Lawrence Berkeley National Laboratory

Title
Early 21st century anthropogenic changes in extremely hot days as simulated by the C20C+ detection and attribution multi-model ensemble

Permalink
https://escholarship.org/uc/item/6x77g805

Authors
Wehner, Michael
Stone, Dáithí
Shiogama, Hideo
et al.

Publication Date
2018-06-01

DOI
10.1016/j.wace.2018.03.001

Peer reviewed

Early 21st century anthropogenic changes in extremely hot days as simulated by the C20C+ detection and attribution multi-model ensemble

Michael Wehner [a,*], Dáithí Stone [a], Hideo Shiogama [b], Piotr Wolski [c], Andrew Ciavarella [d], Nikolaos Christidis [d], Harinarayan Krishnan [a]

[a] Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[b] National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan
[c] Climate Systems Analysis Group, Environmental and Geographical Science, University of Cape Town, Rondebosch, South Africa
[d] Met Office Hadley Centre, Exeter EX1 3PB, United Kingdom

A B S T R A C T

We examine the effect of the 20th and recent 21st century anthropogenic climate change on high temperature extremes as simulated by four global atmospheric general circulation models submitted to the Climate of the 20th Century Plus Detection and Attribution project. This coordinated experiment is based upon two large ensembles simulations for each participating model. The "world that was" simulations are externally forced as realistically as possible. The "world that might have been" is identical except that the influence of human forcing is removed but natural forcing agents and variations in ocean and sea ice are retained. We apply a stationary generalized extreme value analysis to the annual maxima of the three day average of the daily maximum surface air temperature, finding that long period return values have been increased by human activities between 1 and 3 °C over most land areas. Corresponding changes in the probability of achieving long period non-industrial return values in the industrialized world are also presented. We find that most regions experience increases in the frequency and intensity of extremely hot three day periods, but anthropogenic sulfate aerosol forcing changes locally can decrease these measures of heat waves in some models.

1. Introduction

The hottest day of the year has become a standard metric of anthropogenic climate change for both observations and simulations. As one of simplest of the 27 indices defined by the Expert Team on Climate Change Detection Indices (ETCCDI), it is a well behaved indicator of extremely hot weather (Zhang et al., 2011). A sample of the hottest days of the year (denoted herein as *TXx*, following the ETCCDI notation) is drawn from a larger parent sample of hot season daily maximum temperatures.[1] The length of the hot season and hence the size of the parent sample varies depending on location. In the tropics, nearly any day of the year could be the annual maximum but in the mid-latitudes only about 100 days of the year ranging from the late spring to the early fall can be the hottest of the year. From a statistics point of view, *TXx* is a "block maximum" (Coles, 2001) and as the effective size of each block is quite large, *TXx* is well into the tail of the overall distribution. Since it is generally in the "asymptotic regime" of the parent distribution, the distribution of *TXx* can be described by the well-known Generalized Extreme Value (GEV) distribution (Coles, 2001). Long period return values are readily obtained after the three parameters describing the GEV distribution are determined by a fitting procedure.

Surface air temperature was one of the earliest indicators of anthropogenic climate change subject to rigorous "detection and attribution" analyses, and the human influence on its increasing trend since the early 20th century is well established (Houghton et al., 1996). Likewise, *TXx* exhibits a significant increasing trend in both observations and simulations of the recent past. Formal detection and attribution of the human influence on observations of *TXx* has been demonstrated using individual climate models (Christidis et al., 2011, 2014) and the models from the Coupled Model Intercomparison Project (Zwiers et al., 2011; Min et al. 2013; Kim et al., 2016).

Projections of changes in 20 year return values of *TXx* present an alarming picture of very significant increases in severe heat waves with extreme temperature changes substantially larger than projected changes in annual and summer average temperatures (Collins et al., 2013; Kharin et al., 2007, 2013). In these studies, the choice of 20 years as a long period was made due to data constraints. The CMIP models are of limited ensemble size, ranging from 3 to 10 independent realizations of each emissions scenario. Because they are non-stationary due to anthropogenic forcing trends, the authors in these studies used time itself as a

[1] Note: *TXx* is also known as "hot days" because it is defined as the annual maximum of the warmest temperature of the day. Its CMIP5 name is "*tasmax*".

linear covariate (Coles, 2001). Due to the realization that trends may be nonlinear in time, the analyses were confined to two decades. Even with the expansion of the parent datasets by the full use of the available ensembles, extreme sample sizes were still relatively small for some CMIP models.

2. Data and methods

In this study, we examine hot days in the "Climate of the 20th Century Plus Detection and Attribution project" (Stone et al., 2017a). In this project, the experiment design involves the integration of global atmospheric models with prescribed sea surface temperature (SST) and sea ice concentration thus avoiding the computational cost of lengthy full coupled ocean-atmosphere model spin-up simulations. Instead, computational power is directed towards producing larger ensembles than the CMIP simulations. As detailed in Stone et al. (2017a), the C20C+ D&A project numerical experiment consists of two parts. The first, called "All-Hist/est1", is essentially an "AMIP" experiment (Gates, 1992) of the "world that was", where the atmospheric composition, specified SST and sea concentrations are determined from observational products (Stone et al., 2017a). The second part, called "Nat-Hist/CMIP5-est1″, is a counterfactual AMIP experiment of the "world that might have been" with the atmospheric composition, SST and sea concentration modified to remove the anthropogenic effects (Stone and Pall, 2017). SST is modified by subtracting the difference of "historical" and "natural historical" CMIP5 simulations from the observed values in these experiments. Other choices of counterfactual SST distributions are described elsewhere in this special issue (Stone et al., 2017b). Sea ice concentration is modified to maintain consistency with the new SSTs. This is one of many approaches to probabilistic extreme event attribution (National Academies of Sciences, Engineering, and Medicine, 2016) and is designed to aid in separating anthropogenic and natural factors to event likelihood (Pall et al., 2014). The construction of the counterfactual SST and sea ice concentrations in the C20C+ D&A project experiment preserves the timing of observed natural variations in the climate system. For instance, the large El Niño event of 1998 is prescribed in both parts of the experiment. However, since the anthropogenic warming is removed in the counterfactual part, SSTs are somewhat cooler everywhere.

Impacts of heat waves on human health increase as their duration lengthens (Nitschke et al., 2011). Tebaldi and Wehner (2016) used the large ensemble of the Community Earth System Model (Kay et al., 2015) to compare changes in single hot day extrema to changes in the extrema of the running 3 day average of *TXx*, which we denote here as *TX3x*. They found that although 20 year return values of the 3 day averages were slightly less than for daily quantity, *TXx*, 21st century changes of the 3 day averages were considerably larger.

