Journal of Hydrology 525 (2015) 472–485

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

What are the impacts of bias correction on future drought projections?

Fiona Johnson *, Ashish Sharma

University of New South Wales, Sydney, Australia

ARTICLE INFO

Article history:
Received 4 December 2014
Received in revised form 5 March 2015
Accepted 5 April 2015
Available online 9 April 2015
This manuscript was handled by
Konstantine P. Georgakakos, Editor-in-Chief

Keywords:
Drought
General circulation model
Bias correction

SUMMARY

It is expected that anthropogenic greenhouse gas emissions will continue to change our climate, and in turn the characteristics of future drought. Assessments of the risks of future droughts, when at a global or a continental scale, are often based on simulations from General Circulation Models (GCMs). When raw GCM simulations are used, it is assumed that the future deviations from modelled historical climatology represent the future drought. On the other hand, it is known that raw GCM simulations are significantly biased for the variables that affect hydrology and a correction is needed before assessments can be performed. We investigate here whether drought assessments based on raw GCM simulations are biased and the typical extent of this bias. Our assessment is based on the use of monthly precipitation data from 18 CMIP3 GCMs, two popular bias correction alternatives and Australia as the study domain. A number of different precipitation drought attributes have been assessed. These include the Standardized Precipitation Index (SPI), multi-year rainfall statistics and a drought vulnerability statistic that measures the maximum deviation of a time series from its mean.

We find significant differences between droughts assessments using raw GCM simulations and using bias corrected sequences. Large increases in drought frequencies are projected for some parts of Australia. Both bias correction methods moderate these increases. This result is consistent across the three different drought statistics. The bias corrected drought projections also generally have slightly more agreement (smaller range of future changes) across the GCMs compared to the raw projections, which is a promising result for attempting to reduce model structural uncertainty. What this study shows is raw model simulations can lead to incorrect drought assessments even at continental scales and bias corrections should be applied.

1. Introduction

Anthropogenic climate change is now assessed as being "extremely likely" and is expected to lead to detrimental effects for human and natural systems both at regional and global scales (IPCC, 2013). One of the areas of concern is the impact of these changes on water resources systems (Jiménez Cisneros et al., 2014), particularly for parts of the world where water resources are already stressed due to population pressures, technological change and the variability introduced by large scale natural climate drivers (Vörösmarty et al., 2010). General Circulation Models (GCMs) are one method of assessing the likely impacts of increasing greenhouse gas concentrations on natural and human systems. Despite the skill of GCMs at a global scale, there remain concerns about their ability to model regional scale impacts (Fowler et al., 2007a) and to represent certain features of the climate system, such as precipitation (Chan et al., 2013), which form an important input to the modelling of water resources systems (Maraun et al., 2010).

Techniques to overcome the weaknesses of GCMs and which allow scientists to undertake impact assessments, are based on the assumption that some regional scale changes do not create global scale feedbacks and therefore may be corrected outside of the modelling of the GCM without affecting the accuracy of the global scale simulations (Pitman et al., 2012). Methods that correct GCM simulations at a regional scale include dynamical downscaling (regional climate modelling) or statistical downscaling. Given the significant differences in GCM simulated fields compared to observations, there is often a need for a pre-processing step before statistical or dynamic downscaling approaches can be used. This pre-processing step is referred to as bias correction, whereby the GCM or regional climate model results representing the current climate are corrected to match observations. The bias correction model over the historical period is assumed to be the same in future simulations, and can therefore be used to obtain the future projections

* Corresponding author at: School of Civil and Environmental Engineering, University of New South Wales, Kensington, New South Wales 2052, Australia. Tel.: +61 2 9385 9769.
E-mail address: f.johnson@unsw.edu.au (F. Johnson).

http://dx.doi.org/10.1016/j.jhydrol.2015.04.002
0022-1694/

(Buser et al., 2009). Bias correction has been shown to be a simple and effective method that can be quickly be applied over large areas, to multiple models or as pre or post-processing step for more sophisticated downscaling methods (Chen et al., 2013; Gudmundsson et al., 2012; Ines and Hansen, 2006; Johnson and Sharma, 2012; Li et al., 2010; Piani et al., 2010; Teutschbein and Seibert, 2012; Wood et al., 2004).

Hydrologic planning is strongly dictated by the frequency and magnitude of sustained low or high flow anomalies that affect water resources systems. The importance of these anomalies becomes even greater when the system is stressed and water demand is close to the overall availability in the system. Past assessments have evaluated such anomalies based on raw GCM simulations. However, most water resources assessments at a catchment scale require some post-processing or bias correction, before any evaluation (for drought or for other purposes) can be carried out. This contradiction lays the foundation for this research; the key question that is posed is whether bias correction is needed for drought assessment at the scale of a catchment, a region, or an entire continent.

To this end, we assess drought simulations across Australia using monthly precipitation simulations from 22 GCMs of the historical and future climate. We then repeat this assessment using the same GCMs, but after applying two established alternatives for bias correction. One of these alternatives, the Nested Bias Correction (NBC) (Johnson and Sharma, 2012) was designed specifically to correct low frequency variability bias in simulations. The other method, Quantile Mapping (QM) corrects distributional attributes of simulations instead of focusing on persistence related attributes. Our investigation uses these approaches to examine the differences that result with reference to a future climate. We then ask whether raw GCM simulations are indeed different from bias corrected simulations of drought, and if so whether there are clear advantages in using one of the two bias correction alternatives evaluated.

The remainder of this paper is as follows. In Section 2 the bias correction methodologies and the data sources used for the analysis are described. Section 3 discusses the drought and rainfall statistics that have been considered. Section 4 applies the technique to 20th century rainfalls over Australia to compare the modelling of drought using observed, raw GCM outputs and the bias correction techniques. In Section 5 comparisons of projected future changes in drought from the raw and bias-corrected GCM outputs are presented. The final section discusses the implications of the results for future drought assessment and draws conclusions that will be useful for water resources climate change impact assessment.

2. Bias correction of GCM precipitation

2.1. Background

Recent comparisons of bias correction methods (Gudmundsson et al., 2012; Johnson and Sharma, 2011; Teutschbein and Seibert, 2012) have shown that the assumptions of the correction method can affect their performance. Gudmundsson et al. (2012) found that nonparametric methods lead to the lowest errors, whilst Teutschbein and Seibert (2012) recommend power transformations or quantile mapping. Quantile mapping has been found in particular to be useful for correcting high rainfall totals of daily data (Themeßl et al., 2012). If year to year variability in rainfall is an important driver of the behaviour of a water resources system then it is necessary to consider biases in persistence as well as the distribution of the monthly rainfall data (Johnson and Sharma, 2011). The importance of this is highlighted by Rocheta et al. (2014) who have recently shown that the majority of GCMs in the Coupled Model Intercomparison Project 3 (CMIP3) underestimate interannual variability.

In practical terms, the effects of interannual and interdecadal precipitation variability manifest as periods of drought or abnormally wet conditions that can lead to flooding (Kiem et al., 2003; Pui et al., 2011; Verdon et al., 2004). Therefore correctly modelling low-frequency variability is the key to understanding possible changes to future water resources. Future drought assessments have found that larger areas of land are expected to have increased drought frequencies than decreases (Taylor et al., 2013) particularly for the later parts of the 21st century (Burke et al., 2006). Areas of the largest increases in drought frequencies are the Amazon, Central America and South Africa (Dai, 2013; Taylor et al., 2013) which is broadly consistent with the results from other studies (Burke and Brown, 2008; Orlowsky and Seneviratne, 2013) although the studies use different climate models and emission scenarios. These projections are also consistent with observed trends of drought frequency based on soil moisture anomalies (Orlowsky and Seneviratne, 2013).

A recent comprehensive study of global drought (Prudhomme et al., 2014) has been carried out as part of the Inter-Sectoral Impact Model Intercomparison Project (Warszawski et al., 2014). The analyses considered daily runoff and compared it to a daily drought threshold set at the 10th percentile of the daily values from the period 1976–2005. For large parts of the world, large increases in drought frequency were found particularly in Australia. The use of daily drought thresholds does not provide further guidance on longer term droughts due to interseasonal or interannual variability.

One of the problems with previous assessments of future drought frequencies is that generally they have not considered biases in the GCM precipitation simulations (e.g. Burke and Brown, 2008; Taylor et al., 2013). Wang and Chen (2014) used the Bias Correction and Spatial Disaggregation method of Wood et al. (2004) but did not assess the improvements in drought representation in the current climate or the effects of bias correction on the future projections. Prudhomme et al. (2014) also used bias corrected simulations and note that "statistical bias correction can influence the signal of the runoff changes". They go on to state that this uncertainty is believed to be smaller than the structural uncertainty associated with the choice of GCM or global impact model. Our study addresses this assumption directly by considering whether bias correcting GCM simulations improves the representation of historical drought statistics and then secondly if the bias correction leads to differences in the projected frequencies of future droughts. Changes at the individual grid cell are to be expected; specifically we are interested in whether bias correction leads to different estimates of drought at regional or continental scales.

2.2. Bias correction methods

Bias correction techniques have been developed to allow the direct use of GCM outputs for climate change impact assessment applications, whilst accepting that there are problems in the GCM modelling of rainfall (Johnson and Sharma, 2012). Bias correction approaches previously proposed include monthly (Wood et al., 2004) or daily (Ines and Hansen, 2006), quantile matching (Christensen et al., 2008; Gudmundsson et al., 2012) and simple monthly correction factors (Fowler and Kilsby, 2007). One of the weaknesses of all these approaches is that they only consider biases in the distribution of GCM simulations and not the persistence of the simulations. Biases in the representation of persistence translate to a poor characterization of interannual variability which can be particularly important when assessing the impacts of climate change on water resources availability and

infrastructure. The features of interannual variability that are important include the sequencing of wet and dry years and the standard deviations and correlations at time scales longer than a year. Because the focus of our work is on this interannual variability and annual and multi-year droughts, we focus on monthly simulations from the GCMs.

