Figure 1. Please make sure the labels in the plot are not covering the lines and symbols.

Done

Figure 2. Please add a scale-bar (in km) to one of the subplots.

Scale bar added to panel (e)

5   Figure 3. Please add a title and units to the color-scale (here, and in the other figures)

Figure 3,4 colour scale title included, also figures 8, 9.

Figure 5. The legend is not in a readable form, please correct.

Figure redrawn

Figure 8. I suggest withdrawing some of the labels on the scale. In addition, can it be placed to the right of the figure? Same
10   in Fig. 9.

Done

Figure 10. The plot will be clearer if the units ("mm") will be removed and the labels of each line will not cover each other.

Done

Figure 11. The legend is unclear.

15   *Figure redrawn*

Bias in dynamically downscaled rainfall characteristics for hydroclimatic projections

Nicholas J. Potter[1], Francis H.S. Chiew[1], Stephen P. Charles[2], Guobin Fu[2], Hongxing Zheng[1], Lu Zhang[1]

[1]CSIRO Land and Water, Canberra ACT, 2601, Australia
5   [2]CSIRO Land and Water, Floreat WA, 6148, Australia

Correspondence to: Nicholas J. Potter (Nick.Potter@csiro.au)

Abstract. Dynamical downscaling of future projections of global climate model outputs can provide useful information about plausible and possible changes to water resource availability, for which there is increasing demand for regional water resource planning processes. By explicitly modelling climate processes within and across global climate model gridcells for a region, 10   dynamical downscaling can provide higher resolution hydroclimate projections, and independent (from historical timeseries), physically plausible future rainfall timeseries for hydrological modelling applications. However, since rainfall is not typically constrained to observations by these methods, there is often a need for bias correction before use in hydrological modelling. Many bias correction methods (such as scaling, empirical and distributional mapping) have been proposed in the literature, but methods that treat daily amounts only (and not sequencing) can result in residual biases in certain rainfall characteristics, which 15   flow through to biases and problems with subsequently modelled runoff. We apply quantile-quantile mapping to rainfall dynamically downscaled by NARCliM in the State of Victoria, Australia and examine the effect of this on: (i) biases both before and after bias correction in different rainfall metrics; (ii) change signals in metrics in comparison to the bias; and (iii) the effect of bias correction on wet-wet and dry-dry transition probabilities. After bias correction, persistence of wet states is under-correlated (i.e. more random than observations), and this results in a significant bias (underestimation) of runoff using 20   hydrological models calibrated on historical data. A novel representation of quantile-quantile mapping is developed based on lag-one transition probabilities of dry and wet states, and we use this to explain residual biases in transition probabilities. Representing quantile-quantile mapping in this way demonstrates that any quantile mapping bias correction method is unable to correct the underestimation of autocorrelation of rainfall sequencing, which suggests that new methods are needed to properly bias correct dynamical downscaling rainfall outputs.

25   ## 1 Introduction

There is a growing and on-going need for information about plausible and possible changes to water resource availability in the future due to climate change. End users of hydroclimate projections are interested in more spatially detailed information, information on water metrics for low- and high-flow events, and better predicted interdecadal metrics (Potter et al., 2018). Dynamical downscaling (such as provided by the NARCliM project for south-eastern Australia, see Sect. 2.1) has potential to 30   provide this type of information; however, there remain challenges associated with the use of this data. In this paper, we examine the suitability of NARCliM projections for providing hydroclimate projections for south-eastern Australia. Specifically, we look at the extent of biases in rainfall, which necessitate daily bias correction, and the effect of quantile-quantile mapping (QQM) bias correction on rainfall sequencing metrics that are important for runoff generation. Subsequent research in a related paper (Charles et al., 2019) focuses attention on the effect of these biases on runoff.

35   Of particular interest are possible changes to rainfall characteristics that could affect runoff and streamflow. Information on future changes to rainfall are typically derived from ensembles of global climate models (GCMs); however, the spatial resolution of these models is too coarse to provide information at the scale needed for hydrological impact modelling (i.e. catchments or gauges). Downscaling is the process by which finer scale spatial detail is produced from the GCM change information, which is at too coarse a resolution to be usable (Maraun et al., 2010). Many water resource studies use 'empirical

scaling', where historical rainfall observations are scaled (perhaps annually or seasonally) for direct use, or 'statistical downscaling' in which a direct statistical relationship is developed between rainfall and other atmospheric predictors. These methods are relatively simple to use, and results from empirical scaling typically lie in the middle of the range of results from other downscaling methods (Chiew et al., 2010; Potter et al., 2018). However, as empirical scaling methods rely on the

5 historical record of rainfall, future changes in rainfall sequencing (e.g. changes to multi-day accumulations and wet/dry transitions) and consequent effects on runoff cannot be properly modelled. Dynamical downscaling, in which a regional climate model (RCM) of finer spatial resolution than the host GCM, generates rainfall sequences independent from historical observations. However, challenges remain with using dynamical downscaling output since rainfall (and other climate variables) are not explicitly constrained by observations (see, e.g., Piani et al., 2010; Chen et al., 2011; Teutschbein and Seibert,

10 2012). As such, dynamical downscaling outputs typically need to be bias corrected for direct use in hydrological models. In particular, a common feature of dynamical downscaling is the tendency to underpredict the occurrence of zero- and low-rainfall days, which is sometimes known as the drizzle effect (e.g. Maraun, 2013). RCM output has been bias corrected for applications in Australia in Tasmania (Bennett et al., 2014) and central coast of New South Wales (Lockart et al., 2016), but both these studies found residual biases in multi-day rainfall events, dry spell durations, and autocorrelation of rainfall occurrences (see

15 also Themeßl et al., 2012).

