LETTER

Multimodel detection and attribution of changes in warm and cold spell durations

Chunhui Lu1©, Ying Sun1,2 and Xuebin Zhang3

1 National Climate Center, Laboratory for Climate Studies, China Meteorological Administration, Beijing 100081, People’s Republic of China
2 Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, 210044, People’s Republic of China
3 Climate Research Division, Climate Change and Environment Canada, Toronto, ON M3H 5T4, Canada

E-mail: sunying@cma.gov.cn

Keywords: detection and attribution, warm and cold spell duration, human influence

Abstract

The duration of warm and cold spells is measured by the persistence of instances when daily temperatures are greatly above or below their normal values. These spells represent a prolonged period of times when daily temperatures are extreme and can potentially be connected to climate impacts in the agricultural, health, energy and other sectors. This study aims to determine evidence of responses in the durations of warm (WSDI) and cold (CSDI) spells to forcings external to the climate system. We consider the globe, the six continents and China during the period from 1958 to 2010 in this analysis. Here we compare the observed duration indices with those derived from simulated daily temperature by climate models participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5), using an optimal fingerprint method. The results show that, averaged over the global lands, the WSDI has substantially increased by 15 days while the CSDI has decreased by 3 days over the 53 year period. The simulated changes are generally consistent with the observations; models, however, overestimate the observed changes in the WSDI in five of the six continental regions. We consider a signal is detected if it is significant at the 10% level. At the global scale, human influences on both warm and cold spell durations are detected. At the continental scale, an anthropogenic signal is detected in the warm spell durations of most continents. The human influence on cold spell duration is detectable only in the Asian and European continents. In China, human influence can clearly be detected in both the warm and cold spell durations. Responses to natural forcings are generally not detected at the continental or smaller spatial scales.

1. Introduction

Accumulating evidence from observational records has shown remarkable changes in warm and cold extremes at the global and regional scales (Alexander et al 2006, Brown et al 2008, Hartmann et al 2013). The frequency, intensity and duration of extreme temperatures all show clear changes that are consistent with global warming (Donat et al 2013, Sillmann et al 2013a). These changes exert considerable impacts on human society and natural systems. The possible causes of these changes have been explored. Some authors have investigated the effects of external forcings on these changes from the viewpoint of detection and attribution and have found that changes in mean temperatures can be attributed to anthropogenic forcing at global and regional scales (Tett et al 2002, Stott et al 2006, Zwiers and Zhang 2003, Zhang et al 2005, Zhang et al 2006, Bindoff et al 2013, Donat et al 2014, Donat et al 2016). Christidis et al (2005) was the first study presenting a formal detection and attribution analysis of extreme temperatures and found clear anthropogenic influences on the intensity of warm nights and cold days and nights. Subsequent papers have examined possible anthropogenic influences on the intensity (e.g. Zwiers et al 2011, Min et al 2013,
Kim et al (2015, Wen et al 2013) and frequency (e.g. Morak et al 2013, Lu et al 2016) of daily temperature extremes at the global and regional scales. These analyses show that anthropogenic signals can be robustly detected in observed changes in the intensity and frequency of warm and cold extremes (Christidis et al 2011, Zwiers et al 2011, Min et al 2013, Christidis et al 2014, Kim et al 2015) in various regions and in China (Lu et al 2016, Yin et al 2016). Moreover, human activities may also have increased the probability of individual high-impact extreme high temperature events, such as the 2003 European heat wave (Stott et al 2004) and the extreme hot summer in 2013 in eastern China (Sun et al 2014).

The understanding of anthropogenic influence on the duration or persistence of daily temperature extremes has been very limited. Christidis and Stott (2016) studied the duration or persistence of daily temperature extremes around the globe and in Europe. The authors compared a set of 16 observed extreme temperature indices, including those duration indices with those computed from the output of HadGEM3-A simulations. They found that anthropogenic forcings have affected most indices in recent decades with only a few exceptions for a few extreme temperature indices in Europe. In particular, they detected human influence in duration indices in European but not in the global series. As anthropogenic induced warming is everywhere, it is reasonable to assume that most of the changes in these duration indices are also thermal dynamic rather than changes in circulation; one would expect stronger signal-to-noise ratio in the global series than in the European series. Thus the related results reported in Christidis and Stott (2016) do not seem to be robust. This and the fact that the study was based on simulation by one climate model, and the work focused only on the globe and Europe, suggests a need for a more comprehensive analysis with simulations by multiple climate models and for different continents of the earth to determine the robustness of results from the earlier study and evidence of human influence in other continents as well.

