The representation of health-relevant heatwave characteristics in a Regional Climate Model ensemble for New South Wales and the Australian Capital Territory, Australia

Mia H. Gross,a* Lisa V. Alexander,a Ian Macadam,a,b Donna Greena and Jason P. Evansa

a Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, Australia
b Met Office, Exeter, UK

ABSTRACT: Heatwaves have been linked to increased rates of human mortality and morbidity. Given these adverse health impacts, it is crucial to improve our understanding of future changes in these extreme events to inform health impacts studies and adaptation planning. While this information would be most beneficial at a local scale, Global Climate Models provide projections on much coarser resolutions. Regional Climate Models, such as those used in the New South Wales/Australian Capital Territory Regional Climate Modelling (NARClim) project, provide simulations at a finer scale more appropriate for regional assessments. This study uses NARClim output to investigate the ability of a Regional Climate Model ensemble to represent heatwave characteristics through the Excess Heat Factor, an index that includes factors that are known to be important to the heat-health relationship. Both uncorrected and bias-corrected model output is evaluated against observationally-derived heatwave characteristics for the period 1990–2009. The effect of bias-correction on future changes in heatwave characteristics is also assessed. Overall, the simulations provided a good representation of the recent climate and bias-correction did not greatly change simulated heatwave characteristics. Some regions were more affected by bias-correction than others, with bias-correction being most beneficial for coastal regions. We emphasise that these results may not apply to all indices measuring extreme heat, and demonstrate that results for an index based on a fixed absolute temperature threshold are substantially affected when bias-correction is applied. While supporting bias-correction, this study demonstrates that it is not necessarily required when evaluating a relative measure such as the Excess Heat Factor.

KEY WORDS heat extremes; excess heat factor; bias-correction; downscaling; WRF; GCM; model evaluation; Australia

Received 1 December 2015; Revised 8 April 2016; Accepted 10 April 2016

1. Introduction

Heatwaves can have substantial effects on various factors relating to the natural world and society, including human health (Victorian Government Department of Human Services, 2009). For example, it is estimated that between 25,000 and 70,000 deaths were attributable to the European summer heatwave in 2003 (D’Ippoliti et al., 2010). Even in Australia, where hot days are relatively common, runs of days with excess temperatures can still be a problem for human health. In the summer of 2009, 374 excess deaths occurred during a heatwave in the southeast of the continent (Victorian Government Department of Human Services, 2009; PwC, 2011). As climate change is increasing the occurrence of extremely hot temperatures in Australia (Bureau of Meteorology and CSIRO, 2014), and Global Climate Models (GCMs) point to future increases in the duration, frequency and intensity of heatwaves across the continent (Alexander and Arblaster, 2009; Cowan et al., 2014), it would be useful to understand how heatwaves may change at a regional scale. This information would be useful for health policymakers to enable them to better plan for future climate change impacts.

Previous studies using GCMs (e.g. Kiktev et al., 2009; Sillmann et al., 2013) concluded that temperature extremes are simulated with reasonable confidence when compared with observations on a global scale, as well as on continental scales, such as over Australia (Alexander and Arblaster, 2009; Cowan et al., 2014). While GCMs can be useful in such studies, they can lack important regional detail. For this reason, Regional Climate Models (RCMs) are often used to dynamically downscale coarse-resolution GCM output, typically at a 100–300 km grid spacing, to finer resolutions (e.g. 10–50 km) (Evans and McCabe, 2010; Evans et al., 2012). Dynamical downscaling allows for local climate processes to be represented, which can have substantial impacts on aspects of the weather and climate, including heatwaves.

Biases are inherent to the output of dynamical downscaling. GCM output contains biases that are inherited by RCM simulations and RCMs themselves introduce biases. This contributes to uncertainty in dynamically downscaled climate projections. Multi-model ensembles that use a range of different GCMs to force a range of different
RCMs can be used to sample this modelling uncertainty, with the aim of providing information on the robustness of results to model imperfections (Evans et al., 2012). Owing to computational expense, it is impractical to downscale all GCMs with all RCMs, and thus completely sample uncertainties arising from these imperfections. However, models can be carefully selected to maximise the sampling of uncertainty.

Recently, multiple RCM simulations were performed for southeastern Australia as part of the New South Wales/Australian Capital Territory Regional Climate Modelling (NARCliM) project (Evans et al., 2013, 2014). NARCliM used dynamical downscaling to provide regionally detailed data to inform climate impacts and adaptation work across New South Wales (NSW) and the Australian Capital Territory (ACT). The process of model selection for NARCliM was to identify and exclude RCMs and GCMs that perform poorly, as well as selecting those that appear most independent of one another and, as a result, have different imperfections (Evans et al., 2014). GCM performance was assessed across a range of metrics, including those related to the reproduction of observed climatological average conditions over Australia and the simulation of relevant modes of climate variability (e.g. the El Niño-Southern Oscillation) (Evans and Ji, 2012a). An additional condition for GCM selection was that they spanned the range of future changes in temperature and precipitation present in the full GCM ensemble. This led to a wide range of possible outcomes in the climate simulations. RCM performance was assessed by examining the simulation of a set of representative storm systems as well as an extreme fire weather event (Evans and Ji, 2012b).

