Hydrol. Earth Syst. Sci. Discuss., 11, 10683–10724, 2014
www.hydrol-earth-syst-sci-discuss.net/11/10683/2014/
doi:10.5194/hessd-11-10683-2014

Hydrology and Earth System Sciences
Discussions
Open Access

How does bias correction of RCM precipitation affect modelled runoff?

J. Teng[1], N. J. Potter[1], F. H. S. Chiew[1], L. Zhang[1], J. Vaze[2], and J. P. Evans[2]

[1]CSIRO Land and Water Flagship, Canberra, Australia
[2]Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, Australia

Received: 29 August 2014 – Accepted: 29 August 2014 – Published: 23 September 2014

Correspondence to: J. Teng (jin.teng@csiro.au)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Many studies bias correct daily precipitation from climate models to match the observed precipitation statistics, and the bias corrected data are then used for various modelling applications. This paper presents a review of recent methods used to bias correct precipitation from regional climate models (RCMs). The paper then assesses four bias correction methods applied to the weather research and forecasting (WRF) model simulated precipitation, and the follow-on impact on modelled runoff for eight catchments in southeast Australia. Overall, the best results are produced by either quantile mapping or a newly proposed two-state gamma distribution mapping method. However, the difference between the tested methods is small in the modelling experiments here (and as reported in the literature), mainly because of the substantial corrections required and inconsistent errors over time (non-stationarity). The errors remaining in bias corrected precipitation are typically amplified in modelled runoff. The tested methods cannot overcome limitation of RCM in simulating precipitation sequence, which affects runoff generation. Results further show that whereas bias correction does not seem to alter change signals in precipitation means, it can introduce additional uncertainty to change signals in high precipitation amounts and, consequently, in runoff. Future climate change impact studies need to take this into account when deciding whether to use raw or bias corrected RCM results. Nevertheless, RCMs will continue to improve and will become increasingly useful for hydrological applications as the bias in RCM simulations reduces.

1 Introduction

Downscaling is a technique commonly used in hydrology when investigating the impact of climate change. It is a way of bridging the gap between low spatial resolution global climate models (GCMs) and the catchment- or regional-scale hydrological models (Fowler et al., 2007). Dynamical downscaling techniques derive regional-scale

HESSD
11, 10683–10724, 2014

information by using a high-resolution climate model over a limited area and, forcing it with lateral boundary conditions from GCMs or reanalysis products. In brief, it is modelling with a regional climate model, or RCM. With advances in RCMs and the increasing availability of RCM simulations, this type of downscaling is gaining more and more popularity in climate change impact studies (Dosio et al., 2012; Argueso et al., 2013; Seaby et al., 2013; Teutschbein and Seibert, 2010; Maraun et al., 2010; Bennett et al., 2012). A drawback, however, is that precipitation simulations from RCMs are "biased": in addition to errors inherited from the driving GCM, there are systematic RCM model errors, due to imperfect conceptualization and parameterization, inadequate length and quality of reference data sets, and insufficient spatial resolution (Wilby et al., 2000; Wood et al., 2004; Piani et al., 2010b; Chen et al., 2011a; Christensen et al., 2008; Teutschbein and Seibert, 2010). Various "bias correction" methods have been developed in an attempt to minimize these errors (Boe et al., 2007; Piani et al., 2010a; Johnson and Sharma, 2012; Schmidli et al., 2006; Lenderink et al., 2007).

There have been extensive discussions in the climate change literature on the definition of "bias", and some have recommended limiting its use to refer to the correspondence between a mean forecast and the mean of the observations averaged over a certain area and time (Ehret et al., 2012); others have tried to distinguish model biases from model shortcomings and model errors (Teutschbein and Seibert, 2013). For clarity, in this paper we define bias as the systematic distortion of a statistical outcome from the expected value, and we use "error" or "difference" to refer to the discrepancy between a model output and observations.

Many studies have compared and evaluated different bias correction methods; Table 1 summarizes some recent ones and their main conclusions. Most of these studies investigated the impact of bias correction on precipitation and temperature (see column 6, Table 1), while few of them, except for Teutschbein and Seibert (2012) and Chen et al. (2013), tested the effect of bias correction on the outputs of hydrological models. Nearly all studies agree that distribution-based bias correction methods (both parametric and non-parametric) give the best performance in terms of reproducing the

HESSD
11, 10683–10724, 2014

observed climate, whereas means-based methods, in particular linear scaling (LS), are almost always ranked as the least-skilled bias correction method.

Building on the knowledge gained from previous comparison studies, we have assessed in more detail the best performing bias correction technique – distribution mapping – and compared its performance in several forms against the linear scaling (LS) method as a benchmark (other names of the distribution mapping technique include: quantile matching, distribution transformation, probability mapping, and histogram equalisation). Our main interest is to examine the effect of bias correction on downstream hydrological models. The bias correction methods were applied on precipitation, as it is the most critical and difficult-to-model variable in hydrological studies, and evaluated on both precipitation and modelled runoff using a cross-validation method. The raw and bias-corrected precipitation data were used to drive the hydrological models. The key precipitation and runoff characteristics were compared to those of observations to investigate how bias correction affects RCM precipitation, and its follow-on impact on runoff propagating through hydrological models.

Previous studies have shown mixed results in ranking the different types of distribution mapping methods, suggesting that there may be only marginal differences between the methods. For example, some studies have shown that distribution mapping based on theoretical distributions outperforms other bias correction methods (Teutschbein and Seibert, 2013, 2012; Yang et al., 2010). Others have shown that theoretical distribution mapping performs similar to, or only marginally better than, empirical quantile mapping (Berg et al., 2012; Chen et al., 2013). Some studies, on the other hand, show that empirical quantile mapping demonstrates higher skill than theoretical distribution mapping in systematically correcting RCM precipitation (Gudmundsson et al., 2012; Gutjahr and Heinemann, 2013; Li et al., 2010; Lafon, 2013). In view of the discrepancy in the literature, we compared three distribution mapping techniques, each with increasing degree of dependency on the calibration data, in order to evaluate the methods based on both accuracy and robustness.

Berg et al. (2012) found that 30 years of calibration data are required to produce reasonable accuracy for the estimate of precipitation variance. Due to the difficult parameterisation and expensive computational costs associated with RCMs, this requirement is not easily met in impact studies. In the main modelling experiments, we chose two 8 year long periods of RCM data (16 years split in half), with significant difference between them, to examine whether the bias correction method derived from one period works for another period, and if not, what causes it to fail. We then validated the generality of our conclusion using two 30 year long RCM precipitation data (60 years split in half).

Chen et al. (2013) concluded that bias correction performance is location dependent and that virtually no bias correction method succeeds in catchments having low coherence between RCM simulated and observed precipitation sequences. We challenged (and confirmed) this conclusion by evaluating the precipitation sequence simulated by the RCM and quantifying the effect of precipitation sequence on modelled runoff.

The impact of bias correction on the change signals (one period vs. another) in both precipitation and runoff was also explored. There were two possible outcomes from this investigation: if bias correction does not alter the change signals in the hydro-climatic projections, then the use of bias correction should be considered either unnecessary or safe to use, depending on the circumstance (Muerth et al., 2013). If it does alter the change signal, bias correction could be reducing or increasing the errors in the change signals, either way, it introduces an extra level of uncertainty in the modelling chain.

This paper contributes to the present lively discussion on whether bias correction methods should be applied to global and regional climate model data, a conversation initiated by Christensen et al. (2008), stimulated by Ehrel et al. (2012), and continued by more recent studies such as Muerth et al. (2013) and Teutschbein and Seibert (2013).

2 Study area and data

2.1 Study area

The study area was located in the southern Murray–Darling Basin, Australia (Fig. 1). Beginning in the mid-90s, this area experienced a prolonged drought, a so-called “Millennium Drought”, for 10–15 years (Chiew et al., 2010). While the mean annual rainfall over the Millennium Drought was 10–20 % below the long-term mean, in some places the mean annual runoff declined by over 50 %, a reduction unprecedented in historical records (Potter and Chiew, 2011). Eight catchments from the Loddon, Campaspe, and Goulburn River Basins, with areas from 250 to 1033 km^2, were selected for this study. The catchments were mostly unregulated, with continuous climate and streamflow measurements available for 1985–2000, as such the assessment period was chosen. An 8 year period unaffected by the drought (1985–1992) was used as the calibration period, and another 8 year period strongly affected by the drought (1993–2000) was used as the validation period. Subsequently, they were switched for cross-validation.

2.1.1 Observations

Observed daily precipitation data were derived from 0.05° (~ 5 km) gridded climate surfaces and averaged over each catchment. The source of this dataset was the SILO Data Drill (http://www.longpaddock.qld.gov.au/silo) of the Department of Science, Information Technology, Innovation and the Arts, Queensland, Australia (Jeffrey et al., 2001). The SILO gridded climate datasets provide surfaces of daily rainfall and other climate data interpolated from high quality point measurements provided by the Australian Bureau of Meteorology. The daily potential evapotranspiration (PET) sequences used in the hydrological modelling were calculated from SILO climate variables using Morton’s wet environment algorithms (Chiew and McMahon, 1991).

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |
HESSD
11, 10683–10724, 2014

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀ ▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Measured daily streamflow data were sourced from a previous study (Vaze et al., 2010) and used to calibrate the hydrological models.