This paper attempts to fill in this gap. Here, we compare changes in the observed warm spell duration (as measured by the warm spell duration index, WSDI) and cold spell duration (as measured by the cold spell duration index, CSDI) with those simulated by multiple models participating in the Coupled Model Inter-comparison Project Phase 5 (Taylor et al 2012) at both the global and continental scales. These two indices were defined by the ETCCDI (Zhang et al 2011). The WSDI is defined as the annual number of days with at least six consecutive days when the daily maximum temperature is above the 90th percentile. The CSDI is similarly defined as the annual number of days with at least six consecutive days when the daily minimum temperature is below the 10th percentile. The structure of this paper is as follows: section 2 describes the data and detection methods, including the data processing. Section 3 presents the main results. The conclusions of this research and some discussions are given in section 4.

2. Data and methods

2.1. Observations

We consider the WSDI and CSDI over the global land and six continental regions where observation histories are sufficiently long. The six regions are based on the Giorgi regions (Giorgi and Francisco 2000, see figure S1). Additionally, we include China (CN) as a separate region due to our strong interest in this country.

The observational data for the global, continental and China analyses come from the HadEX2 dataset (Donat et al 2013). HadEX2 provides a global gridded land-based dataset at a resolution of 0.5° × 0.5° latitude. The data are based on station observations and cover the period from 1901–2010. Compared with the earlier version of the dataset, HadEX (Alexander et al 2006), HadEX2 provides a better quality global dataset with more stations and a longer observational period. We include only the data after 1958 in our analysis as data coverage was poor prior to this period (Donat et al 2013).

Daily maximum and daily minimum temperatures for 2419 Chinese stations were also used to analyze changes in the WSDI and CSDI in China. These data have been quality controlled and homogeneity adjusted (Xu et al 2013, Wang and Feng 2014) by the China National Meteorological Information Center (NMIC). The station data were aggregated to produce gridded datasets at the HadEX2 grid resolution by averaging indices from all available stations within a grid box. Compared with the Chinese station observations, the HadEX2 dataset can capture the essential features of the observed changes as this is a more comprehensive dataset; however, there are some problems. The Chinese station data show weaker warming in western China than the HadEX2 dataset, though this feature does not lead to clear differences in national averaged extreme temperatures between these two datasets. Additionally, the region with the strongest warming over the western part of the Tibetan Plateau in the HadEX2 dataset does not have underlying station data, even in the 2419 station network (figure S2, lower panel).

2.2. Model simulations

We use indices computed from daily outputs of simulations by the climate models in the Coupled Model Inter-comparison Project Phase 5 (CMIP5; Taylor et al 2012) to estimate climate responses to external forcings and natural internal climate variability. The indices computed by Sillmann et al (2013a, 2013b) are used in this study. We also computed indices for
the model runs for which daily outputs are available but were not used in Sillmann et al (2013a, 2013b). The computational codes that generated the indices in Sillmann et al (2013a, 2013b) are used in these cases. Table S1 lists the model simulations used in this study. Altogether, 96 simulations were conducted with 21 climate models driven by the combined effects of historical natural and anthropogenic forcings (ALL). These simulations ended in 2005, and we used the RCP4.5 simulations for the years 2006–2010 to extend the data to 2010. In total, 32 historical greenhouse gas (GHG) simulations were conducted with nine climate models. Additionally, 36 natural forcing-driven (NAT) simulations were conducted with nine climate models. Both the GHG and NAT simulations ended in 2012. We estimate the response to anthropogenic forcing (ANT) as the differences between the model responses to ALL and NAT forcings assuming linear additivity. Linearly additivity holds for large-scale temperature changes (Meehl et al 2003, Gillett et al 2004). However, non-additive approaches have not been widely adopted. To estimate the natural internal climate variability, we also use index data computed from preindustrial control simulations (CTL) from 28 models. These data consist of 230 pieces of 53 year chunks. The model data are set as missing if the corresponding observation data for the time and space is missing prior to the calculations of the global and regional mean series.

