pattern in the coupled experiment is more heterogeneous than the prescribed multimodel mean response in AMIP and shows a cooling in the North Atlantic and substantial difference along the sea ice edge (Figures 1b and 1c). The differences in the warming patterns over land are mostly less than 0.1–0.2 °C, and thereby often in the range of less than 5–10% of the local warming signal. While the global multimember average warming is very similar in the two experiments, the spread of 10-year mean warming signals, calculated as the differences between all combinations of pairs of present-day and 1.5 °C climates across members, is much larger in the coupled (0.50–0.83 °C, two-sided 95% confidence interval) than in the AMIP experiments (0.62–0.69 °C). The reason is that by prescribing the SST the global mean temperature is strongly constrained as illustrated in Figure 1a (gray lines).

Here we address the question of how ocean coupling affects changes in the probability of extreme temperature anomalies. We use probability ratio (PR), often also referred to as risk ratio, a metric to quantify changes in the odds of extremes (Allen, 2003; Stott et al., 2004). The PR is defined as $PR = P_{1.5°C}/P_{PD}$, where P_{PD} is the probability of exceeding a certain extreme anomaly under present-day conditions and $P_{1.5°C}$ the probability of exceeding the same threshold in a 1.5 °C climate. To quantify $P_{1.5°C}$ and P_{PD}, we combine all years across all members for a given climate, that is, 5,000 (500 × 10) years for the AMIP and 210 (21 × 10) years for the coupled experiment and investigate threshold exceeding extremes. Note that using 21 members induces some sampling uncertainty, which is not systematic.

We find that for a given extreme European temperature anomaly threshold (Figures 2a and 2b, black line), the PR strongly differs between AMIP (Figure 2a) and coupled (Figure 2b) experiments. We calculate the 99th percentile of annual European average temperatures in the AMIP present-day experiment and use this as a fixed temperature anomaly threshold both in the coupled and AMIP experiment. Thereby the temperature threshold represents an extreme anomaly at every grid point and is identical in both experiments, which allows for a direct comparison. For this threshold the AMIP experiments yields a PR of 47.1. This implies that an annual temperature anomaly that on average is expected only once in 100 years in the reference period is expected to occur almost every second year (on average 47.1 times in 100 years) in a 1.5 °C climate. The coupled experiments instead yields a PR of 19.8, which is almost 2.5 times smaller than in the AMIP experiment ($PR = 47.1$) despite a similar regional mean warming (0.73 vs. 0.76 °C) (Figures 2a and 2b). The discrepancies are even larger for East Africa, where the AMIP experiments yields a PR of 60.4, which is more than 12 times larger than in the coupled experiment (Figures 2c and 2d). Again, the regional mean warming is only slightly higher in the AMIP than in the coupled experiment, a difference that cannot account for the major difference in PR. The