2.1.2 RCM data

Most of the analysis in this study was carried out using daily precipitation series for the period 1985–2000, which were simulated by Evans and McCabe (2010) using the weather research and forecasting (WRF) model. Another 60 year long (1950–2009) WRF precipitation dataset (Evans et al., 2014) was used to validate the conclusion reached by using the shorter dataset. For both datasets, WRF was implemented on a 10 km grid using lateral boundary conditions taken from the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis dataset (Kalnay et al., 1996, see http://www.cdc.noaa.gov/cdc/reanalysis). The WRF simulations have been found capable of capturing the drought experienced over the study area in another study (Evans and McCabe, 2010). The daily precipitation series for each catchment were aggregated from the WRF simulation by averaging all the grid cells over the catchment.

3 Method

3.1 Bias correction methods

In this study daily precipitation was the main variable subjected to bias correction. Typically, bias correction methods aim to correct the mean, variance, and/or distribution of the modelled precipitation by using a function *h*:

$$\hat{p}_{\text{obs}} = h(p_{\text{mod}}) \quad (1)$$

so that the transformed precipitation matches the observed data more closely than the modelled precipitation.

HESSD
11, 10683–10724, 2014

3.1.1 Linear scaling (LS)

The simplest choice for h is probably a linear transformation

$$\hat{p}_{\text{obs}} = a p_{\text{mod}} \tag{2}$$

where a is a free parameter that is subject to calibration. This simple form of bias correction is widely used to adjust precipitation from GCMs, RCMs, and statistical downscaling methods (Maraun et al., 2010; Teng et al., 2012a). It can efficiently correct the means but does not account for the higher moments. In this study, this method served as the benchmark as LS has been identified in various studies as the least skilful bias correction method (Gudmundsson et al., 2012; Lafon, 2013; Chen et al., 2013; Teutschbein and Seibert, 2012). The LS parameter a was optimized for each season: DJF (December–February), MAM (March–May), JJA (June–August), and SON (September–November) to account for precipitation seasonality. Similarly, seasonal optimization was also applied for all the other bias correction methods used in this study.

3.1.2 Distribution mapping using the gamma distribution (DMG)

The relation in Eq. (1) can also be modelled so that the distribution of the modelled precipitation matches that of the observations:

$$\hat{p}_{\text{obs}} = F_{\text{obs}}^{-1}(F_{\text{mod}}(p_{\text{mod}})) \tag{3}$$

where F_{mod} is the cumulative distribution function (CDF) of P_{mod} and F_{obs}^{-1} is the inverse CDF corresponding to P_{obs}. These CDFs can be either theoretical distributions fitted to the data, or empirical distributions estimated by sorting the data. The gamma distribution with shape parameter α and rate parameter β (Eq. 4) is often used to represent non-zero precipitation amounts (Piani et al., 2010a; Lafon, 2013), as it has

HESSD
11, 10683–10724, 2014

the ability to approximate the positively skewed distributions (Yang et al., 2010). The probability density function (PDF) for a gamma random variable is given by:

$$f(p) = \frac{\beta^{\alpha} p^{\alpha-1} e^{-\beta p}}{\Gamma(\alpha)} \tag{4}$$

where $\Gamma(\alpha)$ is the gamma function evaluated at α.

When estimating parameters for the gamma distribution, we used the method of maximum-likelihood estimation as it is more accurate (i.e. the standard error of the estimates is lower) compared to the method of moments or least-squares estimation (e.g. Piani et al., 2010a).

Given an occurrence of non-zero precipitation amount $p_i > 0$ for $i = 1, \ldots, n$, the log-likelihood function of the gamma distribution can be written as:

$$l(\alpha, \beta) = \sum_{i=1}^{n} \log f(p_i; \alpha, \beta). \tag{5}$$

The maximum-likelihood estimates for α and β are chosen to maximise this log-likelihood function. To account for dry days, we define the PDF for zero and non-zero precipitation days $f_0(p)$ as a mixed distribution with an atom of probability at $p = 0$ and a gamma distribution for $p > 0$ so that:

$$f_0(p) = \begin{cases} q_0, & p = 0 \\ \frac{(1-q_0)\beta^{\alpha} p^{\alpha-1} e^{-\beta p}}{\Gamma(\alpha)}, & p > 0 \end{cases} \tag{6}$$

The maximum likelihood estimate of q_0 depends only on the relative number of zero-precipitation days (n_0):

$$q_0 = n_0/n. \tag{7}$$

The shape and rate parameters α and β are calculated on the non-zero precipitation amounts.

HESSD
11, 10683–10724, 2014

3.1.3 Distribution mapping using a double gamma distribution (DM2G)

Daily precipitation distributions are typically heavily skewed towards high-intensity values. As a result, when fitting a single gamma distribution, the distribution parameters will be dictated by the most frequently occurring values, but may then not accurately represent the extremes. To capture normal precipitation values as well as extremes, different approaches have been tried, but the most common is to divide the precipitation distribution into segments and fit separate distributions to each segment (Yang et al., 2010; Grillakis et al., 2013; Gutjahr and Heinemann, 2013). Instead of introducing arbitrary cut-offs, we propose what can be interpreted as a two-state distribution. It is a mix of two gamma distributions which can model non-zero precipitation amounts:

$$f(p) = \lambda \frac{\beta_1^{\alpha_1} p^{\alpha_1 - 1} e^{-\beta_1 p}}{\Gamma(\alpha_1)} + (1-\lambda) \frac{\beta_1^{\alpha_2} p^{\alpha_2 - 1} e^{-\beta_2 p}}{\Gamma(\alpha_2)} \quad (8)$$

with $0 < \lambda < 1$. The parameter λ is the relative occurrence of the states, and, fitted correctly, the two gamma distributions represent rainfall occurring in high and low rainfall states. The advantage of this approach compared to segmenting the distribution is that all parameters can be estimated simultaneously using maximum-likelihood estimation. Thus, six parameters – q_0, α_1, β_1, α_2, β_2, and λ – were estimated from observations and from the RCM output for the calibration period; they were then used to correct the RCM output for the validation period.

The Kolmogorov–Smirnov (KS) test (Chakravarti and Laha, 1967) performed on both observations and RCM simulations confirmed that the double gamma distribution gives better fittings compared to the gamma distribution (a table of KS test results is provided as the Supplement). Figure 2 compares the empirical distribution, gamma, and double gamma distribution for one catchment. A significant improvement in fit is achieved by the double gamma distribution compared to the gamma distribution, especially at the higher end.

HESSD

11, 10683–10724, 2014

3.1.4 Empirical quantile mapping

Apart from using theoretical distributions, the empirical CDF is also commonly used to solve Eq. (3) (Themeßl et al., 2011; Gudmundsson et al., 2012; Boe et al., 2007; Bennett et al., 2014). Here the empirical CDFs of observed and modelled precipitation were estimated using empirical percentiles. Values in between the percentiles were approximated using linear interpolation. In cases where new RCM values (such as from the validation period) were larger than the calibration values used to estimate the empirical CDF, a linear regression fit on the last five data points was used to extrapolate beyond the range of observations and allow for possible "new extremes".

3.2 Hydrological modelling (HM)

Two lumped conceptual daily rainfall–runoff models – GR4J (Perrin et al., 2003) and Sacramento (Burnash et al., 1973) – were used to model runoff. The model versions were very similar to those described in foregoing references and in Vaze et al. (2010). The two models have interconnected soil moisture stores and algorithms that mimic the hydrological processes of water moving into and out of soil moisture stores. The choice of models did not have large effect on the conclusions of this study because the errors associated with hydrological models are relatively small (Teng et al., 2012b; Chen et al., 2011b). For the application here, the numbers of parameters calibrated were four for GR4J and 14 for Sacramento. The models were calibrated against observations for the two periods separately, with the model parameters optimised to maximise the NSE-bias objective function; this function is a weighted combination of the Nash–Sutcliffe Efficiency (Nash and Sutcliffe, 1970) and a logarithmic function of bias in the modelled mean annual streamflow (Viney et al., 2009). The models were run at a daily time step. To estimate the impact of bias correction on runoff, the models were driven by WRF precipitation before and after bias correction using the optimised parameters derived from the calibrations described above. The same PET dataset calculated using observed climate variables was used throughout the hydrological modelling.

HESSD
11, 10683–10724, 2014

3.3 Evaluating performance

Comparison of the bias correction methods was based on a split-sample cross-validation approach. The 16 years of data (1985–2000) were split into two periods of 8 years each (1985–1992 and 1993–2000). The bias correction methods were trained using one period and tested against the same period ("same") as well as against the other period ("cross"), and vice versa. Similarly, the hydrological models were calibrated using one period and the parameters were used in the "cross" experiments, treating the validation period as though there were no other information except climate from RCM, like a future period. Compared to studies that have used "odd-year/even-year" or "leave-one-out" validation methods, the design of this experiment puts the bias correction methods to a stricter test so that the impact can be clearly identified.

To gauge the impact of bias correction methods on precipitation, we compared the RCM precipitation before and after bias correction with the observations using salient metrics: annual and seasonal means, 99th percentile precipitation as an indicator of high precipitation events, number of dry days (daily precipitation less than 0.1 mm) per year as an indicator of low precipitation, and 99th percentile of 3 and 5 day cumulative precipitation as indicators of runoff-generating events. The runoff modelled using RCM precipitation before and after bias correction was also evaluated against key runoff characteristics: annual and seasonal means, 99th percentile runoff as an indicator of high-flow events, and number of low-flow days (daily runoff less than 0.01 mm) as an indicator of low-flow conditions. We also looked at the effect of bias correction methods on change signals by comparing the relative difference in precipitation and runoff between the two periods derived from various methods.