We analyze changes in *TX3x* from four different models that contributed to both parts of the C20C+ D&A project experiment.[2] Model #1 is a version of the UK MetOffice atmospheric model contributed by the University of Cape Town, South Africa and is denoted as "UCT-CSAG HadAM3P-N96" (the configuration is similar to that described in Wolski et al. (2014). Model #2 is a more recent version of the UK MetOffice atmospheric model and is contributed by the UK Met Office and denoted as "MOHC HadGEM3-A-N216" (Ciavarella et al., 2017). Model #3 is the MIROC5 model contributed by the National Institute for Environmental Studies, Tsukuba, Japan and denoted as "MIROC5" (Shiogama et al., 2013, 2014). Model #4 is the finite volume dynamical core version of the Community Atmosphere Model contributed by Lawrence Berkeley National Laboratory and denoted as LBNL fvCAM5.1 (Stone et al., 2017a). One of these model output datasets (LBNL fvCAM5.1) together with the similar design ensemble of MRI-AGCM3.2 (Mizuta et al. 2017; Imada et al., 2017) have been previously analyzed to reveal that the chances of

---

[2] All model output used in this study is available at http://portal.nersc.gov/c20c/.

record setting single day temperature minima and maxima and single day precipitation maxima have been altered by the human changes to the climate system (Shiogama et al., 2016). This study expands on Shiogama et al. (2016) by expanding the analysis from two to four of the C20C+simulation datasets and focusing on the high tails of the distribution of temperature maxima rather than the center of the distribution. Model details are provided in Table 1.

As in Tebaldi and Wehner (2016), the available ensemble sizes from the C20C+ D&A experiment are quite a bit larger than from the CMIP5 project. Hence, in this study we draw annual maxima over a short period from each realization for each model. Pooling these block maxima over both years and ensemble members provides relatively large extreme value datasets. The underlying parent data set that the block maxima are drawn from varies according to location. For instance, in the mid-latitudes, there are about 100 days in the summer when the annual maximum of the daily maximum or its running 3 day average temperatures could occur. Slight differences in these periods between models were used to accommodate differences in available data. Because the pooling period is short, a quasi-stationary assumption is made (Wehner, 2004) with the further assumption that extrema across years are independently and identically distributed (i.i.d.). This latter assumption may not strictly hold everywhere (Risser et al., 2017), especially in the tropics. With these assumptions about the nature of the extremal sample, the L-moments fitting procedure for determining the three parameters of the GEV distribution has been demonstrated to work well (Hosking and Wallis, 1997). The large size of the extreme value datasets, shown in the fifth column of Table 1, ensures that the uncertainty from the fitting is small. If the ensemble sizes were not as large, extrema could have been drawn from longer periods and a non-stationary GEV distribution fitted by maximum likelihood methods as an alternative strategy but with the additional uncertainty associated with fitting more parameters. If ensemble sizes were considerably larger, individual years could be analyzed separately without concerns about stationarity since the state of the ocean is prescribed in the C20C+ experiment. The particulars of the details of the GEV analysis used in this study are described in the Supplementary material of Tebaldi and Wehner (2017).

A principal difference between this study and that of Tebaldi and Wehner (2017) stems from the prescribed ocean and sea ice states of the C20C+ models versus the fully prognostic ocean and sea ice of such coupled models. Indeed, most previous GEV analyses of simulated temperature extrema have been performed on fully coupled model output (eg. Kharin et al., 2013). The additional degrees of freedom from prognostic ocean and sea would add to the uncertainty in all projected quantities including extreme temperatures. However, for long period extreme temperature return values this internal uncertainty is considerably less than the model structural uncertainty described in part by the variance across different climate models (Wehner 2010).

3. Results

Once the three parameters of the stationary GEV are determined at each grid point separately for the factual and counterfactual simulations, long period return values are straightforwardly determined and their changes calculated. Fig. 1 shows the difference in the 20 year return value of the 3 day average of annual maximum *TXx* between the All-Hist/est1 and Nat-Hist/CMIP5-est1 simulations for the four C20C+ D&A models. All calculations were performed and are shown on the native model grids, but results should be qualitatively similar to the use of slightly larger regions (Angélil et al., 2017b). Uncertainties in these return value differences due to the fitting procedure were calculated via a bootstrap procedure (Hosking and Wallis, 1997) but found to be miniscule due to the large number of realizations as in Tebaldi and Wehner (2016). However, as is evident in Fig. 1, the results between the four climate models vary substantially including even the sign of the differences between the factual and counterfactual simulations. Outside of Antarctica, three of the models (UCT-CSAG HadAM3P-N96, MOHC

Table 1

Details of the C20C+ D&A models used in this study. The number of realizations is for each part (All-Hist/est1 and Nat-Hist/CMIP5-est1) of the numerical experiment separately as used in this study. For some individual years, additional realizations may be available. The two rightmost columns shows the globally averaged difference between the All-Hist and Nat-Hist hot season temperature and the 20 year return value of the 3 day average daily maximum surface air temperature over land. "Hot season" is defined as the maximum of JJA and DJF averages.

Model	Resolution (#lat X #long)	Number of realizations (All-Hist/Nat-Hist)	Analysis period	Extreme value dataset size	Global land average change in hot season mean temperature (°C)	Global land average change in very extreme temperature (°C)
UCT-CSAG HadAM3P-N96	145X192	50/50	2008–2012	300	1.30	1.29
MOHC HadGEM3-A-N216	324X432	15/15	2002–2011	150	1.11	1.41
MIROC5	128x256	50/50	2011–2015	250	1.03	0.99
LBNL fvCAM5.1	192x288	100/100	2011–2015	500	1.16	1.33
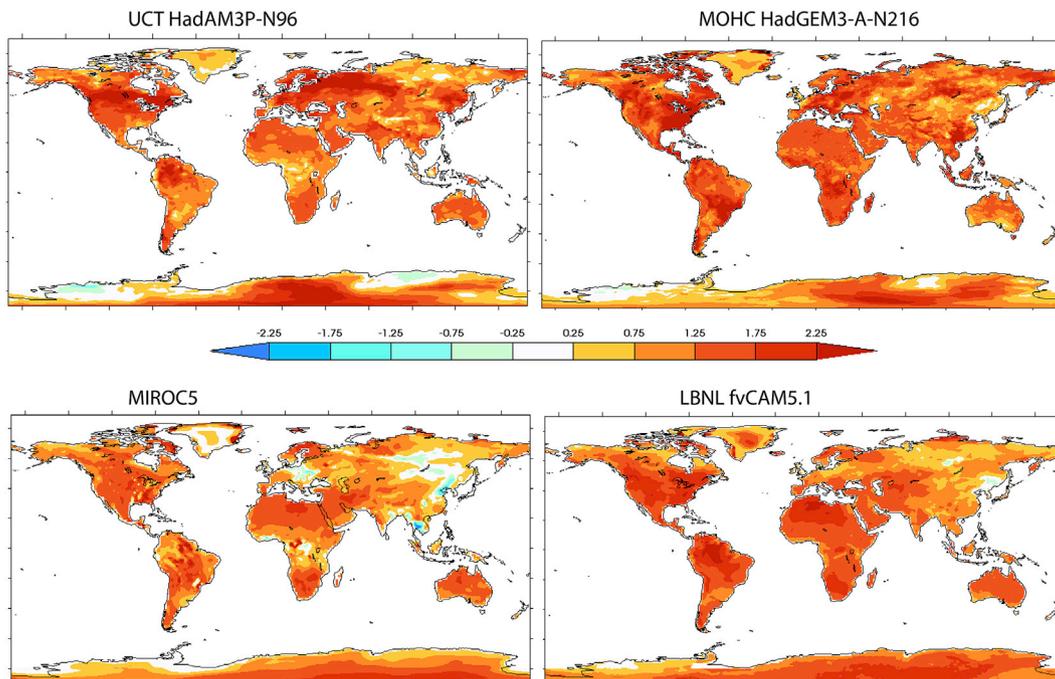
Fig. 1. Change in 20 year return values (°C) between the actual and counterfactual C20C+ D&A simulations. Upper left: UCT-CSAG HadAM3P-N96. Upper right: MOHC HadGEM3-A-N216. Lower left: MIROC5. Lower right: LBNL fvCAM5.1.