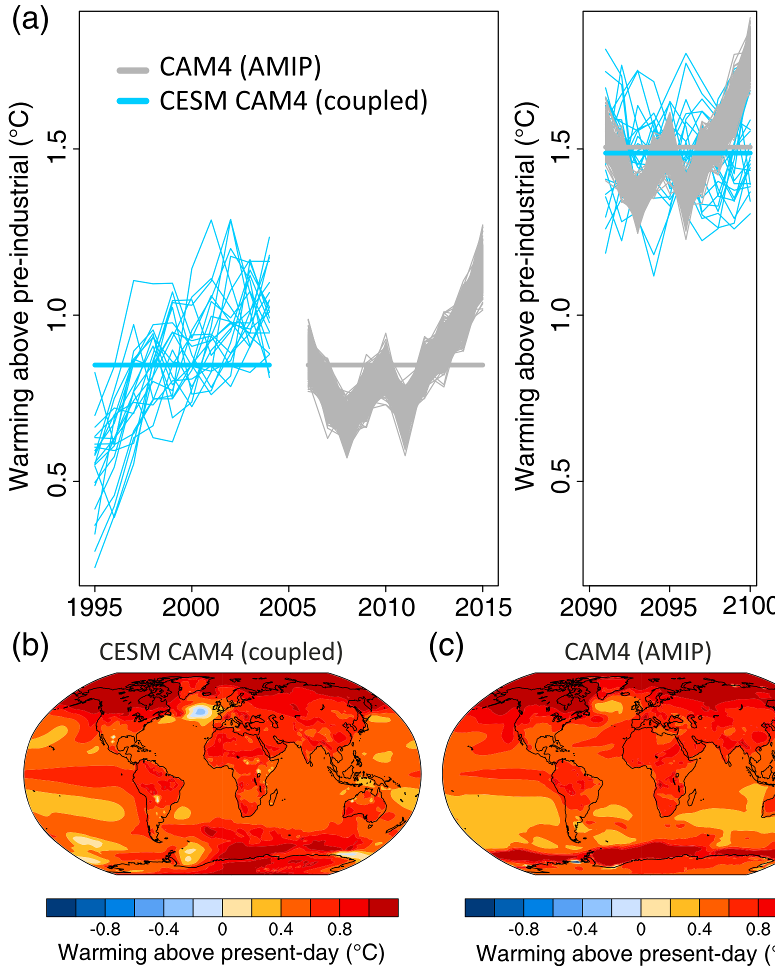

Figure 1. Warming in coupled and AMIP experiment. (a) Global annual mean temperature relative to preindustrial conditions in fully coupled CESM-CAM4 (light blue) and CAM4-AMIP experiment (gray) for the present-day period 1995–2004 and 2006–2015, respectively, and a 1.5 °C climate. (b and c) Multimember mean annual warming between present day and 1.5 °C in fully coupled and AMIP experiment. CAM4 = Community Atmosphere Model version 4; CESM = Community Earth System Model; AMIP = Atmospheric Model Intercomparison Project.

main factor is the variability that is much larger in coupled than in the AMIP experiment. Consequently, the exceedance probability P_{PD}, which by construction equals 1% in the AMIP experiments is much higher in the coupled experiments. As P_{PD}, the denominator in PR, is larger in the coupled experiments, the total estimate of PR is smaller. Figure 2 illustrates that the P_{PD} is larger in the coupled than AMIP experiment by almost the same factor as PR is smaller. This indicates that the lower present-day variability in the AMIP experiments mainly explains the substantially larger PR estimate. The exact differences in PR and P_{PD} needs to be interpreted carefully since there is a sampling uncertainty in the coupled experiments due to the availability of 21 members. However, the fact that there is a substantial difference between the two experimental setups is robust.

To get a more global picture of the relevance of these experimental discrepancies, we repeat the analysis shown in Figure 2 at every land grid point. According to the AMIP experiment annual temperature anomalies that are statistically expected only once in 100 years in present-day conditions increase by a factor of 5–15 over continental northern midlatitudes, 15–40 over parts of Europe and Southeast Asia and by an even higher PR over the tropics. Overall, the PR is substantially larger over the tropics than over the extratropics (Figure S1), which is consistent with a large body of literature (Fischer et al., 2012; Fischer & Knutti, 2015; Harrington et al., 2016; Hawkins & Sutton, 2012; King et al., 2015b; Mahlstein et al., 2011) demonstrating that the signal-to-noise ratio is higher over the tropics than midlatitudes primarily. Consequently, the part of the