4 Results

Figure 3 shows the percentage difference in raw RCM and bias corrected RCM precipitation relative to observations for annual and seasonal means, 99th percentile precipitation, 99th percentile 3 and 5 day cumulative precipitation, and the difference in number of dry days per year. Generally, the raw RCM precipitation exhibits negative errors in annual and seasonal means, with the median errors in raw RCM annual means being −9.1 and −22.5 % for the two periods respectively. There are larger (further away from zero) errors in the drier period (1993–2000). The 99th percentile precipitation is mostly overestimated in one period (1985–1992) and underestimated in the other. The raw RCM performs quite well in reproducing 99th percentile 3 and 5 day cumulative precipitation but slightly overestimates number of low precipitation (< 0.1 mm) days.

4.1 Impact on precipitation

The calibration results in Fig. 3 (denoted by "_same") show that, as expected, all the bias correction methods are able to match the annual and seasonal means of precipitation when validating on the same period as the calibration period, (see LS_same, DMG_same, DM2G_same, and QM_same in the boxplots of annual and seasonal means in Fig. 3). For instance, LS perfectly corrects the median errors in annual means for the two periods (0 and 0 %), followed by DM2G (−0.3 and −0.2 %), QM (0.4 and 0.6 %), and DMG (−0.7 and −0.7 %). Only the distribution mapping methods (DMG, DM2G, and QM) are able to reduce the errors in the high- and low-precipitation characteristics; QM in particular performs exceptionally well in reproducing 99th percentile precipitation and number of dry days per year. LS is not only unable to reduce the errors in high- and low-precipitation characteristics, but also increases the errors in some cases, as seen in the 99th percentile precipitation for 1985–1992 (period I, left panels) and number of dry days per year for 1993–2000 (period II, right panels). This is consistent with the findings from previous studies (Chen et al., 2013; Teutschbein and Seibert, 2012).

By contrast, the cross-validation results (denoted by "_cross") seem to depend on the period, with most of the bias correction methods reducing the raw RCM errors (closer to zero) in period II but increasing the raw RCM errors (away from zero) in period

I. The exception is DJF mean precipitation, where all the bias correction methods increase the errors in both periods. Although DM2G performs better compared to other bias correction methods for nearly all precipitation characteristics (shown by lower median errors given by DM2G_cross in Fig. 3), the difference between the bias correction methods is small compared to the overall large overestimation in period I and underestimation in period II. The cause of this "period dependency" is discussed in Sect. 5.1. It is notable that the errors in DJF and MAM mean precipitation are generally larger than in other seasons; the reason is that the amounts of precipitation in DJF and MAM are smaller in these catchments since they are dominated by winter precipitation, with more than 60 % of the precipitation coming from JJA and SON.

4.2 Impact on runoff

Figure 4 presents the relative differences in runoff characteristics simulated by GR4J using raw RCM and bias corrected precipitation when compared to those modelled using observed precipitation. The layout is the similar as Fig. 3 except, for perspective, two boxes are added to each panel to show the conventional hydrological model errors: "HM_calib", which represents the calibration error (when runoff from GR4J driven by observed precipitation is compared with observed streamflow); and "HM_cross", which represents the cross-validation error (when GR4J runoff driven by observed precipitation and using parameters calibrated to the same period is compared with those using parameters calibrated to a different period).

The errors in runoff show a similar pattern to those for precipitation, but are much larger. They are also considerably larger than the hydrological model errors. For instance, the median errors in mean annual runoff simulated using raw RCM precipitation increase to −33.1 % (period I) and −69.5 % (period II). The calibration results show that LS is no longer able to correct the errors in annual and seasonal mean runoff to zero due to errors in high-percentile precipitation (see the 99th percentile precipitation plot in Fig. 3) and, consequently, in high runoff. QM does not perform very well in correcting the high- and low-runoff characteristics as it was able to do for

HESSD
11, 10683–10724, 2014

the high- and low-precipitation characteristics which may relate to its weakness (as shown in Fig. 3) in reproducing 3 day and 5 day cumulative precipitation. These results highlight the importance of precipitation sequence in runoff production, as discussed in Sect. 5.3.

The cross-validation results show that, after bias correction, the median errors in period I are increased to 62–84 % by various bias correction methods. While the median errors in period II are decreased, an error of −34 to −48 % is still considered large compared to the conventional hydrological model error of less than 10 % as shown by HM_calib and HM_cross.

Figure 5 shows the same results as for Fig. 4 but using the Sacramento hydrological model. Similar observations can be made from this figure but with a larger range of the errors, probably because the Sacramento model errors (HM_calib and HM_cross) are larger in some seasons. Sacramento does better in reproducing observed low flows in period I but slightly worse than GR4J in reproducing high flows. In general, the bias correction affects two hydrological models similarly. Therefore the discussion in the following sections will focus on GR4J.

4.3 Impact on change signals

Figure 6 presents the differences in precipitation change signals when comparing raw RCM simulation and bias corrected RCM simulations to observations. Here, the "change" (ΔP) is defined as the relative difference of various charachteristics between period II and period I (Eq. 9):

$$\Delta P = \frac{P_{\mathrm{II}} - P_{\mathrm{I}}}{P_{\mathrm{I}}} \cdot 100. \tag{9}$$

The baseline in Fig. 6 is the change derived from observations (ΔP_{obs}); the "difference" is between the baseline and the change derived from the raw RCM ($\Delta P_{\mathrm{RCM}} - \Delta P_{\mathrm{obs}}$) and bias-corrected RCM simulations ($\Delta P_{\mathrm{BC}} - \Delta P_{\mathrm{obs}}$). The ΔP_{BC} values used to plot the left panel of each plot in Fig. 6 were derived assuming period I is the validation period

HESSD
11, 10683–10724, 2014

How does bias correction of RCM precipitation affect modelled runoff?

J. Teng et al.

and period II the calibration period:

$$\Delta P_{\mathrm{BC}}^{\mathrm{I}} = \frac{P_{\mathrm{II\_same}} - P_{\mathrm{I\_cross}}}{P_{\mathrm{I\_cross}}} \cdot 100. \tag{10}$$

Similarly, the ΔP_{BC} values used to plot the right panel were derived assuming period II is the validation period and period I the calibration period:

$$\Delta P_{\mathrm{BC}}^{\mathrm{II}} = \frac{P_{\mathrm{II\_cross}} - P_{\mathrm{I\_same}}}{P_{\mathrm{I\_same}}} \cdot 100. \tag{11}$$

For the majority of precipitation characteristics, all bias correction methods seem to produce similar range and median of differences as given by the raw RCM, except for 3 and 5 day cumulative precipitation, where the raw RCM does better than the bias-corrected simulations. To take a closer look, we altered the baseline from change in observations (ΔP_{obs}) to change in raw RCM (ΔP_{RCM}), and the results ($\Delta P_{\mathrm{BC}} - \Delta P_{\mathrm{RCM}}$) are presented in Fig. 7. While the bias correction methods do not seem to affect changes in precipitation means, they do modify changes in high precipitation characteristics as shown in Fig. 7 as a large range of differences given by LS, DMG, DM2G, and QM in 99th percentile precipitation, and 99th percentile 3 and 5 day cumulative precipitation plots.

The follow-on effects on runoff can be seen in Figs. 8 and 9 which show differences in runoff changes (substitute P with Q in Eqs. 9–11) corresponding to Figs. 6 and 7. The differences in runoff changes are much larger compared to those in precipitation changes. The bias correction methods affect change signals in every runoff characteristics (Fig. 9), especially at high flows. This finding is consistent with Hagemann et al. (2011) and Gutjahr and Heinemann (2013), who showed that bias correction can alter climate change signals, a result slightly different from that of Muerth et al. (2013) who concluded that the impact of bias correction on change signals in flow is weak (except for the timing of the spring flood peak).

HESSD
11, 10683–10724, 2014

5 Discussion

5.1 Non-stationarity of the RCM bias

As shown in Figs. 3–5, the cross-validation results are period-dependent. When the errors in the calibration period are larger than, or in a different direction to, the errors in the validation period, all the bias correction methods over-correct the errors in the validation period. When the errors in the calibration period are smaller than, and in the same direction as, the errors in the validation period, all the bias correction methods can reduce errors somewhat even though the under-correction can still be substantial. This is mainly due to the inconsistent errors over time. The large magnitude of errors to be corrected amplifies the differences in the bias correction relationships and results in clear under-correction in one period and over-correction in another.

The differences in errors from the two periods may be a result of insufficient length of data to achieve robust calibration (Berg et al., 2012), or it could be due to the non-stationarity of RCM bias. It is difficult to assess the non-stationarity of biases because time series long enough to achieve robust calibration and validation are rare (Maraun, 2012), and the definition of "long enough" varies for dry and wet regions. However, the probability of bias non-stationarity is high (Ehret et al., 2012). Thus, the results shown here serve as a good indicator for what could happen if bias were to vary over time.

Using a longer record is likely to improve the outcome because it better represents the complete variability, and has less likelihood of calibration and validation periods being very different. To test this, we repeated the same analysis on precipitation using a 60 year long RCM simulation split in half – 30 years for calibration and 30 years for cross-validation. The results (Fig. 10) show improved cross-validation performance across bias correction methods and across characteristics. Nevertheless, the under-correction in the first period (1950–1979), and the over-correction in the second (1980–2009) are still apparent in most of the characteristics. Note that the runoff experiments cannot be repeated using the longer dataset due to limited streamflow data, but it is reasonable to assume that this tendency will have larger manifestation in modelled runoff.