2.3. Detection methods

We perform two types of detection analysis. For the first, we simply ask whether the long-term change in the indices can be expected by chance by performing a simple trend analysis. Here, we compute linear trends in the index series using a least-squares fit. This detection analysis is called a trend analysis in the remaining of the paper. In the second case, we ask a more refined detection and attribution question. Here, we ask whether the expected response to external forcing as simulated by the climate models can be detected in the observations. If the answer is yes, we further ask whether the results are consistent, which allows for the attribution of observed changes to causes.

The detection and attribution analysis is conducted by comparing the time evolutions of the extreme indices in the observations and modeled responses using a standard optimal fingerprinting technique (Allen and Stott 2003). A brief introduction is given in the supporting information available at stacks.iop.org/ERL/13/074013/mmedia. Because of the limited data availability for the estimation of natural internal variability, the fingerprinting method can operate only at relatively small dimensions. Christidis and Stott (2016) used spherical harmonics at a T8 truncation to reduce the dimensionality of the 16 extreme indices at the global scale. Here, we average the data over space and time and conduct the detection and attribution analysis based on global or regional 5 year mean series to reduce the data dimensionality. The optimal fingerprinting method assumes a Gaussian distribution for the residuals. Different methods have been used when the residuals do not follow a Gaussian distribution (Christidis et al 2011, Zwiers et al 2011). As our analyses are performed on space-time averages, we expect the regression residuals to approach a Gaussian distribution. We also use two methods to transform the data such that the distribution of residuals is very close to a Gaussian to examine the sensitivities of the results to the Gaussian assumption.

Two regression analyses are performed, namely, single-signal and two-signal analyses. In the single-signal analysis, the observations are regressed against the multimodel mean responses to an external forcing or combined effect of different forcings (in the case for ALL). This approach provides a quick and simple overview to determine whether the model-simulated response to a specific forcing can be detected in the observed changes. This approach is particularly useful in the case of the ALL signal as it includes the expected response to the combined effect of all known external forcings. In the two-signal analyses, we aim to partition the observed changes into responses to two known external forcings, namely, NAT and ANT, including greenhouse gases and anthropogenic aerosols. The objective is to determine whether the two major external forcings can be detected separately in the presence of the other signal. However, there is a caveat when ALL and NAT are estimated from the different sets of models. The difference between ALL and NAT can also reflect the fact that the models that are used to estimate the signals are different. For this reason, we repeated the detection and attribution analysis using the eight models that provide both ALL and NAT simulations. We found that the detection results were quite similar (figure S7).

The analysis requires two independent estimates of the covariance of natural internal variability. These estimates are achieved by using data from preindustrial control simulations and the residuals of each model member after the removal of the model ensemble mean. We split the control run data and intra-ensemble differences equally into two independent sets with 272 fifty-three-year chunks of noise data in each set. We use these two sets for optimization to obtain the best estimate of scaling factor $\beta$, for estimating the uncertainty range of $\beta$ and for conducting the residual consistency test. The reliability of the detection and attribution analyses depends critically on the model performance in simulating natural internal climate variability. To determine whether model-simulated variability is sufficiently large, we compare the power spectra of the modeled and observed indices. We also conduct a residual consistency test using the implementation of Ribes and Terray (2013).
3. Results

Figure 1 shows the trends in the observed and modeled WSDI and CSDI series. The top panel presents the observed trend based on the HadEX2 data. The trends are characterized by an increase in the duration of warm spells and a decrease in the duration of cold spells, both of which are consistent with global warming. However, regional differences are present in the trends for both indices. WSDI shows the largest increasing trend over India, Southeast Asia and Central Europe and a much weaker positive trend in North and South America. In parts of India, the increase can be as large as 20 days during 1958–2010. There appears to be a negative trend in Greenland, which may be an artifact of data processing as the availability of station data in this region is very limited. A decrease can also be found in South America. CSDI exhibits a strong negative trend for much of Eurasia and a weaker negative trend in eastern Asia and North and South America. The decrease in parts of Eurasia can be as large as 5 days during the 53 year period. A positive trend can be observed in India and some areas in North America. In some areas along the tropics and in South America, a cold spell as defined by ETCCDI is not usually observed (blank areas in the land).