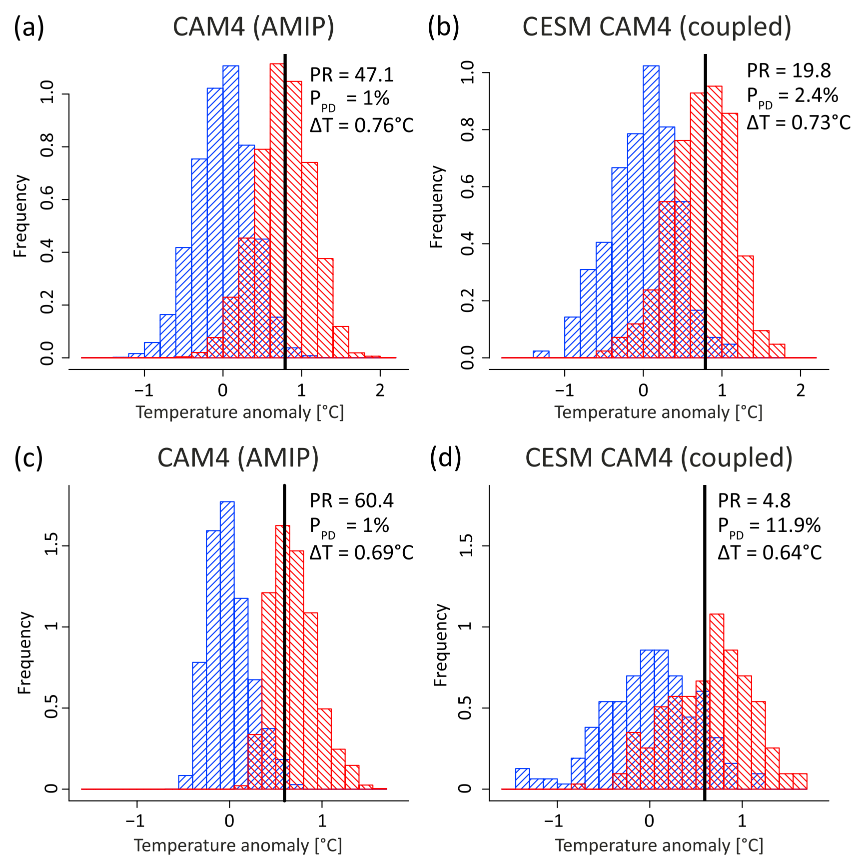

Figure 2. Probability ratios (PRs) for extreme annual mean temperature anomalies. Histogram of annual mean temperatures anomalies (relative to present-day multimember mean) averaged across (a, b) European land region (30–75°N, 10°W to 40°E) and (c, d) East Africa (22–52°E, 12°S to 18°N) for the present-day (blue) and a 1.5 °C climate (red). Histograms sample annual mean temperatures across (a, c) 500 × 10 years in the CAM4-AMIP experiment and (b, d) 21 × 10 years in the fully coupled CESM-CAM4 experiments. The black line marks the event anomaly used to calculate the PR and corresponds to the 99th percentile of the present-day period in the AMIP experiment. CAM4 = Community Atmosphere Model version 4; CESM = Community Earth System Model; AMIP = Atmospheric Model Intercomparison Project.

distribution that exceeds a certain high percentile for any given uniform warming shift of the temperature distribution is substantially larger.

Over most of the tropics including tropical Africa, most of Australia, the north of South America, and Southeast Asia the PR is much higher in the AMIP than in the coupled experiment (Figure 3a). The discrepancies are less pronounced yet still large over parts of western North America and western Europe, parts of the Arctic coasts of Eurasia and North America. On the other hand, the PR is somewhat smaller in the AMIP experiments over substantial areas of continental Eurasia and North America. Note that the two warming patterns are remarkably similar (Figures 1b and 1c) suggesting that, as discussed above for Figure 2, the regionally strong differences in PR arise mainly from variability differences rather than differences in the local mean warming between present day and 1.5 °C. Higher variability in the coupled experiment leads to larger P_{PD} for a given event magnitude and also to a somewhat lower absolute increase in exceedance for a given level of warming, two factors that both contribute to a lower PR for a given level of warming. On the other hand, where the AMIP experiment shows high interannual variability, P_{PD} is sometimes very low in the coupled experiment or even zero for some grid points in the continental climates of Eurasia and North America, meaning that the anomaly corresponding to the local 99th percentile of the AMIP runs is never exceeded in the coupled experiment and PR is undefined. As noted above the limitation to 21 coupled members induces some sampling uncertainty particularly in the tail of the distribution. This problem is alleviated when using the 95th percentiles as thresholds, which leads to lower PR values as documented for any given level of warming (see, e.g., Fischer & Knutti, 2015). Nevertheless, the pattern of the discrepancies between the PR

