The usefulness of different realizations for the model evaluation of regional trends in heat waves

Author:
Perkins, Sarah; Fischer, E

Publication details:
Geophysical Research Letters
v. 40
Chapter No. 21
pp. 5793-5797
0094-8276 (ISSN)

Publication Date:
2013

Publisher DOI:
http://dx.doi.org/10.1002/2013GL057833

The usefulness of different realizations for the model evaluation of regional trends in heat waves

S. E. Perkins[1] and E. M. Fischer[2]

Received 4 September 2013; revised 21 October 2013; accepted 22 October 2013; published 14 November 2013.

[1] The evaluation of a climate model's capability to simulate trends in extreme events is not straightforward. This is due to the role of internal climate variability, the simulated phases of which are unique to each individual model realization. We undertake an assessment of the 21-member Community Earth System Model (CESM) on the basis of its ability to simulate heat wave days frequency over Australia. We employ the extreme heat factor definition to measure heat waves and study events for all summers (November–March) between 1950 and 2005. The spatial pattern, magnitude, and significance of trends in CESM were found to be reasonable since the corresponding observed trends were within the CESM ensemble range. There is a suggestion that the model produces higher interannual variability than what is observed. The trends between realizations of the same model differ strongly, which suggest that internal climate variability can strongly amplify or mask local trends in extreme events.
Citation: Perkins, S. E., and E. M. Fischer (2013), The usefulness of different realizations for the model evaluation of regional trends in heat waves, *Geophys. Res. Lett.*, *40*, 5793–5797, doi:10.1002/2013GL057833.

1. Introduction

[2] There is evidence that extreme temperatures, including heat waves have increased globally [*Diffenbaugh and Ashfaq*, 2010; *Diffenbaugh and Scherer*, 2011; *Hansen et al*., 2012; *Coumou et al*., 2013; *Sillmann et al*., 2013a] and will continue to do so at least until the end of the 21st century [e.g., *Clark et al.*, 2010; *Coumou and Robinson*, 2013; *Sillmann et al*., 2013b]. Before future projections are constructed for any climate field, it is important to provide quantification on the model's ability to simulate the field in question. However, the evaluation of a climate model on its simulation of extreme events can be quite difficult. This is due to numerous reasons, including a lack of data (extremes are rare by definition), the complexity of the driving mechanisms, and the sometimes intricate methods used to derive these events [see *Kharin et al*., 2007]. In the case of heat waves, this may be further complicated by the range of impacting characteristics (e.g., the frequency, intensity, and duration) and subjective nature of the events themselves [*Perkins and Alexander*, 2013].

[3] Even more challenging than evaluation of average statistics or climatologies of heat waves is the evaluation of the trends. We should not expect a single realization of a climate model to reproduce exactly the observed heat wave trends, due to the role of the internal climate variability. The different initial conditions of each model realization lead to a unique time evolution of variability, even though the underpinning physical representation and processes are identical. This is especially true for fully coupled climate models, where the modes of variability in ocean, sea ice, and atmosphere may be out of phase between simulations, and each simulation should be treated as one possible realization given the prescribed forcing. This theory can be extended to observations—the variability and corresponding impact on certain climate fields that have been recorded are just one outcome from a range of possible outcomes under a given forcing. Taking all this into account, it is therefore almost impossible for a climate model to replicate observed changes in extremes, at least in a single realization.

[4] In this study, we employ a large ensemble of a single coupled climate model to determine (1) the *overall* capability of the model to simulate observed heat wave trends and (2) how different could the observations have been given the same forcing. Here we employ the Community Earth System Model (CESM) over Australia. The CESM experiment consists of 21 members, unique only in their initial conditions. Since this is an exercise in evaluation, the historical simulations are used, which are forced with natural components and observed historical emissions from when the simulations commence in January 1950. Section 2 discusses the data and methodology, section 3 presents the results, and section 4 provides a short discussion and concluding remarks.