The middle panel of figure 1 shows the trend in the multimodel ensemble mean under ALL forcing. WSDI increases almost everywhere. The spatial pattern of trends is in generally consistent with the observed trend pattern. The magnitude over Europe and South America is reasonably consistent with that of the observed trend, while the small observed trend in North America, Australia and Africa could not be well simulated in the models. The models also underestimate the observed changes in Asia. CSDI decreases everywhere, which is largely consistent with the observations. The lower panel of figure 1 shows the trends in the multimodel ensemble mean under NAT forcing. The NAT trends are generally weak, much weaker than the observations for WSDI, suggesting that NAT forcing alone cannot contribute to the observed changes in warm spell duration. The NAT trend in the CSDI is also very weak and of opposite sign to that of the observations, indicating a negligible role of natural forcing in the CSDI decrease.

Table 1 shows trends in the global and regional averages of the WSDI and CSDI during 1958–2010 in the observations and the model simulations. Averaged over the globe, WSDI increased by approximately 15 days, while CSDI decreased by approximately 3 days. As shown in figure 1, the observed WSDI trend...
Table 1. Trends (days over 53 years) in the WSDI and CSDI during 1958–2010 computed with observations and the multimodel ensemble simulations.

|       | GLB | ASI | EUR | NAM | SAM | AUS | AFR | CN |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| OBS   | 14.62 | 16.48 | 10.33 | 5.94 | 15.42 | 7.74 | 7.16 | 11.61 |
| ALL   | 13.72 | 11.82 | 11.89 | 13.13 | 17.4 | 11.42 | 12.87 | 12.71 |
| NAT   | 3.33  | 3.12  | 3.23  | 3.44  | 3.39  | 2.80  | 3.28  | 3.02  |

|       | GLB | ASI | EUR | NAM | SAM | AUS | AFR | CN |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| OBS   | –2.81  | –3.18  | –2.97  | –1.48  | –2.07  | –1.17  | –0.80  | –3.34  |
| ALL   | –3.86  | –3.99  | –3.50  | –3.14  | –3.11  | –2.08  | –1.88  | –3.49  |
| NAT   | 1.21  | 1.43  | 1.32  | 1.16  | 0.90  | 0.79  | 0.79  | 1.16  |

is of a similar magnitude to the modeled trend under ALL forcing, but the modeled trend under NAT forcing is much weaker. The modeled CSDI trend under ALL forcing is of the same sign as the observed trend but of a larger magnitude. However, the modeled CSDI trend under NAT forcing is of the opposite sign as that observed.

Figure 2 displays the 5 year mean anomaly (relative to 1961–1990 mean) of the WSDI and CSDI at the global scale from the observations and model simulations. We consider 5 year mean values as a compromise between removing natural inter-annual variability and maintaining responses to natural external forcing such as volcanic activity. The trends in the global mean series are also shown. The observed WSDI is characterized by little change in the early period and an obvious increase after the late 1970s. This feature is successfully reproduced by the ALL simulations: the observed series match the ensemble mean series and fall within the range of the model simulations. The response to GHG displays similar temporal variations but has a larger trend, greater than that of ALL and the observations, reflecting a stronger warming response to GHG forcing. The observations are not consistent with the NAT simulations and are well above the range of the NAT simulations starting in the early 1980s. CSDI is characterized by a small decreasing trend. Again, the observations are consistent with the model simulations under ALL and GHG forcing but are not consistent with the model simulations under NAT forcing. As the period 1961–1990, in which temperature is lower, is used as base period, and as the value of the CSDI cannot be smaller than zero, there is very little room for CSDI to decrease. However, there is no such constraint for an increase in the WSDI. This characteristic may explain the large difference in the magnitudes of the trends of the WSDI and CSDI series in both the observations and model simulations under ALL forcing. This difference may also explain the larger variability in the WSDI compared to that in
Figure 3. Power spectral density (PSD) of the WSDI (a) and CSDI (b) computed from relevant global mean values from 1958 to 2005 for the observations (black) and model simulations. The pink and blue shadings show the 5%–95% ranges of the PSD computed from individual members of ensemble runs forced with historical ALL and historical NAT.