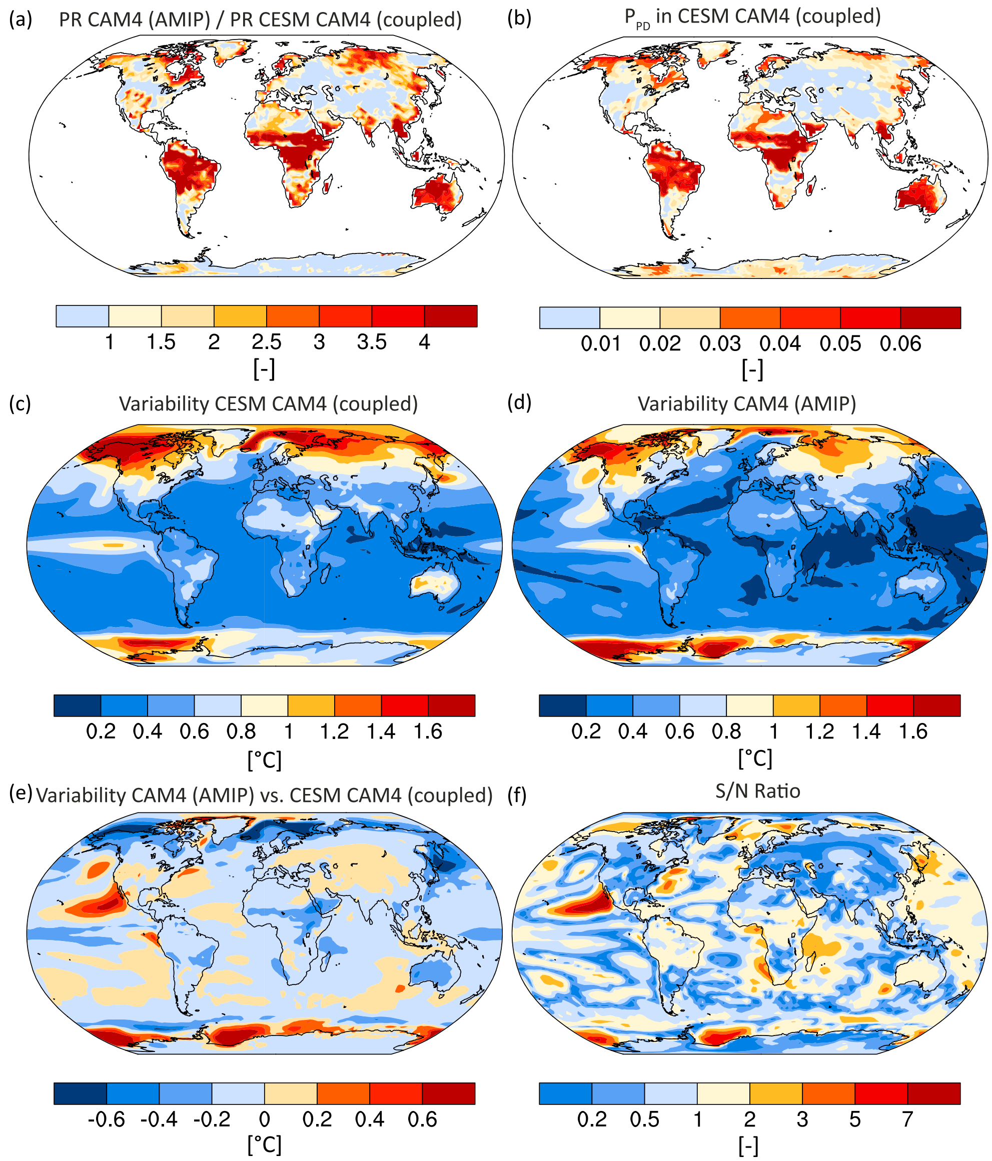

Figure 3. PR and variability in coupled versus uncoupled simulations. (a) Sensitivity of the probability to experimental setup illustrated as the ratio of PR derived from AMIP and fully coupled experiment. A factor of 2 suggests that the increase in hot extremes as a result of mean warming simulated in the AMIP is twice the one simulated in the fully coupled experiment setup. (b) Present-day exceedance of the extreme annual mean temperature anomaly CAM4-AMIP threshold but in coupled CESM-CAM4 simulations. (c, d) Interannual variability expressed as one standard deviation across local annual mean temperatures in all ensemble members. (e) Difference between variability shown in (c) and (d). (f) Ratio of difference (shown in e) and standard deviation across the interannual variability of the 10-year periods, that is, standard deviation across 21 members of the standard deviations across individual 10-year periods. . CAM4 = Community Atmosphere Model version 4; CESM = Community Earth System Model; AMIP = Atmospheric Model Intercomparison Project. PR = probability ratio.

pattern from the AMIP and coupled experiments remain the same (Figure S2). Likewise, at seasonal time scales for June–August the overall patterns look similar (Figure S3).