Some applications of RCM output, such as for forcing hydrological and agricultural models in climate change impacts studies, require that biases in the output be corrected (Ashfaq et al., 2010). Biases can be characterised by assessing raw model output against observations, allowing adjustments to be made accordingly (Ashfaq et al., 2010). By design, bias-correction increases the resemblance of RCM output to observations, and this resemblance is important for many impacts studies (e.g. Ashfaq et al., 2010; Piani et al., 2010; PalMazumder and Done, 2015; Macadam et al., 2016). Although there are limitations associated with bias-correction, for example, the reliance on high-quality observations and uncertainties in bias-corrected output for future projections (Piani et al., 2010), previous work has found bias-corrected model output to be robust in simulating extremes (e.g. PalMazumder and Done, 2015). Recognising the utility of bias-corrected data for climate impacts studies, the NARCliM project produced bias-corrected datasets for temperature and precipitation (Evans et al., 2014).

A number of indices related to temperature extremes and heatwaves have been developed. The Excess Heat Factor (EHF) (Nairn and Fawcett, 2013, 2015) is a percentile-based index that can be used to describe the severity of a heatwave event with respect to the burden on human health (Langlois et al., 2013; Nairn and Fawcett, 2015). It provides an effective measure of the heat–health relationship by accounting for acclimatisation and heat stress, and has recently been shown to outperform other heat stress indices in determining heat-health relationships in Perth, Australia (Scalley et al., 2015). More traditional indices based on absolute, or fixed, thresholds, have also been used to assess extreme heat events. SU30 (Alexander and Herold, 2015), is an index which measures the number of days where temperatures exceed 30°C annually. It is also relevant to the human health sector (WMO, 2012), although this index does not explicitly include factors that are known to be important to the heat–health relationship, such as those included in the EHF, which take into account acclimatisation and consecutive days of hot night-time temperatures (Nairn and Fawcett, 2015).

This paper assesses how suitable uncorrected and bias-corrected RCM output are for providing information on health-relevant heatwave characteristics. We evaluate EHF data derived from uncorrected and bias-corrected NARCliM output against corresponding data derived from observations of the recent climate in NSW and the ACT. SU30 data are similarly derived to compare with EHF results. Further, we compare future changes in heatwave characteristics derived from uncorrected and bias-corrected model output to assess the impact of bias-correction on projected future changes in heatwaves.

2. Data and methods

2.1. Observational data

We used the Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) dataset (Trewin, 2013) to assess the ability of an RCM ensemble to simulate the recent climate. This data contains daily maximum and minimum temperature observations for 112 meteorological observing stations across Australia that have been quality controlled and adjusted to account for inhomogeneities through time. Although this does not mean that ACORN-SAT is free of errors, it does represent the highest quality station temperature dataset available for Australia. Of the 112 ACORN-SAT stations, 25 are located in NSW and the ACT (see Figure 1). To aid interpretation of the results, the study region was divided into three sub-regions: coastal, central-northern and central-southern.

Point data, such as the ACORN-SAT dataset, may not be characteristic of the area represented by gridded climate model output (Zhang et al., 2011) and so we also consider gridded data from the Australian Bureau of Meteorology operational daily temperature analyses known as the Australian Water Availability Project (AWAP) (Jones et al., 2009). This includes a 0.05° × 0.05° (5 × 5 km) gridded product containing daily maximum and minimum temperature. We extracted data from the dataset for the grid points closest to the ACORN-SAT stations.

2.2. Climate model output

Daily maximum and minimum temperatures for the NSW and ACT region were extracted from the NARCliM
simulations in a similar way to the gridded AWAP dataset, taking data from the land grid point closest to each of the ACORN-SAT stations. In order to preserve the magnitude of temperature extremes, interpolation between grid points was not performed, as this would have smoothed the data.

NARCliM uses a 12-member GCM/RCM ensemble. Four GCMs from the Coupled Model Inter-comparison Project phase 3 (CMIP3) (i.e. CCCMA3.1, CSIRO-MK3.0, ECHAM5, MIROC3.2) (Meehl et al., 2007) are downscaled using three different configurations of the Weather Research and Forecasting (WRF) RCM (Evans et al., 2014), designated ‘R1’, ‘R2’ and ‘R3’. Each configuration uses different combinations of parameterisation schemes for cumulus convection, cloud microphysics, radiation and the boundary layer (Evans and Ji, 2012b). NARCliM uses WRF to simulate the climate of the region for three 20-year time slices, 1990–2009 (recent climate), 2020–2039 (near-future) and 2060–2079 (far-future) (Evans et al., 2014). Projections for the two future periods are based on the Intergovernmental Panel on Climate Change’s Special Report Emissions Scenarios (SRES) A2 high emissions scenario (Nakićenović and Swart, 2000). We recognise that the GCMs may have improved between the older generation CMIP3 models and the newer CMIP5 model ensembles (Sillmann et al., 2013), and so a successor project to NARCliM may benefit in using the most updated GCMs.