2. Data and Methods

2.1. Model Data

[5] The simulations are performed with the Community Earth System Model (CESM) version 1.0.4 including the Community Atmosphere Model version 4 and fully coupled ocean, sea ice, and land surface components [*Gent*, 2011]. All simulations are driven with historical forcing also used in the Climate Model Intercomparison Project phase 5 (CMIP5). On 1 January 1950, small random perturbations on the order of 10^{-13} are imposed to the atmospheric temperature initial condition field of the reference run to produce a 21-member ensemble. The perturbation is in the range of unit round-off perturbations and is not intended to represent the internal variability itself but acts as a trigger to initiate the model into a different mode through the chaotic nature of the climate system. See *Fischer et al*. [2013] for more

[1]ARC Centre of Excellence for Climate System Science, Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia.
[2]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland.

Corresponding author: S. E. Perkins, ARC Centre of Excellence for Climate System Science, Climate Change Research Centre, University of New South Wales, Sydney, NSW 2052, Australia.
(sarah.perkins@unsw.edu.au)

0094-8276/13/10.1002/2013GL057833

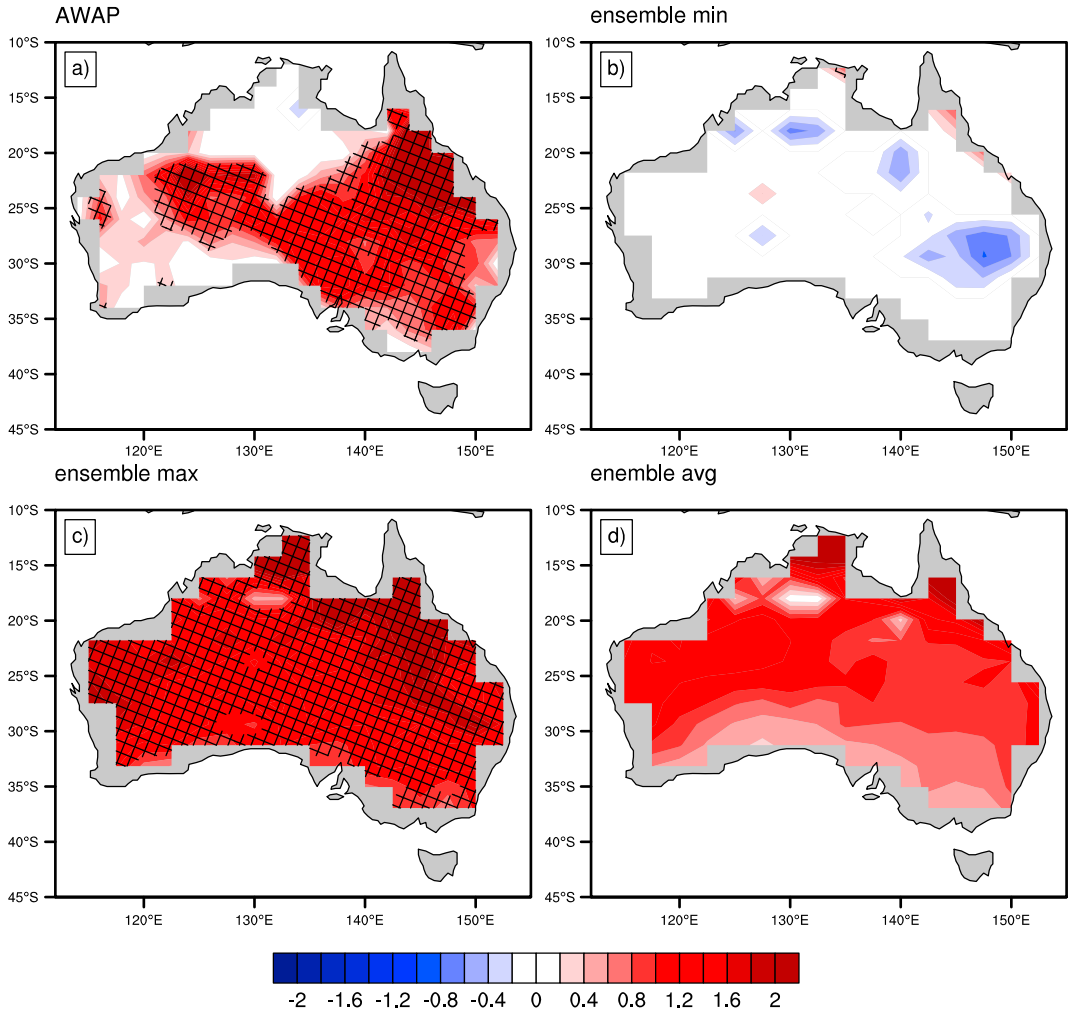

Figure 1. Decadal trends in the number of heat wave day frequency (HWF) measured by the EHF index. HWFs are calculated per summer season as the percentage of days during November–March, where each day belongs to a 3 day (or greater) period where the EHF (equation (3)) is positive. Trends are calculated by the nonparametric Sen's slope estimator. (a) Trend in the AWAP observational data set, (b) smallest trend at each grid box from the 21-member CESM ensemble historical simulation, (c) highest trend at each grid box from the 21-member CESM ensemble and (d) trend calculated from the CESM ensemble average. Although not displayed, almost all trends are statistically significant in the ensemble average. All trends are from the period 1951–2005. Statistical significance at the 5% level is displayed with hatching in Figures 1a and 1c. No trends were significant in Figure 1b.