As discussed above, much of the difference in the PR estimates between AMIP and coupled experiments relates to the interannual variability of temperature, here expressed as one standard deviation of annual mean temperatures across all years and members (Figures 3c and 3d). Variability is substantially lower in the AMIP experiment over tropical land and ocean areas, along the sea ice edges and in the North Atlantic (Figure 3e). These are regions that experience substantial ocean and sea ice internal variability including low-frequency variability (Deser et al., 2012; Manabe & Stouffer, 1996) that is underestimated by prescribing 10 years of observed SST variability. It is worth noting that the HAPPI experiment assesses predominantly a negative phase of the Pacific Decadal Oscillation and includes a strong El Niño event only at the very end of the experiment. Therefore, HAPPI misses out on some of the variability of the leading mode over the tropical regions. On the other hand, there are areas such as the extended tongue in the eastern Pacific off the U. S. West Coast, off the Australian west coast, or in the western North Atlantic where variability is larger in the AMIP runs (Figures 3c and 3d). Potential reasons for this somewhat counterintuitive behavior are (a) that the prescribed observed years were a decade of unusually high variability, (b) the low-frequency variability is very small, or (c) that the coupled model exhibits systematically lower variability than the real world in general, or than observed in that particular decade. To quantify whether (a) or (c) are important factors, we compared the absolute value of the difference shown in Figure 3c with the standard deviation (across members) of the standard deviations across the 10 years within each of the members. The ratio of the two is a metric, similar to a signal-to-noise ratio, quantifying whether any of the members in the coupled model simulates a decade with local variability similar to the one in the prescribed observed decade. We find that the coupled experiment tends to systematically underestimate interannual variability off the North American west and east coasts, where the ratio takes values of 2–3 or more (Figure 3f). On the other hand, there are few areas where the coupled model seems to somewhat underestimate variability. Note that the exact magnitude of differences between AMIP and coupled experiments is model dependent, strongly depends on the magnitude of low-frequency variability and should be tested by other modeling groups. However, here we primarily want to highlight that there is a difference that in some cases may be very pronounced.

The discrepancies between coupled and AMIP experiments are substantially smaller for PR of daily (Figure S4) or multiday anomalies. Consistent with seasonal and annual anomalies, PR for daily temperature extremes are substantially higher in AMIP experiments over tropical Africa, South America, southern Asia, and Australia. Over large parts of North America, Eurasia, and southern Africa on the other hand, daily PR levels are equivalent or smaller in the AMIP experiment. The results suggests that for daily extremes over midlatitudinal land there may be a benefit of using AMIP experiments to increase the sample size and better sample tails of the distribution.

Computational costs are one main motivation to run AMIP instead of coupled experiments. Our findings suggest that there is a trade-off between sample size that can be enhanced with AMIP experiments and sampling of ocean internal variability that is not fully captured in AMIP experiments. Whether one or the other dominates depends on the area and event time scales of interest. For annual and seasonal mean anomalies and particularly for the tropics, and also for parts of the midlatitudes, the AMIP experiments tend to yield substantially larger PR than coupled experiments, which are consistent with findings in Risser et al. (2017). As outlined above, this difference is due to a different framing of the projections. The AMIP experiments yield projections conditional on the observed realization of ocean variability; that is, they quantify the change in the odds of extremes given that, for example, the exact same sequence of El Niño–Southern Oscillation would occur at end of the 21st century in a warmer background climate. This conditionality is rarely communicated, which leads to estimates in changes of extremes being systematically biased high and overconfident for unconditional projections for a specific decade at the end of the 21st century. The relative magnitude and significance level of the changes in temperature extremes in such experiments is larger than in the real world.

The AMIP setup on is based on the assumption that the nature of ocean variability, for instance, the sequence, intensity, and spatial structure of El Niño or Atlantic Multidecadal Variability, is unaffected by climate change beyond any prescribed mean response. While some findings suggest that this might not always be the case (Cai et al., 2014), further research is needed to rigorously test the validity of these assumptions and to understand (a) to what extent the characteristics of ocean variability are affected by climate change and (b) to what extent ocean-sea ice-atmosphere interactions need to be accounted for particularly during extreme anomalies. Regarding the overconfidence of the results, experiments prescribing SST responses from individual models following the HAPPI TIER2 protocol (Mitchell et al., 2017) would be beneficial.