Johnson and Sharma (2012) proposed the NBC method to correct the representation of variability and persistence in the GCM outputs at multiple time scales. The approach was developed based on the rationale in Koutsoyiannis (2001) which allows two time series of different temporal resolutions to be combined, maintaining their individual stochastic properties whilst ensuring that they are consistent with each other. Full details of the method can be found in Johnson and Sharma (2012) with a summary of the main ideas presented below.

The NBC method post-processing methodology is sufficiently general so that it can correct distributional and persistence attributes on multiple time scales of interest. For the general implementation of the nesting algorithm, the bias-corrected time series can be expressed as shown in Eqs. 1–3.

$$\tilde{P}^m_{i,j} = g_i\left(P^m_{i,j} | P^h_{i,j}, \theta_i\right) \tag{1}$$

$$\tilde{P}^m_{j} = \sum_i \tilde{P}^m_{i,j} \tag{2}$$

$$\tilde{P}^m_{j} = g_j\left(\tilde{P}^m_{j} | P^h_{j}, \theta_j\right) \tag{3}$$

where $g_{()}$ and $\theta_{()}$ represent the transformation model and associated parameters for each of the time scales of interest which are monthly (i) and annual (j) for the purposes of drought assessment. P^m is the modelled output at monthly or annual time scale (i.e. for the ith month and jth year or just the jth year in the case of the annual sequence), $\tilde{P}'^m$ is the aggregated partially corrected model output after one level of nesting and $\tilde{P}^m$ is the corrected output and P^h is the observations over the historical period. The aim is to transform the modelled outputs such that $\tilde{P}^m$ exhibits the same distributional and persistence attributes as P^h across the nesting time scales used.

The parameters θ to be corrected are the mean, standard deviation and lag-1 autocorrelation coefficient although in principle any statistics could be chosen (Johnson and Sharma, 2012).

To achieve this, the corrections at the monthly level are first calculated for each month by finding the mean, standard deviation and autocorrelation and applied to the model simulations (Eq. (1)). These corrected simulations are then aggregated to the annual level (Eq. (2)) and the next level of corrections is applied (Eq. (3)). The annual mean is already corrected by the monthly mean correction but the annual variability and persistence will not necessarily be correct until Eq. (3) is applied. A further extension of NBC was presented as the Recursive NBC (Mehrotra and Sharma, 2012) that repeats the nesting process several times to further reduce the biases.

Quantile mapping has also been used to post-process the GCM outputs to assess drought modelling for the historical and future climates. The idea behind quantile mapping is to use either an empirical distribution or fitted probability distribution to match the observed and modelled rainfall quantiles (Gudmundsson et al., 2012; Maurer et al., 2013; Teutschbein and Seibert, 2012; Wood et al., 2004). Using similar notation as for NBC the quantile mapping is described in Eq. (3) where q is the quantile of interest and F is the empirical cumulative distribution function with F^{-1} its inverse.

$$\tilde{P}^m_{i,q} = F^{-1}_h\left(F_m\left(P^m_{i,q}\right)\right) \tag{3}$$

In the implementation here, we use the empirical cumulative distribution for each month to do the correction. The mapping can then be used for the future rainfall simulations. In practice the rainfall amount for month i and year j from the GCM is ranked against all rainfall totals for month i. The rainfall total with the same rank in the observed data is adopted as the corrected amount. This lookup table can then be used for the future simulations by assuming that the bias is multiplicative and that the ratio of the observed to modelled rainfall can be applied to correct the future simulation.

The advantage of quantile mapping is that it can correct the full distribution of rainfall amounts including the extreme wet and dry

Table 1
Details of GCMs used.

Modelling group	Name	Available ensembles	Atmosphere resolution
Bjerknes Centre for Climate Research, Norway	bccr_bcm2_0	1	T63 (∼1.9°)
Canadian Centre for Climate Modelling and Analysis, Canada	cccma_cgcm3_1	5	T47 (∼2.8°)
Météo-France/Centre National de Recherches Météorologiques, France	cnrm_cm3	1	T63 (∼1.9°)
Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atmospheric Research, Australia	csiro_mk3_5	1	T63 (∼1.9°)
U.S. Department of Commerce/National Oceanic and Atmospheric Administration (NOAA)/Geophysical Fluid Dynamics Laboratory (GFDL), USA	gfdl_cm2_0	1	2° × 2.5°
U.S. Department of Commerce/National Oceanic and Atmospheric Administration (NOAA)/Geophysical Fluid Dynamics Laboratory (GFDL), USA	gfdl_cm2_1	1	2° × 2.5°
NASA/GISS, USA	giss_model_e_r	1	4° × 5°
INGV, National Institute of Geophysics and Volcanology, Italy	ingv_echam4	1	T106 (∼1.125°)
Institute for Numerical Mathematics, Russia	inmcm3_0	1	4° × 5°
Institut Pierre Simon Laplace, France	ipsl_cm4	1	2.5° × 3.75°
Center for Climate System Research (University of Tokyo), National Institute for Environmental Studies and Frontier Research Center for Global Change (JAMSTEC), Japan	miroc3_2_medres	3	T42 (∼2.8°)
Meteorological Institute of the University of Bonn, Meteorological Research Institute of the Korea Meteorological Administration (KMA), and Model and Data Group, Germany/Korea	miub_echo_g	3	T30 (∼3.9°)
Max Planck Institute for Meteorology, Germany	mpi_echam5	3	T63 (∼1.9°)
Meteorological Research Institute, Japan	mri_cgcm2_3_2a	5	T85 (∼1.4°)
National Center for Atmospheric Research, USA	ncar_ccsm3_0	4	T85 (∼1.4°)
National Center for Atmospheric Research, USA	ncar_pcm1	4	T42 (∼2.8°)
Hadley Centre for Climate Prediction and Research/Met Office, UK	ukmo_hadcm3	1	2.5° × 3.75°
Hadley Centre for Climate Prediction and Research/Met Office, UK	ukmo_hadgem1	1	1.3° × 1.9°

tails (Themeßl et al., 2012). Biases in the skewness of the rainfall distribution are also addressed through this correction method. An important disadvantage in the context of drought modelling is that the persistence in the GCM time series (i.e. the relationship between adjacent months or years) cannot be addressed. It is therefore of interest how the different bias correction methods behave for both the current climate and the future simulations.

2.3. Data

In this study, the two bias correction methods have been applied to the simulations of precipitation from the Coupled Model Intercomparison Project (CMIP3) with the intention of assessing the impact they have on drought simulations for the future. GCM simulations at monthly time steps were used and all available ensemble members were corrected. All GCMs outputs

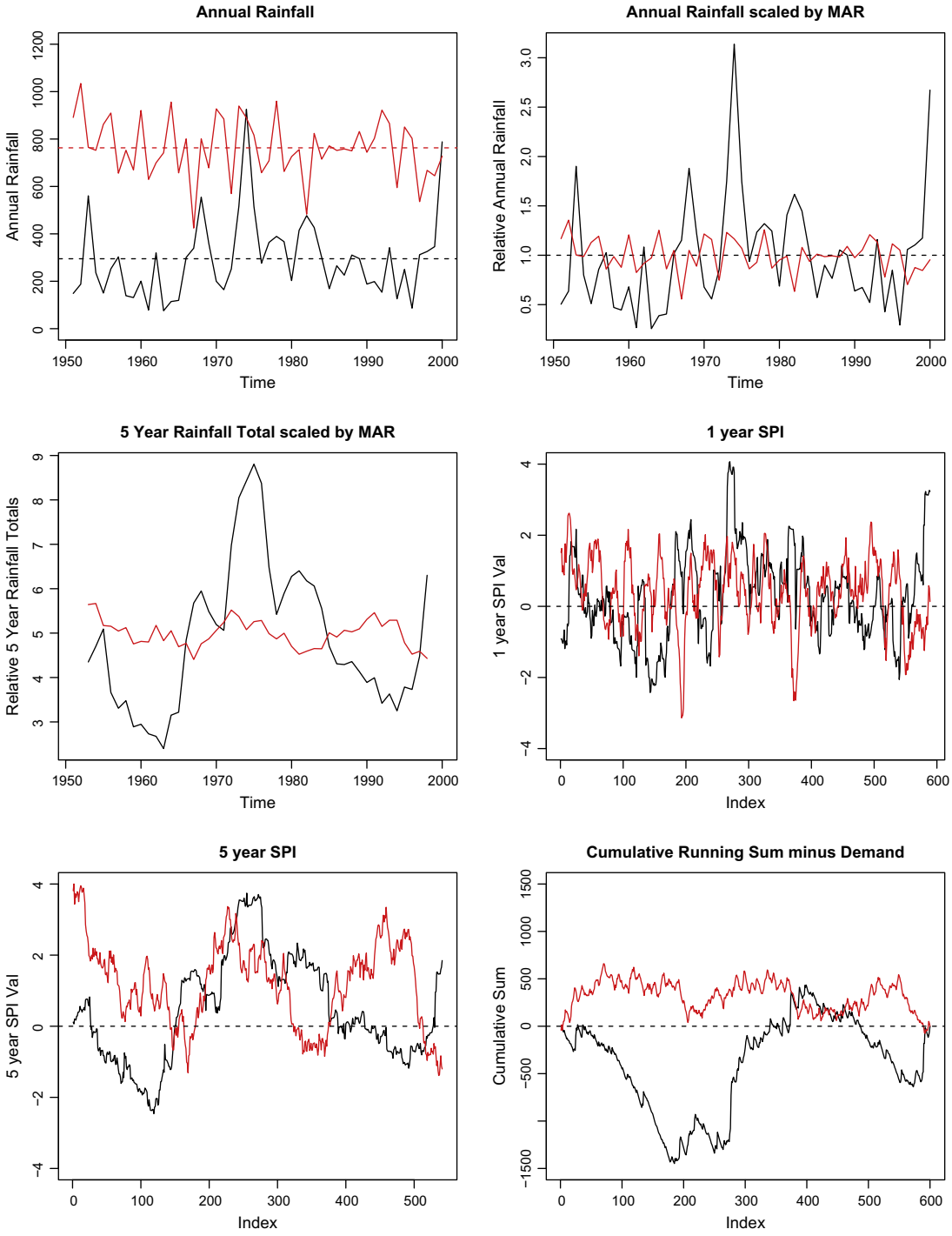

Fig. 1. Time series of annual rainfall and drought estimates for two locations in Australia showing high interannual variability (black) and low interannual variability (red) for (a) the annual rainfall time series, (b) the annual rainfall scaled by MAR, (c) the running sum of 5 year rainfalls scaled by MAR, (d) 1 year SPI time series, (e) 5 year SPI time series and (f) the residual mass curve of running sums minus demand. Demand is defined to be equal to the mean monthly rainfall. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

were interpolated to a common T63 grid to allow the results to be compared across space. For the historical climate, the 20th century simulations were used from 1900 to 1999 and for the future simulations the SRESA2 scenario was adopted for a 30 year period at the end of the 21st century. To test the modelling of drought for the historical climate, a split sample approach was used with first half of the record used to calibrate the bias correction and the second half of the record used for an independent validation. Table 1 lists the GCMs that have been used in the study.