details. All simulations share the same model version, emission scenario, and initial conditions except for the atmosphere. After a period of a few years, the simulations differ strongly in their phases of modes of variability of ocean and atmosphere.

2.2. Observational Data

[6] The Australian Water Availability Project (AWAP) data set is employed to evaluate CESM's simulation of historical heat wave day frequencies (HWFs). AWAP uses empirical interpolation and function fitting to construct daily gridded observations from *all* available stations that are managed by the Australian Bureau of Meteorology across the Australian continent [*Jones et al.*, 2009]. Since the resolution of CESM is $1.875° \times 2.5°$, AWAP was interpolated to the same resolution via a bilinear method. AWAP and CESM were compared for 1950–2005, which equates to 55 Austral summers.

2.3. Measuring Heat Waves

[7] *Perkins and Alexander* [2013] describe a comprehensive framework in which to measure heat waves considering multiple definitions and characteristics, this study focuses on just one of each. Here we use the extreme heat factor (EHF), from which the number of heat wave day frequency (HWF) are examined. Initially described by *Nairn and Fawcett* [2013], the EHF is based on two excess heat indices:

$$\mathrm{EHI(accl.)} = (T_i + T_{i\text{-}1} + T_{i\text{-}2})/3 - (T_{i\text{-}3} + \ldots + T_{i\text{-}32})/30 \quad (1)$$

$$\mathrm{EHI(sig.)} = (T_i + T_{i\text{-}1} + T_{i\text{-}2})/3 - T_{90} \quad (2)$$

[8] Where T_i is the average daily temperature for day i, and T_{90} is the calendar day 90th percentile. The average daily temperature is defined as the average T_{min} and T_{max} within

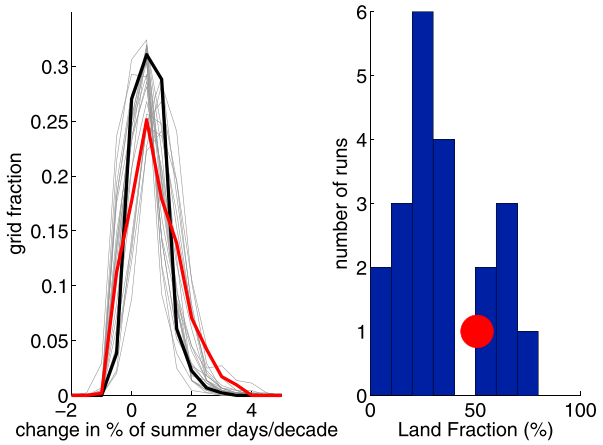

Figure 2. Probability density functions of trend magnitude for the Australian continent (left). Individual ensemble members are grey, the ensemble average is black, and the observed is red. Histogram showing the number of CESM members with significant trends for various land fractions (right). The red dot is the observed, where 50% of Australia displays significant (increasing) trends. The CESM ensemble average (not shown) is approximately 95%.