HadGEM3-A-N216 and LBNL fvCAM5.1) produce warmer extreme temperatures everywhere when the human changes to the climate system are considered. However, MIROC5 has several spatially coherent areas of statistically significant decreases.

These large return value decreases in the MIROC5 simulations are most likely due to the methods used to treat sulfate aerosols. All four models vary the sulfate aerosol concentration in the factual simulations but not in an identical manner. Most importantly, MIROC5 and HadGEM3-A-N216 use a prognostic aerosol formulation with prescribed emissions, whereas the other two models use prescribed aerosol concentrations based on offline simulations (and HadAM3P-N96 retains anthropogenic aerosols in its Nat-Hist/CMIP5-est1 simulations). A prognostic formulation allows the aerosol concentration to directly interact with the simulated weather. During a heat wave, air masses often stagnate. Hence, daily aerosol concentrations can become much higher than the monthly average especially in the lower portions of the atmosphere. Such localized air pollution is highly reflective to solar radiation and can result in cooler temperatures than would occur under equivalent clear sky conditions. Return value decreases in the MIROC5 model are most pronounced in eastern China, and eastern Europe due to this effect. Similar decreases in the Congo and southeast Asia are likely the result of aerosol emissions due to fires (H. Shiogama and O. Angélil, personal communications).

The LBNL fvCAM5.1 simulations also exhibit a region of cooling in *TXx* in eastern China but of much smaller extent than the MIROC5 simulations. The LBNL fvCAM5.1 was deliberately configured with prescribed bulk aerosol concentrations of Kiehl et al. (2000) generated from previously performed interactive calculations to save compute resources. In these simulations, daily atmospheric aerosol concentrations are interpolated from monthly climatologies and are smoothly varying in time with no relationship to the actual simulated meteorological conditions. Hence, in the temporarily stagnant air masses associated with some extreme temperature events, the aerosol concentrations would lower than if an interactive treatment were used. Nonetheless, the combined direct and indirect cooling effects of increased atmospheric aerosols in the LBNL fvCAM5.1 are still large enough to cause limiting cooling in *TXx* in the actual simulations compared to the counterfactual simulations. UCT-CSAG HadAM3P-N96 used the same time-invariant aerosol concentrations for both the actual and counterfactual simulations, resulting in no localized areas of cooling. Aerosols in the MOHC HadGEM3-A-N216 were calculated with a prescribed emissions protocol (Martin et al., 2006). Although this model shows a hint of a cooling in very small portions of eastern China, it is less than the cooling in both MIROC5 (with prescribed emissions) and LBNL fvCAM5.1 (prescribed concentrations) demonstrating the sensitivity to details of the aerosol formulation.

As mentioned above, the reference periods are defined slightly differently for the models shown in Fig. 1 in order to reduce uncertainty in GEV parameter estimation. However, this could add to the model structural uncertainty due the known influences of the ocean surface state on extreme temperatures over land. We examined this by repeating the analysis of LBNL fvCAM5.1 for the longest period, 2002-2015 (not shown). Despite differences in the 20 year return values averaged over the periods, the average changes from the Nat-Hist to the All-Hist configurations are remarkably independent as to the choice of reference period. This is a consequence of the careful construction of the prescribed Nat-Hist sea surface temperatures which preserve the patterns of the All-Hist values.

The change in the probability of an extreme event of an estimated magnitude is often used to convey the anthropogenic effect on actual individual extreme events (Stone and Allen, 2005). A "risk ratio" (actually a ratio of probabilities) can be simply defined as

$$RR = \frac{P_a}{P_c} = \frac{RT_c}{RT_a}$$

where P_n is the estimated probability of an event of a given magnitude in the counterfactual world that might have been and P_a is the estimated probability of an event of that same magnitude in the actual world that was. As the probability of an event of a certain magnitude occurring at least once in a given year is simply the inverse of the return time (RT) of that magnitude, the evaluation of risk ratio from the fitted GEV distributions used to calculate the return value changes in figure is straightforward. Fig. 2 shows the natural logarithm of the risk ratio for simulated counterfactual 20-year return values of daily maximum temperatures occurring in the actual world simulations. In this calculation, RT_c is by definition 20 years and RT_a is calculated using the fitted GEV distributions from the actual world simulations and the estimated counterfactual 20-year return values. In our calculations, we limited RT_a to be no less than 1 year, hence the maximum risk ratio is 20. Positive values of $ln(RR)$ indicate that the anthropogenic influence on the climate system increased the likelihood of counterfactual extreme temperatures, while negative values indicate that the likelihood is decreased. If there were no

influence, the two probabilities would be equal leading to a risk ratio of unity and its natural logarithm would be zero. Not surprisingly the differences in this risk ratio between models is similar to the differences in the return value changes shown in Fig. 1.

The larger ensemble size (~100) of the LBNL fvCAM5.1 model permits evaluation of risk ratio for longer counterfactual return periods without significant increase in uncertainty from the fitted GEV distribution. Fig. 3 shows the changes in daily maximum temperature return values from this model for 20, 50, 100 and 200-year return periods. These changes are essentially the same for all of these long period return values indicating that there is little change in the shape of the far tails of the fitted GEV probability density distributions. Fig. 4 shows the associated risk ratio for this selection of counterfactual return values indicating substantial increases in risk ratio as event magnitudes become rarer. In areas such as the Sahara, the risk ratio is at or near its maximum value, $1/RT_c$. However, even in most other regions where the counterfactual extremes occur less frequently that once per year in the actual simulations, the risk ratio also increases with event rarity.