Gridded monthly rainfall data over Australia from the Bureau of Meteorology (BOM) was used as the observational data set (Grant et al., 2008). This product was created as part of the Australia Water Availability Project (AWAP) from daily rainfall stations and Barnes successive correction and three-dimensional smoothing splines to interpolate the rainfall onto a 0.05° grid (Grant et al., 2008). Data already aggregated to monthly time series was provided by the Bureau of Meteorology. Data from 1900 to 1999 was used for the analyses to match the 20th century simulations from the GCMs. Area averaging was used to re-grid the data to the same resolution as the GCM outputs.

3. Drought statistics

There are a number of ways to assess the severity of drought and even the definition of drought has a range of meanings. For this work meteorological drought has been considered as the assessment is based on rainfall sequences. The timescales of interest are those more generally considered with hydrological drought i.e. deficits in the order of a year to multiple years or decades. In other parts of the world with less inherent climate variability drought conditions can result from deficits as short as three to six months (Fowler et al., 2007b).

The SPI was developed to provide a simple calculation of drought (Guttman, 1999). An incomplete gamma distribution is fitted to running totals of rainfall, and then transformed to a standard normal distribution. The quantiles of the normal distribution are used to assess the severity of the drought. Negative values of the index occur during dry periods, with positive values indicating wet conditions. The SPI can be calculated for varying intervals; intervals of 1, 2 and 5 years have been assessed in this study. The SPI has been used for future drought assessments by Burke and Brown (2008). Mpelasoka et al. (2008) note that because the SPI can be calculated on different time scales it can account for the effects of rainfall deficits of varying lengths on separate parts of the hydrologic cycle. As in the work by Burke and Brown (2008) a severe drought threshold has been set at the 5th percentile of the SPI distribution at each location. Future changes can be assessed by counting the number of exceedances of this threshold. If the future is drier, then exceedances of the threshold will occur more frequently than in the historic climate (i.e. 5% of all time periods). The converse is also true for wetter future conditions. The spell length of the maximum deficit period in the SPI sequences has also been calculated.

A number of other statistics of rainfall that have been used for drought or water resources assessments have been considered to understand the ability of GCMs to represent interannual variability correctly and the impacts of bias correction on the current and future projections. The first rainfall statistic that has been considered is the coefficient of variation of multi-year rainfall sequences (2 year and 5 year specifically). A similar statistic (the minimum of running sums of 2 year and 5 year rainfall totals) was used by Srikanthan and Pegram (2009) to evaluate stochastically generated rainfall sequences. Johnson and Sharma (2011) found that the minima of the time series could be too noisy and instead adopted the standard deviation. The standard deviations are scaled by the annual mean to allow comparisons across space and are therefore equivalent to the coefficient of variation.

The rescaled range is calculated from a residual mass curve, which measures the cumulative departure of a time series from its mean. The difference between the maximum and minimum of the residual mass curve is the range, which can be scaled by the standard deviation of all deficits/surpluses. The rescaled range is used in water resources management as a surrogate of reservoir storage. It is a useful statistic for assessing GCM simulations of drought as storage calculations integrate dependence and trends that may be present in the time series (Koutsoyiannis, 2002).

Fig. 1 shows the annual rainfall time series and estimates of SPI and rescaled range for two locations in Australia. The first location

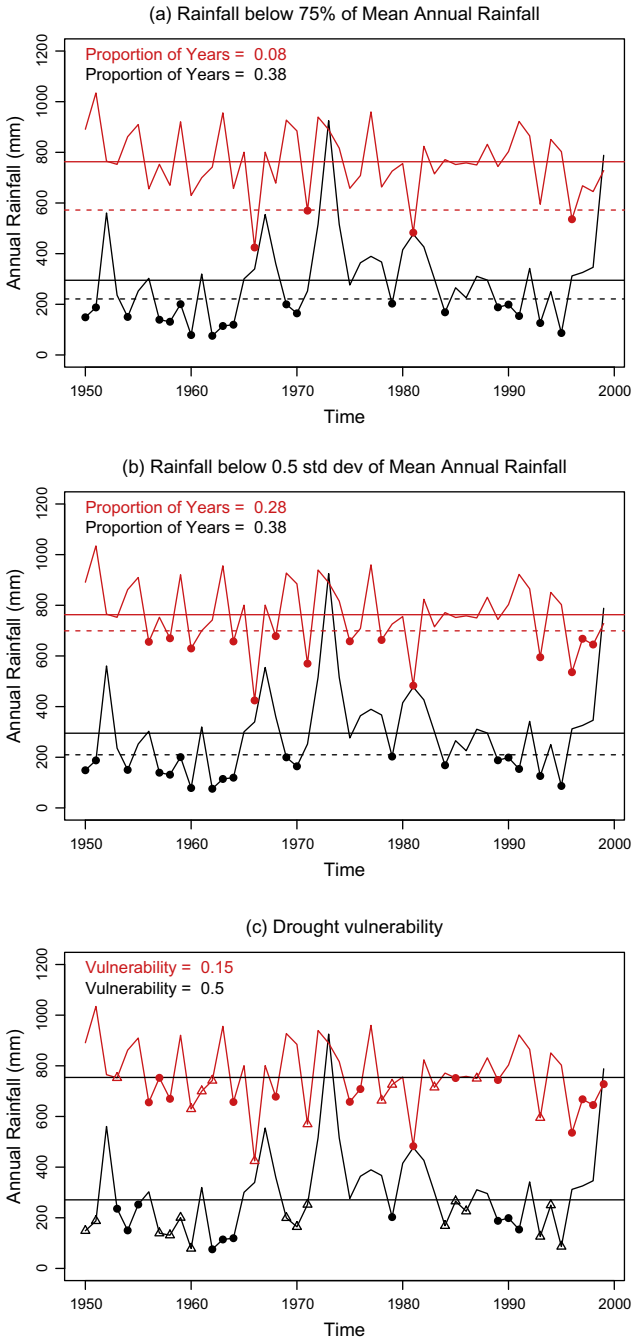

Fig. 2. As for the locations in Fig. 1, time series of annual rainfall showing the calculation of MAR75, estimates that are more than 0.5 standard deviations below the mean, and Peel drought vulnerability.

has higher rainfalls (shown in red in Fig. 1a) but lower multiyear variability. The second location is more arid with mean annual rainfall (MAR) of the order of 300 mm but there is large year-to-year variability. This difference in variability is evident when the annual rainfall time series are scaled by the MAR. The majority of years at the high rainfall site are within 50% of the mean, whilst at the low variability site, the annual distribution is much more skewed with a number of very dry years and several very wet years with rainfalls more than double the MAR. Using multiple year rainfall totals smooths the curves in Fig. 1b and emphasizes the persistence at the site with high variability with a number of consecutive dry years or wet years. Next, the SPI values are calculated for the two sites for 1 year and 5 year drought estimates. Interestingly the difference between the two time series in terms of their multi-year persistence is not reflected in the SPI series. This is because at each location the rainfall series are transformed via a truncated gamma distribution to a normal distribution. Thus in all locations the range of values is approximately between ±4 and therefore the SPI values cannot be used to discriminate between sites with low or high interannual variability. This means that the SPI is useful for comparing how future droughts might vary at one location but is not effective for comparing drought severity at different locations.

Spells of dry months can be used to consider persistence and drought characteristics of time series. The number of consecutive months when rainfall or streamflow is below some threshold is called the "Run Length". Peel et al. (2005) propose a drought vulnerability statistic by comparing run lengths of rainfall below the median. The run lengths can also be calculated for any other threshold of interest such as the mean or a value with a more frequent exceedance probability. The cumulative sum of the deficits below the threshold is calculated for every dry spell. The drought vulnerability is then defined as the maximum deficit below the chosen threshold and it is standardized across space by dividing by the mean annual rainfall. Readers interested in this concept are referred to Peel et al. (2005) for more information on these statistics.