at 24 h cycle (9 A.M.–9 A.M.). Equations (1) and (2) are then combined to derive EHF:

$$\mathrm{EHF} = \max[1, \mathrm{EHI(accl.)}] \times \mathrm{EHI(sig.)} \quad (3)$$

[9] Equation (1) represents an acclimatization factor where a 3 day average is compared to the previous 30 days, and equation (2) compares the same 3 day average to a climatological threshold. Note that this 3 day average is moving and is unique to each day. See *Nairn and Fawcett* [2013] and *Perkins and Alexander* [2013] for a full explanation. Also note that here a 15 day window calendar day 90th percentile (based on 1961–1990) was used for equation (2) and not the original climatological 95th percentile, as used by *Nairn and Fawcett* [2013]. This means that each day of the year has a unique 90th percentile value, calculated from all respective dates between 1961 and 1990, ± a 7 day window around this date. For each austral summer (November–March), HWF was calculated by counting the number of days where positive EHF conditions exist for 3 or more consecutive days. The summation of all days was then converted to a percentage of heat wave days relative to all summer days during the November–March period.

[10] Trends in HWF were calculated per decade using Sen's Kendall slope estimator, which is nonparametric, robust against outliers, and nonnormally distributed data [*Sen*, 1968; *Zhang et al*., 2005]. HWF trends were calculated per grid box, per model run, and for the AWAP data. The statistical significance of trends was computed at the 5% significance level using the nonparametric Mann-Kendall test.

3. Results

3.1. Spatial Trend Patterns

[11] Figure 1 presents the spatial decadal trend of HWF for AWAP, the lowest and highest trend per grid box across the 21-member CESM ensemble, and the CESM ensemble average trend. In this study, ensemble average refers to the mean HWF calculated from all 21 realizations, per year, and grid box, from which the trend is calculated. It should be made clear that the spatial patterns in Figures 1b and 1c are not physically plausible, as the lowest (highest) trend for one grid box may occur in a different ensemble member than the lowest (highest) trend for a grid box elsewhere over the continent. What is demonstrated, however, is the range of trends the CESM model projects at the grid box level. This reflects the local uncertainty in the trends due to internal climate variability. The lowest trend per grid box (Figure 1b) is between −1% and 1% per decade, thereby the model suggests that despite increasing greenhouse gas concentrations, the overall trend to more HWF could have been masked or even reversed by climate variability. Conversely, Figure 1c shows that the highest trends for the CESM model are positive and statistically significant for almost all of Australia. This implies that the internal climate variability may also have amplified the changes in certain members, leading to pronounced trends that exceed the ensemble average (Figure 1d) by almost a factor 2. The ensemble average trends, in this study, refer to the mean HWF calculated from all 21 realizations, per year, and grid box, from which the trend is calculated. This ensemble average trends, which represents an overall estimate of the models' forced response, display almost identical spatial patterns compared to Figure 1c, although trend magnitudes are generally 0.8% per decade smaller (i.e., regional trends are 1.2 days per decade smaller in Figure 1d than in Figure 1c).

[12] HWF trends in AWAP (Figure 1a) for most of Australia are comparable to either the highest, or average trends from the CESM ensemble. For regions where observed trends are smaller than those presented in Figures 1c and 1d (northwest/southwest Australia), they are greater than the lowest CESM trends presented in Figure 1b. This means that, while it is highly unlikely that one single CESM member can very closely replicate observed conditions, the spatial pattern of the trend present in AWAP falls *within* the range of trends simulated by CESM.

[13] The ensemble average presented in Figure 1d also displays the trend that exists in the absence of noise generally caused by the internal climate variability of each run, assuming that the ensemble of 21 members is large enough. Therefore, Figure 1d also displays the forced signal in HWF, the CESM model projects over Australia and is what would be expected if the variability of the climate system did not influence the occurrence of heat wave days. While the area of increasing trends over central western Australia has been highlighted as an area of caution in the AWAP data set [*King et al*., 2013; *Perkins and Alexander*, 2013], the CESM ensemble suggests that increasing trends in observed HWF are plausible over this area.