Otto et al. (2012) pointed out that small changes in temperature return value and large changes to risk ratios are not inconsistent. In fact, the behavior shown in Figs. 3 and 4 is a consequence of the strong boundedness of the fitted GEV cumulative distributions characteristic of short term temperature extrema. Fig. 6 shows a typical plot of return value as a function of return period for the fitted GEV distribution of a representative mid-latitude land grid cell from the two C20C+ D&A configurations. The dotted red vertical line, representing the 20 year return period, intersects the counterfactual (red) return time curve at a value of about 312.1 K. The dotted black horizontal line, drawn from that temperature value, intersects the actual (black) curve at a return period of approximately 2.5 years, resulting in a risk ratio of approximately 8 for this value of temperature. At larger return periods, return values increase asymptotically towards the bounds determined by the fitted GEV distributions. As a result, the difference in return values also tends towards an asymptotic magnitude as seen in Fig. 3. (The short blue lines, representing differences at 20, 50, 100, and 200 year return periods are all roughly equal at about 2 K.) However, return periods in the actual
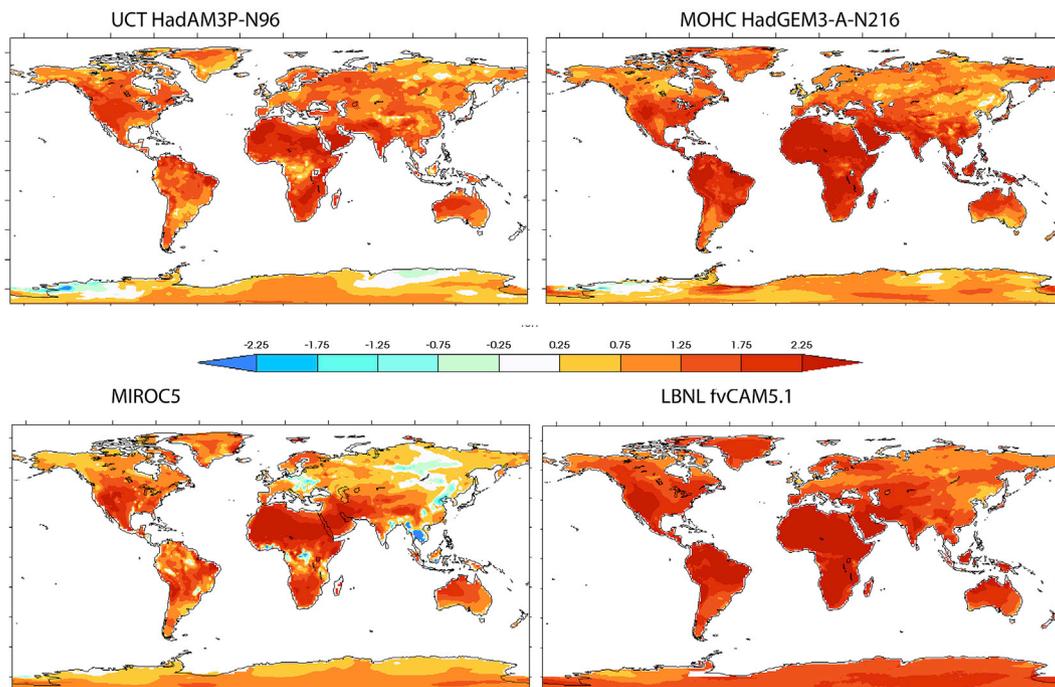

Fig. 2. The natural logarithm of the Risk Ratio, *ln(RR),* between the actual and counterfactual C20C+ D&A simulations for the temperature values associated with the 20 year return value of the counterfactual simulations. Upper left: UCT-CSAG HadAM3P-N96. Upper right: MOHC HadGEM3-A-N216. Lower left: MIROC5. Lower right: LBNL fvCAM5.1. Note that max(RR) = 20 and ln(20)~3.
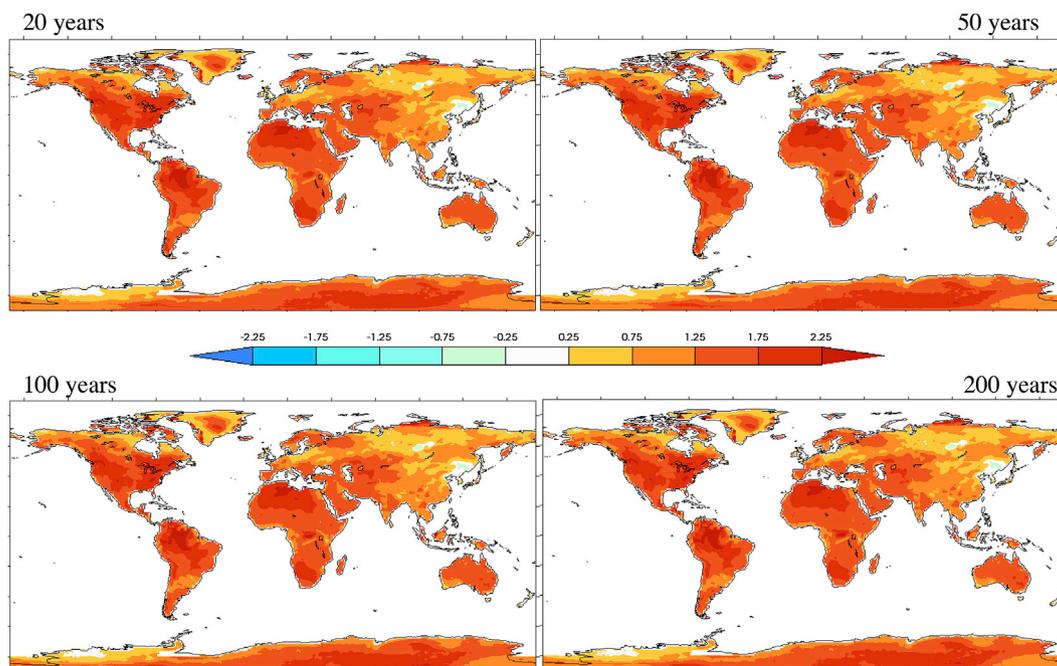
Fig. 3. Change in return value for different fixed return periods using the LBNL fvCAM5.1 simulations. Upper left: 20 years. Upper right: 50 years. Lower left: 100 years. Lower right: 200 years.
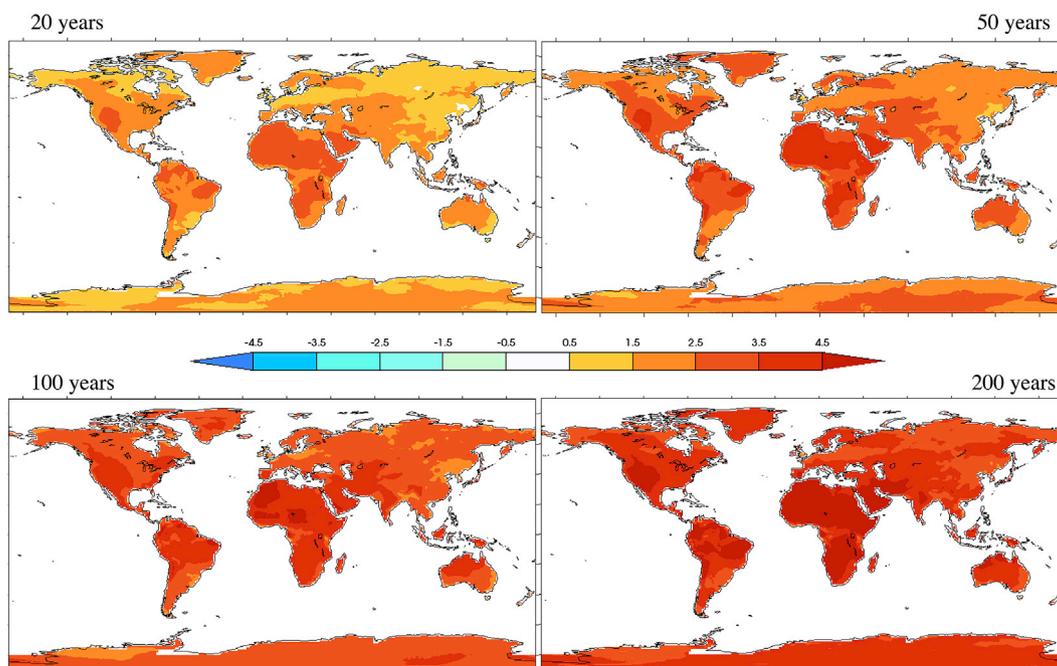