Counts of annual rainfall totals below the mean or median can also be constructed in other ways. Ault et al. (2014) have assessed long-term sequences of flows by counting the number of times that the flow/rainfall falls below 0.5 standard deviations from the mean. Depending on the coefficient of variation of the data, this test can yield quite similar results to considering rainfalls that are more than 75% below MAR (denoted as MAR75 in the remainder of the paper). On the other hand when the coefficient of variation is small, then there will be far few instances of rainfall less than 75% of the mean than when using the 0.5 standard deviation threshold. Fig. 2 shows the relationship of the Peel drought vulnerability with the two methods of counting rainfall below a threshold for the same two locations as shown in Fig. 1. For the site with high rainfall and relatively low variability (red lines), the difference between using MAR75 threshold (Fig. 2a) to the 0.5 standard deviation is quite evident, whereas for the site with low variability the thresholds are quite similar. The MAR75 statistic therefore appears to have better

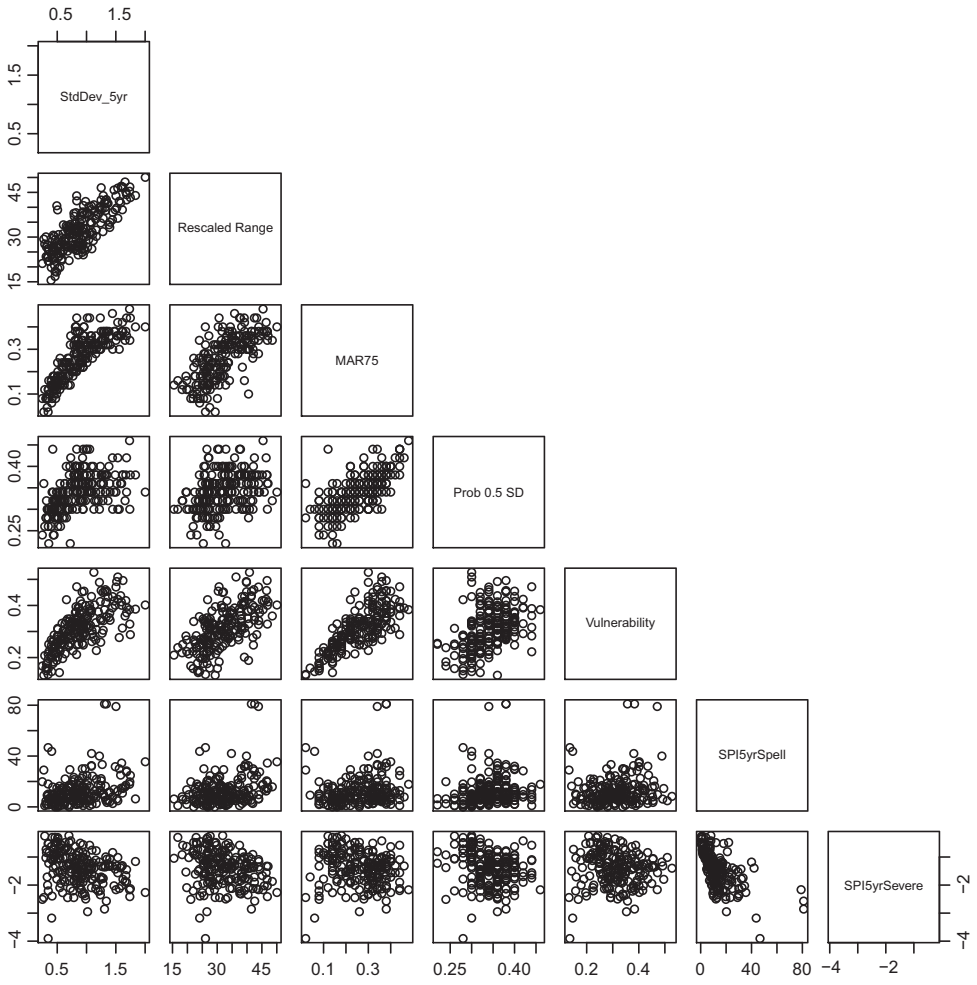

Fig. 3. Relationships between observed drought statistics. Each plot contains 201 estimates, representing the range of drought values across Australia.

discrimination between sites with high and low interannual variability than using the standard deviation measure of Ault et al. (2014). In terms of drought vulnerability (Fig. 2c), there is quite a large difference between the estimates at the two sites. The high variability site has large vulnerability due to the large deficit that occurs around 1960. All but one of the spells below the median last more than one year at this location, compared to only 30% (5 of 16) of the runs at the high rainfall, low variability site.

In Fig. 3 the observation based estimates of each of the rainfall and drought statistics at all 201 grid cells in Australia are compared. The statistics are calculated for the period from 1950 to 1999 and the SPI severe drought frequencies are calculated relative to the threshold calculated for the earlier part of the record. This gives a sense of the performance that could be expected from the GCMs and the bias correction methods if there were no biases. It can be seen that there is high correlation between most of the statistics. The probability of annual rainfall being more than 0.5 standard deviations below the mean is less well correlated with the other rainfall statistics, in particular the Peel vulnerability and the rescaled range. As discussed for Fig. 1 there is less agreement between the SPI and other statistics due to the standardization in the SPI calculations.

4. Historical climate drought representation in GCMs

Using the same statistics as the previous section, the ability of the CMIP3 GCMs to capture the important features of rainfall

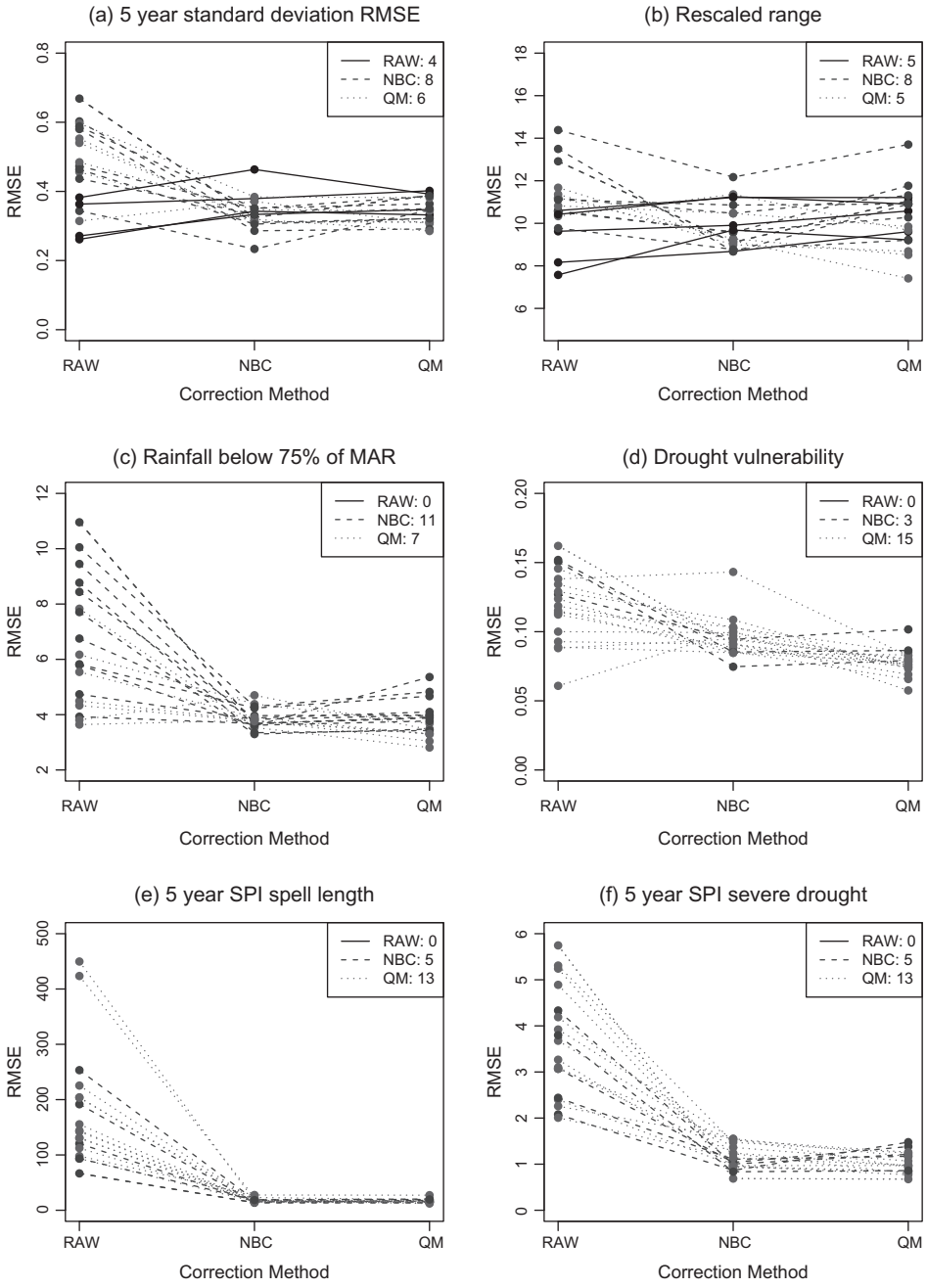

Fig. 4. Errors for each GCM in drought estimates, averaged over space for (a) 5 year rainfall time series standard deviation, (b) rescaled range, (c) MAR75, (d) drought vulnerability, (e) SPI 5 year spell length and (f) SPI 5 year severe drought threshold. The method that leads to the smallest RMSE from the raw, NBC and QM results for each GCM has been used to pick the line type.

persistence for water resources management is now assessed. The Root Mean Square Error (RMSE) has been calculated for each GCM by comparing drought statistics based on the raw or bias-corrected simulations to the observations for the period 1950–1999. The RMSE reflects the average error across all grid cells in the domain. In Fig. 4 the RMSEs for six different drought statistics are shown for all of the first ensemble members from the 18 GCMs. Unsurprisingly in the majority of cases errors are much larger for the raw GCM simulations than either of the bias-corrected alternatives. For the first three rainfall-based statistics, the NBC method leads to smaller errors than the QM method for the majority of the GCMs. For the Peel drought vulnerability and the SPI statistics, the raw GCM simulations never have the smallest error and the QM is the preferred method in around 70% of the cases. This suggests that different drought statistics capture varying characteristics of the rainfall time series. In the case of the SPI and drought vulnerability, it appears that biases in the rainfall distribution are more critical in correctly representing the drought than biases in the persistence. As shown in Fig. 1, the rescaled range and multi-year rainfall totals are strongly affected by persistence and therefore the NBC will perform better in correcting biases in their representation.