3.2. Continental Trends and Time Series

[14] Figure 2 (left) presents probability density functions (PDFs) of the frequency of trend magnitude of HWF over Australia for each individual CESM member (grey), the ensemble average (black), and AWAP (red) for each grid box as a fraction of the total number of grid boxes over Australia. It is notable that despite discrepancies between the model members at the grid point level, the members tend to agree remarkably well in this spatial PDF perspective. All members simulate positive trends at the majority of the land fraction and trends larger than 1% over a substantial fraction of the continent. Similar to Figure 1, it is clear that AWAP

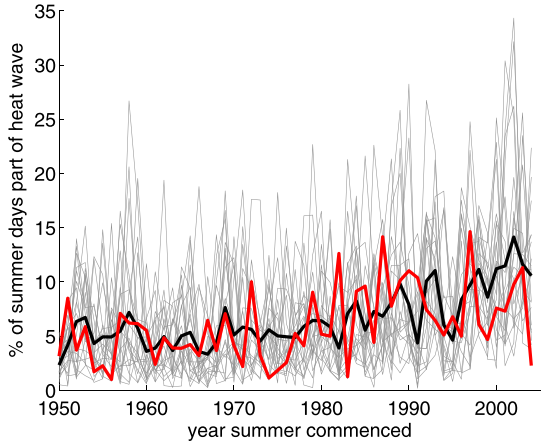

Figure 3. Continentally averaged HWF time series for each CESM ensemble member (grey), the CESM ensemble average (black), and the observations (red).

lies *within* the range of the CESM members in terms of trend magnitude.

[15] Figure 2 (left) also illustrates that the frequency of AWAP trends across Australia is comparable to each CESM realization, as well as the ensemble average. Using the nonparametric Mann-Wilcoxon test [*Mann and Whitney*, 1947] and a 5% significance level, the AWAP PDF is not significantly different from any individual realization or the ensemble PDF. This means that at the continental scale, CESM is producing plausible simulations of both trend magnitude of HWF and the spatial fraction where these trends occur. Individually however, Figure 2 (left) shows that almost all realizations are biased toward either slightly lower or slightly higher mean magnitudes than what is observed in AWAP. This could be due to slightly different mean trends, or the random effect of internal variability in these two groups of realizations; however, the exact quantification of this requires further study. Furthermore, the CESM PDFs, inclusive of the ensemble PDF, are too narrow resulting in overconfident trends that is consistent with the assessment of global mean trends by *van Oldenborgh et al.* [2013].

[16] Figure 2 (right) displays the percentage of land with significant trends in HWF. Ten CESM members produce significant trends for 20%–40% of Australia, and between 50% and 80% of trends are significant for another six members. The remaining five members produce the smallest number of significant trends, which is between 0% and 20%. Similar to Figure 2 (left), if a single CESM simulation or a random subset was chosen, it is likely that AWAP would compare less favorably. Yet this is reduced when considering all 21 members. Just over 50% of grid points in AWAP possess significant trends, once again placing AWAP *within* the range of the CESM ensemble, which is validated by AWAP not being statistically significantly different from the CESM sample. It is interesting to note that for the CESM ensemble, there are two groups of trend significance, those below 40% of the overall land fraction and those above 50%. This could be due to an undersampling of realizations, higher mean warming in the latter group, or different spatial patterns in trends (and the physical causes therein) for each different realization. However, it is not within the scope of the current study to explore these explanations and will be addressed in future research.