The results suggest that non-stationarity of the RCM bias is one of the main obstacles preventing bias correction from achieving good outcomes, which makes the choice of bias correction method a secondary issue. When applying bias correction to a future period (as in most climate change impact studies), it is better to calibrate using a long dataset (30 years or more), or at least a data period that best reflects the future (e.g. calibrate over a dry period and apply to a dry future RCM simulation, and vice versa). As the bias correction relationship is unlikely to be the same for two periods, the more different the periods are (different means, extremes, low-frequency variability, etc.), and the larger the magnitude of bias to be corrected, the smaller the chance of getting satisfactory results from bias correction.

5.2 Performance of the bias correction methods

Figure 11 shows a selected example comparing daily CDF of the four bias correction methods (LS, DMG, DM2G, and QM) for calibration and cross-validation experiments. The LS performs poorly in both calibration and cross-validation as it under-estimates small and medium rainfall values (< 95th percentile) and over-estimates the very high rainfall values (> 95th percentile). The DMG performs significantly better than the LS because it attempts to correct the distribution rather than simply scaling the data with one factor. The DM2G performs better than DMG for its better representation of distribution, especially at high end (as shown in Fig. 2). By definition, the QM will always give perfect results in calibration but the over-fitting can lead to poorer performance in cross-validation, particularly when the errors in the two periods are very different.

In general, the best results are produced by either QM or DM2G in this study. The non-parametric QM fits every part of the entire distribution and performs the best when the errors in the two periods are similar. When the errors are different in the two periods, the DM2G is likely to be more robust (theoretical distribution with 6 parameters) and has less chance of over-fitting like QM is liable to do. Nevertheless, the difference between the three distribution mapping methods are very small in our modelling experiments (and as reported in the literature) because of the large corrections required which are then amplified by the inconsistent errors in different periods, as discussed in Sect. 5.1.

5.3 Importance of precipitation sequence

The errors in the bias corrected precipitation are significantly amplified in modelled runoff. The choice of hydrological models does not have big impact because of the relatively small errors associated with hydrological models. Although the RCM precipitation can be bias corrected to practically match the observed precipitation means and high precipitation amounts (see calibration results in Fig. 3), there can still be considerable errors in the modelled runoff (see calibration results in Fig. 4).

Apart from precipitation intensity, other aspects of precipitation can also affect runoff. Precipitation sequence is one of them as runoff generation is driven by high precipitation events that last over several days, and preceding events influence runoff by changing soil moisture content. The importance of precipitation sequence can be quantified in our modelling experiments by analysing calibration results for the QM. As shown in calibration plot (left) in Fig. 11, QM corrects the RCM daily precipitation to perfectly match the observed daily precipitation distribution; therefore the errors in the modelled runoff should mainly reflect the differences in the precipitation sequences between RCM and observations. To examine whether this is the case, we compared the wet spell histograms of observations, RCM and QM corrected RCM precipitation. Figure 12 shows the results for one of the study catchments. Compared to observations, the raw RCM simulation shows a lack of short events and an excess of very long events. This is the widely reported “drizzle effect” – RCMs simulate too many low-intensity precipitation events and too few high-intensity precipitation events (Gutowski et al., 2003). Although the QM is able to break long events into many shorter ones by reducing the “drizzles” with intensity below probability q_0 (Eq. 7) to zero, as shown by the increased number of short events, there are still differences in the wet spell frequencies. These differences are due to the lack of short wet spells followed by

HESSD
11, 10683–10724, 2014

long dry spells, as seen in the scatter plots in Fig. 13, which show length of dry spells on the y-axes and the following wet spells on the x axes.

The runoff errors in the QM calibration results, for the eight catchments in the two calibration periods, range from −17 to 24 % (median of 5 %) for the 99th percentile runoff and −7 to 7 % (median of 1 %) for mean annual runoff ("QM_same" in Fig. 4). These errors are significant considering that the precipitation distribution perfectly matches that of the observations. But they are relatively small compared to the errors in runoff from bias correction for cross-validation periods.

These results show that, the bias correction methods tested here are unable to overcome the discrepancy in the precipitation sequence. It is important for the RCMs to better simulate the number and length of storms, and the dry periods that intersperse them. After all, the ultimate approach to reduce errors in models is to improve the models themselves. This will require better process descriptions and implementations, higher spatiotemporal resolution, and perhaps using multi-model/multi-physics ensembles, as seen in recent development in this area (Ji et al., 2014; Evans et al., 2012; Flaounas et al., 2011).

6 Conclusions

This paper reviewed recent studies comparing various bias correction methods as applied to RCM simulations. The distribution mapping techniques were selected to remove errors (relative to observations) in daily precipitation series simulated by the weather research and forecasting (WRF) model for eight catchments in southeast Australia. The performance of three different techniques – DMG, DM2G, QM – and a linear scaling method (LS) as a benchmark, was evaluated with the focus on the follow-on impact on runoff modelling.

The results confirm the relatively higher skill of the distribution-based methods, compared to the linear scaling method, in correcting key precipitation characteristics. The best results are produced by either QM or DM2G. The non-parametric QM fits every part of the entire distribution and performs the best when the errors in the

HESSD
11, 10683–10724, 2014

calibration and validation periods are similar. When errors in the two periods are different, DM2G is more robust as it has a smaller number of parameters and so there is less chance of over-fitting, like QM is liable to do. However, the difference between the distribution mapping methods tested here is small because of the large corrections required and the inconsistent errors in the calibration and validation periods (non-stationarity).

The errors remaining in bias corrected precipitation lead to amplified errors in modelled runoff. The choice of hydrological models does not seem to matter because of the relatively small errors associated with hydrological models. The bias correction methods tested here cannot overcome limitations of RCM in simulating all precipitation features that influence runoff, in particular, precipitation sequence. The errors in modelled runoff are strongly influenced by the inconsistent RCM errors over time, even though this can be partially overcome by using a long calibration dataset.

Results further show that whereas bias correction does not seem to affect the change signals in precipitation means, it can introduce extra uncertainty to the change signals in high precipitation amounts, and consequently, in runoff. Future climate change impact studies need to take this into account when deciding whether to use raw or bias corrected RCM results. Nevertheless, the bias in RCM simulations will continue to reduce as RCM accuracy improves and RCMs will become increasingly useful for hydrological studies.

The Supplement related to this article is available online at doi:10.5194/hessd-11-10683-2014-supplement.

Acknowledgements. This work was carried out in the CSIRO Land and Water Flagship and was funded by the Victorian Climate Initiative (VicCI). J. P. Evans was funded through the Australian Research Council as part of Discovery Project DP0772665 and Future Fellowship FT110100576. We thank James Bennett and Steve Charles for comments on an earlier version of this manuscript.

HESSD
11, 10683–10724, 2014

References

Argüeso, D., Evans, J. P., and Fita, L.: Precipitation bias correction of very high resolution regional climate models, Hydrol. Earth Syst. Sci., 17, 4379–4388, doi:10.5194/hess-17-4379-2013, 2013.

Bennett, J. C., Ling, F. L. N., Post, D. A., Grose, M. R., Corney, S. P., Graham, B., Holz, G. K., Katzfey, J. J., and Bindoff, N. L.: High-resolution projections of surface water availability for Tasmania, Australia, Hydrol. Earth Syst. Sci., 16, 1287–1303, doi:10.5194/hess-16-1287-2012, 2012.

Bennett, J. C., Grose, M. R., Corney, S. P., White, C. J., Holz, G. K., Katzfey, J. J., Post, D. A., and Bindoff, N. L.: Performance of an empirical bias-correction of a high-resolution climate dataset, Int. J. Climatol., 34, 2189–2204, doi:10.1002/joc.3830, 2014.

Berg, P., Feldmann, H., and Panitz, H. J.: Bias correction of high resolution regional climate model data, J. Hydrol., 448–449, 80–92, doi:10.1016/j.jhydrol.2012.04.026, 2012.

Boe, J., Terray, L., Habets, F., and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, Int. J. Climatol., 27, 1643–1655, doi:10.1002/joc.1602, 2007.

Burnash, R. J. C., Ferral, R. L., and McGuire, R. A.: A Generalised Streamflow Simulation System – Conceptual Modelling for Digital Computers, Technical Report, Joint Federal and State River Forecast Center, Sacramento, 204 pp., 1973.

Chakravarti, I. M. and Laha, R. G.: Handbook of Methods of Applied Statistics, John Wiley & Sons, New York, 460 pp., 1967.

Chen, J., Brissette, F. P., and Leconte, R.: Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, J. Hydrol., 401, 190–202, doi:10.1016/j.jhydrol.2011.02.020, 2011a.

Chen, J., Brissette, F. P., Poulin, A., and Leconte, R.: Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed, Water Resour. Res., 47, W12509, doi:10.1029/2011wr010602, 2011b.

Chen, J., Brissette, F. P., Chaumont, D., and Braun, M.: Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America, Water Resour. Res., 49, 4187–4205, doi:10.1002/wrcr.20331, 2013.

Chiew, F. H. S. and McMahon, T. A.: The applicability of morton and penman evapotranspiration estimates in rainfall–runoff modeling, Water Resour. Bull., 27, 611–620, doi:10.1111/j.1752-1688.1991.tb01462.x, 1991.

Chiew, F. H. S., Young, W., Cai, W., and Teng, J.: Current drought and future hydroclimate projections in southeast Australia and implications for water resources management, Stoch. Env. Res. Risk A., 25, 601–612, doi:10.1007/s00477-010-0424-x, 2010.

Christensen, J. H., Boberg, F., Christensen, O. B., and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, Geophys. Res. Lett., 35, L20709, doi:10.1029/2008gl035694, 2008.

Dosio, A., Paruolo, P., and Rojas, R.: Bias correction of the ENSEMBLES high resolution climate change projections for use by impact models: analysis of the climate change signal, J. Geophys. Res., 117, D17110, doi:10.1029/2012jd017968, 2012.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 16, 3391–3404, doi:10.5194/hess-16-3391-2012, 2012.