The GCM simulations are first downscaled to approximately 50 km grid spacing over the whole of Australia and the surrounding ocean. The 50 km WRF simulations are then used to provide boundary conditions for 10 km WRF simulations covering NSW and ACT. For this paper we chose to analyse these ~10 km gridded simulations for the climate of the recent past and far-future periods.

The NARCliM project bias-corrected daily temperature data for each individual model grid point separately, using the AWAP dataset (Evans and Argüeso, 2014). Note that in theory, the ACORN-SAT dataset could have been used; however, other applications of NARCliM output required spatially complete bias-corrected temperature and precipitation data, which could not have been delivered with the ACORN-SAT data. An inverse distance weighting method was used to resample the 5 km gridded AWAP data to the ~10 km NARCliM grid. Cumulative distribution functions (CDFs) were fitted to daily temperatures from the simulations and from the observations for the period 1990–2009 assuming a Gaussian distribution. Following other studies (e.g. Piani et al., 2010), the NARCliM project chose a Gaussian distribution to fit the temperature data, which is both simple and could be applied to the entire NARCliM domain, for the entire year. We recognise that this may not be the most suitable distribution for all seasons and locations across the study region, and further work on its applicability is recommended. Transfer functions mapping the simulated CDFs to the observed CDFs were derived and used to correct the NARCliM data for all three time slices. An underlying assumption is that biases do not change fundamentally in nature between the recent and future climate time periods (Evans et al., 2013).
2.3. Calculation of extreme heat indices

We calculate the EHF from the observational data and NARCliM output. The EHF combines two excess heat indices (EHIs). EHI\text{\textsubscript{sig}}, the difference between a 3-day average of daily mean temperature and a high percentile over a climatological reference period, expressed as an anomaly. Traditionally, the 95th percentile is used (e.g., Perkins and Alexander, 2013; Nairn and Fawcett, 2015), however in this study, we use the 90th percentile following Perkins \textit{et al.} (2015), also implemented in the software used for the index calculations. The percentile is seasonally varying, so when we use the term ‘heatwave’, we are referring to heat relative to the time of year. By using the whole year, more robust statistics could be calculated, while the sample size only using summer months would have been too small, with very few heatwave-worthy days (i.e., \(\sim 7.5\) days for a 150-day summer). The second index, EHI\text{\textsubscript{accl}}, calculates the difference between daily temperature over a 3-day average period and the mean daily temperature for the preceding 30 days (Nairn \textit{et al.}, 2009; Nairn and Fawcett, 2013). EHI\text{\textsubscript{accl}} also expressed as an anomaly, represents expected acclimatisation by considering the immediate days preceding the heatwave event (Langlois \textit{et al.}, 2013). The EHIs are derived as follows:

\[
\text{EHI}_{\text{sig}} = \left( T_i + T_{i-1} + T_{i-2} \right) / 3 - T_{90}\% \quad (1)
\]

\[
\text{EHI}_{\text{accl}} = \left( T_i + T_{i-1} + T_{i-2} \right) / 3 - \left( T_{i-3} + \ldots + T_{i-30} \right) / 30 \quad (2)
\]

where \(T_i\) is the daily mean temperature for day \(i\), calculated by taking the average of daily maximum and minimum temperature, and \(T_{90}\%\) is the 90th percentile of daily mean temperature for the time period being assessed calculated using a 15-day moving window centred on day \(i\). Both (1) and (2) are calculated in \(^\circ\text{C}\). The EHF is the product of (1) and (2), and is thus calculated in \(^\circ\text{C}^2\):

\[
\text{EHF} = \text{EHI}_{\text{sig}} \times \max \left( 1, \text{EHI}_{\text{accl}} \right) \quad (3)
\]

Nairn and Fawcett (2013, 2015) define heatwaves when the EHF is positive. We extend the Nairn and Fawcett definition in order to be able to describe various heatwave characteristics such as intensity, frequency and duration, described below. Following Perkins and Alexander (2013), for a positive EHF value to occur on day \(i\), the average temperature over a 3-day period, i.e., day \(i\), day \(i-1\), day \(i-2\), must be above \(T_{90}\%), which does not necessarily imply that all days within the 3-day period will be ‘extreme’. For a heatwave to be defined, at least 3 consecutive days must have EHF > 0. This means that for a 3-day heatwave to occur, for example, day \(i\), day \(i+1\) and day \(i+2\) must all have positive EHF values.

Similar to the methods adopted by Perkins and Alexander (2013), based on Fischer and Schär (2010), we use the EHF data to calculate annual values of five heatwave characteristics:

1. HWA: the peak EHF value from the hottest heatwave event of the year.
2. HWM: average annual magnitude of heatwaves.
3. HWN: annual number of heatwave events.
4. HWF: annual frequency of days contributing to heatwave events.
5. HWD: the duration of the longest annual heatwave event.

To facilitate subsequent discussion, these metrics have been grouped into intensity indices (HWA and HWM) and duration/frequency indices (HWN, HWF and HWD).