In Fig. 5 the spatial patterns of errors across Australia are shown for the raw GCM simulations and the bias-corrected outputs. In this case, error is defined at each grid cell as the RMSE of the estimate from the 18 GCMs compared to the observation at that location. The patterns of errors in the raw simulations are reasonably consistent for the three chosen drought statistics (left column), with the largest errors in the centre of the country and relatively smaller errors in the southeast. Both bias correction methods substantially reduce the errors. For the Peel drought vulnerability and MAR75 statistics, the pattern of errors after bias correction broadly follows the raw simulations, with the smallest errors tending to occur in the southeast and around the coast. For the SPI Severe Drought the patterns of errors are relatively uniform after bias correction, particularly when compared to the other drought statistics. The slight spatial variation in the errors shows that they are reversed after bias correction with the smallest errors in the north and central parts of Australia and relatively higher errors occurring in the southeast. As discussed earlier the SPI is standardized across space so the impact of the magnitude of rainfall or the interannual variability at any particular location is lessened. For all three drought statistics there is little difference in the performance of the two bias correction methods when aggregated over all the models.

The results presented thus far have been aggregated across space or models. In Figs. 6 and 7 results for two GCMs are shown to illustrate how the errors in raw and bias corrected GCM vary at individual grid cells. In Fig. 6, the results from CSIROMk3.5 are presented. In this GCM the NBC method consistently outperforms the QM results, due to the specific structure of the biases in the GCM simulation. For this GCM it is found that the rainfall

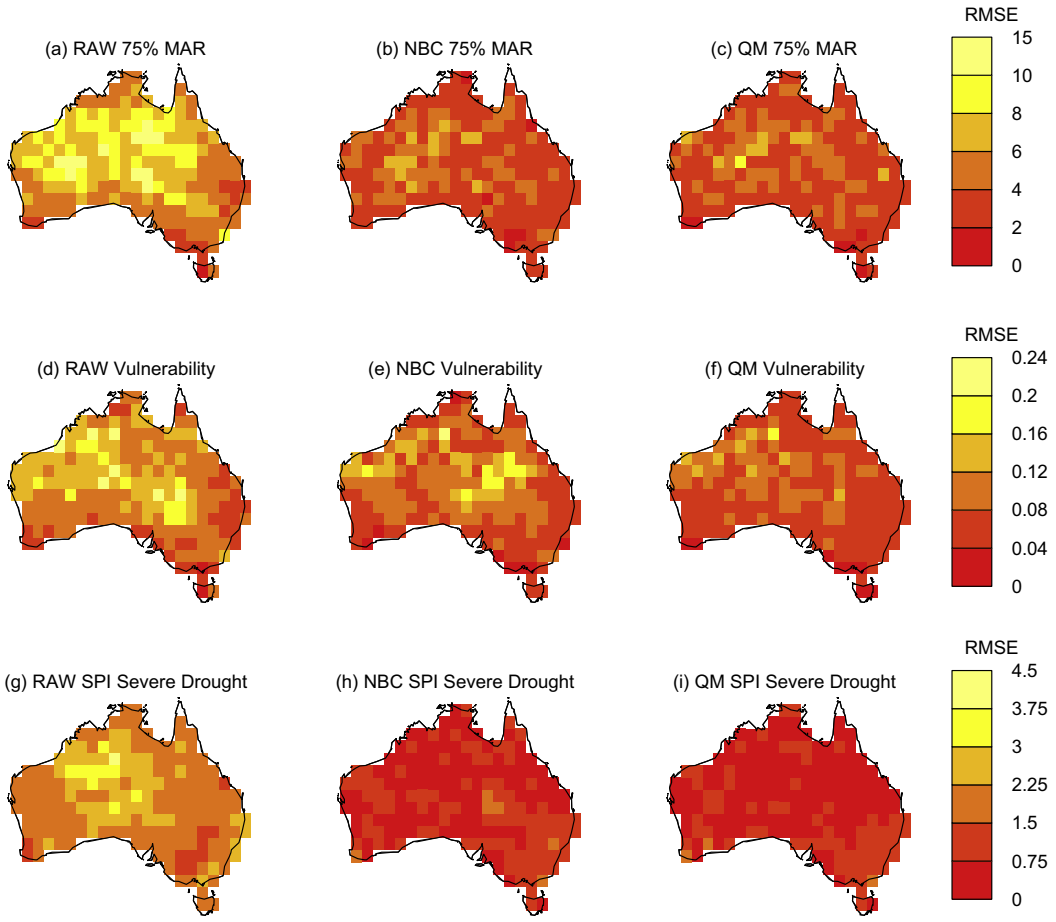

Fig. 5. Error in drought estimates at each grid cell calculated as the average across all 18 GCMs of RMSE between the observations and the GCM estimates for MAR75 (top row), drought vulnerability (middle row) and for 5 year SPI Severe Drought frequency (bottom row). Raw and bias corrected results are shown as the columns. The smallest errors are in red colors whilst larger errors are shown in lighter colors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distributions generally have a similar shape to the observed distributions and therefore corrections of the first and second moment are sufficient without the need to correct the full distribution (Johnson and Sharma, 2012). Under these circumstances the autocorrelation corrections are effective and the NBC method improves the representation of persistence in the model and therefore also the drought statistics.

In Fig. 7, similar results are presented for INGV_ECHAM4 where it is clear that the raw GCM simulations are much too wet with all drought statistics substantially underestimated. For this GCM the QM method generally leads to smaller errors. The NBC overcorrects the persistence leading to slightly larger drought estimates than the QM method although the results are much improved in both cases compared to the raw GCM simulations. For example in Fig. 7d two grid cells have been labelled and it can be seen that in these locations the drought vulnerability is close to double the observed vulnerability using the NBC method whilst the QM method has matched the observations reasonably well. Reviewing the time series of monthly rainfall, the NBC method has produced more zero rainfall months than the observations and several multi-month spells of zero rainfall that are not seen at all in the observations. The QM has slightly fewer dry months than the observations but does not lead to any overly long dry spells and hence the vulnerability is represented better.

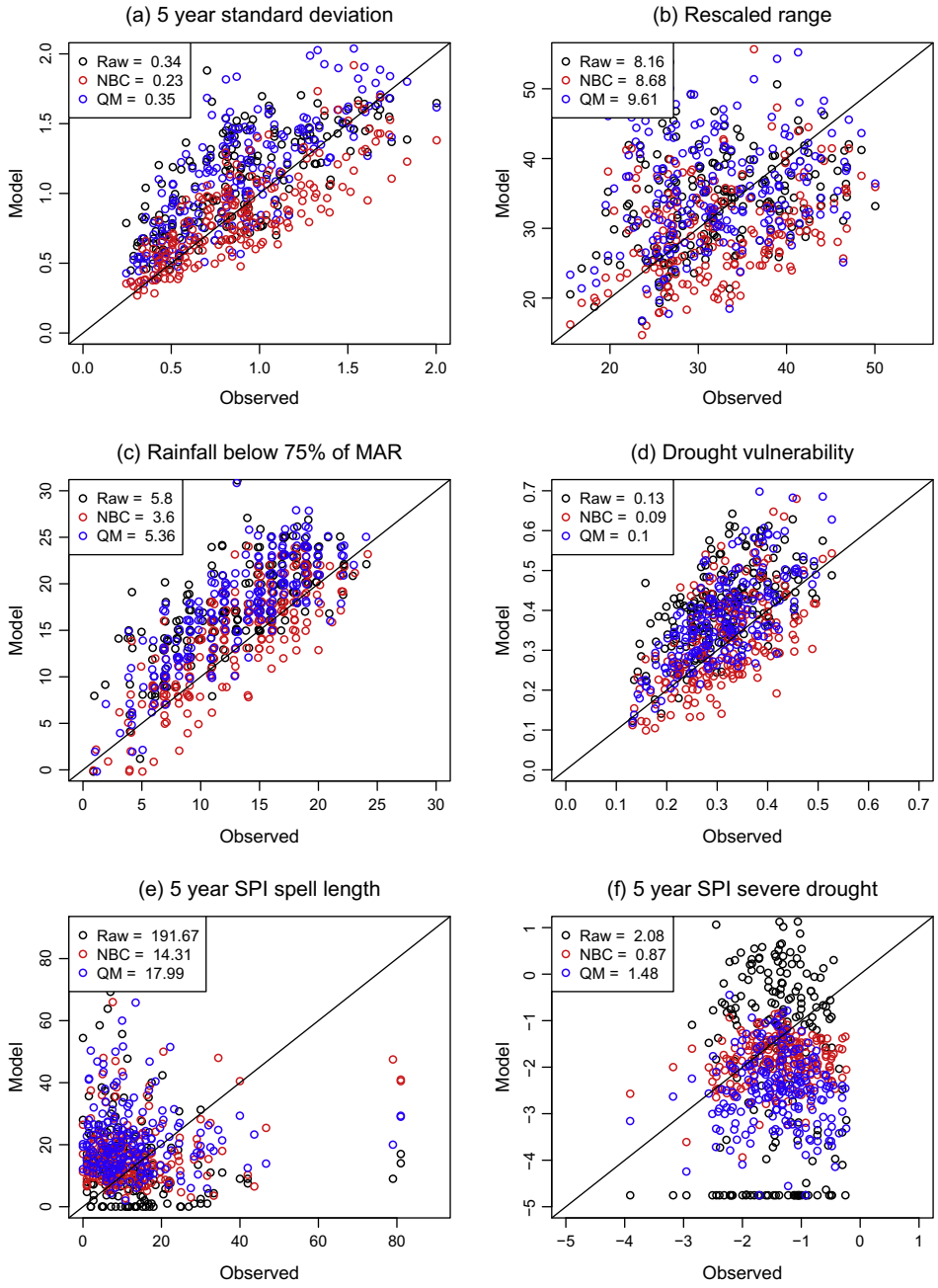

Fig. 6. Comparison of observed and modelled drought statistics for the CSIRO Mk3.5 GCM for (a) 5 year rainfall time series standard deviation, (b) rescaled range, (c) MAR75, (d) drought vulnerability, (e) SPI 5 year spell length and (f) SPI 5 year severe drought threshold. Raw GCM simulations are shown in black, NBC in red and QM in blue. The RMSE of the model estimates is shown in each panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Drought projections for 2080

In the previous section, bias correction approaches were shown to improve the representation of drought for the historical climate significantly compared to raw GCM simulations. Now it is of interest how drought frequencies may change in the future. Previous studies have considered future drought frequencies as changes relative to current climate drought thresholds (Burke and Brown, 2008). The important implication of this approach is that any biases in the current climate simulations are firstly ignored and secondly assumed to remain constant in the future. In this study, biases in the current climate are explicitly calculated and assessed. However, as for any bias correction application the assumption that biases are stationary in time is still required for the bias correction to be valid in the future (Maurer et al., 2013; Teutschbein and Seibert, 2012). Haerter et al. (2011) showed that future changes from bias corrected simulations can be different from those calculated using raw climate model simulations and the differences depend on the assumptions of the bias correction method. Here we further consider this point in the context of drought statistics and compare the bias-corrected future drought frequencies to raw simulations from GCMs to identify the effects of bias correction.