Water resources in Victoria are shared by urban users, irrigators, industry and the environment. Long-term water strategies and shorter term sustainable water strategies are required for Victoria's water regions by the Victorian Water Act. A key aspect of these water planning processes is accounting for scenarios of climate change as determined by the available science. Cool-season rainfall in Victoria since 2000 has averaged 15% less than the long-term average during the 20th century (Hope et al.,

20 2017). This reduction has been linked to the observed expansion of Hadley Cell circulation (Post et al., 2014). The median scenario of climate change for Victoria typically has reduced rainfall and runoff later in the century, with slightly larger percentage declines in the western parts of the State (Post et al., 2012; Potter et al., 2016). Providing better information to improve water planning processes includes developing finer spatial resolution projections and different metrics of daily rainfall amounts and occurrences.

25 Most bias-correction methods alter daily amounts with the application of distributional mappings. The temporal structure of the occurrences is most often unaltered. Further, bias correction can and does affect the magnitude of change signals (Hagemann et al., 2011; Gutjahr and Heinemann, 2013; Dosio, 2016). The underlying assumption of bias correction is that the RCM output faithfully represents the climate processes responsible for rainfall, although the amounts themselves may not be accurate. Water resource projection modelling is concerned with future changes, and so an argument could be made that

30 although the rainfall amounts are biased for hindcast (historical) simulations, they will presumably be equally biased for future simulations. In this way, changes can be inferred from comparing biased historical and future rainfall and runoff. However, the sensitivity of runoff to rainfall means that biased rainfall can have large effects on the change signal of runoff (Teng et al., 2015). Furthermore, hydrological models are calibrated to historical rainfall and runoff sequences, and since the distribution of runoff is usually highly skewed, using biased rainfall sequences can distort the distribution of runoff, thus creating large

35 biases in high and low runoff amounts. These biases in runoff make inferences on the changes to runoff characteristics highly uncertain when biased rainfall inputs are used.

Bias correction identifies a relationship or mapping between observed historical rainfall and modelled historical rainfall (in this case hindcast RCM rainfall). This mapping when applied to hindcast RCM rainfall results in a distribution of rainfall identical (or very similar, depending on the methods) to the historical observations. This mapping can then be applied to future

40 RCM rainfall, resulting in unbiased future rainfall sequences. Of course, applying the relationship into the future assumes the bias in RCM rainfall does not change into the future or for different (wetter or drier) climate periods. Bias-correction methods (see Schmidli et al., 2006; Boé et al., 2007; Lenderink et al., 2007; Christensen et al., 2008; Piani et al., 2010; Themeßl et al., 2011; Teng et al., 2015) fall into three main categories:

- Scaling or change-factor methods;
- Non-parametric (empirical) quantile-quantile mapping (QQM); and
- Parametric (distributional) QQM.

Scaling methods simply consider the change in mean and apply a constant factor to correct bias in RCM rainfall. Quantile-
5   quantile mapping matches each quantile (or a selection of quantiles) of the two distributions. This can be done using the
empirical cumulative density or fitting a distribution to both observed and hindcast RCM daily rainfall amounts.

Teng et al. (2015) demonstrated that representing daily rainfall distributions with double-gamma distributions (see also, e.g.,
Yang et al., 2010) was largely identical to empirical QQM, implying that distributional and empirical approaches give similar
results so long as the distribution is sufficiently flexible. Arguably, the choice between non-parametric (empirical) or
10 parametric (distributional) mapping is a representation of the bias-variance trade-off problem. Empirical mapping will reduce
bias to zero, but at the cost of increasing the variance of predictions, since the mapping will be very sensitive to individual
amounts. Distributional mapping fits the data across the entire rainfall distribution but can result in the hindcast RCM rainfall
not being mapped exactly to the historical distribution. For this study we apply empirical quantile-quantile mapping for each
season across integral percentiles as described below. Overall there is a small but relatively unimportant difference between
15 different methods for QQM, at least in the Australian context studied by Teng et al. (2015).

Whereas quantile-quantile mapping can effectively reduce historical error in daily rainfall amounts to zero, albeit with some
of the caveats already mentioned, the bias corrected rainfall timeseries could still harbour biases and unrealistic characteristics
that will result in runoff biases after being routed through a rainfall-runoff model. Specifically, QQM bias correction cannot
remove biases in rainfall sequencing and multi-day accumulations that might not be readily apparent when considering only
20 the daily distribution of rainfall amounts (e.g. Addor and Seibert, 2014; Li et al., 2016; Ines and Hansen, 2006). However,
Terink et al. (2010) and Rajczak et al. (2016) contend that bias correction produces reasonable transition probabilities and spell
durations.

Unfortunately, it is not easy to tell exactly which characteristics of rainfall drive runoff generation, and in general the sensitivity
will depend on catchment physical characteristics, storm type and intensity, as well as antecedent moisture and groundwater
25 stores (Goodrich and Woolhiser, 1991; Bell and Moore, 2000; Beven, 2001). Spectral and multifractal approaches (e.g. Milly
and Wetherald, 2002; Matsoukas et al., 2000; Tessier et al., 1996) show that rainfall variability at shorter timescales is by and
large incorporated into soil moisture buffers thus dampening runoff variability at these timescales. However, over timescales
of several days and greater, variability in runoff matches variability in rainfall more and more closely. As such, it is evident
that large, intense rainfall events (measured perhaps by the upper tail of the rainfall distribution), more seasonal rainfall regimes
30 (Wolock and McCabe, 1999), relatively larger variability of rainfall (Potter and Chiew, 2011), and large multi-day
accumulations of rainfall are most important for runoff generation (Addor and Seibert, 2014), particularly for high flow events
(Jaun et al., 2008), and we focus on these kinds of rainfall metrics in this study.