The five EHF-heatwave characteristics were calculated using the ClimPACT software package (Alexander and Herold, 2015), in which some additional quality control was performed on the ACORN-SAT data. The 1991–2008 climatological reference period was used to calculate 90th percentiles of daily mean temperature. These were calculated separately for the ACORN-SAT and AWAP datasets, and each of the NARCliM 1990–2009 simulations. For each individual NARCliM GCM/RCM combination, the 90th percentile values for the 1990–2009 simulation were used in the calculation of EHF values for both the 1990–2009 and 2060–2079 simulations. Implicit in this choice is the assumption that current levels of acclimatisation to temperature extremes will persist into the future. We recognise that increased acclimatisation and adaptation to temperature extremes could occur in the future as the climate warms, but consideration of this is beyond the scope of this study.

A statistic was sought to summarise the heatwave characteristics so that the NARCliM simulations and observations could be compared, and changes in heatwaves between recent and future NARCliM time periods characterised. The mean of the annual index values over the 20-year periods analysed was used for this purpose. In ClimPACT, all negative values are set to zero (Alexander and Herold, 2015). The mean provides a simple, yet robust, analysis of the limited samples of data available. In addition, the median was calculated to assess the robustness of the obtained mean values to outliers.

In evaluating the performance of the 12 NARCliM runs against the observations, errors were calculated by subtracting the ACORN-SAT mean values from the model mean values. Analysis was only undertaken for stations for which at least 80% of the data were available within ACORN-SAT. Seven stations were therefore excluded for analysis of the recent climate due to data incompleteness. Differences in the mean values between 1990–2009 and 2060–2079 were calculated to assess future changes in heatwaves. All stations were included when calculating future changes, as these calculations were not limited by the availability of observations.

An additional index using a fixed absolute temperature threshold was included to compare against the EHF. Annual values of the SU30 index were calculated from daily maximum temperatures. As for the EHF, stations missing more than 20% of data were omitted from the analysis.
3. Results

3.1. Observations

Figure 2 displays annual values of heatwave characteristics averaged over the 1990–2009 period calculated from the available stations in the ACORN-SAT daily temperature dataset. The central-northern and central-southern regions (see Figure 1) generally show more intense heatwaves (HWA and HWM, Figure 2(a) and (b)) than coastal stations, with certain inland stations (e.g. Deniliquin, station 25 in Figure 2) experiencing substantially hotter heatwaves. Most of the state and territory experienced an average of greater than 1.2 annual heatwave events (HWN, Figure 2(c)). Southern coastal stations tended to experience a greater number of heatwave events than more northern stations. Southern coastal stations also show a greater total number of heatwave days than northern coastal stations (HWF, Figure 2(d)). Mean heatwave frequency values range from 6 to 7 heatwave days annually for southern coastal stations, such as Moruya Heads (station 1 in Figure 2) and Sydney (station 4), to 4–5 heatwave days for northern coastal stations, such as Port Macquarie (station 7) and Coffs Harbour (station 8). The average duration of the longest heatwave event in each year (HWD, Figure 2(e)) is shorter along the coast than inland, with the average duration of the longest heatwaves being less than 2 days in Williamtown (station 6). Though the EHF calculation defines heatwaves as being at least 3 days long, 1990–2009 mean values of HWD of less than three are possible as there are years in which no heatwaves occur.

Overall, results indicate more intense and longer heatwaves for central regions of NSW and the ACT than the coast, with southern coastal stations generally showing greater values across all metrics than northern coastal stations.

3.2. Uncorrected NARCliM simulations for the recent climate

The subsequent discussion refers to the uncorrected data for the 1990–2009 NARCliM time-slice. Figure 3 shows the 1990–2009 mean values of each index for the 12 model runs (solid lines) compared to ACORN-SAT observations (thick dashed line) and AWAP 5 km gridded observations (dotted line). Solid lines of the same colour represent the three RCM runs forced with the same GCM. RCM runs using the same GCM tend to cluster together, while the four GCMs sometimes follow different paths, indicating that mean values for each of the heatwave characteristics differ more between GCMs than between RCMs.

ACORN-SAT and AWAP observations agree relatively well for both intensity-related metrics (Figure 3(a) and (b)). Model output for these metrics show a tendency towards overestimation for the coastal region when compared with observations. The central regions do not show an overall positive intensity bias. However, ECHAM5 consistently displays more intense heatwaves on average than the observations. For example, the greatest mean value in ECHAM5 for HWA (Figure 3(a)) is 31.8°C², occurring in the central-southern station of Wyalong (station 21), compared with much lower values in ACORN-SAT and AWAP, 19 and 23.7°C², respectively.

Except for the most southern coastal stations, AWAP values are greater than ACORN-SAT values for the duration/frequency-related metrics (Figure 3(c) and (e)). For these metrics, model output tends to agree more closely with ACORN-SAT than AWAP, although the models generally give lower values than both observational datasets. This is discussed further in Section 4.1. ECHAM5 generally displays the highest mean values compared with the other GCMs for these indices, often overestimating ACORN-SAT values for some stations, particularly in the north.