For each of the drought statistics presented in Section 4 the future drought frequencies relative to the historical climate have been calculated. For the raw GCM simulations, the historical climate case is the 20th century GCM simulations without any corrections. The bias-corrected future sequences are compared to

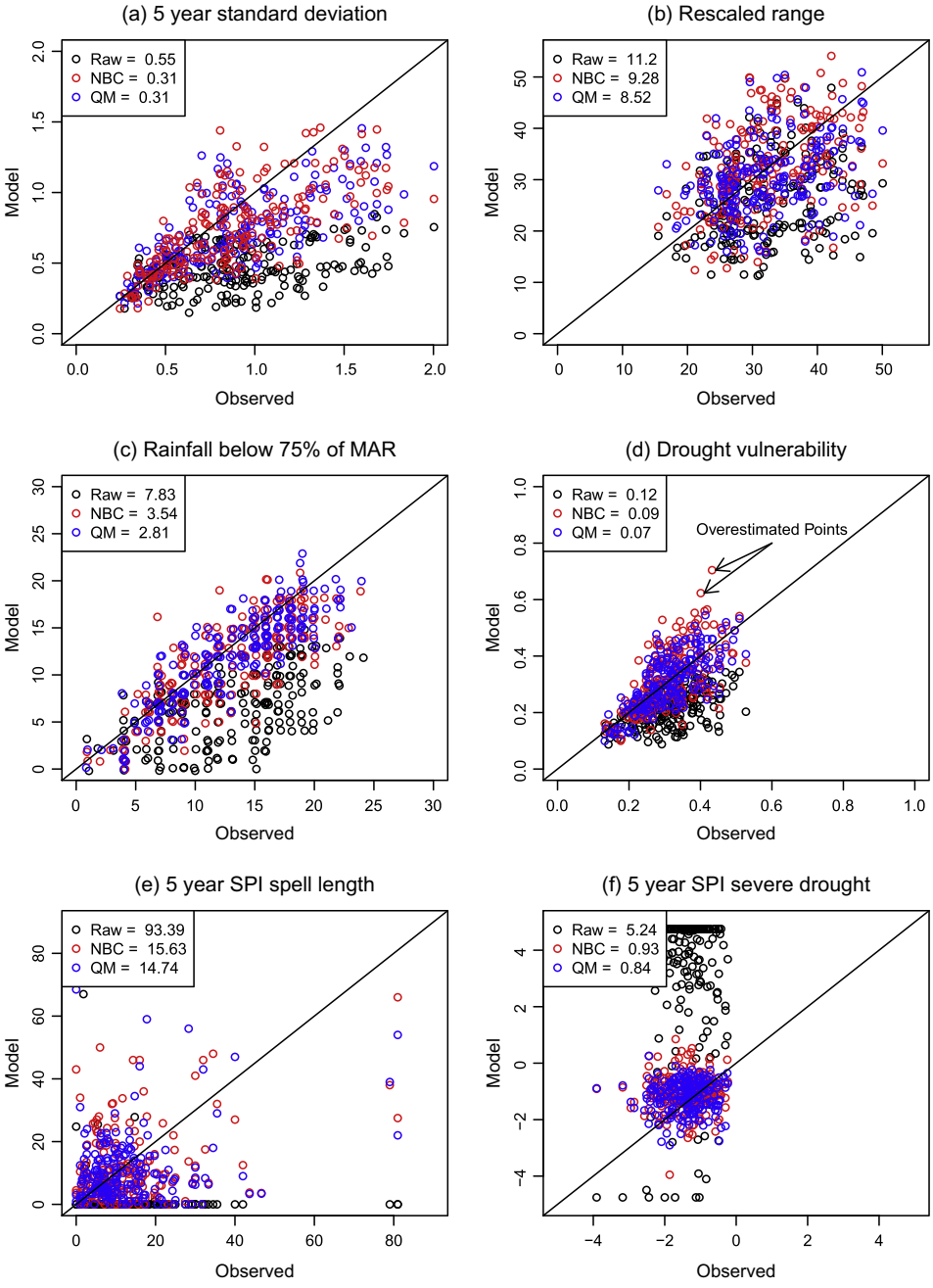

Fig. 7. As for Fig. 6 but showing results for INGV_ECHAM4.

the observed drought frequencies for 1970–1999, acknowledging that the bias correction is not perfect so there will be some errors in this comparison.

In Fig. 8 the raw GCM future drought frequencies for each grid cell and each model are compared to the future bias corrected drought frequencies. A local polynomial regression is fitted to the data in each panel to explore the relationship between drought frequencies with and without bias correction. Three drought statistics have been considered.

The first statistic presented in Fig. 8 is the severe drought thresholds from the SPI as defined by Burke and Brown (2008) (top row of Fig. 8). When future conditions are drier the historical severe (5th percentile) drought threshold will be exceeded more frequently. For example, a future drought frequency of 20% means that the event that currently only occurs with a return period of 1 in 20 will occur in the future with a return period of 1 in 5. Such conditions are evident in Fig. 8 where approximately 50% of grid cells and models project drier conditions and the frequency of exceeding the current drought thresholds increases. These large changes tend to be found in the southern parts of the country and are evident whether raw or bias corrected simulations are used.

When the raw GCM estimates are compared to the bias-corrected estimates, it can be seen that the fitted regression line lies below the one to one line at the highest frequencies. This means that bias correction tends to lead to smaller future increases in drought frequency than using raw GCM simulations. For both bias correction approaches the threshold for this moderation starts at raw drought frequencies of around 30%. This response is consistent for both the moderate and severe drought thresholds and for both bias correction approaches.

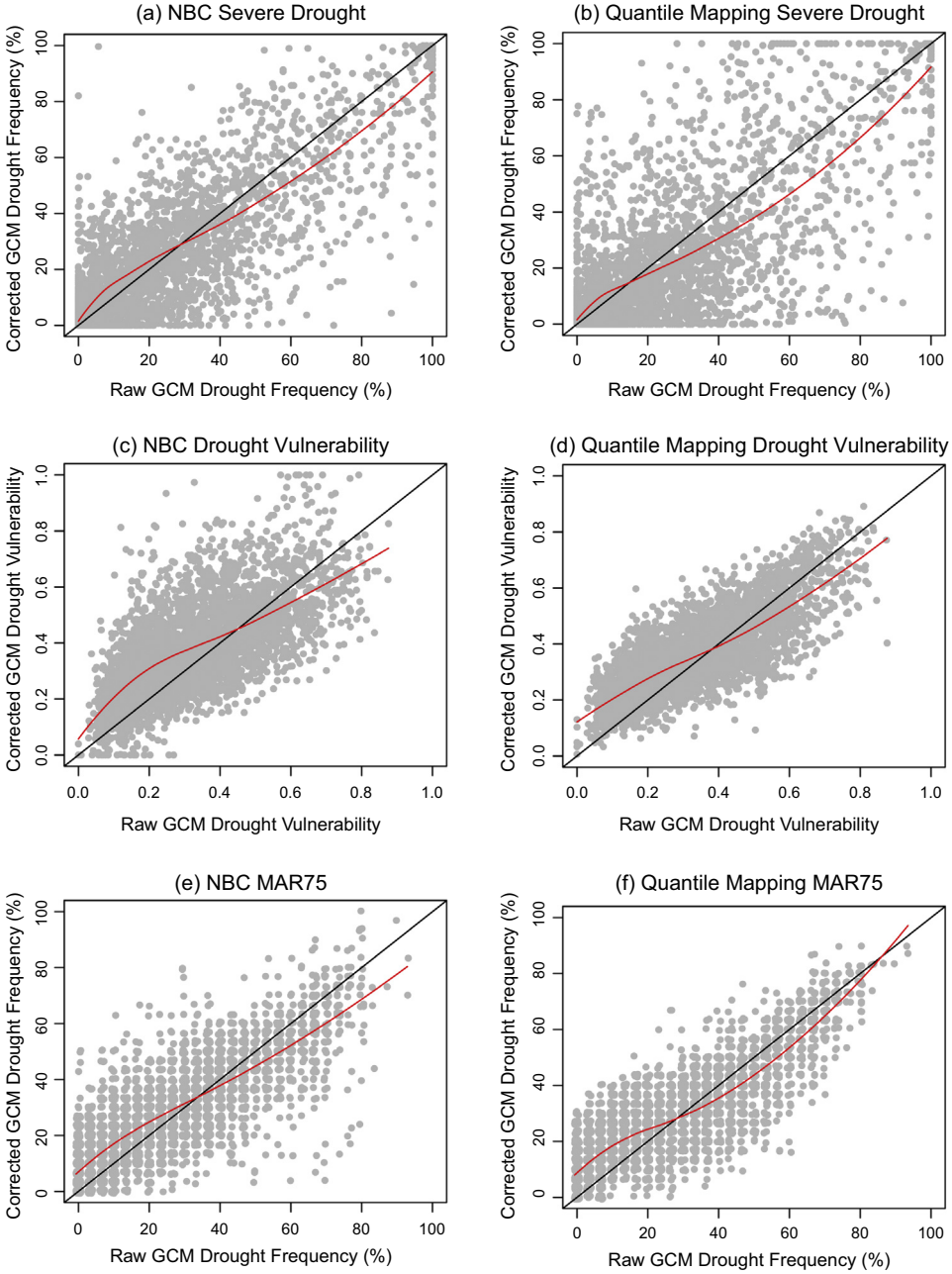

Fig. 8. Comparisons of raw and bias corrected projections of future drought for all grid cells and all GCMs for 5 year SPI Severe Drought frequency (top row), drought vulnerability (middle row) and MAR75 (bottom row). A local polynomial regression has been fitted in each case to summarize the relationship between raw and bias-corrected estimates.