4. Conclusions

There is great interest in understanding the likelihoods and associated risks of potential future climate extremes, especially at the Paris Agreement global warming targets of 1.5 and 2 °C warming above preindustrial conditions. In this study, we assessed the implications of the model setup for the quantification of changes in the odds of temperature extremes between different levels of warmings. Our analysis illustrates the strong sensitivity in the outcomes of such analyses related to the use of different model experiments. We demonstrate that despite using the exact same GCM the projected changes in the probability of extreme annual temperature anomalies for a climate consistent with a 1.5 °C warming target are in some cases much larger if SSTs are prescribed over a decade (AMIP configuration) rather than if the model is run in a fully coupled configuration. For the same extreme annual temperature anomaly, the AMIP experiments yield changes in the probability between a 1.5 °C warming target and present-day climate that are larger by a factor of 2.5 over Europe and even by a factor of more than 12 over East Africa. Even though the local warming is almost identical, the changes in the odds of extremes are much larger in the AMIP experiment particularly over tropical Africa, South America, Australia, and also over parts of Europe and North America. We demonstrate that this is due to the substantially higher variability in the coupled experiments, which leads to a higher probability of exceeding threshold under present-day conditions. The variability is smaller in the AMIP experiments since by prescribing only 10 years of SSTs, not all potential extreme El Niño or La Niña states or different phases of the Atlantic Multi-decadal Oscillation and Pacific Decadal Oscillation consistent with a certain level of warming are sampled. Thus, particularly over land areas that are directly affected by modes of ocean variability, the changes in the probability of extremes between two climate states may be substantially overestimated by using AMIP experiments. This could partly be overcome by prescribing SSTs for several decades like in the C20C+ experiments (Wehner et al., 2018) with the challenge that the reference period would then be highly transient in itself. Furthermore, since the prescribed SST response pattern is identical in all members, the AMIP experiments tend to be overconfident in the response leading to a too high significance level particularly in areas that are directly affected by SST and sea ice variability. An extension beyond prescribing the multimodel mean CMIP5 response (TIER1) and sampling the multimodel uncertainty in the SST response (TIER2) would help reducing overconfidence. Note that for extreme daily anomalies over continental middle to high latitudes the difference between the change in the odds of extremes are equivalent or even smaller in the AMIP experiments. This is primarily because, due to larger sample size, AMIP experiments better sample high-frequency variability in the atmosphere-land surface system, which is dominant in these regions.

The two experimental designs are also widely used in extreme event attribution, where they are suited to address distinctly framed attribution questions (Christidis et al., 2018). AMIP experiments allow to quantify human influence on the odds of extremes given the observed SST and sea ice conditions for a given year or decade (Pall et al., 2011; Schaller et al., 2016), whereas fully coupled transient experiments quantify the human influence between different states including all compatible realizations of variability (King et al., 2015a; Lewis & Karoly, 2013; Stott et al., 2004). The discrepancies between the setups are regionally large (Risser et al., 2017). Thus, recent efforts to compare different event attribution approaches for the same event (Otto et al., 2018; Uhe et al., 2016) are desirable and can ideally help disentangling the role of the different setups.

In terms of projections the interpretation of the two model setups is less straightforward. It is important to consider the strength and weaknesses of the setups already in the design phase and compare AMIP to fully coupled projections taking into account the different framing of the setups, as well as carefully interpreting the findings in light of the methodological choices made.

For locations and time scales where atmospheric variability dominates, an AMIP ensemble with a very large sample size is beneficial in allowing to assess very rare extremes. However, the fundamental assumptions that ocean and sea ice variability and its effect on atmospheric dynamics in midlatitudes remains unchanged need to be clearly stated and their validity need to be revisited in detailed process studies. For regions and time scales for which extremes are strongly affected by ocean variability, such as seasonal or annual anomalies in Australia or tropical Africa, a coupled model setup is more suitable for real-world projections for a given period at the end of the 21st century.

Moreover, it is important to take into account that in contrast to fully coupled transient projections, AMIP experiments yield projections conditional on the observed realization of SST in the reference period being exactly repeated at the end of the 21st century in a warmer future background climate. This is often poorly communicated. If AMIP experiments are instead interpreted directly as a projection for a given decade at the end of the 21st century independent of ocean variability, they regionally lead to estimates of changes in extremes that are systematically biased high and overconfident.

Acknowledgments
Andrew D. King received funding from the Australian Research Council (DE180100638 and CE170100023). Carl-Friedrich Schleussner acknowledges support by the German Federal Ministry of Education and Research (01LS1613A). The output of the AMIP simulations is accessible through http://portal.nersc.gov/c20c/data.html.

References

Allen, M. (2003). Liability for climate change. *Nature*, *421*(6926), 891–892. https://doi.org/10.1038/421891a

Arblaster, J. M., & Alexander, L. V. (2012). The impact of the El Nino-Southern Oscillation on maximum temperature extremes. *Geophysical Research Letters*, *39*, L20702. https://doi.org/10.1029/2012gl053409

Barsugli, J. J., & Battisti, D. S. (1998). The basic effects of atmosphere-ocean thermal coupling on midlatitude variability. *Journal of the Atmospheric Sciences*, *55*(4), 477–493. https://doi.org/10.1175/1520-0469(1998)055

Cai, W., Borlace, S., Lengaigne, M., Van Rensch, P., Collins, M., Vecchi, G., et al. (2014). Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change*, *4*(2), 111–116. https://doi.org/10.1038/NCLIMATE2100