The power spectra of the observed and model-simulated global indices are shown in figure 3. The power spectra for the modeled indices are computed using time series from individual model runs. The power spectra of the observed WSDI lie within the range of the ALL simulations on all the time scales examined. This result suggests that the model-simulated...
variability is also consistent with that of the observations, which indicates that detection and attribution analyses based on these model simulations should be credible. The power spectra of the NAT series are more evenly distributed compared with those of the ALL series and are larger at higher frequencies but smaller at low frequencies, corresponding to a lack of a long-term trend in the NAT responses. The power spectra of the observed CSDI are generally consistent with those of the ALL simulations but are at the lower edge of the modeled power spectra at low frequencies. This result means that the modeled variability at low frequency, which is the most relevant to detection and attribution, may be too large. The implication for detection and attribution analysis is that the confidence interval of the scaling factor could be too large, making it hard to detect the signal. The observed power spectra are larger than those of the model simulations at high frequencies, but this relation has little effect on detection and attribution analysis because high frequency variability is filtered out in the 5 year mean series. The comparisons of the power spectra for the six continents and China yield similar results (figures S5 and S6). We therefore conclude that the modeled variability of the CSDI is generally sufficient for credible detection analyses.

The best estimates and the 90% confidence intervals of the scaling factors from the two-signal analysis are summarized in figures 4(c) and (d). For the WSDI, ANT and NAT are separately detected in the observed series for global land (GLB), the continents of Europe (EUR), North America (NAM), Australia (AUS) and China (CN), while ANT is additionally detected in the continents of Asia (ASI), South America (SAM), and Africa (AFR). The 90% confidence interval of their scaling factors include 1, indicating that the observed changes in the WSDI and CSDI are generally consistent with the expected response to the combined effect of anthropogenic and natural external forcings as simulated by the CMIP5 models. The scaling factors of GHG, ANT, and NAT for the WSDI are significantly greater than zero at the 5% level; however, the residual consistency tests failed, meaning that the regression fit was poor. For the CSDI, the GHG, ANT and NAT signals are not detected individually, indicating that responses to these individual forcings alone are not sufficient to explain the observed changes.

At the continental scale, the detection results for the WSDI series are generally similar to those of the global scale, but the consistency between the modeled responses and the observed changes are poor. For example, the modeled response to ALL forcing is too large compared with the observed changes in most regions except Asia and China. This difference may have some relevance for climate model evaluation. The modeled response to external forcing is not detected in the CSDI for most of the regions, except in Asia and Europe. In fact, even in the perfect model world, the signal-to-noise ratios for the CSDI in these regions are low (table S2). These results highlight the challenge of detecting and attributing changes in the CSDI.

The best estimates and the 90% confidence intervals of the scaling factors from the two-signal analysis are summarized in figures 4(c) and (d). For the WSDI, ANT and NAT are separately detected in the observed series for global land (GLB), the continents of Europe (EUR), North America (NAM), Australia (AUS) and China (CN), while ANT is additionally detected in the continents of Asia (ASI), South America (SAM), and Africa (AFR). The 90% confidence interval of their scaling factors include 1, indicating that the observed changes in the WSDI and CSDI are generally consistent with the expected response to the combined effect of anthropogenic and natural external forcings as simulated by the CMIP5 models. The scaling factors of GHG, ANT, and NAT for the WSDI are significantly greater than zero at the 5% level; however, the residual consistency tests failed, meaning that the regression fit was poor. For the CSDI, the GHG, ANT and NAT signals are not detected individually, indicating that responses to these individual forcings alone are not sufficient to explain the observed changes.

At the continental scale, the detection results for the WSDI series are generally similar to those of the global scale, but the consistency between the modeled responses and the observed changes are poor. For example, the modeled response to ALL forcing is too large compared with the observed changes in most regions except Asia and China. This difference may have some relevance for climate model evaluation. The modeled response to external forcing is not detected in the CSDI for most of the regions, except in Asia and Europe. In fact, even in the perfect model world, the signal-to-noise ratios for the CSDI in these regions are low (table S2). These results highlight the challenge of detecting and attributing changes in the CSDI.

The best estimates and the 90% confidence intervals of the scaling factors from the two-signal analysis are summarized in figures 4(c) and (d). For the WSDI, ANT and NAT are separately detected in the observed series for global land (GLB), the continents of Europe (EUR), North America (NAM), Australia (AUS) and China (CN), while ANT is additionally detected in the continents of Asia (ASI), South America (SAM), and Africa (AFR). The 90% confidence interval of their scaling factors include 1, indicating that the observed changes in the WSDI and CSDI are generally consistent with the expected response to the combined effect of anthropogenic and natural external forcings as simulated by the CMIP5 models. The scaling factors of GHG, ANT, and NAT for the WSDI are significantly greater than zero at the 5% level; however, the residual consistency tests failed, meaning that the regression fit was poor. For the CSDI, the GHG, ANT and NAT signals are not detected individually, indicating that responses to these individual forcings alone are not sufficient to explain the observed changes.