The main aim of the paper is to investigate the effect of bias correction on rainfall characteristics relevant to runoff generation.
Given that QQM bias-correction approaches are specifically tailored for correcting daily amounts while retaining the existing
35 sequencing of occurrences, we seek to understand the effect of bias correction on rainfall sequencing and transition
probabilities. Specifically, we investigate whether key rainfall metrics contain biases in dynamically downscaled GCM
hindcasts relative to observations. We then examine if bias correction acts to either enhance or moderate any such biases, and
whether bias correction affects change signals (i.e. downscaled GCM future relative to downscaled GCM historical). Section
2 describes the data and methods used in this study, Sect. 3 presents results, and Sect. 4 and 5 contain discussion and
40 conclusions.

2 Data and Methods

2.1 The NARCliM Project

The NSW and ACT Regional Climate Modelling Project (NARCliM; https://climatechange.environment.nsw.gov.au/Climate-projections-for-NSW/About-NARCliM; Evans et al., 2014) is a regional climate change project to deliver high resolution
5 dynamically downscaled climate change projections. As noted previously, GCMs run at a spatial resolution that is unable to provide meaningful information for decision makers at catchment and basin scale. To resolve subgrid processes, the NARCliM project uses the 'Weather Research and Forecasting' (WRF) regional climate model (specifically the Advanced Research WRF version 3; Evans and McCabe, 2010), which is a mesoscale atmospheric model with many applications both in numerical weather prediction and climate projections (Skamarock and Klemp, 2008), forced by atmospheric variables output from GCMs.
10 The NARCliM modelling domain is most of south-eastern Australia and the neighbouring Pacific Ocean at 10km × 10km resolution, nested within the larger CORDEX Australasian region (Giorgi et al., 2009) at 50km × 50km resolution (Evans et al, 2014). WRF reads in output from a GCM along its lateral and lower boundaries and simulates the climate on a finer resolution within those boundaries (Skamarock et al., 2008). Out of the 23 CMIP3 (Meehl et al., 2007) GCMs available at the time the NARCliM project started, four were eventually chosen for downscaling: MIROC3.2 (medres), ECHAM5, CCCM3.1
15 and CSIRO-Mk3.0 (Evans et al., 2013). Initially, GCMs that did not adequately represent climate dynamics of the region were eliminated. Based on a meta-analysis, GCMs were then ranked, and independence of model error and future changes, were assessed to represent the range of model ensemble (Evans and Ji, 2012a). A similar procedure was applied to select the configurations of WRF used to downscale each host GCM (Evans and Ji, 2012b; see also Ji et al., 2016; Olson et al., 2016). WRF models were evaluated against 2-week heavy rainfall events with an intent to select the best possible RCM configurations
20 for rainfall generation while also accounting for model uncertainty (Olson et al., 2016).
Overall, there is reasonable confidence in NARCliM projections generally, for both rainfall and temperatures (Evans et al., 2012; Olson et al., 2016; Ji et al., 2016), particularly at daily scale for rainfall (Gilmore et al., 2016), although NARCliM has a quantitative cold and wet bias generally (Ji et al., 2016). The underlying performance of the RCM component of WRF has been tested extensively (Evans et al., 2012; Andrys et al., 2016; Ekström, 2016; Olson et al., 2016; Ji et al., 2016; Gilmore et
25 al., 2016). The three WRF physics configurations are labelled R1, R2 and R3 (see Evans et al., 2014, Table 1), detailed information is provided elsewhere (Evans et al., 2012; Ekström, 2016; Gilmore et al., 2016; Ji et al., 2016). The full NARCliM model ensemble thus consists of twelve members (four selected GCMs, each downscaled by the three RCM configurations). Ji et al. (2016) evaluated model output against AWAP (see below) rainfall, demonstrating good representation of precipitation processes at a wide range of timescales, but conclude that bias correction is still needed before applying the data. For
30 hydrological applications, the R2 configuration is recommended (Olson et al., 2016). Although we consider the entire modelling ensemble in this paper, in a related paper (Charles et al., 2019) we use only that specific RCM physics scheme for modelling runoff. We discuss the physical credibility of the NARCliM ensemble in greater detail in Sect. 4 below.

2.2 Daily Rainfall Data

Daily accumulated precipitation from the NARCliM 12-member WRF-downscaled ensemble (referred to from here onwards
35 as 'NARCliM rainfall') is produced at approximately 10km × 10km grids, corresponding to the NARCliM domain covering south-eastern Australia (Evans et al., 2014; Ji et al., 2016), which was bilinearly interpolated to a regular grid using Climate Data Operators (CDO; https://code.mpimet.mpg.de/projects/cdo). NARCliM provides bias corrected data (Evans and Argüeso, 2014), which applies a parametric gamma distribution quantile-quantile mapping procedure, which is similar in many respects to the non-parametric procedure we apply to the raw NARCliM data in this paper (Teng et al., 2015).
40 The historical (baseline) period for NARCliM projections is 1990–2009, and we relate NARCliM historical rainfall features to observations and NCEP/NCAR reanalysis over the same period in this paper. Realisations of future rainfall follow the A2

Scenario of the Special Report on Emissions Scenarios (SRES) (Nakićenović et al., 2000) at 2060–2079. Change signals presented later are thus averages over 2060–2079 compared to 1990–2009. Observed rainfall data is obtained from the Australian Water Availability Project (AWAP; Jones et al., 2009). This is a 0.05°×0.05° gridded dataset interpolating observations from point rainfall records from the Australian Bureau of Meteorology. The AWAP rainfall observations are projected over a regular latitude-longitude grid, hence the need for interpolation of NARCliM data to be aligned with this. The AWAP rainfall data is then regridded by using weighted averages of AWAP gridcells overlapping the 0.1°×0.1° AWAP grid.

