How does bias correction of regional climate model precipitation affect modelled runoff?

J. Teng¹, N. J. Potter¹, F. H. S. Chiew¹, L. Zhang¹, B. Wang¹, J. Vaze¹, and J. P. Evans²

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

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

Received: 29 August 2014 – Published in Hydrol. Earth Syst. Sci. Discuss.: 23 September 2014
Revised: 11 December 2014 – Accepted: 8 January 2015 – Published: 4 February 2015

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 differences between the methods are small in the modelling experiments here (and as reported in the literature), mainly due to the substantial corrections required and inconsistent errors over time (non-stationarity). The errors in bias corrected precipitation are typically amplified in modelled runoff. The tested methods cannot overcome limitations of the 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 regional-, catchment- or point-scale hydrological models (Fowler et al., 2007). Dynamical downscaling techniques derive regional-scale 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 hydrological impact studies (Dosio et al., 2012; Argüeso 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 conceptualisation and parameterisation, 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 minimise 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 av-
compared three distribution mapping techniques, each with different 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. 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 similarly 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.

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 modelled runoff. The bias correction methods were applied on modelled precipitation, as it is the most critical and difficult-to-model variable in hydrological studies (Vaze et al., 2011), and evaluated on both precipitation and 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.

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 Ehret 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-1990s, 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 km², 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. The observations and RCM simulations are aggregated to each catchment and compared at this level.

### 2.1.1 Observations

Observed daily precipitation data were derived from 0.05° (~5 km) gridded climate surfaces and averaged over each

---

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

<table>
<thead>
<tr>
<th>Study</th>
<th>Study area</th>
<th>Spatial resolution</th>
<th>Validation period</th>
<th>Number of bias correction methods assessed on precipitation</th>
<th>Variable(s) evaluated</th>
<th>Statistics evaluated</th>
<th>Metric(s) used</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thernfell et al. (2011)</td>
<td>Domain covering the Alpine region including Austria</td>
<td>Modeled at 10 km resolution. Validated at station scale</td>
<td>11 years of data (1981–1990), 1999 were used in a 11-fold “leave one out” cross-validation</td>
<td>Seven</td>
<td>Precipitation modeled by RCM MM5 forced with the ERA-40 reanalysis data</td>
<td>Median, variability and indicators for extremes</td>
<td>Bias</td>
<td>Quantile mapping shows the best performance, particularly at high quantiles.</td>
</tr>
<tr>
<td>Berg et al. (2012)</td>
<td>Domain covering the entirety of Germany and its near surrounding areas</td>
<td>Modeled at 7 km resolution. Validated against 1 km gridded observations</td>
<td>30 years of data (1971–2000). Two realisations were used for calibration and validation respectively</td>
<td>Three</td>
<td>Precipitation, temperature modelled by RCM COSMO-CLM driven by a GCM (ECHAM5-MPIOM)</td>
<td>Mean and variance of temperature and precipitation</td>
<td>Bias</td>
<td>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 a 30-year-long calibration period is required for a reasonable approximation.</td>
</tr>
<tr>
<td>Teutschbein and Seibert (2012, 2013)</td>
<td>Five catchments in Sweden</td>
<td>Modeled at 25 km resolution. Validated at catchment scale</td>
<td>10 years of data (1991–2000), with each year used in a 10-fold “leave one out” cross-validation</td>
<td>Six</td>
<td>Precipitation, temperature modelled by 11 RCMs driven by different GCMs. Streamflow simulated by hydrological model HBV</td>
<td>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.</td>
<td>Mean absolute error (MAE) on temperature and precipitation CDFs</td>
<td>Distribution mapping performs 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.</td>
</tr>
<tr>
<td>Gutmansson et al. (2012)</td>
<td>Domain covering Norway and Nordic Arctic</td>
<td>Modeled at 25 km resolution. Validated at station scale</td>
<td>41 years of data (1960–2000) split into 10 sub-samples for a 10-fold “leave one out” cross-validation</td>
<td>11</td>
<td>Precipitation modeled by RCM HIRHAM forced with the ERA-40 reanalysis data</td>
<td>Precipitation at 0.1, 0.2, … 1.0 percentile cross-validation</td>
<td>Mean absolute errors (MAEs) at equally spaced probability intervals</td>
<td>Nonparametric methods perform the best in reducing systematic errors, followed by parametric transformations with three or more free parameters, with the distribution derived transformations ranked the lowest.</td>
</tr>
<tr>
<td>Lafon et al. (2013)</td>
<td>Seven catchments in Great Britain</td>
<td>Modeled at 25 km resolution. Validated at catchment scale</td>
<td>40 years of data (1961–2000) split into moving window of 10-year sub-samples for a 31-fold “leave one out” cross-validation</td>
<td>Four</td>
<td>Precipitation modeled by RCM HadRM3-PPE-UK driven by a GCM (HadCM3)</td>
<td>Mean, standard deviation, coefficient of variation, skewness, kurtosis</td>
<td>Average of the relative differences (ARD)</td>
<td>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.</td>
</tr>
<tr>
<td>Chen et al. (2013)</td>
<td>10 catchments in North America</td>
<td>Modeled at 50 km resolution. Validated at station scale</td>
<td>20 years of data (1981–2000) split into odd years and even years for cross-validation</td>
<td>Six</td>
<td>Precipitation modeled by four RCMs (CRCM, HRM3, RCM3 and WRF0) driven by NCEP reanalysis data. Flow discharge simulated using the hydrological model HYDAS</td>
<td>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</td>
<td>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</td>
<td>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.</td>
</tr>
<tr>
<td>Gutjahr and Helmichmann (2013)</td>
<td>A German state and its surrounding areas</td>
<td>Modeled at 4.5 km resolution. Validated at station scale</td>
<td>10 years of data (1991–2000), with each year used in a “leave one out” cross-validation resulting in 81 combinations</td>
<td>Three</td>
<td>Precipitation, temperature modelled by RCM COSMO-CLM driven by a GCM (ECHAM5)</td>
<td>Precipitation at 0.1, 0.2, … 1.0 percentile</td>
<td>Mean absolute errors (MAEs) at equally spaced probability intervals</td>
<td>The empirical method outperforms both parametric alternatives.</td>
</tr>
</tbody>
</table>