Evans, J. P. and McCabe, M. F.: Regional climate simulation over Australia's Murray–Darling basin: a multitemporal assessment, J. Geophys. Res., 115, D14114, doi:10.1029/2010jd013816, 2010.

Evans, J. P., Ekström, M., and Ji, F.: Evaluating the performance of a WRF physics ensemble over South-East Australia, Clim. Dynam., 39, 1241–1258, doi:10.1007/s00382-011-1244-5, 2012.

Evans, J. P., Ji, F., Lee, C., Smith, P., Argüeso, D., and Fita, L.: Design of a regional climate modelling projection ensemble experiment – NARCliM, Geosci. Model Dev., 7, 621–629, doi:10.5194/gmd-7-621-2014, 2014.

Flaounas, E., Bastin, S., and Janicot, S.: Regional climate modelling of the 2006 West African monsoon: sensitivity to convection and planetary boundary layer parameterisation using WRF, Clim. Dynam., 36, 1083–1105, doi:10.1007/s00382-010-0785-3, 2011.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, Int. J. Climatol., 27, 1547–1578, doi:10.1002/joc.1556, 2007.

Grillakis, M. G., Koutroulis, A. G., and Tsanis, I. K.: Multisegment statistical bias correction of daily GCM precipitation output, J. Geophys. Res., 118, 3150–3162, doi:10.1002/jgrd.50323, 2013.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, Hydrol. Earth Syst. Sci., 16, 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

Gutjahr, O. and Heinemann, G.: Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM, Theor. Appl. Climatol., 114, 511–529, doi:10.1007/s00704-013-0834-z, 2013.

Gutowski, W. J., Decker, S. G., Donavon, R. A., Pan, Z. T., Arritt, R. W., and Takle, E. S.: Temporal-spatial scales of observed and simulated precipitation in central U.S. climate, J. Climate, 16, 3841–3847, doi:10.1175/1520-0442(2003)016<3841:tsooas>2.0.co;2, 2003.

Jeffrey, S. J., Carter, J. O., Moodie, K. B., and Beswick, A. R.: Using spatial interpolation to construct a comprehensive archive of Australian climate data, Environ. Modell. Softw., 16, 309–330, doi:10.1016/S1364-8152(01)00008-1, 2001.

Ji, F., Ekström, M., Evans, J., and Teng, J.: Evaluating rainfall patterns using physics scheme ensembles from a regional atmospheric model, Theor. Appl. Climatol., 115, 297–304, doi:10.1007/s00704-013-0904-2, 2014.

Johnson, F. and Sharma, A.: A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations, Water Resour. Res., 48, W01504, doi:10.1029/2011wr010464, 2012.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, doi:10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2, 1996.

Lafon, T.: Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods, Int. J. Climatol., 33, 1367–1381, doi:10.1002/joc.3518, 2013.

Lenderink, G., Buishand, A., and van Deursen, W.: Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach, Hydrol. Earth Syst. Sci., 11, 1145–1159, doi:10.5194/hess-11-1145-2007, 2007.

Li, H., Sheffield, J., and Wood, E. F.: 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, doi:10.1029/2009jd012882, 2010.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, doi:10.1029/2009rg000314, 2010.

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velázquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., and Turcotte, R.: On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, Hydrol. Earth Syst. Sci., 17, 1189–1204, doi:10.5194/hess-17-1189-2013, 2013.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Perrin, C., Michel, C., and Andreassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275–289, doi:10.1016/s0022-1694(03)00225-7, 2003.

Piani, C., Haerter, J. O., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, Theor. Appl. Climatol., 99, 187–192, doi:10.1007/s00704-009-0134-9, 2010a.

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., and Haerter, J. O.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, J. Hydrol., 395, 199–215, doi:10.1016/j.jhydrol.2010.10.024, 2010b.

Potter, N. J. and Chiew, F. H. S.: An investigation into changes in climate characteristics causing the recent very low runoff in the southern Murray–Darling Basin using rainfall–runoff models, Water Resour. Res., 47, W00G10, doi:10.1029/2010wr010333, 2011.

Schmidli, J., Frei, C., and Vidale, P. L.: Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods, Int. J. Climatol., 26, 679–689, doi:10.1002/joc.1287, 2006.

Seaby, L. P., Refsgaard, J. C., Sonnenborg, T. O., Stisen, S., Christensen, J. H., and Jensen, K. H.: Assessment of robustness and significance of climate change signals for an ensemble of distribution-based scaled climate projections, J. Hydrol., 486, 479–493, doi:10.1016/j.jhydrol.2013.02.015, 2013.

Teng, J., Chiew, F. H. S., Timbal, B., Wang, Y., Vaze, J., and Wang, B.: Assessment of an analogue downscaling method for modelling climate change impacts on runoff, J. Hydrol., 472–473, 111–125, doi:10.1016/j.jhydrol.2012.09.024, 2012a.

Teng, J., Vaze, J., Chiew, F. H. S., Wang, B., and Perraud, J.-M.: Estimating the relative uncertainties sourced from GCMs and hydrological models in modeling climate change impact on runoff, J. Hydrometeorol., 13, 122–139, doi:10.1175/jhm-d-11-058.1, 2012b.

Teutschbein, C. and Seibert, J.: Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies, Geography Compass, 4, 834–860, doi:10.1111/j.1749-8198.2010.00357.x, 2010.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods, J. Hydrol., 456–457, 12–29, doi:10.1016/j.jhydrol.2012.05.052, 2012.

Teutschbein, C. and Seibert, J.: Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions?, Hydrol. Earth Syst. Sci., 17, 5061–5077, doi:10.5194/hess-17-5061-2013, 2013.

Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, Int. J. Climatol., 31, 1530–1544, doi:10.1002/joc.2168, 2011.

Vaze, J., Chiew, F. H. S., Perraud, J., Viney, N., Post, D. A., Teng, J., Wang, B., Lerat, J., and Goswami, M.: Rainfall–runoff modelling across southeast Australia: datasets, models and results, Australian Journal of Water Resources, 14, 101–116, 2010.

Viney, N. R., Perraud, J.-M., Vaze, J., Chiew, F. H. S., Post, D. A., and Yang, A.: The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments, 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Cairns, Australia, 13–17 July 2009, 3421–3427, 2009.

Wilby, R. L., Hay, L. E., Gutowski, W. J., Arritt, R. W., Takle, E. S., Pan, Z., Leavesley, G. H., and Clark, M. P.: Hydrological responses to dynamically and statistically downscaled climate model output, Geophys. Res. Lett., 27, 1199–1202, doi:10.1029/1999gl006078, 2000.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Climatic Change, 62, 189–216, doi:10.1023/B:CLIM.0000013685.99609.9e, 2004.

Yang, W., Andreasson, J., Graham, L. P., Olsson, J., Rosberg, J., and Wetterhall, F.: Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, Hydrol. Res., 41, 211–229, doi:10.2166/nh.2010.004, 2010.

HESSD
11, 10683–10724, 2014

How does bias correction of RCM precipitation affect modelled runoff?

J. Teng et al.

Table 1. Recent studies comparing different RCM bias correction methods.

Study	Study area	Spatial resolution	Validation period	Number of bias correction methods assessed on precipitation	Variable(s) assessed	Statistics evaluated	Metric(s) used	Conclusion
Themeßl et al. (2011)	Domain covering the Alpine region including Austria	Modelled at 10 km resolution Validated at station scale	11 years of data (1981–1990, 1999) were used in a 11-fold "leave one out" cross-validation	Seven	Precipitation modelled by RCM MM5 forced with the ERA-40 reanalysis data	Median, variability, and indicators for extremes	Bias	Quantile mapping shows the best performance, particularly at high quantiles.
Berg et al. (2012)	Domain covering entire Germany and its near surrounding areas	Modelled at 7 km resolution Validated against 1 km gridded observations	30 years of data (1971–2000) Two realisations were used for calibration and validation respectively	Three	Precipitation, temperature modelled by RCM COSMO-CLM driven by a GCM (ECHAM5-MPIOM)	Mean and variance of temperature and precipitation	Bias	Histogram equalisation (HE) method corrects not only means but also higher moments, but approximations of the transfer function are necessary when applying to new data. About 30 year long calibration period is required for a reasonable approximation.
Teutschbein and Seibert (2012, 2013)	Five catchments in Sweden	Modelled at 25 km resolution Validated at catchment scale (taken from one grid cell)	40 years of data (1961–1990) were used for calibration and validation in the first study and split into warm/cold years and dry/wet years for cross-validation in the second study	Six	Precipitation, temperature modelled by 11 RCMs driven by different GCMs Streamflow simulated by hydrological model HBV	Mean, standard deviation, 10th and 90th percentile daily temperature during summer and winter Mean, standard deviation, coefficient of variation, 90th percentile, probability of wet days, and average intensity of wet days during summer and winter Mean monthly streamflow, Spring flood peak, Autumn flood peak, total flows, and annual 15 day low flows.	Mean Absolute Error (MAE) on temperature and precipitation CDFs	Distribution mapping perform the best for both climate projections and hydrological impact qualifications. It performs especially well in terms of the simulation of hydrological extremes. It also shows the best transferability to potentially changed climate conditions.
Gudmundsson et al. (2012)	Domain covering Norway and Nordic Arctic	Modelled at 25 km resolution validated at station scale	41 years of data (1960–2000) split into 10 subsamples for a 10-fold "leave one out" cross-validation	11	Precipitation modelled by RCM HIRHAM forced with the EAR40 reanalysis data	Precipitation at 0.1, 0.2,... 1.0 percentile	Mean Absolute Errors (MAEs) at equally spaced probability intervals	Nonparametric methods perform the best in reducing systematic errors, followed by parametric transformations with three or more free parameters, with the distribution derived transformations rank the lowest.