The median was also assessed to account for possible outliers affecting calculation of the mean (Figure S1, Supporting Information). Results indicated that the median values behaved in a similar way to the mean values, suggesting that outliers did not affect the results.

Figure 3 shows the similarities in results between the different intensity-related metrics and between the different duration/frequency-related metrics. For the sake of brevity, therefore, the amplitude of the hottest day in the hottest yearly heatwave event (HWA) and the annual sum of heatwave days (HWF) were chosen as representative measures for the subsequent sections.

Figures 4 and 5 show the magnitude of error for all 12 model runs against the ACORN-SAT observational data for HWA and HWF. As also shown in Figure 3, the three RCM runs forced with each GCM are reasonably similar, at least in terms of sign, while the four GCMs indicate greater variation.

Figure 4 shows that the models generally overestimate HWA along the coast. Some simulations underestimate the observations for central regions (e.g. CSIRO-MK3.0 R1, R2 and MIROC3.2 R1, R2, R3). The ECHAM5-forced simulations overestimate the observations, and generally indicate the greatest magnitude of error, with coastal stations showing an error of around 10 to 20°C². Overall, greater positive errors occur in coastal stations compared with the rest of NSW and the ACT, indicating that the models are simulating more intense heatwaves along the coast compared with observations. CSIRO-MK3.0 and MIROC3.2 simulations mostly have the smallest errors in coastal stations, of less than 5°C² in some instances.

Figure 5 shows that, in general, the models underestimate HWF compared with the observations. However, the ECHAM5 simulations show widespread positive biases in the north. For most simulations, excluding the ECHAM5 simulations, southern coastal stations (e.g. Moruya Heads, station 1) indicate greater errors than northern coastal stations (e.g. Coffs Harbour, station 8). Some simulations (e.g. ECHAM5 R1, R2, R3 and CSIRO-MK3.0 R1, R2) slightly overestimate heatwave frequency in these northern coastal stations, while all 12 model runs underestimate the observations for southern coastal stations.

Figure 2. Mean values of EHF-heatwave characteristics for ACORN-SAT stations where at least 80% of daily maximum and minimum temperature is available over the period 1990–2009. The numbers on the map are the station numbers assigned in Figure 1.

© 2016 Royal Meteorological Society

Int. J. Climatol. (2016)
Figure 3. Line plots of mean values of each EHF-heatwave metric for the period 1990–2009 calculated from the uncorrected ~10 km NARCliM dataset, with ACORN-SAT observations (thick dashed line) and AWAP 5 km gridded observations (dotted line). Station numbers correspond with numbers from Figure 1.
3.3. Comparison of uncorrected and bias-corrected datasets

Comparisons between the uncorrected and bias-corrected datasets inform discussion of the value of bias-correction in Section 4.2. The subsequent figures represent data for bias-corrected NARClIM output plotted against data for uncorrected output. Each of the available stations is represented by a symbol indicating the GCM used. Equivalent figures for the heatwave characteristics not shown are included as Figures S2–S4.

Figure 4. Model error in the peak EHF value from the hottest heatwave event of the year, calculated by subtracting ACORN-SAT mean values from uncorrected NARClIM data. Station numbers correspond with numbers from Figure 1.
3.3.1. Heatwave amplitude

The scatter plots in Figure 6 show the model error against ACORN-SAT observations for the recent climate (a–c) and the corresponding future changes (d–f) for the hottest day of the hottest yearly heatwave event, separated by region. While many of the data points cluster around the diagonal, indicating similar values between bias-corrected and uncorrected datasets, some stations and simulations show differences between the datasets.

It is clear that bias-correction reduces the error for the hottest day of the hottest yearly heatwave event for coastal stations (Figure 6(a)) more than inland
Figure 6. Scatter plots of the peak EHF value from the hottest heatwave event of the year, comparing bias-corrected and uncorrected datasets of (a–c) the recent climate model errors (i.e. model mean – ACORN-SAT mean) and (d–f) future changes (2060–2079 relative to 1990–2009) by region.

stations (Figure 6(b) and (c)). This could imply that the bias-corrected output for future changes is more reliable, where coastal stations show smaller changes in heatwave amplitude when bias-corrected (Figure 6(d)). In this case, the most negative biases in the recent climate also produce the smallest increase in the future, with bias-correction tending to narrow the distribution. Central-northern stations are largely unaffected by bias-correction for the recent climate (Figure 6(b)). However, future changes in this region (Figure 6(e)) show that bias-correction does affect model output for some GCMs. For example, some runs within CCCMA3.1 and MIROC3.2 are projecting larger increases in heatwave amplitude when the data is bias-corrected, compared with uncorrected output. Unlike
coastal stations where the distribution is narrowed, the CCCMA3.1 model produces the lowest errors during the recent climate, but largest increases for future changes. In this case, bias-correction tends to enhance some of the future changes, expanding the distribution. Recent climate data for the central-southern stations (Figure 6(c)) are also mostly unaffected by bias-correction. However, bias-corrected output in some ECHAM5 simulations slightly reduces the error when the models overestimate the observations. Future changes simulated by ECHAM5 indicate that bias-correction produces slightly smaller increases than uncorrected output (Figure 6(f)). This might suggest that uncorrected values slightly overestimate future changes in heatwave amplitude.