---

www.hydrol-earth-syst-sci.net/19/711/2015/  
Hydrol. Earth Syst. Sci., 19, 711–728, 2015
catchment. The source of this data set 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 data sets 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). 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 data set (Evans et al., 2014) was used to validate the conclusion reached by using the shorter data set. For both data sets, 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 data set (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 \( h \):

\[
\hat{p}_{\text{obs}} = h(p_{\text{mod}})
\]

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

3.1.1 Linear scaling (LS)

The simplest choice for \( h \) is probably a linear transformation

\[
\hat{p}_{\text{obs}} = a p_{\text{mod}},
\]

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 skillful bias correction method (Gudmundsson et al., 2012; Lafon, 2013; Chen et al., 2013; Teutschbein and Seibert, 2012). The LS parameter \( a \) was optimised for each season: DJF (December–February), MAM (March–May), JJA (June–August) and SON (September–November) to account for precipitation seasonality. Similarly, seasonal optimisation 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:

\[
\hat{p}_{\text{obs}} = F_{\text{obs}}^{-1}(F_{\text{mod}}(p_{\text{mod}})),
\]

Where...
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 either be theoretical distributions fitted to the data, or empirical distributions estimated by sorting the data. The gamma distribution with shape parameter $\alpha$ and rate parameter $\beta$ (Eq. 4) is often used to represent non-zero precipitation amounts (Piani et al., 2010a; Lafon, 2013), as it has the ability to approximate the positively skewed distribution, or empirical distributions estimated by sorting the data, (blue and yellow) fitted to the same precipitation data depicted in black.

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.

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

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

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

$$q_0 = n_0/n.$$  

The shape and rate parameters $\alpha$ and $\beta$ are calculated on the non-zero precipitation amounts.