2.3 Quantile-Quantile Mapping Bias Correction

Quantile-quantile mapping bias correction works by estimating the cumulative density function for observed and modelled historical observed (here AWAP) daily rainfall amounts: F_o and F_m. These are then combined to produce a mapping function:

$$P_o = F_o^{-1} \circ F_m \, (P_m)$$

(Fig. 1). The map $F_o^{-1} \circ F_m$ thus returns the observed daily distribution (or approximately so, depending on the method) when applied to the modelled historical timeseries, and a bias corrected future timeseries when applied to the modelled future timeseries. We use the R package 'qmap' (Gudmundsson et al., 2012), to estimate the cumulative density functions for the mapping function. This package estimates quantiles for both observed and modelled non-zero rainfall at integral percentiles including 0 (minimum) and 100 (maximum). The quantile for a particular daily rainfall amount is then estimated using linear interpolation between percentiles, and linear extrapolation in case of future modelled rainfall lying outside the historical distribution. Compared to using the empirical distributions directly, differences between the bias-corrected modelled rainfall distribution and the distribution of observations can occur (of the order of 2–3%) because the interpolation between large rainfall percentiles (particularly 99 to 100) will not match observed percentiles exactly. As noted by Teng et al. (2015), sufficiently flexible approaches to bias correction give very similar results. QQM bias correction in this way was applied separately to each three-month season (i.e. DJF, MAM, JJA and SON) in each grid cell independently, using the full historical period (1990–2009) as calibration period, to both historical (1990–2009) and future (2060–2079) periods.

2.4 Defining Transition Probabilities

Whereas QQM bias correction can correct the daily distribution exactly, daily bias correction is not set up to correct sequences and accumulations (Addor and Seibert, 2014). To this end, we consider not only how bias correction affects daily metrics of rainfall, but also the sequencing of wet days that produce runoff. One way of measuring this is through transition probabilities of wet and dry sequences. To this end, we consider a simple two-state Markov Chain rainfall occurrence model. Here, the probability of a wet or dry day depends on whether the previous day was wet or dry. A 'dry' day can be defined as either zero rainfall or rainfall below a given threshold (such as 1 mm). Define the wet-to-wet transition probability (i.e. the probability of a wet day following a wet day) as $w = \Pr\left(W_n | W_{n-1}\right)$ and the corresponding dry-to-dry transition probability as $d = \Pr\left(D_n | D_{n-1}\right)$. These determine the Markov Chain since $\Pr(D_n | W_{n-1}) = 1 - w$ and $\Pr(W_n | D_{n-1}) = 1 - d$. Further, the probability of the occurrence of a dry day, p, is fully determined by the w and d parameters, as given by Cox and Miller (1965):

$$p = \frac{1 - w}{2 - d - w}$$

Equivalently,

$$w = 1 - p \frac{1 - d}{1 - p}$$

This relationship can be plotted for a range of probabilities p (dotted lines in Fig. 10). Note that if a series of occurrences of dry and wet days has zero autocorrelation (i.e. the state probability is independent of the rainfall state in the previous day), then it follows that $\Pr(D_n | D_{n-1}) = \Pr(D_n | W_{n-1}) = \Pr(D_n) = p$. As such, the diagonal line where $p = d$ (dashed line in Fig.

5

10) corresponds to an independent (i.e. zero autocorrelation) series of occurrences. The area above and to the right of the dashed line corresponds to a series of occurrences with positive autocorrelation (i.e. $\Pr(D_n|D_{n-1}) > p$, so that dry sequences are more likely to persist), whereas the area below and to the left of the dashed line corresponds to series with negative autocorrelation. The framework developed above is used in Sect. 3.3 as a novel way to represent the relationships between state-transition probabilities and rainfall quantiles to investigate the effect of bias correction on transition probabilities.

3 Results

3.1 Assessment of Regional Performance of Modelled Rainfall

Modelled rainfall is compared to observed (AWAP) rainfall in Figure 2. ECHAM5-R1 was chosen as a representative GCM-RCM ensemble member here as it had the median historical regional rainfall across Victoria from the NARCliM model ensemble. The spatial rainfall fields downscaled from both the reanalysis and GCM both show reasonable agreement with the spatial pattern of the observed mean annual rainfall, with relatively larger rainfall across the mountain ranges to the east of the State and across the southern coast, and less rainfall to the more arid north-west region. However, both rainfall fields are evidently positively biased with around 100–200 mm of excess rainfall consistently across most of the State (Fig. 2f, g), relatively more for the far eastern part of Victoria for the reanalysis data and across the mountain ranges for the GCM data, interspersed with small patches of negative bias (observations larger than downscaled data). Averaged across Victoria, the mean absolute bias is approximately 26% for the downscaled reanalysis and 43% for the downscaled GCM data. In comparison, the absolute change signal averages 3%, rising to around 10% in the eastern part of Victoria (Fig. 2h).