3.3.2. Heatwave frequency

Figure 7 shows bias-corrected and uncorrected model output for heatwave frequency, for both model error against ACORN-SAT observations in the recent climate (a) and future changes (b). In Figure 7(a) almost all data points lie close to the diagonal, indicating that bias-correction makes little difference in error for many stations and simulations for the recent climate. The largest difference between the datasets shows bias-correction reducing the model error by 0.9 days for the northern coastal station of Port Macquarie (station 7) (ECHAM5, R3, upper right-hand quadrant), compared with the equivalent uncorrected simulation. The second largest difference between datasets occurs for the central-northern station of Inverell (station 10) (CCCMA3.1, R1, bottom left-hand quadrant), where the bias-corrected simulation shows a slight increase in error compared with the uncorrected run, of less than 1 day.

As for the recent climate errors, future changes in heatwave frequency (Figure 7(b)), are largely unaffected by bias-correction, as shown by clustering around the diagonal. Bias correction results in some small increases in future increase in heatwave frequency for certain CCCMA3.1 simulations. The largest difference occurs for the northern coastal station of Port Macquarie (station 7), with the bias-corrected CCCMA3.1 (R1) simulation projecting an increase of 41.6 heatwave days, compared with the equivalent uncorrected run projecting an increase of 37.1 days. The second largest difference, also for Port Macquarie, shows a projected increase of 3.45 more

Figure 7. Scatter plots of the annual sum of heatwave days, comparing bias-corrected and uncorrected datasets of (a) the recent climate model errors (i.e. model mean – ACORN-SAT mean) and (b) future changes (2060–2079 relative to 1990–2009), and similarly, (c) recent climate model errors and (d) future changes for the number of days where maximum temperatures exceed 30 °C.
days in the bias-corrected simulation compared with the uncorrected run (CCCMA3.1, R2). Bias-correction has minimal impact on the two simulations showing the largest increases in heatwave frequency. These correspond with the central-northern stations of Inverell (station 10) and Moree (station 11), projecting increases of 51.35 and 52.95 days in uncorrected simulations, and 51.80 and 53.15 days for the equivalent bias-corrected simulations.

3.4. Evaluation of uncorrected and bias-corrected datasets for a fixed-threshold index

While the results in the previous section show that the EHF heatwave characteristics derived from model output are mostly unaffected by bias-correction, it is important to acknowledge that other types of indices may be strongly influenced when bias-correction is applied. Figure 8 shows a comparison between uncorrected and bias-corrected output for the fixed-threshold index, SU30. For both the recent climate model errors (Figure 7(c)) and future changes (Figure 7(d)), there is a clear distinction between bias-corrected simulations and uncorrected simulations.

Overall, Figure 7(c) shows that model error for the recent climate tends to be closer to zero when bias-correction is applied compared with the uncorrected output. In some cases the errors are approximately the same for both the bias-corrected and uncorrected simulations, but the majority of simulations show that bias-correction produces values that resemble the ACORN-SAT observations more closely. There are also some instances where bias-correction increases the error compared with uncorrected simulations, as seen in certain CSIRO-MK3.0, ECHAM5 and MIROC3.2 simulations in the upper right-hand quadrant above the diagonal.

Figure 7(d) shows that a future increase in mean SU30 is projected for all stations by all simulations. Many stations and simulations show different magnitudes of increases when model output is biased. Although bias-correction often results in smaller increases in SU30, the largest differences between bias-corrected and uncorrected data occur when larger increases are projected by the bias-corrected data. The largest increases in magnitude of change occur for data from the CSIRO-MK3.0 and MIROC3.2 simulations. Although bias-correction affects some simulations more than others (e.g. most of the ECHAM5 data lie close to the diagonal, indicating that bias-correction does not greatly affect the model output), it is apparent that bias-correction can substantially affect SU30 values derived from climate model output.

4. Discussion

4.1. Climate model evaluation

Projections of heatwaves and other climate extremes rely on the ability of climate models to realistically simulate the observed climate. The NARCliM-derived EHF-heatwave indices for the recent climate were therefore evaluated against indices calculated from the ACORN-SAT and gridded AWAP observational temperature datasets.

No individual NARCliM simulation perfectly represents the EHF-heatwave metrics calculated from the ACORN-SAT observations. This finding supports the use of a multi-model ensemble, as some models better resemble the observations than others, depending on the metric and region. This is shown through the different magnitudes of model error between the different GCMs (see Figures 4 and 5), and supports the selection of models that appear most independent of each other and therefore have different imperfections (Evans et al., 2012). Overall, across all metrics, model output shows larger errors in coastal stations than more inland stations. This might be expected due to the more complex topography and complicated processes that occur in coastal regions, such as sea breezes. These processes are difficult for RCMs to simulate (Evans and McCabe, 2010), even at &sim;10 km grid spacing.