Camargo, S. J., Emanuel, K. A., & Sobel, A. H. (2007). Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *Journal of Climate*, *20*(19), 4819–4834. https://doi.org/10.1175/Jcli4282.1

Chen, H., Schneider, E. K., Kirtman, B. P., & Colfescu, I. (2013). Evaluation of weather noise and its role in climate model simulations. *Journal of Climate*, *26*(11), 3766–3784. https://doi.org/10.1175/Jcli-D-12-00292.1

Christidis, N., Ciavarella, A., & Stott, P. A. (2018). Different ways of framing event attribution questions: The example of warm and wet winters in the UK similar to 2015/16. *Journal of Climate*. https://doi.org/10.1175/JCLI-D-17-0464.1

Christidis, N., Jones, G. S., & Stott, P. A. (2015). Dramatically increasing chance of extremely hot summers since the 2003 European heatwave. *Nature Climate Change*, *5*(1), 46–50. https://doi.org/10.1038/Nclimate2468

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*(3–4), 527–546. https://doi.org/10.1007/s00382-010-0977-x

Fischer, E. M., Beyerle, U., & Knutti, R. (2013). Robust spatially aggregated projections of climate extremes. *Nature Climate Change*, *3*(12), 1033–1038. https://doi.org/10.1038/nclimate2051

Fischer, E. M., & Knutti, R. (2015). Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature Climate Change*, *5*(6), 560–564. https://doi.org/10.1038/nclimate2617

Fischer, E. M., Oleson, K. W., & Lawrence, D. M. (2012). Contrasting urban and rural heat stress responses to climate change. *Geophysical Research Letters*, *39*, L03705. https://doi.org/10.1029/2011gl050576

Fischer, E. M., Sedlacek, J., Hawkins, E., & Knutti, R. (2014). Models agree on forced response pattern of precipitation and temperature extremes. *Geophysical Research Letters*, *41*, 8554–8562. https://doi.org/10.1002/2014GL062018

Gershunov, A., & Barnett, T. P. (1998). ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results. *Journal of Climate*, *11*(7), 1575–1586. https://doi.org/10.1175/1520-0442(1998)0112.0.Co;2

Harrington, L. J., Frame, D. J., Fischer, E. M., Hawkins, E., Joshi, M., & Jones, C. D. (2016). Poorest countries experience earlier anthropogenic emergence of daily temperature extremes. *Environmental Research Letters*, *11*(5). https://doi.org/10.1088/1748-9326/11/5/055007

Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*, L01702. https://doi.org/10.1029/2011gl050087

Hurrell, J. W., Holland, M., Gent, P., Ghan, S., Kay, J., Kushner, P., et al. (2013). The community Earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1

King, A. D., Donat, M. G., Fischer, E. M., Hawkins, E., Alexander, L. V., Karoly, D. J., et al. (2015b). The timing of anthropogenic emergence in simulated climate extremes. *Environmental Research Letters*, *10*(9). https://doi.org/10.1088/1748-9326/10/9/094015

King, A. D., Karoly, D. J., & Henley, B. J. (2017). Australian climate extremes at 1.5°C and 2°C of global warming. *Nature Climate Change*, *7*(6), 412. https://doi.org/10.1038/Nclimate3296

King, A. D., van Oldenborgh, G. J., Karoly, D. J., Lewis, S. C., & Cullen, H. (2015a). Attribution of the record high Central England temperature of 2014 to anthropogenic influences. *Environmental Research Letters*, *10*(5). https://doi.org/10.1088/1748-9326/10/5/054002

Kopparla, P., Fischer, E. M., Hannay, C., & Knutti, R. (2013). Improved simulation of extreme precipitation in a high-resolution atmosphere model. *Geophysical Research Letters*, *40*, 5803–5808. https://doi.org/10.1002/2013GL057866

Lewis, S. C., & Karoly, D. J. (2013). Anthropogenic contributions to Australia's record summer temperatures of 2013. *Geophysical Research Letters*, *40*, 3705–3709. https://doi.org/10.1002/grl.50673

Mahlstein, I., Knutti, R., Solomon, S., & Portmann, R. W. (2011). Early onset of significant local warming in low latitude countries. *Environmental Research Letters*, *6*(3). https://doi.org/10.1088/1748-9326/6/3/034009

Manabe, S., & Stouffer, R. J. (1996). Low-frequency variability of surface air temperature in a 1000-year integration of a coupled atmosphere-ocean-land surface model. *Journal of Climate*, *9*(2), 376–393. https://doi.org/10.1175/1520–0442(1996)0092.0.Co;2

Massey, N., Aina, T., Rye, C., Otto, F., Wilson, S., Jones, R., & Allen, M. (2012). Have the odds of warm November temperatures and of cold December temperatures in Central England changed. *Bulletin of the American Meteorological Society*, *93*, 1057–1059.