3.2 Bias Correction of NARCliM

Consistent with Fig. 2, the raw NARCliM rainfall is mostly wetter (positive bias) across Victoria (Fig. 3a), except for a tendency towards underprediction in the south-east coast for some models. The quantile-quantile bias correction method is formulated to correct historical quantiles of rainfall exactly (when applied to the same time period used for calibration) so that bias-corrected mean annual rainfall, as well as any quantiles, is approximately equal (Fig. 3b). However, the method used here does not correct high rainfall quantiles (e.g. P99 and above) exactly, due to the interpolation between quantiles as described in Sect. 2.3 (Fig. 4b). This residual bias appears to be randomly distributed spatially and results in bias-corrected mean annual rainfall not being exactly corrected. However, this effect is generally less than 5% of the mean annual rainfall. This effect can be removed entirely by using empirical density functions rather than interpolated values, but with the effect of increasing prediction uncertainty.

Figure 5 shows the distribution of bias for different rainfall metrics before bias correction and the residual bias after QQM bias correction. The bias in raw NARCliM ranges from 5% to 50% for all percentiles (Fig. 5a), increasing as the percentile increases (i.e. more relative bias for higher rainfall amounts). As with Fig. 3, bias at all percentiles is effectively reduced to zero after bias correction, although the residual bias is relatively larger at larger percentiles (higher rainfall), similarly to Fig. 4b. NARCliM rainfall is overestimated at all seasons and months before bias correction (Fig. 5b), although winter rainfall is relatively less biased than summer rainfall. Bias correction reduces bias to zero annually and seasonally, since QQM is applied to each season separately. Since the intra-seasonal relative monthly rainfall amounts are not exactly equal to the observed amounts, seasonal bias correction occasionally overcorrects bias, particularly in February, April, May and June, with the bias corrected rainfall in these months being less than observed whereas they were overpredicted before bias correction; this could be reduced to zero using monthly correction factors, although at the potential cost of overfitting. Overall though, the absolute relative bias is reduced and closer to zero in all months compared to the raw RCM monthly amounts.

Figure 5c shows relative bias in rainfall sequencing related metrics. Autocorrelation of rainfall amounts is underpredicted before bias correction, and the magnitude of this bias actually increases after bias correction. Whereas QQM reduces bias in

6

dry-dry transition probabilities (i.e. the probability of dry sequences persisting), bias in wet-wet transition probabilities increases after bias correction so that, similarly to autocorrelation, the probability of wet spells persisting is considerably underpredicted after bias correction. We examine this in more detail in Sect. 3.3 using the transition probability framework developed in Sect. 2.4. Since dry spells are directly related to dry-dry transition probabilities, the bias in mean and maximum

5 dry spells is well corrected, whereas maximum 3-day rainfall accumulation and wet-spell occurrences all have negative bias after QQM bias correction. Different percentiles of 3-day rainfall accumulation (calculated as percentiles of a 3-day moving sum of rainfall timeseries) have different residual bias (Fig. 6). Three-day accumulation percentiles below the 80th are all slightly overestimated after bias correction, but above the 80th percentile, a large residual underestimation is present. This underestimation reduces at around the 99th percentile but is moderated somewhat for the 3-day maximum (i.e. 100th

10 percentile). As noted in section 2.1, WRF models were selected according to their skill in reproducing selected 2-week periods of heavy rainfall. This provides a potential explanation for the smaller bias in 3-day maxima relative to 3-day 99% rainfall.

It is likely that underpredicting wet spell occurrences and persistence (Fig. 5c and Fig. 6) will result in runoff from the bias corrected rainfall being underpredicted too. To explore this, runoff was modelled from bias corrected rainfall and observed potential evapotranspiration data using GR4J (see Charles et al., 2019 for more details). Figure 7 plots the percentage difference

15 in bias-corrected ensemble-median mean annual runoff for each 0.1°×0.1° cell compared to mean annual runoff modelled using AWAP observed rainfall. The ensemble median of runoff across Victoria is underpredicted by between 10–20% across almost all of Victoria, which suggests that the residual bias in wet spell occurrences and persistence is problematic for runoff modelling. Whereas the smallest percentage biases appear to be over the high-runoff producing region (Fig. 7c), this region has the highest absolute biases with bias in runoff of more than −20 mm (Fig. 7b). Characteristics and biases of runoff from

20 bias corrected NARCliM rainfall is explored in more detail by Charles et al. (2019).

3.3 Residual Bias in Rainfall State Transition Probabilities

The results in this study demonstrate that dry-dry transition probabilities for NARCliM have low residual bias (possibly due to the emphasis in QQM on preserving zero-rain occurrences), but that wet-wet transition probabilities have more bias (i.e. they are closer to zero, see Fig. 5c) after QQM bias correction. These residual biases in wet-wet transitions result in the

25 persistence of wet spells being underestimated even though the volumetric amount of rainfall is, by design of QQM bias correction, equal to observed rainfall at any gridpoint.

After bias correction, dry-dry transition probabilities for a 1 mm threshold are reduced, but still have a small negative bias (Fig. 5c). Both the bias corrected WRF-downscaled reanalysis and bias corrected WRF-downscaled GCM results (Fig. 8) show a spatial pattern very similar to observations with higher dry-dry transition probabilities to the northwest of the State, and at

30 similar places along the southern coastline. However, the reanalysis and more so the GCM result has a lower dry-dry transition probability across almost all of the region (Fig. 8). As such, dry spells from the bias-corrected model output are likely to be shorter in duration and less common than that from the observed rainfall (although bias correction does reduce the bias in dry spells somewhat compared to the bias in the raw data as seen in Fig. 5c).