ACORN-SAT indicates that central regions have generally been experiencing more intense events compared with the coast (see Figure 2). Model output for central regions agrees relatively well with the observations, with some models performing better than others depending on the station. For example, the ACORN-SAT and AWAP observational data for Inverell (station 10) and Moree (station 11) are represented well by all simulations except those forced with ECHAM5 (see Figure 3(a)). Most of the models tend to overestimate heatwave amplitude for coastal stations, although the magnitude of error varies across the GCMs, with ECHAM5 indicating the largest overall error along the coast (see Figure 4).

ACORN-SAT observations show that coastal stations have generally been experiencing shorter and less frequent events than more inland stations (see Figure 2). Some of the NARCliM simulations resemble ACORN-SAT observations closely, with central-northern stations generally showing the smallest model error for heatwave frequency, relative to the rest of the region (see Figure 5). Most of the model simulations underestimate the observational data along coast, particularly for southern stations, although the CSIRO-MK3.0 simulations resemble both observational datasets closely for heatwave duration for all coastal stations (see Figure 3(e)).

Generally, ACORN-SAT and AWAP values for the intensity-related metrics agree relatively well, suggesting that the calculation of these indices is robust to the use of both point-based and high-resolution gridded observational datasets. For the duration/frequency-related metrics, ACORN-SAT mostly shows less frequent and shorter heatwaves than AWAP. The model output for these metrics compare better with ACORN-SAT than with the gridded AWAP dataset. This would be a surprising result if the differences between the two datasets were entirely due to the effects of spatial interpolation of station-based data, as many studies conclude that gridded model output more closely represents gridded observations than point observations (Zhang et al., 2011). However, differences between the ACORN-SAT and AWAP datasets...
also arise through inclusion of data from a larger set of stations in the AWAP dataset. There were also station moves to out-of-town sites that occurred during the 1990s in NSW, where minimum temperatures in the AWAP dataset were tracking 0.2 °C cooler than ACORN-SAT data (B. Trewin, 2016; personal communication). The greater degree of quality control of the station data comprising the ACORN-SAT dataset may also contribute to the differences, although in this study, this effect may be small as we have applied additional quality control to both datasets using the ClimPACT software. What might be surprising is, given that NARClIM performs bias-correction using...
AWAP data, that the model simulations for these metrics do not resemble AWAP as well as ACORN-SAT. At this point, we cannot offer an explanation as to why this is the case, and so further investigation is required. In any case, the recent climate results indicate that an RCM ensemble can represent high-quality point-based observations reasonably well.

4.2. The value of bias-correction

Bias-corrected model output aims to decrease uncertainties in climate model simulations and provide a more realistic depiction of the climate than raw model output (Piani et al., 2010; Evans et al., 2012). The results presented here demonstrate that, for heatwaves, whether or not this is achieved is somewhat dependent on the heatwave characteristic being assessed and how it is calculated.

4.2.1. Excess heat factor and heatwave characteristics

The results indicate that the intensity-related metrics show differences between bias-corrected and uncorrected datasets for certain stations and simulations, for both the recent climate and future changes (see Figure 6 and Figure S2). This suggests that bias-correction could improve the ability of climate model output to reproduce observed EHF-heatwave metrics related to magnitude and amplitude. In particular, differences between the datasets for heatwave amplitude are apparent for the CCCMA3.1 and ECHAM5 simulations. When bias-correction is applied, some of the positive errors for these simulations are reduced. The results in Figure 6 for the central-northern region demonstrate that bias-correction can have an impact on future projections, even if there is no obvious effect on the recent climate. Further work to explore possible explanations for this is recommended. However, the result as it stands emphasises the importance of assessing the effects of bias-correction on future changes in impacts-relevant climate variables, not just on recent climate values.

Recent climate values and future changes in the EHF-derived duration/frequency-related metrics are mostly unaffected by bias-correction (see Figure 7(a) and (b)) and Figures S3–S4). The few instances in which bias-correction is affected by model output occur for coastal stations, and are only noticeable in ECHAM5 and CCCMA3.1 simulations. The effect of bias-correction on the intensity-related metrics, but the lack of effect on the duration/frequency-related metrics from the uncorrected and corrected datasets, is related to the fact that the bias-correction only corrects the distribution, and not the temporal sequencing of data (Evans and Argüeso, 2014; Evans et al., 2014).

While coastal stations indicate a higher sensitivity to bias-correction, inland stations are largely unaffected. This might be because the models have larger errors along the coast due to complex fine-scale processes. Further work is recommended to completely understand what might be affecting these stations and simulations.

4.2.2. Comparison of the excess heat factor with a fixed-threshold index

The EHF index evaluated in this study is a relative measure, calculated with respect to a climatology. This means that the EHF-defined heatwave metrics for climate model simulations are calculated relative to a model-specific threshold. This reduces the effect of biases in absolute temperatures on the metrics and likely limits the effect of bias-correction. Indices that are calculated using absolute, or fixed thresholds, as is the case for more traditional extremes indices, are likely to be more affected by model biases and bias-correction.