At the continental scale, the detection results for the WSDI series are generally similar to those of the global scale, but the consistency between the modeled responses and the observed changes are poor. For example, the modeled response to ALL forcing is too large compared with the observed changes in most regions except Asia and China. This difference may have some relevance for climate model evaluation. The modeled response to external forcing is not detected in the CSDI for most of the regions, except in Asia and Europe. In fact, even in the perfect model world, the signal-to-noise ratios for the CSDI in these regions are low (table S2). These results highlight the challenge of detecting and attributing changes in the CSDI.
and Africa (AFR) when the effect of NAT is taken into account. These results indicate that both anthropogenic forcing and natural external forcing have detectable responses in the observations at the global and continental scales. As NAT has a very small long-term trend, it follows that the ANT response dominates the observed increase in the WSDI. For the CSDI, ANT is detected in the presence of NAT in the GLB, ASI, EUR, and CN series but is not detected in the other regions. NAT is not detected for the globe or for all the continents. These results indicate that anthropogenic forcing has contributed to the observed changes in the CSDI at the global and some regional scales.

We also conducted detection and attribution analyses over China based on the HadEX2 dataset (figure S8). The results are very similar to those of analyses based on more comprehensive datasets that contain more than 2000 stations. While HadEX2 may have missed important regional details over China, this dataset is able to capture essential large-scale information, at least for the WSDI and CSDI.

Figure 5 shows the observed trends in the WSDI and CSDI and those attributable to various external forcings. The global averaged WSDI increased by approximately 15 days over 1958–2010. The best estimate of the increase attributable to all external forcing is approximately 14 days (with a 90% confidence interval of 12–16 days). The best estimate of the attributable increase due to anthropogenic forcing (ANT) is approximately 13 days (with a 90% confidence interval of 11 to 15 days). For the CSDI, the observed decrease attributable to all external forcing is approximately 3 days (with a 90% confidence interval of 1–5 days). The decrease that is attributable to anthropogenic forcing is approximately 4 days (with a 90% confidence interval of 1–7 days).

Because the WSDI and CSDI are both bounded by zero, the time series for the WSDI and CSDI are unlikely to follow Gaussian distributions. This property may have an effect on the estimation of the confidence intervals of the scaling factors as the regression procedure for the detection and attribution analysis assumes a Gaussian distribution. We examined the robustness of our results to this potential violation of the distributional assumption by taking the cubic root for both the observational series and the modeled series and repeating the detection and attribution analysis. The results are very similar to those reported above. We also applied a natural logarithm transformation to the data, and the detection and attribution results are also similar. We therefore conclude that our detection results are robust to this particular assumption.

4. Conclusions and discussion

In this study, we focused on possible causes for the observed long-term changes in the duration of warm spells and cold spells. We found a substantial increase in warm spell duration and a decrease in cold spell duration, both of which are consistent with warming. The climate models generally successfully reproduce the observed changes in the WSDI and CSDI at the global scale. However, the models overestimate the observed changes in the WSDI for five continents and exhibit poor performance in simulating the CSDI in American regions. We also found that much of the observed changes in the WSDI and CSDI can be attributed to external forcing and anthropogenic forcing at the
global scale and for many continents. Christidis and Stott (2016) also analyzed the Wędzi and CSDI along with many other temperature indices using simulations conducted with one global model. The authors detected anthropogenic influences in the Wędzi and CSDI over Europe but not in the global series. Their results are difficult to interpret as one would generally expect more robust detection at the global scale than over a continent for temperature-related indicators. Our analysis must have benefited from the use of multimodel ensemble data. The larger datasets enabled us to produce more robust estimates for the modeled responses to external forcings and the natural internal variability, both of which are important for robust detection and attribution analyses. The results from our study show strong consistencies between trends in the observations and the model simulations as well as a strong consistency of the detection results under different analysis settings.

Anthropogenic influence has resulted in hotter annual hottest daily temperatures and warmer annual coldest daily temperatures (Zwiers et al 2011, Min et al 2013). This influence has also contributed to more frequent moderate hot temperature extremes and less frequent cold temperature extremes (Morak et al 2013, Lu et al 2016). Here, we found that anthropogenic influence contributes to longer warm spells and shorter cold spells. Together, these findings provide clear evidence of anthropogenic influences on extreme temperature and show that anthropogenic forcing has altered different aspects of temperature extremes, including their magnitude, frequency, and duration. Extreme hot temperatures have clearly become hotter, more frequent, and longer lasting, while extreme cold temperatures have become less cold, less frequent and more short-lived. Because extreme temperature has impacts on both human society and natural systems, and because warming is projected to continue, rapid action is required to adapt to the new climate.