HESSD
11, 10683–10724, 2014

Table 1. Continued.

Study	Study area	Spatial resolution	Validation period	Number of bias correction methods assessed on precipitation	Variable(s) assessed	Statistics evaluated	Metric(s) used	Conclusion
Lafon et al. (2013)	Seven catchments in Great Britain	Modelled at 25 km resolution Validated at catchment scale	40 years of data (1961–2000) split into moving window of 10 year subsamples for a 31-fold "leave one out" cross-validation	Four	Precipitation modelled by RCM HadRM3-PPE-UK driven by a GCM (HadCM3)	Mean, standard deviation, Coefficient of variation, skewness, kurtosis	Average of the Relative Differences (ARD)	If both precipitation data sets (modelled and observed) can be approximated by a gamma distribution, the gamma-based quantile mapping method offers the best combination of accuracy and robustness. Otherwise, the nonlinear method is more effective at reducing the bias. The empirical quantile mapping method can be highly accurate, but results are very sensitive to the choice of calibration time period.
Chen et al. (2013)	10 catchments in North America	Modelled at 50 km resolution Validated at catchment scale	20 years of data (1981–2000) split into odd years and even years for cross-validation	Six	Precipitation modelled by four RCMs (CRCM, HRM3, RCM3, and WRFG) driven by NCEP reanalysis data Flow discharge simulated using the hydrological model HSAMI	Mean, standard deviation, and 95th percentile wet-day precipitation Mean daily discharge, the mean of 95th percentile spring high flow, and the mean of 5th percentile summer low flow	Absolute Relative Error (ARE) on precipitation and discharge Nash–Sutcliffe model Efficiency coefficient (NSE), Root Mean Square Error (RMSE), and Transformed Root Mean Square Error (TRMSE) for daily discharge	The performance of bias correction is location dependent. The distribution-based methods are consistently better than the mean-based methods for both precipitation projections and hydrological simulations.
Gutjahr and Helnemann (2013)	A German state and its surrounding areas	Modelled at 4.5 km Validated at station scale	Ten years of data (1991–2000), with each year used in a "leave-one-out" cross-validation resulting in 81 combinations	Three distribution-based methods	Precipitation, temperature modelled by RCM COSMO-CLM driven by a GCM (ECHAM5)	Precipitation at 0.1, 0.2,. . . 1.0 percentile	Mean Absolute Errors (MAEs) at equally spaced probability intervals	The empirical method outperforms both parametric alternatives.

HESSD
11, 10683–10724, 2014

How does bias correction of RCM precipitation affect modelled runoff?

J. Teng et al.

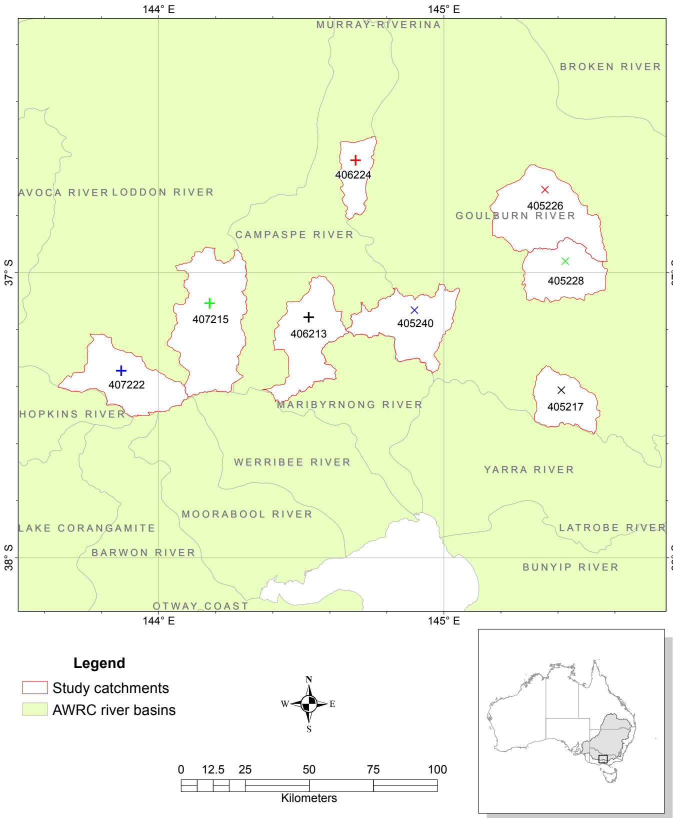

Figure 1. Map showing the eight study catchments (white with red outline, large map), and their locations within the major river basins devised by the Australian Water Resources Council (AWRC). The unique symbols in each catchment identify the catchments and will be used in later figures.

HESSD
11, 10683–10724, 2014

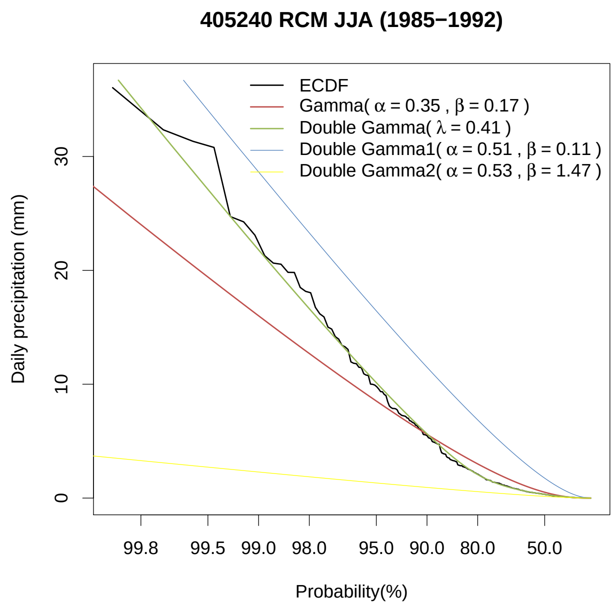

Figure 2. CDF plot comparing a gamma distribution (red) and a double gamma distribution (green), which consists of two gamma distributions (blue and yellow), fitted to the same precipitation data for one study catchment. The empirical distribution is shown in black.

HESSD
11, 10683–10724, 2014

How does bias correction of RCM precipitation affect modelled runoff?

J. Teng et al.

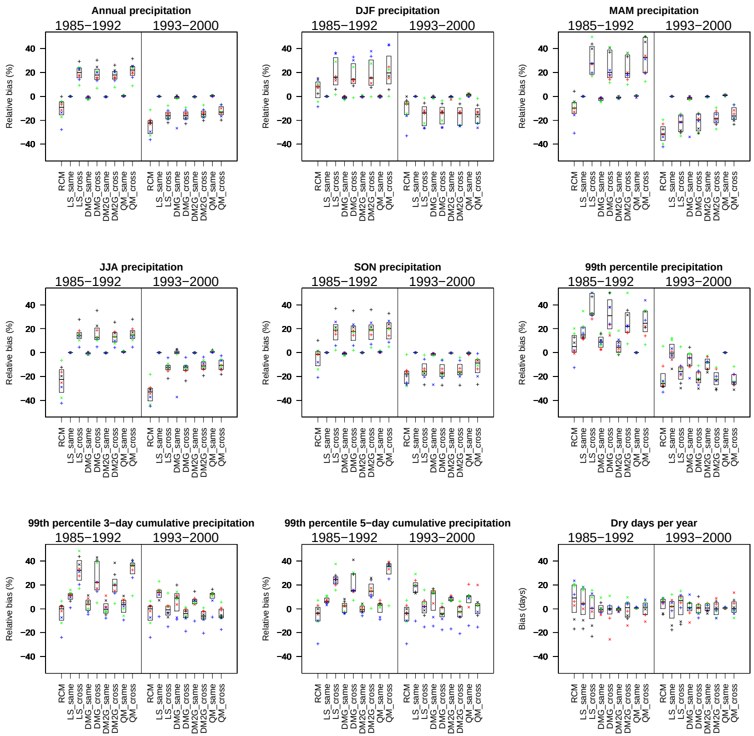

Figure 3. Relative bias of precipitation characteristics, expressed as percentage differences relative to observations between raw RCM and bias corrected RCM precipitation. Each panel displays a different characteristic (title at top of panel) and the percentages were calculated after applying four different bias correction methods (key at bottom) to eight catchments over two periods (1985–1992, left; 1993–2000, right). The bias correction methods were LS, DMG, DM2G, and QM (see text), and the "_same" suffix denotes calibration and the "_cross" refers to cross-validation. Boxes indicate interquartile range; markers indicate the numbers from each catchment (markers are constrained to the edges of the plotting area if the values exceed the range of plotting); the symbol for each catchment can be found in Fig. 1.