The results in Section 3.4 illustrate these differences using SU30. The advantage of bias-correction is to remove biases inherent in the models (Piani et al., 2010; Evans and Argüeso, 2014), and in this case, the biases are indeed removed when corrected. To emphasise this, and illustrate the difference in the behaviour of an analogous EHF-derived index, Figure 8 shows the uncorrected and bias-corrected mean values for the recent climate for both HWF and SU30. As already discussed, the EHF-derived HWF values are largely unaffected by bias-correction, with uncorrected (Figure 8(a)) and bias-corrected data (Figure 8(b)) being very similar. For the threshold-based SU30 index, uncorrected model simulations show a wide range of values for all stations that experience days with temperatures above 30 °C (Figure 8(c)). Bias-correction forces the 12 model runs to cluster together, much closer to the observationally derived data (Figure 8(d)). Although the data for coastal stations is improved, the observations for these stations are represented with slightly less accuracy than more inland stations. This could be a result of uncorrected biases that arise from the complex processes in these areas that models find challenging to represent. Further work to explain this would be useful. In addition, it is worth noting that the NARClim data have been bias-corrected towards AWAP observations. This may be why, in some instances (e.g. Bathurst, station 20), the bias-corrected data agree more with AWAP than ACORN-SAT.

5. Conclusions

EHF-derived heatwave characteristics in the recent observational record are represented reasonably well by the uncorrected NARClim simulations. In general, the models perform better for more inland stations compared with coastal stations. This is likely due to the more complicated climatic processes that occur along the coast that RCMs find difficult to simulate.

Bias-corrected datasets do not greatly increase the resemblance of the NARClim-derived EHF indices to observation-derived indices. The lack of effect from bias-correction is particularly apparent for duration/frequency-related metrics, where model error mostly remains similar to the uncorrected model output, and future changes are unaffected. This is likely because the bias-correction technique does not correct the
temporal sequencing of daily data. The intensity-related metrics, although not substantially affected, do show reductions in error for some stations and simulations when bias-correction is applied. For example, coastal stations are often more affected by bias-correction than inland stations. The results for future changes in the intensity-related metrics are generally consistent with this, showing bias-correction to have a greater effect on coastal stations. Depending on the region and model, there are differences in future heatwave characteristics when bias-corrected, particularly for the intensity-related metrics. For example, increases in heatwave amplitude for coastal stations tend to be reduced when output is corrected, while more central regions, particularly in the north, show larger increases.

The EHF-heatwave metrics are calculated as relative measures and, importantly, extreme heat-related indices that are calculated from fixed temperature thresholds may be more affected by bias-correction. We demonstrated that this is indeed the case for SU30, or the days where maximum temperatures exceed 30 °C.

These findings suggest that assessing relative indices, such as the EHF, may not necessarily require the use of bias-corrected datasets. However, while the benefits of bias-correction appear to be less for the EHF-derived metrics than for fixed-threshold indices, there is evidence to suggest that bias-correction is beneficial for EHF heatwave intensity-related metrics for some simulations. There appears to be little benefit in bias-correcting the distribution of daily temperature data for EHF-derived frequency/duration-related metrics, at least for the bias-correction method and simulations from the NAR CliM project. Further work investigating other regions, bias-correction methods and RCM simulations would be required to determine if these results are universally true. This study has demonstrated it may be feasible to use RCMs to generate future projections of EHF indices that could be used to inform future health policy and health planning.

Acknowledgements

We acknowledge the use of data products from the NSW/ACT Regional Climate Modelling Project (NARClM) funded through a consortium of project partners, namely NSW Office of Environment and Heritage, Climate Change Research Centre, University of New South Wales, ACT Environment and Sustainable Development Directorate, Sydney Water, Sydney Catchment Authority, Hunter Water, Transport for NSW, NSW Department of Primary Industry and NSW Office of Water. We thank Markus Donat of the Climate Change Research Centre for help in extracting AWAP data for this study. We also thank Sarah Perkins-Kirkpatrick and Nicholas Herold of the Climate Change Research Centre for their advice on the use and definition of the Excess Heat Factor and use of the ClimPACT software. This study was supported by the Australian Research Council (ARC) Centre of Excellence for Climate System Science (grant CE110001028).

Supporting Information

The following supporting information is available as part of the online article:

Figure S1. Line plots of median values of each EHF-heatwave metric for the period 1990–2009 calculated from the uncorrected ~10 km NARClM dataset, with ACORN-SAT observations (thick dashed line) and AWAP 5 km gridded observations (dotted purple line). Station numbers correspond with numbers from Figure 1.

Figure S2. Scatter plots of the average annual magnitude of heatwaves, comparing bias-corrected and uncorrected datasets of (a) the recent climate model errors (i.e. model mean – ACORN-SAT mean) and (b) future changes (2060–2079 relative to 1990–2009).

Figure S3. As Figure S2, but for the annual number of heatwave events.

Figure S4. As Figure S2, but for the length of the longest annual heatwave event.

References


Evans JP, Ji F. 2012b. Choosing the RCMs to perform the downscaling. NARClM Technical Note 2. NARClM Consortium, Sydney, Australia.