Whereas the dry-dry transition probabilities were largest in the north-west, drier, part of Victoria, the wet-wet transition

35 probabilities are largest over the high-runoff producing region (Fig. 9), which corresponds to the high-relief, high-altitude part of the State. As with the dry-dry transition probabilities, both bias-corrected downscaled reanalysis and bias-corrected downscaled GCMs reproduce the spatial pattern of wet-wet transition probabilities, but there is considerable residual bias in these probabilities across the entire region. The residual bias in WRF-downscaled GCM transition probabilities is over 10% over most of Victoria. This residual bias results in underestimation of wet spell occurrences and durations and multiday

40 accumulations of rainfall (Fig. 5c). The bias in wet-wet transition probabilities is more problematic for modelling runoff than the bias in dry-dry transition probabilities, not only because it is of larger magnitude, but because runoff is sensitive to multiday

7

wet spells, and the larger wet-wet probabilities occur in high runoff producing areas, which we would like to model correctly for regional water availability projections.

Figure 10 shows rainfall percentiles for a sample grid cell overlaid on the transition-probability space developed in Sect. 2.4. Other grid cells and GCMs show very similar responses (as can be seen in the low spread of results for d and w in Fig. 5c).

5 Quantile-quantile mapping bias correction equates the quantiles for each rainfall amount such that equal rainfall amounts for observations and bias-corrected rainfall occur on the same probability contours (dotted lines in Fig. 10), with raw values translated along the rainfall amount curve. That is, values on the blue line in Fig. 10 map to corresponding values on the red line; the slight variation between the lines is due to different bias corrections in each season. As such, wet-wet and dry-dry transition probabilities for a given rainfall threshold (e.g. 1 mm) for bias-corrected rainfall are equal to the transition

10 probabilities for the corresponding amount in the raw data. For example, in Fig. 10, the exceedance probability for 1 mm in the observed data (green line) is 0.774. The corresponding quantile in the raw data (blue line) is 2.675 mm. This amount is mapped to 1 mm in the bias-corrected data (red line), and the corresponding wet-wet and dry-dry transition probabilities for 2.675 mm are identical to the transition probabilities for 1 mm in the bias-corrected data. Recall that points above and to the right in Fig. 10 correspond more (positive) serial correlation, which means that the observed rainfall timeseries contains more

15 correlation structure in the sequence of wet and dry spell occurrences than the modelled rainfall sequence, and that QQM bias correction cannot rectify this since daily QQM retains the autocorrelation structure of the raw time series since daily amounts are simply rescaled. We surmise that, as bias correction is ultimately intended to produce physically plausible rainfall, bias correction methods that adjust occurrences are needed to properly correct biases for hydrological modelling.

3.4 Change signals

20 Here we examine change signals in rainfall metrics (i.e. percentage difference in RCM future relative to RCM historical averages) specifically looking at whether bias correction alters the change signals. Figure 11a shows the change signals in different rainfall percentiles. For the raw data, there is a small decrease in low to moderate rainfall amounts less than the non-zero 40th percentile and a future increase in non-zero percentiles above 50%. The nature of QQM bias correction means that raw and bias corrected equal percentiles cannot be meaningfully compared. Nevertheless, a similar pattern is found with the

25 bias corrected data, namely that larger rainfall amounts have larger relative changes than smaller rainfall amounts.

Figure 11b shows change signals in mean annual, seasonal and monthly average rainfall. The magnitude of the median change signal in mean annual rainfall is around –5%, and seasonal changes are comparable to the annual change except for SON rain. Rain in SON has a decrease projected by the NARCliM ensemble of around 20% (found to be statistically significant by Olson et al., 2016). Compared to the raw bias in mean and seasonal rainfall (Fig. 5b) of between 25%–50%, these change signals are

30 between one-half and one-tenth of the bias. After bias correction, there is little difference in the magnitude and direction of change in seasonal and monthly averages. However, the mean annual rainfall change is moderated somewhat, and this is problematic since mean annual changes are most often considered in regional projection applications (a discussion on bias correction effects on change signals is considered in the next section, and Charles et al. (2019) discuss this in the context of the present study in more detail).

35 Although the residual bias in rainfall sequencing metrics is not eliminated, and in some cases (e.g. wet-wet transition probabilities) is actually increased, after bias correction (Fig. 5c), Fig. 11c shows that the change signals in rainfall sequencing metrics is largely unaffected by bias correction. With the exception of maximum 3-day rainfall accumulation, the distributions of sequencing change signals are largely identical before and after bias correction. The difference in change signal for maximum 3-day rainfall accumulation is presumably related to the relatively larger increase in large rainfall events (e.g. P99

40 in Fig. 11a).

4 Discussion

Individual percentiles and seasonal totals are, by design, effectively reduced to zero using QQM. This elimination of bias is due to the common calibration and validation period (1990–2009) used for bias correction. For applications, different calibration and validation periods should be used to accurately estimate the bias from QQM. Nevertheless, some interpolation

5 and extrapolation occurs in the approach used here, so there is some random residual bias in higher percentiles (i.e. high rainfall amounts). This residual bias can be eliminated altogether by using the exact empirical density functions, but at the cost of increased predictive uncertainty. Using empirical densities also raises problems with extrapolation past historical amounts. Although Teng et al. (2015) demonstrated that distributional and empirical approaches give similar results if the distribution is sufficiently flexible, the choice of method should be considered carefully in relation to the end products required.

10 We demonstrate in Sect. 3.4 that change signals (future mean relative to historical) in rainfall metrics can be considerably smaller than the bias (modelled historical relative to observed historical). On one hand, this seems problematic since biases in processes can be considered so large that the changes are insignificant. On the other hand, there is no particular legitimacy for this viewpoint, certainly not from a statistical sense. The magnitude of bias does not provide any sort of confidence level in changes to rainfall metrics. However, given such relatively large biases, it is reasonable to assume that there are some errors

15 in the way particular climate processes are modelled, either through the host GCM or the RCM. It would be desirable to understand the reasons and climatic process responsible for biases and assess whether these processes are unrealistic, in addition to whether these biases render the changes physically implausible.