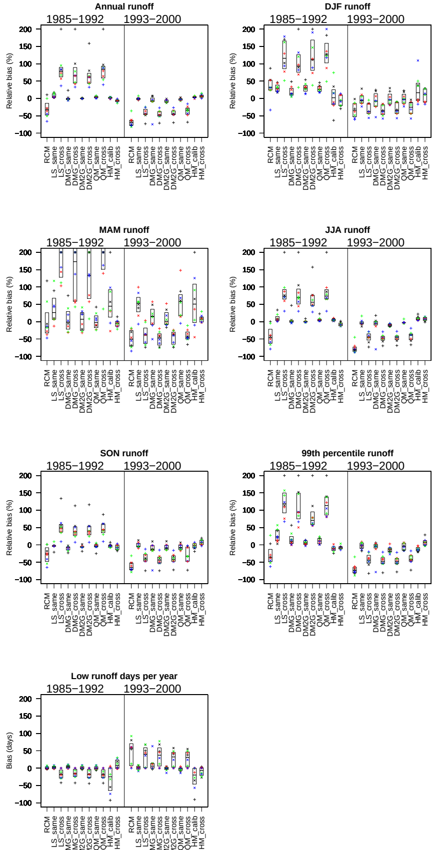

Figure 4. Relative bias in runoff characteristics derived from precipitation-driven hydrological model GR4J. Values are percentage differences, relative to runoff modelled from observed precipitation, when GR4J was driven by raw RCM precipitation and bias corrected RCM precipitation. Same layout as Fig. 3, with additional HM_calib and HM_cross, which represent calibration errors and cross-validation errors from GR4J alone.

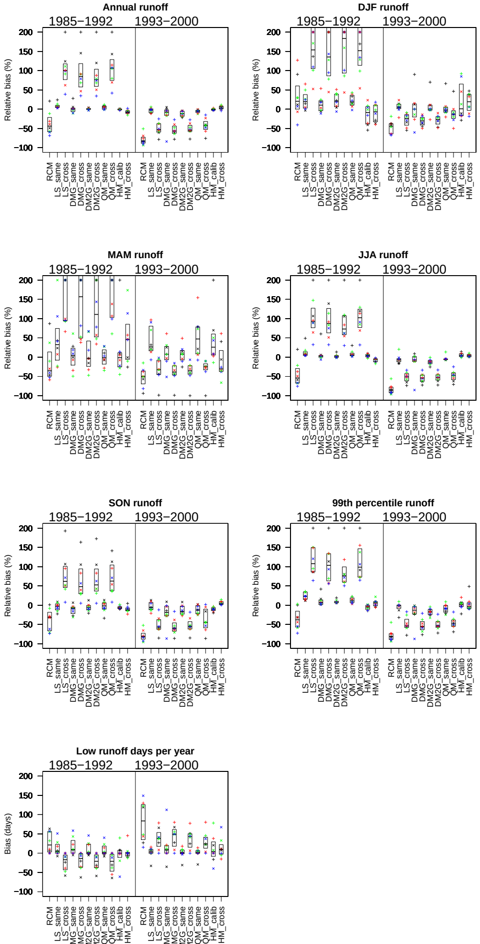

Figure 5. As for Fig. 4 but using results from hydrological model Sacramento.

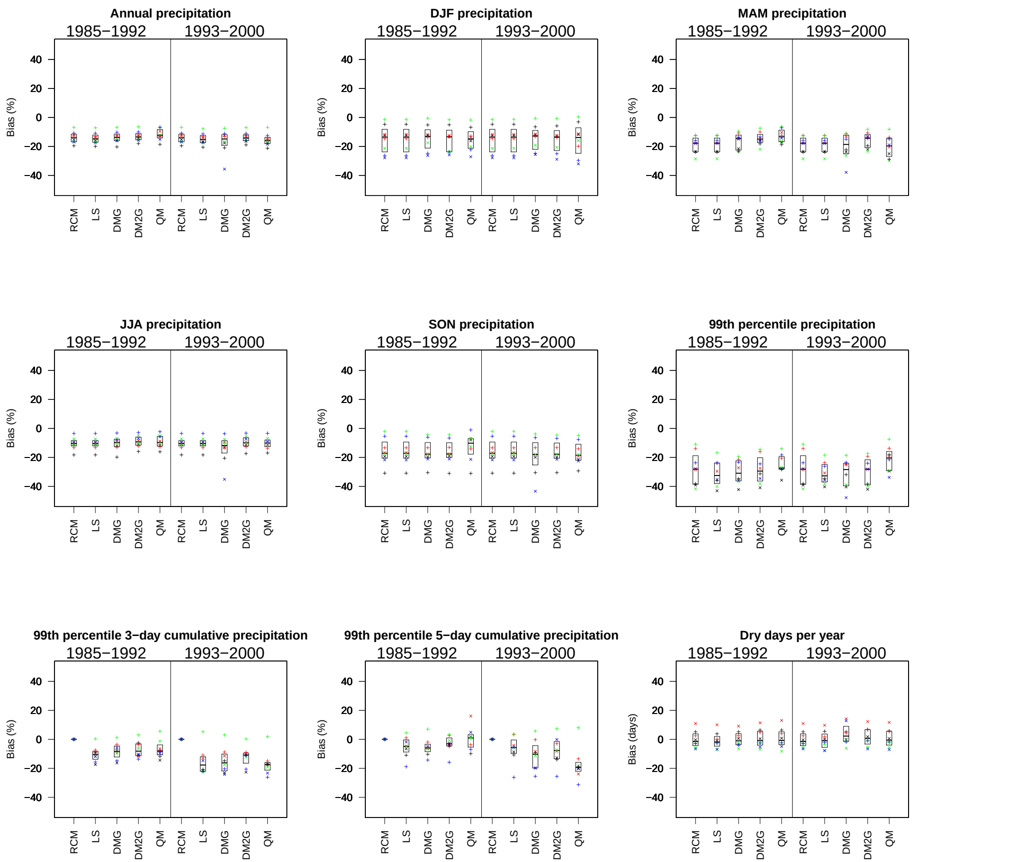

Figure 6. Differences between RCM simulations and observations in change signals in precipitation characteristics between periods 1985–1992 and 1993–2000. The left and right panels indicate the validation periods in each case.

HESSD

11, 10683–10724, 2014

How does bias correction of RCM precipitation affect modelled runoff?

J. Teng et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀▌ | ▶▌
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

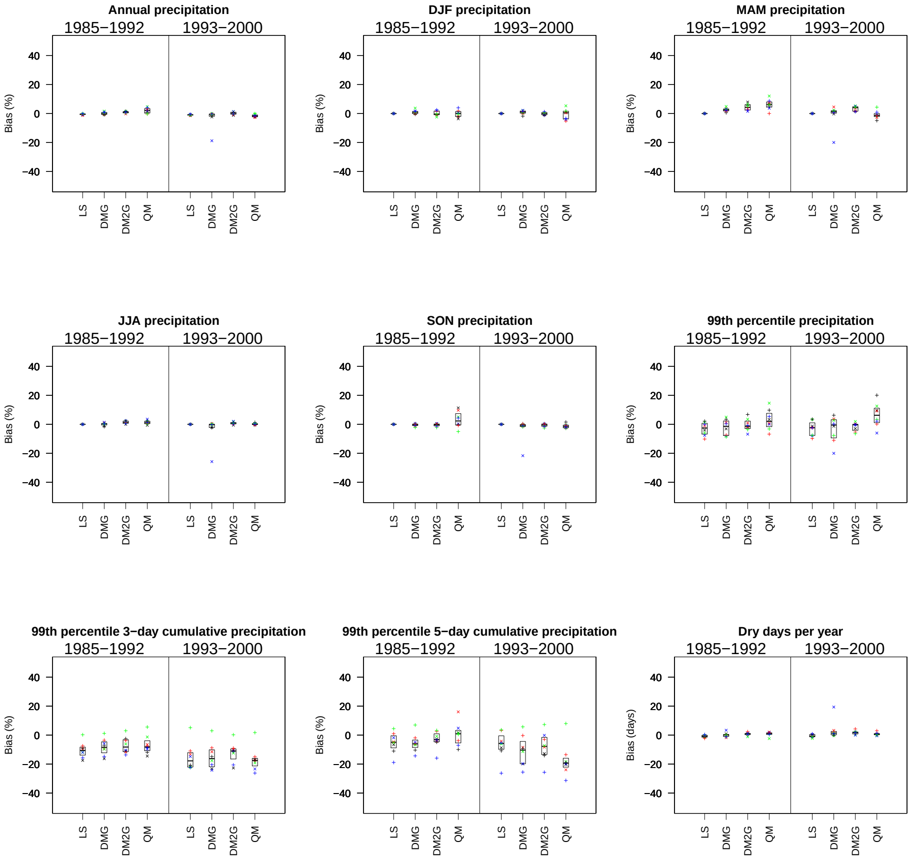

Figure 7. Differences between bias-corrected RCM and raw RCM simulations in change signals in precipitation characteristics between periods 1985–1992 and 1993–2000. The left and right panels indicate the validation periods in each case.

HESSD
11, 10683–10724, 2014

How does bias correction of RCM precipitation affect modelled runoff?

J. Teng et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

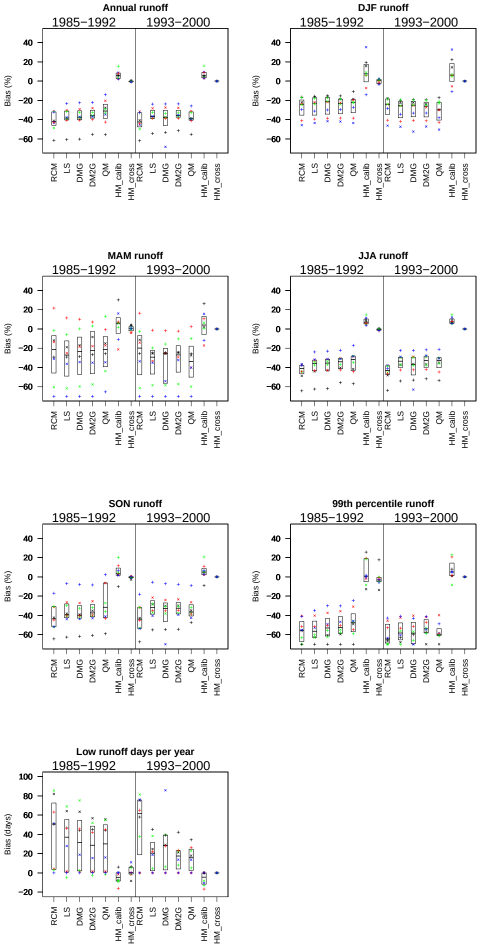

Figure 8. As for Fig. 6, but showing runoff characteristics modelled by GR4J.