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; Smith et al., 2014). 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:

$$f(p) = \lambda \beta_{1}\alpha_1 p^{\alpha_1 - 1} e^{-\beta_{1} p} + (1 - \lambda) \beta_{2}\alpha_2 p^{\alpha_2 - 1} e^{-\beta_{2} p}$$

with $0 < \lambda < 1$. The parameter $\lambda$ 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 – $q_0$, $\alpha_1$, $\beta_1$, $\alpha_2$, $\beta_2$ and $\lambda$ – 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 high end.
3.1.4 Empirical quantile mapping (QM)

Apart from using theoretical distributions, the empirical CDF is also commonly used to solve Eq. (3) (Themell 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). Both 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 a large effect on the conclusions of this study because the errors associated with hydrological models are relatively small compared to errors in GCM/downscaling (Teng et al., 2012b; Chen et al., 2011b). In this study, 4 and 14 parameters were calibrated for GR4J and Sacramento respectively. 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 data set calculated using observed climate variables was used throughout the hydrological modelling. By keeping PET the same, the possible impact of PET and the correlation between RCM precipitation and PET was not considered in this study to isolate the impact of precipitation.

### 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 (contrasting between wet and dry) so that the impact of different climatic conditions can be clearly identified.

---

**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.
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 the 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 precipi-
tion, with more than 60 % of the precipitation coming from JIA 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 similar to 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 is compared with observed streamflow). 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 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

Figure 7. Differences between bias-corrected RCM and raw RCM simulations in change signals in precipitation characteristics between the periods 1985–1992 and 1993–2000. The left and right panels indicate the validation periods in each case.
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, although the magnitude of impact can be different. As the focus of this study is on the impact of bias correction method, only results from GR4J are presented and discussed in the following sections.

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” ($\Delta P$) is defined as the relative difference of various characteristics between period II and period I (Eq. 9):

$$\Delta P = \frac{P_{II} - P_{I}}{P_{I}} \cdot 100\%.$$  \hspace{1cm} (9)

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

$$\Delta P_{II}^{I} = \frac{P_{II \text{ same}} - P_{I \text{ cross}}}{P_{I \text{ cross}}} \cdot 100\%.$$  \hspace{1cm} (10)

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

$$\Delta P_{BC}^{II} = \frac{P_{I \text{ cross}} - P_{I \text{ same}}}{P_{I \text{ same}}} \cdot 100\%.$$  \hspace{1cm} (11)
For the majority of precipitation characteristics, all bias correction methods seem to produce a 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 ($\Delta P_{obs}$) to change in raw RCM ($\Delta P_{RCM}$), and the results ($\Delta P_{BC} - \Delta P_{RCM}$) 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 characteristic (Fig. 9), especially high flows. This finding is consistent with Hagemann et al. (2011), Cloke et al. (2013) 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).

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
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 (note that the observed precipitation is completely hidden by QM in this plot) and the right plot shows cross-validation results.
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 data set (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. These problems have implications on the application of bias correction to climate model outputs in hydrological impact studies and related sectors (even more so at extremes like floods). Projections derived from bias corrected climate input should therefore be interpreted cautiously and/or combined with other approaches (Cloke et al., 2013; Smith et al., 2014).

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 the 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 six parameters) and has less chance of over-fitting like QM is liable to do. Nevertheless, the difference between the three distribution mapping methods is 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

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

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 data set due to limited streamflow data, but it is reasonable to assume that this tendency will have a larger manifestation in modelled runoff.

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 the 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 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 developments 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 simulated precipitation. 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 calibration and validation periods are similar. When errors in the two periods are different, DM2G can be more robust as it has a smaller number of parameters and so there is less chance of over-fitting.
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 in bias corrected precipitation lead to amplified errors in modelled runoff. The bias correction methods tested here cannot overcome the limitations of the RCM in simulating all precipitation features that influence runoff, in particular, daily precipitation sequence. The errors in modelled runoff are strongly influenced by the inconsistent RCM errors over time, although this can be partially overcome by using a long calibration data set.

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.

These problems associated with bias correction in general have implications on its application to climate model outputs in hydrological impact studies. Projections derived from bias corrected climate input should therefore be interpreted with caution and/or combined with other approaches. 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/hess-19-711-2015-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, as well as the anonymous HESS reviewers for their constructive feedback.

Edited by: H. Cloke

References


J. Teng et al.: How does bias correction of RCM precipitation affect modelled runoff?