[17] Figure 3 displays continentally averaged time series in HWF for each CESM member (grey), the ensemble average (black), and AWAP (red). Similar to Figures 2a and 2b, the AWAP time series is bounded by the CESM realizations and is not statistically distinguishable from any of them. The interannual variability of some CESM realizations, however, is much larger than in AWAP, corresponding to multiple occurrences of HWF percentages in CESM that are much higher than the maximum values in AWAP. Over the period of 1950–2005, Australia-averaged HWF in AWAP ranges from 2% to 15%, whereas the realizations from less than 1% to almost than 35%. This may be interpreted as an indication that substantially more severe heat waves could have been possible than what was observed. This result could also point toward a deficiency of the model in capturing the correct variability in HWF (thereby generating more heat waves at yearly time scales). However, we do not expect this to be the case at the grid point level where the range is rather conservative, since the observational PDF in Figure 2 (left) is wider than for the individual members. Overall, CESM produces a similar change in HWF compared to what has been observed—the ensemble average of CESM (as well as each individual member) produces a similar change (increase) in their occurrence compared to what has been observed since 1975.

4. Discussion and Conclusion

[18] This study evaluated the 21-member CESM model on its capability to simulate summertime heat wave day frequency (HWF) over Australia. CESM simulations with historical anthropogenic and natural forcings were evaluated with observed fields based on the AWAP data set, over the time period 1950–2005. The spatial pattern, magnitude, and significance of trends in CESM were found to be reasonable since the corresponding AWAP trends were *within* the ensemble range, as well as CESM not demonstrating any systematic biases or large variation among ensemble members. However, although AWAP falls within the range of the continentally averaged CESM time series, a consistent overestimation of maximum HWF values and large yearly fluctuations by multiple realizations suggest that the model produces higher interannual variability than what is observed. Moreover, the narrow PDFs of CESM in Figure 2 (left) suggest that the model is overconfident in the estimation of local trends and does not produce enough spatial variability. This could be to some extent a result that the analysis starts at the time of the initial perturbation to the ensemble; however, we find that the spread between ensemble members reaches the full extent over the extratropics within only 2–3 years. There is a marked trend toward increasing HWF values in all simulations and AWAP from 1975 onward.

[19] Having multiple members of a single model allows us to detect this underlying trend, the CESM model projects—the more members that are included, the clearer the underlying trend becomes as the influence of internal climate variability is further reduced. If the CESM model produced no trend in HWF in the absence of this variability, it would be clear in the ensemble average trend (i.e., no trend), and adding additional members would only enhance this result. Although there was no significant difference between AWAP and each CESM realization in Figures 2a and 3, the use of a single realization may result in mistaken conclusions on the capability of the model to simulate changes in heat

wave days. For example, if a realization where continentally averaged HWF is regularly above 15% were randomly selected as an overall representative of the CESM model, one would likely conclude that the model grossly overestimates HWF interannual variability in comparison to AWAP (see Figure 3). Looking at a larger selection of simulations, we see that while there is a tendency for larger variability, it is not as prominent in every realization. This is an important point, since it is quite rare for a large number of single-model realizations to be available. We stress here that while it is important to use all available realizations from one model, if this is only a single or handful of simulations, then the expectation placed on the quantitative evaluation of extreme events should be reduced.

[20] There are other interesting findings from this study. First, the forced (anthropogenic) signal in CESM is toward increasing trends (Figure 1d) in the number of heat wave days. Second, the range of trends detected spatially across the CESM members illustrates the influences internal climate variability has on the occurrence of heat wave days—each realization has its own internal climate variability, resulting in different fluctuations in both space and time (Figures 1 and 3). Third, this range of variability among the CESM realizations does not have to be viewed as a deficiency of the model. Rather it suggests that at many grid points, the trends could have been substantially larger or smaller than observed as a result of climate variability, that is, the occurrences of heat waves observed over Australia during 1950–2005 are likely to have been different, should different phases of climate variability have occurred. This result is not endemic to extreme events and is true for other climate variables, such as mean temperature [*Foster and Rahmstorf*, 2011] This has strong implications for the future as it indicates that heat waves may locally not increase for decades or alternatively change dramatically at a rate much greater than expected from the long-term trend. Moreover, the CESM ensemble provides a platform to investigate the role of various modes of internal climate variability on heat waves, since multiple realizations are not possible from observations. This work is intended for future research. The model further suggests that future summers with substantially more heat wave days are possible than observed so far.