Similar results are found when the Peel drought vulnerability and MAR75 statistic are considered (middle and bottom rows of Fig. 8). For these cases, the future drought estimates are calculated by using the median or mean annual rainfall from historical climate with the future climate time series. For example the MAR75 future drought statistic calculates the percentage of years in the future where the rainfall falls below 75% of the 1970–1999 mean annual rainfall. Even in the historical climate the value of this statistic will vary depending on the variance and skewness of the annual rainfall time series and therefore it is not obvious from Fig. 8 which estimates are wetter or drier than the historical climate except for the SPI statistic. This is the clear advantage of using the predefined thresholds of the SPI, which by definition are known drought frequencies. The disadvantage of the threshold exceedances is that they give no measure of the relative wetness or dryness of different locations unlike the MAR75 and Peel drought vulnerability.

How do drought frequencies change at individual locations and is there a strong spatial pattern to these changes? Fig. 9 provides the consensus position from all the GCMs for raw and bias-corrected changes in drought frequency. In the top row, the drought frequency using the SPI 5 year severe drought threshold is assessed, whilst the bottom row shows the results using the Peel drought vulnerability. Increased (decreased) drought frequency is shown in red (blue) when two-thirds of the GCMs project drier (wetter) conditions. The areas shown in grey occur when the GCMs are more evenly split in whether wetter or drier conditions will occur in the future. In the case of the SPI there are two clear signals whether raw or bias corrected results are used. Wetter conditions are projected for northern Australia and drier conditions in the south, particularly in southwestern Western Australia and along the eastern coast of the Great Australian Bight. Interestingly in the southeast less frequent droughts are projected if bias-corrected simulations are used whereas the raw GCMs show mixed signals. After bias correction at these locations a small number of GCMs (around 2–5 models) have lower drought frequencies than found in the raw simulations and this is enough to then cross the threshold of two-thirds of models agreeing on the direction of the change. It is not always the same GCM at each grid cell where the drought frequencies change direction.

There is considerable uncertainty in the projections of the future drought frequency that is masked by the use of mean changes or GCM consensus results. At each grid cell the minimum and maximum future drought frequencies are found for all 38 model realizations for the 5 year SPI severe drought threshold. The model spread is the difference between the maximum and minimum values. In Fig. 10, empirical cumulative distribution functions are shown for the raw and bias-corrected model spread. These distributions represent the model spread at the 201 grid cells. At approximately 10–15% of grid cells the model spread is 100% (the maximum possible) and bias correction does not improve this situation. The NBC method decreases the range of all the remaining grid cells, implying a decrease in model structural uncertainty following bias correction. The QM method leads to an increase in model spread at about 30% of grid cells but for the remainder of locations the range is substantially decreased compared to the raw simulations. Decreases in uncertainty are found mainly in the northern parts of Australia. Areas with the highest uncertainty are in the southeastern part of the country and this is where uncertainty is not improved after bias-correction.

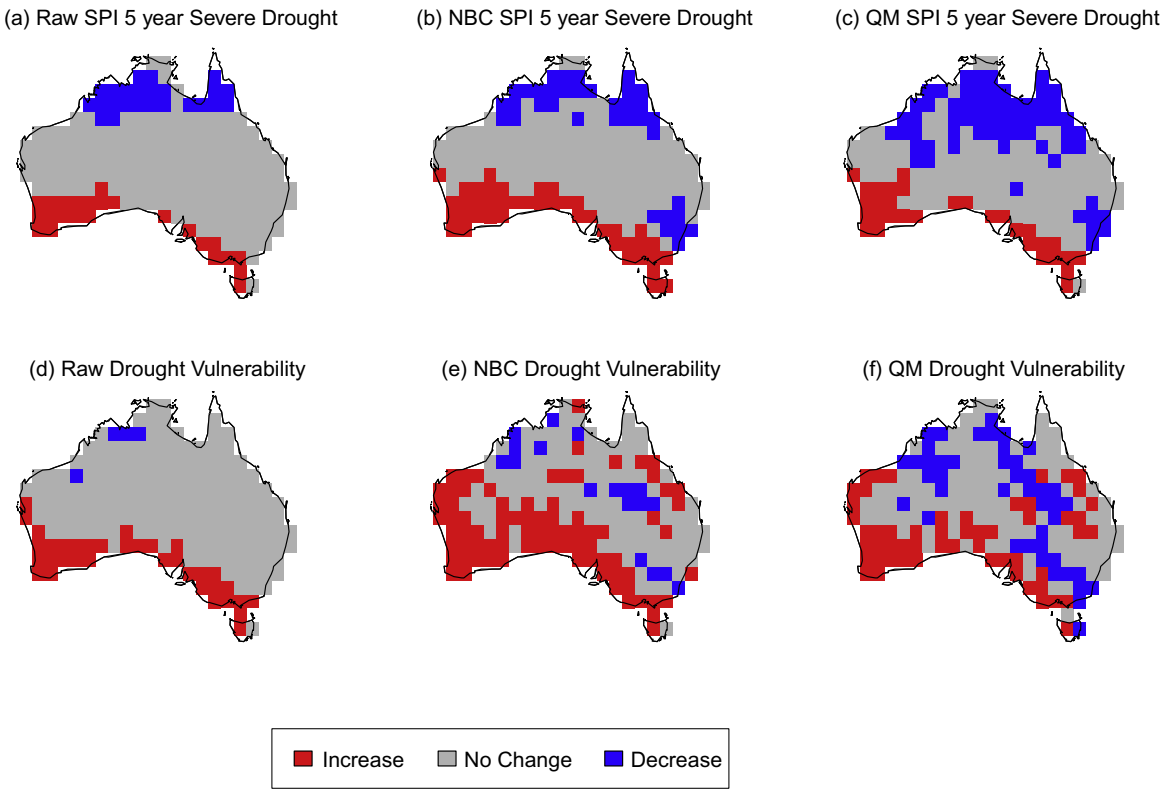

Fig. 9. Spatial variation of model consensus in future changes in drought frequency for SPI 5 year severe drought (top row) and drought vulnerability (bottom row) for raw, NBC and QM simulations. Areas shown in red (blue) have at least two-thirds of models projecting increased (decreased) drought frequency in the future. Areas shown in grey have limited model consensus on the direction of change in future droughts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

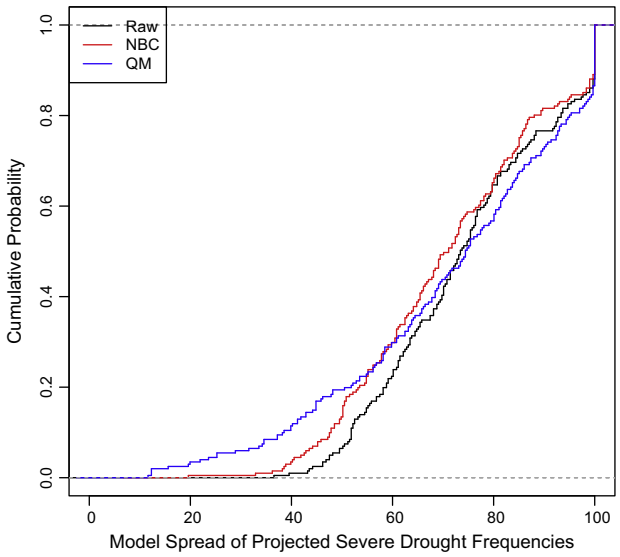

Fig. 10. Empirical cumulative frequency distributions of the model spread across all GCMs of the future drought frequency (5 year SPI) for raw and bias corrected simulations. Uncertainty in the model projections is smaller when the spread is smaller.

6. Discussion and conclusions

In this paper we have sought to consider the impact of bias correction of GCM precipitation simulations on future drought frequencies. We find that future drought frequencies for Australia are expected to increase but bias correction tends to moderate these increases and in some instances improves the consensus amongst a large set of GCMs.

One of the limitations of the study is that the drought assessments were only based on precipitation simulations rather than drought measures that consider the full water balance. Different drought projections would be expected if a combined drought index such as the Palmer Drought Severity Index (PDSI) (Palmer, 1965) or the Reconnaissance Drought Index (RDI) (Zarch et al., in press) was used. Both of these approaches would require joint bias correction of precipitation and potential evaporation (Piani and Haerter, 2012). This joint correction is simple to formulate within the NBC framework, and amounts to the use of a Multivariate Autoregressive Model instead of a simple Autoregressive Model in the bias correction method (Mehrotra and Sharma, 2015). It is important to jointly correct the temperature/evaporation and precipitation rather than implementing independent corrections so that the dynamic relationships between the two can be maintained.

An alternative to the joint modelling of precipitation and evaporation for drought is to assess drought through analysis of soil moisture derived from the GCMs directly (Sheffield and Wood, 2008). The disadvantage of this approach is that soil moisture is available for fewer models than precipitation. There are also problems with attempting to bias correct soil moisture projections from GCMs due to the sparse coverage of soil moisture observation networks and additional problems with the scale discrepancy between the point measurements and the GCM grids. Future work could investigate the utility of satellite soil moisture products for this purpose.