Acknowledgments

We appreciate the constructive comments from John Mitchell, an anonymous reviewer and an ERL board member. We acknowledge the Program for Climate Model Diagnosis and Intercomparison and the World Climate Research Programme’s Working Group on Coupled Modeling for their roles in making the WCRP CMIP multimodel datasets available. This study was supported by the National Key R&D Program of China (2018YFA0605604), the National Natural Science Foundation of China (41675074 and 41775082) and two Climate Change Projects (CCSF201704 and CCSF201805).

ORCID iDs

Chunhui Lu @ https://orcid.org/0000-0003-0335-8465

References

Alexander L V et al 2006 Global observed changes in daily climate extremes of temperature and precipitation J. Geophys. Res. 111 D05109
Allen M R and Stott P A 2003 Estimating signal amplitudes in optimal fingerprinting part 1: theory Clim. Dyn. 21 477–91
Brown S J, Caesar J and Ferro C A 2008 Global changes in extreme daily temperature since 1950 J. Geophys. Res. 113 D05115
Bindoff N L et al 2013 Detection and attribution of climate change: from global to regional Climate Change 2013: The Physical Science Basis ed T F Stocker et al (Cambridge: Cambridge University Press) pp 867–952
Christidis N, Stott P A, Brown S, Hegerl G C and Caesar J 2005 Detection of changes in temperature extremes during the second half of the 20th century Geophys. Res. Lett. 32 L20716
Christidis N, Stott P A and Brown S J 2011 The role of human activity in the recent warming of extremely warm daytime temperatures J. Clim. 24 1922–30
Christidis N, Stott P A and Zwers F W 2014 Fast-track attribution assessments based on pre-computed estimates of changes in the odds of warm extremes Clim. Dyn. 45 1547–64
Christidis N and Stott P A 2016 Attribution analyses of temperature extremes using a set of 16 indices Weather Clim. Extremes 14 24–35
Donat M G et al 2013 Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset J. Geophys. Res. Atmos. 118 2098–118
Donat M G et al 2014 Changes in extreme temperature and precipitation in the Arab region: long-term trends and variability related to ENSO and NAO Int. J. Climatol. 34 581–92
Donat M G, Alexander I V, Herold N and Dittus A J 2016 Temperature and precipitation extremes in century-long gridded observations, re-analyses, and atmospheric model simulations J. Geophys. Res. Atmos. 121 11174–89
Giori F and Francisco R 2000 Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM Clim. Dyn. 16 169–82
Gillett N P, Wehner M F, Tett S F B and Weaver A J 2004 Testing the linearity of the response to combined greenhouse and sulfate aerosol forcing Geophys. Res. Lett. 31 L14201
Hartmann D L et al 2013 Observations: atmosphere and surface Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change ed T F Stocker et al (Cambridge: Cambridge University Press)
Kim Y H, Min S K, Zhang X, Zwiers F, Alexander I V, Donat M G and Tung Y S 2015 Attribution of extreme temperature changes during 1951–2010 Clim. Dyn. 46 1769–82
Lu C, Sun Y, Wan H, Zhang X and Yin H 2016 Anthropogenic influence on the frequency of extreme temperatures in China Geophys. Res. Lett. 43 6311–18
Meehl G A, Washington W M, Wigley T M L, Arblaster J M and Dai A 2003 Solar and greenhouse gas forcing and climate response in the 20th century J. Clim. 16 426
Min S K, Zhang X, Zwiers F W, Shogama H, Tung Y S and Wehner M 2013 Multi-model detection and attribution of extreme temperature changes J. Clim. 26 7430–51
Morak S, Hegerl G C and Christidis N 2013 Detectable changes in the frequency of temperature extremes J. Clim. 26 1561–74
Ribes A and Terray L 2013 Application of regularised optimal fingerprint to attribution. Part II: application to global near-surface temperature Clim. Dyn. 41 2837–53
Stott P A, Stone D A and Allen M R 2004 Human contribution to the European heat wave of 2003 Science 342 610–4