We acknowledge that the underlying performance of GCMs in accurately simulating climate dynamics of the region under consideration is extremely important. In the context of the study area in the current manuscript, South-Eastern Australia, the

20 selection of GCMs representing important climate processes has been an on-going research strand both for the CMIP3 climate model ensemble (Smith and Chandler, 2010; Kirono and Kent, 2011; Kent et al., 2013; CSIRO, 2012; Evans and Ji, 2012a; McMahon et al., 2015), which is used for the current NARCliM dataset, as well as the CMIP5 ensemble (see CSIRO and Bureau of Meteorology, 2015; Hope et al., 2016). As mentioned in the introduction, the RCM component of NARCliM, WRF, has been tested extensively, including the general cold and wet bias over south-eastern Australia. It is suggested that the wet

25 bias (Fig. 3a) is related to subgrid cloud cover representations (see, e.g., García-Díez et al., 2015; Di Virgilio et al., 2019) and correction of this is the subject of current, ongoing research (Di Virgilio et al., 2019). Current work (NARCliM1.5) is continuing to develop the modelling framework at a higher spatial resolution (5km × 5km), using CMIP5 models and improved RCM configuration (Downes et al., 2019).

The use of models that produce plausible climate dynamics is of course desirable; however, in practice it is not necessarily

30 always possible. Apart from the fact that the 'best' models identified by the above references differ according to the criteria used, using a dynamical downscaling ensemble for hydrological applications is an opportunistic endeavour, relying largely on existing data products, which have been prepared with many applications in mind, not just hydrological applications. As such, it is not always practical to choose the GCMs and RCMs that best represent climate dynamics important for hydrological applications; many studies have also contended that accuracy in representing historical conditions is no guarantee that future

35 changes are correctly modelled (e.g. Knutti et al., 2010; Racherla et al, 2012).

As mentioned in the introduction. Evans and McCabe (2010) examined the RCM component (WRF) of NARCliM, concluding that the El Niño-Southern Oscillation, the chief climate process modulating interannual variability of rainfall (Power et al., 1999), was well modelled over south-eastern Australia. Evans and McCabe (2010) also concluded that the severity and duration of recent prolonged droughts over south-eastern Australia were also captured, although the spatial pattern was not characterised

40 exactly. The sub-tropical ridge, which determines the seasonal positioning of storm tracks over southern Australia, was less well represented by WRF (Andrys et al., 2016). Based on the results cited here, we have confidence in the modelling setup of NARCliM to represent atmospheric circulation for southern Australia reasonably well, although we acknowledge that bias correction of NARCliM for end-user applications should consider model skill in atmospheric circulation.

In general, bias correction does not tend to alter the change signals in rainfall metrics (with the exception of 3-day accumulation and low rainfall percentiles). Nevertheless, small differences in rainfall metrics can result in large differences in runoff metrics and other water availability measures (e.g. low flows and high flows). High runoff and even average runoff amounts can be very sensitive to 3-day rainfall accumulation, which we saw can be altered through daily bias correction. It is recommended

5   that the effects of bias correction are included in any uncertainty analysis undertaken.

Section 3.4 of the manuscript shows that bias correction can affect change signals (future relative to historical) of different hydroclimatic metrics (see also Hagemann et al., 2011; Gutjahr and Heinemann, 2013; Dosio, 2016). Under the assumption that bias is time invariant, Gobiet et al. (2015) argue that bias correction improves the accuracy of climate change signals. Cannon et al. (2015), however, argue that trend-preserving methods should be used (see also Li et al., 2010; Wang and Chen,

10   2014). Maraun (2016) and Maraun et al. (2017) summarise the debate surrounding the use of trend-preservation methods and conclude that the decision should be informed by the credibility or otherwise of the GCMs in representing the processes driving the changes. This debate further highlights the need for informed selection and screening of GCMs at the start of the modelling process. However, we argue that there is value in reporting both pre- and post-bias correction future changes in light of the difficulties involved in model selection and assessment, particularly in the case of pre-existing and computationally expensive

15   projections such as dynamically downscaled ensemble such as NARCliM.

The simple QQM method used here does not consider spatial correlation between rainfall gauges or gridcells at all. Maintaining spatial correlations is clearly important for runoff generation and neglecting this can lead to 'inflation'. Inflation refers to a phenomenon in bias correction (Maraun, 2013) or statistical downscaling (von Storch, 1999) where an unmeasured predictand variable is estimated using the predicted values from a statistical model. Since models contain error, the variance of a timeseries

20   of predicted values is expected to be less than the variance of the true time series of the variable. In the present context of bias correcting rainfall from RCMs, Maraun (2013) demonstrates that bias correction reduces subgrid spatial heterogeneity compared to actual precipitation, and that this is particularly problematic when GCM or RCM resolution is much greater than that of observations. In this case, the spatial correlation between gauges is increased. As a result, large rainfall amounts become overestimated and low amounts underestimated. Preserving the correct spatial correlation between gauges or gridpoints is an

25   important issue, and the issue of unintended spatial effects of (temporal) bias correction is compounded by applying bias correction independently to each gridcell, as we have done here – although this tends to reduce subgrid spatial correlation (see Hnilica et al., 2017). Maraun (2013) recommends aggregating catchment rainfall prior to bias correction to reduce the issue of inflation, and Charles et al. (2019) examine this in more detail in relation to catchment runoff production. Variance inflation due to differing grid cell sizes (Maruan, 2013) is less an issue for the current study, as NARCliM grid cell size is comparable