Correctly understanding drought likelihood can allow society to prepare and manage water resources systems to minimize the impacts of these devastating events. How drought frequencies may change in the future due to anthropogenic climate change is a question of interest to many sections. Answering this question is difficult due to the uncertainties in climate model projections of rainfall in particular. Most previous drought assessments have focused on the change in drought frequencies without explicitly considering the biases in the rainfall and therefore in the drought estimates for the current climate. In this paper it is shown that bias correcting precipitation simulations prior to calculating drought changes can lead to different conclusions on future drought frequencies. A number of drought statistics were considered including the Standardized Precipitation Index, statistics of the rainfall time series on annual and inter-annual timescales and a drought vulnerability statistic. Two different bias correction techniques were investigated that corrected the monthly GCM rainfall simulations, namely Nested Bias Correction and Quantile Mapping. Both were shown to reduce the prediction errors when the simulations are compared to drought frequencies calculated from observed data.

Future drought frequencies were assessed using raw GCM outputs, which assumes that the biases in the GCMs are additive and can therefore be removed by taking differences between the future and current projections. The bias corrected GCM outputs using Quantile Mapping and Nested Bias Correction were also assessed. It was found for the CMIP3 GCMs assessed in this study that bias correction tended to moderate the largest future increases in drought frequency suggested when using the raw GCM outputs. It is thus concluded that raw model simulations can lead to incorrect drought assessments even at continental scales and bias corrections need to be applied. These results are consistent across models and across the continent. A slight improvement in the agreement in the number of models projecting drier or wetter conditions was also found after bias correction, although for many parts of Australia there is still substantial uncertainty across the projections.

Acknowledgments

We acknowledge the modelling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modeling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. We acknowledge the Australian Research Council and the Sydney Catchment Authority for partial funding to support this work.

References

Ault, T.R., Cole, J.E., Overpeck, J.T., Pederson, G.T., Meko, D.M., 2014. Assessing the risk of persistent drought using climate model simulations and paleoclimate data. J. Clim.

Burke, E.J., Brown, S.J., 2008. Evaluating uncertainties in the projection of future drought. J. Hydromet. 9 (2), 292–299.

Burke, E.J., Brown, S.J., Christidis, N., 2006. Modeling the recent evolution of global drought and projections for the twenty-first century with the Hadley Centre climate model. J. Hydromet. 7 (5), 1113–1125.

Buser, C., Künsch, H., Lüthi, D., Wild, M., Schär, C., 2009. Bayesian multi-model projection of climate: bias assumptions and interannual variability. Clim. Dyn. 33 (6), 849–868.

Chan, S. et al., 2013. Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation? Clim. Dyn. 41 (5–6), 1475–1495.

Chen, J., Brissette, F.P., Chaumont, D., Braun, M., 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. Water Resour. Res. 49 (7), 4187–4205.

Christensen, J.H., Boberg, F., Christensen, O.B., Lucas-Picher, P., 2008. On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys. Res. Lett. 35.

Dai, A., 2013. Increasing drought under global warming in observations and models. Nat. Clim. Change 3 (1), 52–58.

Fowler, H.J., Kilsby, C.G., 2007. Using regional climate model data to simulate historical and future river flows in northwest England. Clim. Change 80 (3), 337–367.

Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007a. Linking climate change modelling to impact studies: recent advances in downscaling techniques for hydrological modelling. Int. J. Climatol. 27 (12), 1547–1578.
Fowler, H.J., Kilsby, C.G., Stunell, J., 2007b. Modelling the impacts of projected future climate change on water resources in north-west England. Hydrol. Earth Syst. Sci. 11 (3), 1115–1124.
Grant, I., Jones, D., Wang, W., Fawcett, R., Barratt, D., 2008. Meteorological and remotely sensed datasets for hydrological modelling: A contribution to the Australian Water Availability Project, Catchment-scale Hydrological Modelling & Data Assimilation (CAHMDA-3) International Workshop on Hydrological Prediction: Modelling, Observation and Data Assimilation, Melbourne. Citeseer.
Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T., 2012. Technical note: downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. Hydrol. Earth Syst. Sci. 16 (9), 3383–3390.
Guttman, N.B., 1999. Accepting the standardized precipitation index: a calculation algorithm. J. Am. Water Resour. Assoc. 35 (2), 311–322.
Haerter, J., Hagemann, S., Moseley, C., Piani, C., 2011. Climate model bias correction and the role of timescales. Hydrol. Earth Syst. Sci. 15 (3), 1065–1079.
Ines, A.V.M., Hansen, J.W., 2006. Bias correction of daily GCM rainfall for crop simulation studies. Agric. For. Meteorol. 138 (1–4), 44–53.
IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535pp.
Jiménez Cisneros, B.E. et al., 2014. Freshwater resources. In: Field, C.B. et al. (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 229–269.
Johnson, F., Sharma, A., 2011. Accounting for interannual variability: a comparison of options for water resources climate change impact assessments. Water Resour. Res. 47 (4), W04508.
Johnson, F., Sharma, A., 2012. A nesting model for bias correction of variability at multiple times scales in general circulation model precipitation simulations. Water Resour. Res. 48, W01504.
Kiem, A.S., Franks, S.W., Kuczera, G., 2003. Muti-decadal variability of flood risk. Geophys. Res. Lett. 30 (2), 1035.
Koutsoyiannis, D., 2001. Coupling stochastic models of different timescales. Water Resour. Res. 37 (2), 379–391.
Koutsoyiannis, D., 2002. The Hurst phenomenon and fractional Gaussian noise made easy. Hydrol. Sci. J. – J. Des Sci. Hydrol. 47 (4), 573–595.
Li, H., Sheffield, J., Wood, E.F., 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. J. Geophys. Res. 115, D10101.
Maraun, D. et al., 2010. Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. 48 (RG3003), 34.
Maurer, E.P., Das, T., Cayan, D.R., 2013. Errors in climate model daily precipitation and temperature output: time invariance and implications for bias correction. Hydrol. Earth Syst. Sci. 17 (6), 2147–2159.
Mehrotra, R., Sharma, A., 2012. An improved standardization procedure to remove systematic low frequency variability biases in GCM simulations. Water Resour. Res. 48 (12).
Mehrotra, R., Sharma, A., 2015. Correcting for systematic biases in multiple raw GCM variables across a range of timescales. J. Hydrol. 520, 214–223.
Mpelasoka, F., Hennessy, K., Jones, R., Bates, B., 2008. Comparison of suitable drought indices for climate change impacts assessment over Australia towards resource management. Int. J. Climatol. 28 (10), 1283–1292.
Orlowsky, B., Seneviratne, S.I., 2013. Elusive drought: uncertainty in observed trends and short- and long-term CMIP5 projections. Hydrol. Earth Syst. Sci. 17 (5), 1765–1781.
Palmer, W.C., 1965. Meteorological Drought. Research Paper No. 45.
Peel, M.C., McMahon, T.A., Pegram, G.G., 2005. Global analysis of runs of annual precipitation and runoff equal to or below the median: run magnitude and severity. Int. J. Climatol. 25 (5), 549–568.
Piani, C., Haerter, J.O., 2012. Two dimensional bias correction of temperature and precipitation copulas in climate models. Geophys. Res. Lett. 39 (20), L20401.
Piani, C., Haerter, J., Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. Theoret. Appl. Climatol. 99 (1), 187–192.
Pitman, A.J., Arneth, A., Ganzeveld, L., 2012. Regionalizing global climate models. Int. J. Climatol. 32 (3), 321–337.
Prudhomme, C. et al., 2014. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. Proc. Natl. Acad. Sci. 111 (9), 3262–3267.
Pui, A., Lall, A., Sharma, A., 2011. How does the Inter-decadal Pacific Oscillation affect Design Floods in Australia? Water Resour. Res. 47, W05554.
Rocheta, E., Sugiyanto, M., Johnson, F., Evans, J., Sharma, A., 2014. How well do general circulation models represent low-frequency rainfall variability? Water Resour. Res. 50 (3), 2108–2123.
Sheffield, J., Wood, E.F., 2008. Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven Simulations of the terrestrial hydrologic cycle. J. Clim. 21 (3), 432–458.
Srikanthan, R., Pegram, G.G.S., 2009. A nested multisite daily rainfall stochastic generation model. J. Hydrol. 371 (1), 142–153.
Taylor, I. et al., 2013. The impact of climate mitigation on projections of future drought. Hydrol. Earth Syst. Sci. 17, 2339–2358.
Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. J. Hydrol. 456, 12–29.
Themeßl, M.J., Gobiet, A., Heinrich, G., 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. Clim. Change 112 (2), 449–468.
Verdon, D.C., Wyatt, A.M., Kiem, A.S., Franks, S.W., 2004. Multidecadal variability of rainfall and streamflow: Eastern Australia. Water Resour. Res. 40 (W10201).
Vörösmarty, C.J. et al., 2010. Global threats to human water security and river biodiversity. Nature 467 (7315), 555–561.
Wang, L., Chen, W., 2014. A CMIP5 multimodel projection of future temperature, precipitation, and climatological drought in China. Int. J. Climatol. 34 (6), 2059–2078.
Warszawski, L. et al., 2014. The inter-sectoral impact model intercomparison project (ISI–MIP): project framework. Proc. Natl. Acad. Sci. 111 (9), 3228–3232.
Wood, A.W., Leung, L.R., Sridhar, V., Lettenmaier, D.P., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Clim. Change 62 (1–3), 189–216.
Zarch, M.A.A., Sivakumar, B., Sharma, A., in press. Droughts in a warming climate: a global assessment of Standardised precipitation index (SPI) and Reconnaissance drought index (RDI). J. Hydrol. http://dx.doi.org/10.1016/j.jhydrol.2014.09.071 (in press).