Fig. 4. The natural logarithm of the Risk Ratio, *ln(RR),* for different counterfactual return periods using the LBNL fvCAM5.1 simulations. Upper left: 20 years. Upper right: 50 years. Lower left: 100 years. Lower right: 200 years.

simulations drawn from the base of those blue lines are all very close to 2.5 years dictating that risk ratio would increase as a function of counterfactual return period as seen in Fig. 4.

Uncertainty can be quantified by the method of Hosking and Wallis (1997). In their method, an ensemble of GEV parameters can be generated by first fitting the original data to a GEV distribution then generating random samples distributed according to that distribution and a random number generator. Although not a classical bootstrap, the standard deviation of this ensemble may be interpreted in the same manner as a standard error. Fig. 5 shows the standard error in *ln(RR)* of the 20 year

Nat-Hist return value for the C20C+ models. Note that in comparing the compressed color scale of Fig. 5 to *ln(RR)* in Fig. 2, there few regions where the sign of *RR* would change. In fact, this measure of uncertainty is highest in the few models that there are regions where the likelihood of extreme temperature is decreased (presumably due to aerosol forcing increases). As a result, event attribution statements about high temperature events are largely unaffected by the uncertainty in GEV parameter estimation.

For the C20C+ D&A experiment, Risk Ratios for temperatures thresholds drawn from a counterfactual simulation will always be
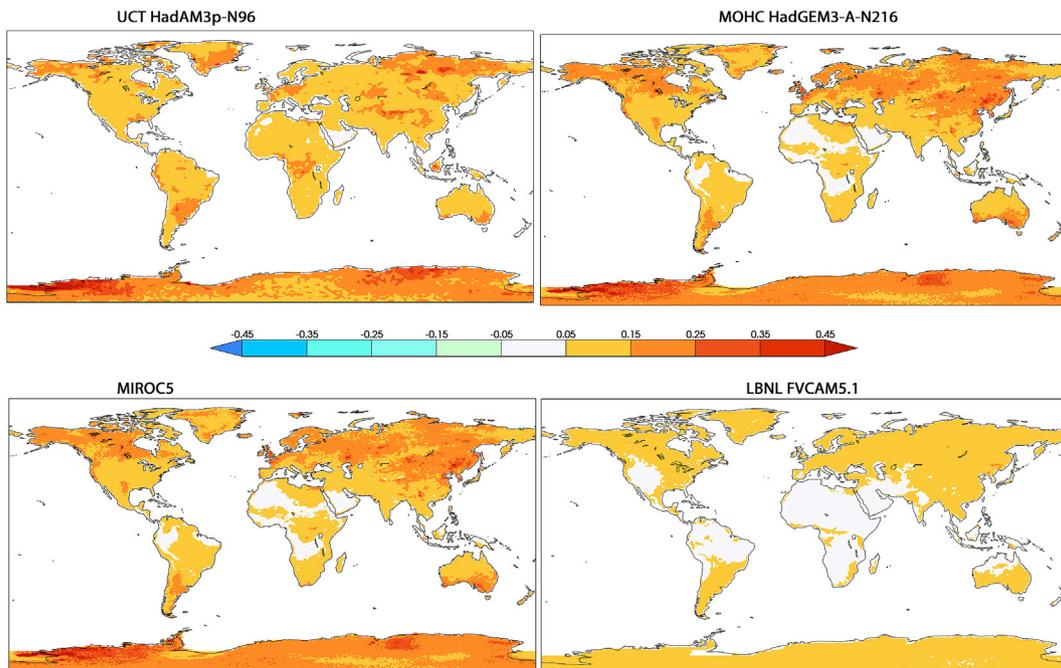
Fig. 5. Standard error estimates of the natural logarithm of the Risk Ratio, *ln(RR),* for the temperature values associated with the 20 year return value of the counterfactual simulations. Upper left: UCT-CSAG HadAM3P-N96. Upper right: MOHC HadGEM3-A-N216. Lower left: MIROC5. Lower right: LBNL fvCAM5.1.
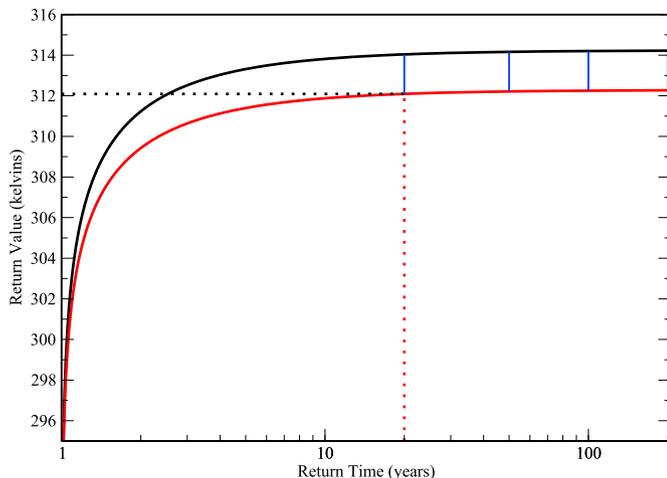