Meehl, G. A., Tebaldi, C., Teng, H., & Peterson, T. C. (2007). Current and future US weather extremes and El Niño. *Geophysical Research Letters*, *34*, L20704. https://doi.org/10.1029/2007gl031027

Mitchell, D., AchutaRao, K., Allen, M., Bethke, I., Beyerle, U., Ciavarella, A., et al. (2017). Half a degree additional warming, prognosis and projected impacts (HAPPI): Background and experimental design. *Geoscientific Model Development*, *10*(2), 571–583. https://doi.org/10.5194/gmd-10-571-2017

Otto, F. E. L., Massey, N., van Oldenborgh, G. J., Jones, R. G., & Allen, M. R. (2012). Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophysical Research Letters*, *39*, L04702. https://doi.org/10.1029/2011gl050422

Otto, F. E. L., van der Wiel, K., van Oldenborgh, G. J., Philip, S., Kew, S. F., Uhe, P., & Cullen, H. (2018). Climate change increases the probability of heavy rains in Northern England/Southern Scotland like those of storm Desmond—A real-time event attribution revisited. *Environmental Research Letters*, *13*(2). https://doi.org/10.1088/1748-9326/aa9663

Pall, P., Aina, T., Stone, D. A., Stott, P. A., Nozawa, T., Hilberts, A. G. J., et al. (2011). Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, *470*(7334), 382–385. https://doi.org/10.1038/nature09762

Risser, M. D., Stone, D. A., Paciorek, C. J., Wehner, M. F., & Angelil, O. (2017). Quantifying the effect of interannual ocean variability on the attribution of extreme climate events to human influence. *Climate Dynamics*, *49*(9–10), 3051–3073. https://doi.org/10.1007/s00382-016-3492-x

Sanderson, B. M., Xu, Y., Tebaldi, C., Wehner, M., O'Neill, B., Jahn, A., et al. (2017). Community climate simulations to assess avoided impacts in 1.5 and 2 °C futures. *Earth System Dynamics*, *8*(3), 827–847. https://doi.org/10.5194/esd-8-827-2017

Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., et al. (2016). Human influence on climate in the 2014 southern England winter floods and their impacts. *Nature Climate Change*, *6*(6), 627–634. https://doi.org/10.1038/Nclimate2927

Schiemann, R., Demory, M. E., Shaffrey, L. C., Strachan, J., Vidale, P. L., Mizielinski, M. S., et al. (2017). The resolution sensitivity of Northern Hemisphere blocking in four 25-km atmospheric global circulation models. *Journal of Climate*, *30*(1), 337–358. https://doi.org/10.1175/Jcli-D-16-0100.1

Schleussner, C. F., Rogelj, J., Schaeffer, M., Lissner, T., Licker, R., Fischer, E. M., et al. (2016). Science and policy characteristics of the Paris Agreement temperature goal. *Nature Climate Change*, *6*(9), 827–835. https://doi.org/10.1038/Nclimate3096

Stott, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J. P., van Oldenborgh, G. J., et al. (2016). Attribution of extreme weather and climate-related events. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(1), 23–41. https://doi.org/10.1002/wcc.380

Stott, P. A., Stone, D. A., & Allen, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature*, *432*(7017), 610–614. https://doi.org/10.1038/nature03089

Uhe, P., Otto, F. E. L., Haustein, K., van Oldenborgh, G. J., King, A. D., Wallom, D. C. H., et al. (2016). Comparison of methods: Attributing the 2014 record European temperatures to human influences. *Geophysical Research Letters*, *43*, 8685–8693. https://doi.org/10.1002/2016gl069568

Wehner, M., Stone, D., Shiogama, H., Wolski, P., Ciavarella, A., Christidis, N., & Krishnan, H. (2018). Early 21st century anthropogenic changes in extremely hot days as simulated by the C20C+ detection and attribution multi-model ensemble. *Weather and Climate Extremes*, *20*, 1–8. https://doi.org/10.1016/j.wace.2018.03.001

Wehner, M. F., Smith, R. L., Bala, G., & Duffy, P. (2010). The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model. *Climate Dynamics*, *34*(2–3), 241–247. https://doi.org/10.1007/s00382-009-0656-y