[21] **Acknowledgment.** This project was funded by the ARC Centre of Excellence for Climate System Science, grant number CE110001028.

[22] The Editor thanks two anonymous reviewers for their assistance in evaluating this manuscript.

References

Clark, R. T., J. M. Murphy, and S. J. Brown (2010), Do global warming targets limit heatwave risk?, *Geophys. Res. Lett.*, *37*, L17703, doi:10.1029/2010GL043898.

Coumou, D., and A. Robinson (2013), Historic and future increase in the global land area affected by monthly heat extremes, *Environ. Res. Lett.*, *8*, 0–6, doi:10.1088/1748-9326/8/3/034018.

Coumou, D., A. Robinson, and S. Rahmstorf (2013), Global increase in record-breaking monthly mean temperatures, *Clim. Change*, *118*(3-4), 771–782.

Diffenbaugh, N. S., and M. Ashfaq (2010), Intensification of hot extremes in the United States, *Geophys. Res. Lett.*, *37*, L15701, doi:10.1029/2010GL043888.

Diffenbaugh, N. S., and M. Scherer (2011), Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries, *Clim. Change*, *107*, 615–624, doi:10.1007/s10584-011-0112-y.

Fischer, E. M., U. Beyerle, and R. Knutti (2013), Spatial aggregation reveals robust projections in climate extremes, *Nat. Clim. Change*, doi:10.1002/2013GL057833, accepted.

Foster, G., and S. Rahmstorf (2011), Global temperature evolution 1979–2010, *Environ. Res. Lett.*, *6*, 44022, doi:10.1088/1748-9326/6/4/044022.

Gent, P. (2011), The Community Climate System Model version 4, *J. Clim.*, *24*, 4973–4991, doi:10.1175/2011JCLI4083.1.

Hansen, J., M. Sato, and R. Ruedy (2012), Perception of climate change, *Proc. Natl. Acad. Soc.*, *109*(37), E2415–E2423.

Jones, D. A., W. Wang, and R. Fawcett (2009), High-quality spatial climate data-sets for Australia, *Aust. Meteorol. Oceanogr. J.*, *58*, 233–248.

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

King, A. D., L. V. Alexander, and M. G. Donat (2013), The efficacy of gridded data to examine extreme rainfall characteristics: A case study for Australia, *Int. J. Climatol.*, *33*, 2376–2387.

Mann, H. B., and D. R. Whitney (1947), On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Stat.*, *18*, 50–60.

Nairn, J., and R. Fawcett (2013), Defining heatwaves: Heatwave defined as a heat impact event servicing all community and business sectors in Australia. CAWCR Technical Report No. 060.

Perkins, S. E., and L. V. Alexander (2013), On the measurement of heatwaves, *J. Clim.*, *26*, 4500–4517, doi:10.1175/JCLI-D-12-00383.1.

Sen, P. K. (1968), Estimates of the regression coefficient based on Kendall's tau, *J. Am. Stat. Assoc.*, *63*, 1379–1389.

Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, and D. Bronaugh (2013a), Climate extreme indices in the CMIP5 multi-model ensemble. Part 1: Model evaluation in the present climate, *J. Geophys. Res. Atmos.*, *118*, 1716–1733, doi:10.1002/jgrd.50203.

Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, and D. Bronaugh (2013b), Climate extreme indices in the CMIP5 multi-model ensemble. Part 2: Future climate projections, *J. Geophys. Res. Atmos.*, *118*, 2473–2493, doi:10.1002/jgrd.50188.

van Oldenborgh, G. J., F. J. Doblas Reyes, S. S. Drijfhout, and E. Hawkins (2013), Reliability of regional climate model trends, *Environ. Res. Lett.*, *8*, 014055, doi: 10.1088/1748-9326/8/1/14055.

Zhang, X., et al. (2005), Trends in Middle East climate extremes indices during 1950–2003, *J. Geophys. Res.*, *110*, D22104, doi:10.1029/2005JD006181.