Fig. 6. Return value of 3 day average daily maximum temperature (kelvins) versus return time (years) for a representative mid latitude grid point from the actual (black) and counterfactual (red) LBNL fvCAM5.1simulations. See the text for discussion. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

bounded since the counterfactual probability of such temperatures is nonzero by definition and the factual probability is generally higher. However, when considering temperature thresholds realized in the factual simulations, any values above the counterfactual bound would result in infinite risk ratios, that is, such high temperatures could not be realized in the absence of the simulated human interference in the climate system. For a conservative attribution statement about actual extreme temperatures realized, it is often preferable to base the statement on an estimate of the lower bound of the risk ratio with some high confidence (e.g. 5% statistical significance level). In the case shown in Fig. 6, temperatures between 312.3 K and 314.2 K are possible in the actual simulations but not in the counterfactual simulations. For temperatures close to or above the counterfactual bound, conventional bootstrap methods to estimate confidence intervals of the risk ratio may

result in infinite risk ratios for some bootstrap elements due to random exceedances of the bound. In Jeon et al. (2016), we presented the statistical formalism to construct one-sided uncertainty estimates by inverting a likelihood ratio test when conventional bootstrap methods produce infinities.[3] We refer interested readers to that study.

4. Conclusions

We present changes in 20 year return values and associated risk ratios of 3 day heat wave temperatures calculated from four models submitted to the Climate of the Twentieth Century Plus (C20C+) Detection and Attribution project. A principal finding is that model treatments of atmospheric aerosols, which is not specified in the C20C+ experimental design, can have large localized influences on the magnitude and even sign of the anthropogenic change in extreme high temperature events. Statements based on models that prescribe atmospheric aerosol concentrations rather than emissions may erroneously conclude that the human interference in the climate system increased the risk and magnitude of an actual observed heat wave when in fact the net effect of greenhouse gas (GHG) and aerosol increases were to decrease the risk and magnitude due to rapid increase of localized aerosol concentrations in stagnant conditions. However it is important to remark that such cooling of extreme temperatures is very limited in extent to those regions with significant air pollution and that there remain substantial uncertainties in modeling aerosol processes. Rigorous quantification of this effect is not possible with the current C20C+ database as the protocols did not specify that aerosol forcing changes be saved. Further experimentation with a single model with multiple aerosol package options (such as the LBNL fvCAM5.1) would be the most controlled class of simulations in this regard.

Most other land areas of the planet are simulated to have exhibited an increase of between 1 and 3 °C for long period return values of the 3 day average of the daily maximum temperature as a result of anthropogenic climate change. While the changes in return value are stable as the rarity

---

[3] Even if the best estimate of the risk ratio is infinite, a finite lower bound may exist.

of the counterfactual event is increased, relative changes in frequency are not. This consequence of the boundedness of the Generalized Extreme Value (GEV) distributions fitted to the annual maxima causes the so-called Risk Ratio to increase with rarity in those areas where the anthropogenic effect is a warming of extreme temperatures.

Differences in land surface temperatures across models are due both to such non-GHG forcing differences and to differences in the models' sensitivity to the GHG increase itself. Since the ocean state of each model has been changed in the same way for the C20C+ experiment, surface air temperature differences over land and sea ice are the result of local differences in radiative forcing responses as well non-local changes in the general circulation of the atmosphere. The final two columns of Table 1 show the differences between the All-Hist and Nat-Hist simulations globally averaged for both the average of the "hot season" and very extreme surface air temperatures revealing variation in model sensitivity to increases in anthropogenic forcing over land. Hot season here is defined as the maximum of either June-July-August (JJA) or December-January-February (DJF) seasonal averages. The extreme temperature differences shown in Table 1 are calculated from the 20-year return values of the three day average of daily maximum temperature. However, from both Figs. 3 and 5, these values are good approximations to extreme temperatures at longer return periods. For some locations, depending on the climate model, extreme temperature increases can be greater than the corresponding increases in average temperature during the season that they are most likely to occur in. While temperatures are generally still best described as bounded distributions, their tails are slightly heavier as temperatures warm in these locations.

Pooling across a relatively small number of years produces fitted GEV distributions with very low uncertainties. Most colored regions in Figs. 1–4 would remain the same color if the lower bound of the changes in return values or return times were used. This uncertainty from the statistical fitting procedure is largest for the MOHC HadGEM3-A-N216 simulations as it contains only 15 realizations for each of the C20C+ D&A experiment configurations. However, by pooling a few more years, say a decade, this source of uncertainty would be reduced to a safely negligible amount if the probability of the event is not strongly conditional on interannual sea surface temperature anomalies.

The original design of the C20C+ D&A experiment is targeted towards making attribution statements of individual events in specific years. The prescribed ocean element of this design allows consideration of the effect of the state of the ocean on such statements by considering the extreme events occurring only in the relevant year. However, uncertainties in the estimates of longer period return values for individual years from the extreme value distribution fitting procedures may be significantly larger than the pooled results presented in Figs. 1–4. On the other hand, known teleconnections between surface air temperature and the state of the ocean are well documented (Santoso et al., 2015) and uncertainties in long period return values from pooling years may also be non-neglible. A more detailed analysis of such interannual variability is deferred to a later study.

Careful attention would need to be paid to the uncertainties in single year analyses of MIROC5 and UCT-CSAG HadAM3P-N96, each with 50 realizations per configuration and even LBNL fvCAM5.1with ~100 realizations. The fifteen realizations of MOHC HadGEM3-A-N216 used here are not enough for single year analyses of this type. However, 105 simulations are available in the C20C+ archive for the years 2014 and 2015 and even larger ensembles for the most recent years have recently been performed. Although not used in this study, a limited number of years (2011-2013) were simulated 400 times for both the factual and counterfactual parts of the C20C+ D&A experiment (Angélil et al., 2017a) and would provide attribution statements with very low uncertainties from the statistical fits.