HESSD
11, 10683–10724, 2014

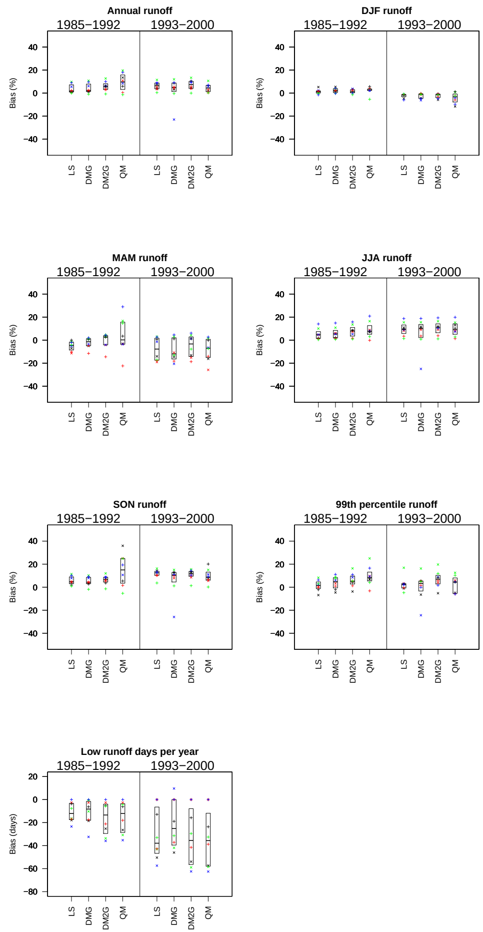

Figure 9. As for Fig. 7, but showing runoff characteristics modelled by GR4J.

HESSD
11, 10683–10724, 2014

How does bias correction of RCM precipitation affect modelled runoff?

J. Teng et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀I ▶I
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

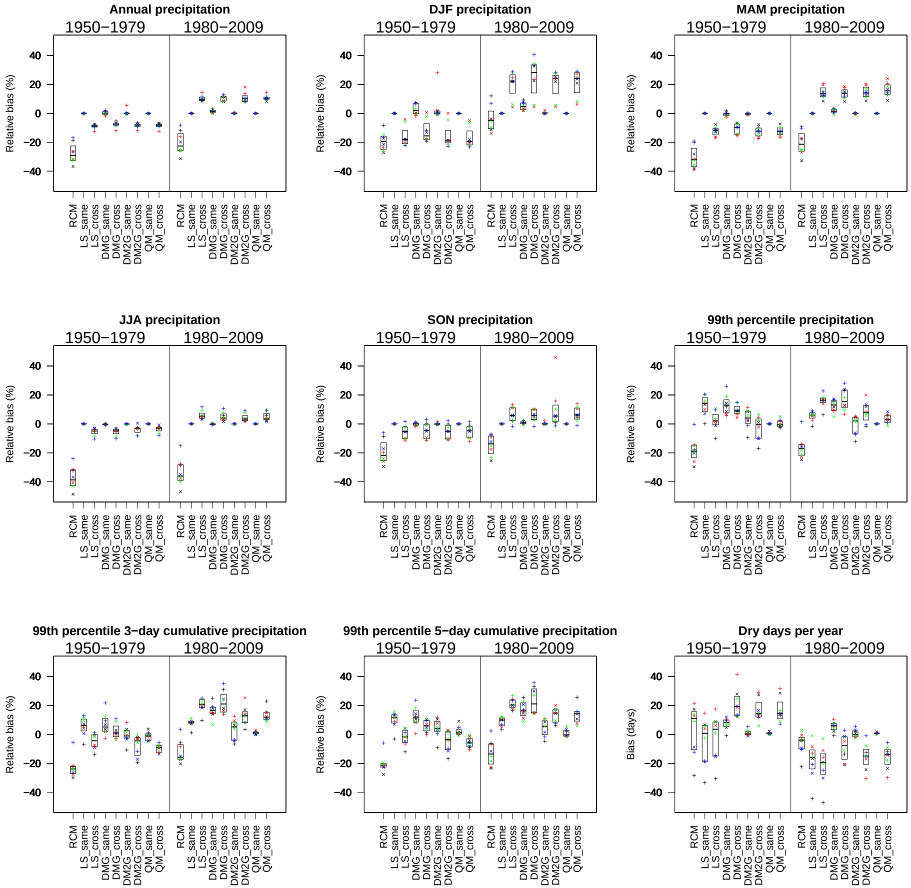

Figure 10. As for Fig. 3, but showing results from the long-term (two 30 year long) experiments.

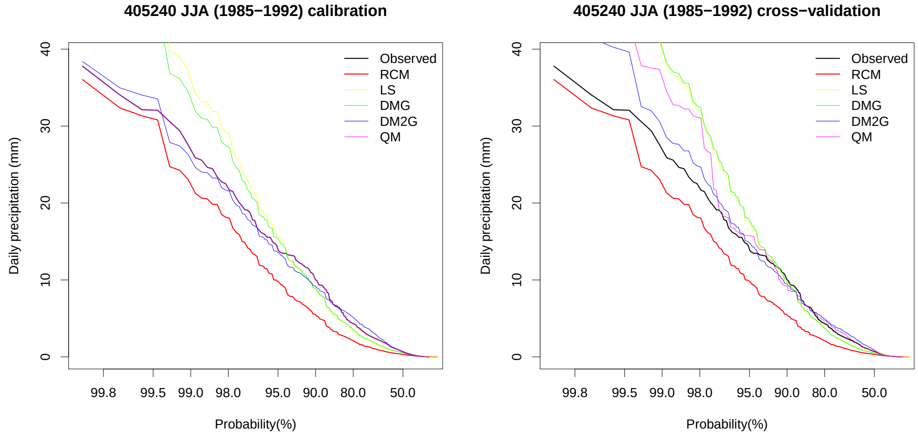

Figure 11. Comparison of CDFs derived from observed, raw RCM, and bias-corrected RCM daily precipitation data for one study catchment. The left plot shows calibration results and the right plot shows cross-validation results.

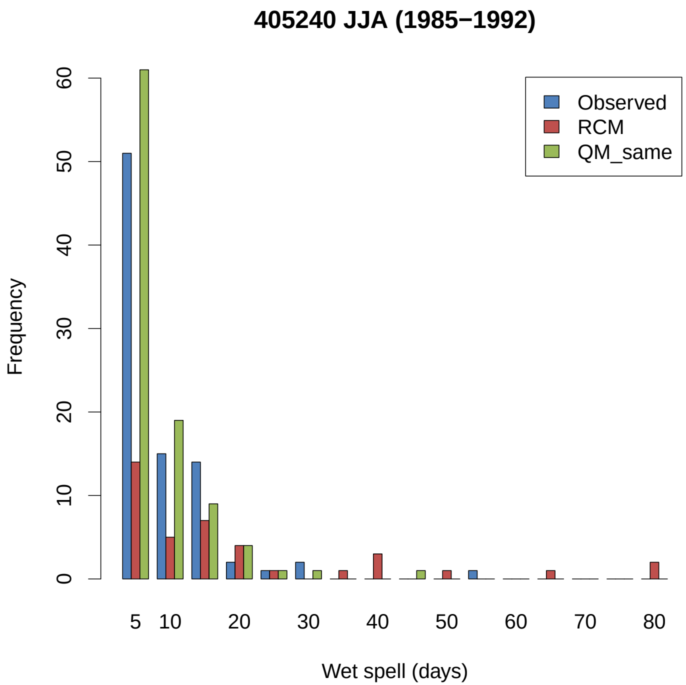

Figure 12. Histograms of consecutive wet days for observed, raw RCM, and QM bias-corrected RCM precipitation for one study catchment.

HESSD
11, 10683–10724, 2014

How does bias correction of RCM precipitation affect modelled runoff?

J. Teng et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

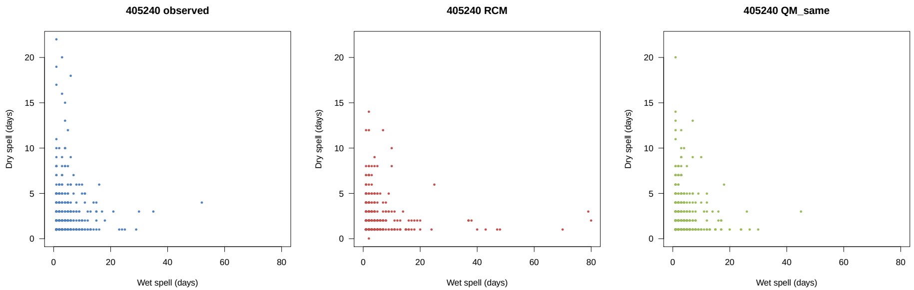

Figure 13. Scatter plots showing the length of each dry spell on y axes and the length of the following wet spell on x axes for observed, raw RCM, and QM bias-corrected RCM precipitation for an example catchment.

HESSD
11, 10683–10724, 2014