Three of the four models used in this study have been compared to reanalyses and found to have significant biases in mean and extreme seasonal temperatures (Angélil et al., 2016). Bias correction techniques have generally involved uniformly adding the difference between observed and model simulations of the recent past (in this case All-Hist) to all simulations whether they be of the past, present or future. While bias correction often is done for the mean climate, Jeon et al. (2016) presented a quantile bias correction technique specifically targeted towards refined estimates of the risk ratio. Indeed, errors in the simulated mean climate may be unrelated to those in the tail of the distribution due to differences in the relevant physical processes. Such uniform bias correction of long period return values only shifts the curves of Fig. 5 up or down, preserving the differences at fixed return period (i.e. the length of the short blue lines does not change). The differences between long period return values in bounded distributions such as in Fig. 5 converge as values get near the bound. Hence, extreme temperature bias correction factors are largely insensitive to return period even for only modestly rare events. As a result, "mechanistic" attribution statements (Pall et al., 2014; Easterling et al., 2016) that state the effect of anthropogenic climate change on the observed magnitude of an extreme event of fixed rarity (such as Figs. 1 and 3) are insensitive both to errors in the estimation of the observed return period as well as the model's mean climatological bias, as long as both the factual and counterfactual estimate of return period are high enough to be near the bound. This is not the case for "probabilistic" attribution statements that state the effect of anthropogenic climate change on the chances of an event of fixed magnitude (via the risk ratio). Even if the value of the estimated return period in the factual simulation does not change (i.e. the location of the dashed red line of Fig. 5) because it is determined by observations (as in the approach described in Jeon et al., 2016), the estimated return period in the counterfactual simulations would change (i,e. the length of the black dotted line) and the estimate of risk ratio would be altered by bias correction.

As Hansen et al. (2012) has pointed out, the shift in the distribution of surface air temperature since the mid 20th century has been profound. Temperatures that were rare prior to the 1980's now occur with regularity. In such cases, estimating return periods and probabilities via extreme value techniques in counterfactual simulations or in the early part of the observed temperature record may be appropriate. However, extreme value techniques may not be appropriate in the actual simulation or in the present day part of the observational record if temperature values are no longer in the tail of the distribution and other methods of estimating probabilities should be used (O'Brien et al., 2016). Also in such cases, application of the same bias correction factor to both factual and counterfactual simulations may be inappropriate as very different portions of the distribution are of interest. Rather, judicious choice of separate quantile bias corrections would be advisable.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.wace.2018.03.001.

References

Angélil, O., Perkins-Kirkpatrick, S., Alexander, L., Stone, D., Donat, M., Wehner, M., Shiogama, H., Ciavarella, A., Christidis, N., 2016. Comparing regional precipitation and temperature extremes in climate model and reanalysis products. Weather and Climate Extremes 13, 35–43. https://doi.org/10.1016/j.wace.2016.07.001.

Angélil, O., Stone, D., Wehner, M., Paciorek, C.J., Krishnan, H., Collins, W., 2017a. An independent assessment of anthropogenic attribution statements for recent extreme weather events. J. Climate 30, 5–16. https://doi.org/10.1175/JCLI-D-16-0077.1.

Angélil, O., Stone, D., Perkins-Kirkpatrick, S., Alexander, L., Wehner, M., Shiogama, H., Wolski, P., Ciavarella, A., Christidis, N., 2017b. On the nonlinearity of spatial scales in extreme weather attribution statements, early on release. Clim. Dynam. https://doi.org/10.1007/s00382-017-3768-9.

Christidis, N., Stott, P.A., Brown, S.J., 2011. The role of human activity in the recent warming of extremely warm daytime temperatures. J. Clim. 24, 1922–1930.

Christidis, N., Stott, P.A., Zwiers, F.W., 2014. Fast-track attribution assessments based on pre-computed estimates of changes in the odds of warm extremes. Clim Dyn 45, 1547. https://doi.org/10.1007/s00382-014-2408-x.

Ciavarella, A., Christidis, N., Andrews, M., Groenendijk, M., Rostron, J., Elkington, M., Burke, C., Lott, F., Stott, P., 2017. HadGEM3-A based system for probabilistic attribution of extreme weather and climate events. Weather and Climate Extremes (in preparation for this special issue).

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer.

Collins, M., Knutti, R., Arblaster, J.M., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W.J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A.J., Wehner, M., 2013. Long-term climate change: projections, commitments and irreversibility. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Easterling, D.R., Kunkel, K.E., Wehner, M.F., Sun, L., 2016. Detection and attribution of climate extremes in the observed record. Weather Clim. Extrem. 11, 17–27. https://doi.org/10.1016/j.wace.2016.01.001.

Gates, L., 1992. AMIP: the atmospheric model Intercomparison project. Bull. Amer. Meteor. Soc. 73, 1070–1962.

Hansen, J., Sato, M., Ruedy, R., 2012. Perception of climate change. Proc. Nat. Acad. Sci 109 (37), E2415–E2423. https://doi.org/10.1073/pnas.1205276109.

Hosking, J.R.M., Wallis, J.R., 1997. Regional frequency analysis: an approach based on L-moments. Cambridge University Press, 244 pages.

Houghton, J.T., Meira Filho, L.G., Callander, B.A., Harris, N., Kattenberg, A., Maskell, K. (Eds.), 1996. Climate Change 1995: the Science of Climate Change, Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, ISBN 0-521-56433-6.

Imada, Y., Maeda, S., Watanabe, M., Shiogama, H., Mizuta, R., Ishii, M., Kimoto, M., 2017. Recent enhanced seasonal temperature contrast in Japan from large ensemble high-resolution climate simulations. Atmosphere, 2017 8 (3), 57. https://doi.org/10.3390/atmos8030057.

Jeon, S., Paciorek, C.J., Wehner, M.F., 2016. Quantile-based bias correction and uncertainty quantification of extreme event attribution statements. Weather and Climate Extremes 12, 24–32. https://doi.org/10.1016/j.wace.2016.02.001.

Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J., Bates, S., Danabasoglu, G., Edwards, J., et al., 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, 1333–1349.

Kharin, V.V., Zwiers, F.W., Zhang, X., Hegerl, G.C., 2007. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. J. Climate 20, 1419–1444.

Kharin, V.V., Zwiers, F.W., Zhang, X., Wehner, M., 2013. Changes in temperature and precipitation extremes in the CMIP5 ensemble. Climatic Change 119, 345–357. https://doi.org/10.1007/s10584-013-0705-8.

Kiehl, J.T., Schneider, T.L., Rasch, P.J., Barth, M.C., Wong, J., 2000. Radiative forcing due to sulfate aerosols from simulations with the national center for atmospheric Research community climate model. Version 3 J. Geophys. Res. Atmos. 105 (D1), 1441–1457. https://doi.org/10.1029/1999JD900495.

Kim, Y.H., Min, S.K., Zhang, X., Zwiers, F., Alexander, L.V., Donat, M.G., Tung, Y.S., 2016. Attribution of extreme temperature changes during 1951–2010. Clim. Dyn 46 (5–6), 1769–1782.

Martin, G.M., Ringer, M.A., Pope, V.D., Jones, A., Dearden, C., Hinton, T.J., 2006. The physical properties of the atmosphere in the new Hadley Centre Global Environmental Model, HadGEM1. Part 1: model description and global climatology. J. Clim. 19 (7), 1274–1301. https://doi.org/10.1175/JCLI3636.1.

Min, S.-K., Zhang, X., Zwiers, F., Shiogama, H., Tung, Y.-S., Wehner, M., 2013. Multi-model detection and attribution of extreme temperature changes. J. Clim. 26, 7430–7451. https://doi.org/10.1175/JCLI-D-12-00551.1.

Mizuta, R., Murata, A., Ishii, M., Shiogama, H., Hibino, K., Mori, N., Arakawa, O., Imada, Y., Yoshida, K., Aoyagi, T., Kawase, H., Mori, M., Okada, Y., Shimura, T., Nagatomo, T., Ikeda, M., Endo, H., Nosaka, M., Arai, M., Takahashi, C., Tanaka, K., Takemi, T., Tachikawa, Y., Temur, K., Kamae, Y., Watanabe, M., Sasaki, H., Kitoh, A., Takayabu, I., Nakakita, E., Kimoto, M., 2017. Over 5000 years of ensemble future climate simulations by 60 km global and 20 km regional atmospheric models. Bull. Amer. Meteor. Soc. July 2017, 1383–1398. https://doi.org/10.1175/BAMS-D-16-0099.1.

National Academies of Sciences, Engineering, and Medicine, 2016. Attribution of Extreme Weather Events in the Context of Climate Change. The National Academies Press. https://doi.org/10.17226/21852.

Nitschke, M., Tucker, G., Hansen, A., Williams, S., Zhang, Y., Peng, B., 2011. Impact of two recent extreme heat episodes on morbidity and mortality in Adelaide, South Australia: a case-series analysis. Environ. Health 10 (1), 42–51.

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. Geophys. Res. Let. 39 (L04702).

O'Brien, T.A., Kashinath, K., Cavanaugh, N.R., Collins, W.D., O'Brien, J.P., 2016. A fast and objective multidimensional kernel density estimation method: fastKDE. Comput. Stat. Data Anal. 101, 148–160.

Pall, P., Wehner, M., Stone, D., 2014. In: Grojahn, Richard, Li, Jianping, Swinbank, Richard, Volkert, Hans (Eds.), Probabilistic Extreme Event Attribution in Dynamics and Predictability of Large-scale, High-impact Weather and Climate Events. Cambridge University Press, ISBN 978-1-107-07142-1, pp. 37–46.

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. Clim. Dyn. https://doi.org/10.1007/s00382-016-3492-x.

Santoso, A., Cai, W., Collins, M., McPhaden, M., Jin, F.-F., Guilyardi, E., Vecchi, G., Dommenget, D., Wang, G., 2015. ENSO extremes and diversity: dynamics, teleconnections, and impacts. Bull. Amer. Meteor. Soc. 96, 1969–1972. https://doi.org/10.1175/BAMS-D-15-00141.1.

Shiogama, H., Watanabe, M., Imada, Y., Mori, Y.M., Ishii, M., Kimoto, M., 2013. An event attribution of the 2010 drought in the South Amazon region using the MIROC5 model. Atmos. Sci. Lett. 14, 170–175.

Shiogama, H., Watanabe, M., Imada, Y., Mori, Y.M., Kamae, Y., Ishii, M., Kimoto, M., 2014. Attribution of the June-July 2013 heat wave in the southwestern United States. SOLA 10, 122–126. https://doi.org/10.2151/sola.2014-025.

Shiogama, H., Imada, Y., Mori, M., Mizuta, R., Stone, D., Yoshida, K., Arakawa, O., Ikeda, M., Takahashi, C., Arai, M., Ishii, M., Watanabe, M., Kimoto, M., M, 2016. Attributing historical changes in probabilities of record-breaking daily temperature and precipitation extreme events. SOLA 12, 225–231. https://doi.org/10.2151/sola.2016-045.

Stone, D.A., Allen, M.R., 2005. Attribution of global surface warming without dynamical models. Geophys. Res. Lett. 32, L18711 https://doi.org/10.1029/2005GL023682.

Stone, D.A., Pall, P., 2017. A benchmark estimate of the effect of anthropogenic emissions on the ocean surface. Geosci. Model Dev. (GMD) (In review).

Stone, D.A., Christidis, N., Folland, C., Perkins-Kirkpatrick, S., Perlwitz, J., Shiogama, H., Wehner, M.F., Wolski, P., Cholia, S., Krishnan, H., Murray, D., Ang'elil, O., Beyerle, U., Ciavarella, A., Dittus, A., Quan, X.-W., 2017a. Experiment design of the international CLIVAR C20C+ detection and attribution project. Weather and Climate Extremes (in preparation for this special issue).

Stone, D.A., Risser, M.D., Angélil, O.M., Wehner, M.F., Cholia, S., Keen, N., Krishnan, H., O'Brien, T.A., Collins, W.D., 2017b. A basis set for exploration of sensitivity to prescribed ocean conditions for estimating human contributions to extreme weather in CAM5.1-1degree. Weather and Climate Extremes (submitted for this special issue).

Tebaldi, C., Wehner, M., 2016. Benefits of mitigation for future heat extremes under RCP4.5 compared to RCP8.5. Climatic Change. https://doi.org/10.1007/s10584-016-1605-5.

Wehner, M.F., 2004. Predicted 21st century changes in seasonal extreme precipitation events in the Parallel Climate Model. J. Climate 17, 4281–4290.

Wehner, M.F., 2010. Sources of uncertainty in the extreme value statistics of climate data. Extremes 13, 205–217. https://doi.org/10.1007/s10687-010-0105-7.

Wolski, P., Stone, D., Tadross, M., Wehner, M., Hewitson, B., 2014. Attribution of floods in the Okavango basin, southern Africa. J. Hydrol. 511, 350–358. https://doi.org/10.1016/j.jhydrol.2014.01.055.

Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Tank, A.K., Peterson, T.C., Trewin, B., Zwiers, F.W., 2011. Indices for monitoring changes in extremes based on daily temperature and precipitation data. WIREs Clim Change 2, 851–870. https://doi.org/10.1002/wcc.147.

Zwiers, F.W., Zhang, X.B., Feng, Y., 2011. Anthropogenic influence on long return period daily temperature extremes at regional scales. J. Clim. 24, 881–892.