30   to that of the gridded rainfall observations, and the next generation of dynamically downscaled climate projections (Downes et al., 2019) is to be provided at 0.05°×0.05° resolution, identical with the gridded rainfall products used in Australia. However, the issue of using a bias correction methodology that corrects daily amounts (and more generally temporal structure) while preserving spatial structure across catchments and basins remains a challenge and is a direction for further research. Equally, Bárdossy and Pegram (2012) noted that RCM rainfall was considerably less spatially correlated than observations in the Rhine

35   basin, even after QQM bias correction. Lower modelled spatial correlation in rainfall would lead to underestimated extreme flow events (even at the level of 1-year return period events), with a consequent underestimation in areally averaged inflows. Bárdossy and Pegram (2012) recommend pre-correcting spatial correlation using matrix recorrelation methods or by using a sequential recorrelation techniques, and this should be taken into consideration when applying bias correction for projection applications.

40   Although in this study, rainfall alone is bias corrected; for runoff applications, temperature or a suitable representation of potential evapotranspiration is needed. Methods exist for correcting rainfall and temperature simultaneously (e.g. Hoffmann and Rath, 2012; Piani and Haerter, 2012; Mehrotra et al., 2018). However, potential evapotranspiration has a second-order effect on runoff compared to rainfall (Chiew, 2006; Potter et al., 2011), and bias correction was shown to not significantly

affect the inherited relationships between rainfall and temperature (Wilcke et al., 2013). Certainly, the host GCM and RCM should correctly represent relationships between atmospheric variables in the study region, further highlighting the need for climate model assessment in construction of the model ensemble.

Another important consideration is the relevant metrics to be considered by end users. Bias correction by season, for example, can alter change signals annually (Haerter et al., 2011), and care must be taken as to which metrics are of interest, and which are the most appropriate bias correction methods to apply in order to properly account for the metrics of interest. Certainly, caution must be applied when considering rainfall and runoff metrics that were not considered when applying bias correction to projections. Low flow metrics are particularly problematic (Potter et al., 2018), where different downscaling and bias correction methods can give very different answers.

Although daily bias correction methods as outlined in this paper tend to result in residual bias in multi-day metrics, generally change signals in transition probabilities are very similar before and after bias correction. This information could thus potentially be extracted from RCMs to drive local weather generation or stochastic methods to provide future rainfall projections that can be suitable for local hydrological projections. Maintaining interannual and multi-decadal correlations and spatial correlations between rainfall gauges, remains a challenge for stochastic methods, however.

5 Conclusions

The relative magnitude of change signals (future RCM to historical RCM) of the different rainfall metrics examined here is typically less than the magnitude of the bias. Mean annual rainfall change is an order of magnitude smaller than the bias in mean annual rainfall but seasonal changes are closer to half of the bias in seasonal averages. Although this might call into question the validity of the change signal, one approach is to assume that the magnitudes of the changes are responsive to changing greenhouse gas emissions, insofar as the changing atmospheric processes are realistically modelled by the RCM. Indeed, this is the basic premise behind empirical scaling, i.e. that the change is the authentic signature of the climate modelling especially since the RCMs are not explicitly tuned to observed rainfall.

Individual percentiles and seasonal totals are, by design, effectively reduced to zero using QQM. Monthly totals retain some residual bias because of compensating biases within each season due to small errors in rainfall seasonality by the RCMs.

Metrics associated with rainfall sequencing (e.g. serial correlation, wet-wet and dry-dry state transition probabilities and quantiles of 3-day accumulation) all have significant residual bias, particularly so for wet-wet state transition probabilities in which the magnitude of bias in raw RCM historical runs is amplified after bias correction. This residual bias leads to a considerable underestimation of mean annual runoff after rainfall is routed through a hydrological model because runoff is very sensitive to multiday accumulations of rainfall and sequencing of wet spells in particular.

An analysis of the lag-one transition probabilities (i.e. wet state to wet state and dry state to dry state) showed that NARCliM rainfall had transitions to different states that are more random (i.e. more independent) compared to observed rainfall. QQM bias correction is unable to correct these transition probabilities as QQM retains the transition probabilities for any particular quantile. Since persistence of wet spells is critical for runoff generation, a different approach to bias correction is needed to successfully use NARCliM for runoff projections that can correct rainfall sequencing to better represent the observed correlation structure in wet and dry occurrences.

Change signals in annual, seasonal and monthly average rainfall and rainfall sequencing metrics are largely preserved after bias correction, with the exception of maximum 3-day rainfall accumulation. Rainfall sequencing metrics (such as state transition probabilities and daily rainfall autocorrelation) are largely unchanged by bias correction suggests the possibility of using this information to drive either weather-generation models or stochastic/resampling-based bias correction methods to produce hydrologically realistic rainfall sequences for hydroclimate projection applications.

11

References

Addor, N., and Seibert, J.: Bias correction for hydrological impact studies – beyond the daily perspective, Hydrol. Process., 28, 4823–4828, https://doi.org/10.1002/hyp.10238, 2014.

Andrys, J., Lyons, Thomas, J. L., and Kala, J.: Evaluation of a WRF ensemble using GCM boundary conditions to quantify mean and extreme climate for the southwest of Western Australia (1970–1999). Intl. J. Climatol., https://dx.doi.org/10.1002/joc.4641, 2016.

Bárdossy, A., and Pegram, G.: Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. Water Resour. Res., 48, W09502, https://doi.org/10.1029/2011WR011524, 2012.

Bell, V. A., and Moore, R. J.: The sensitivity of catchment runoff models to rainfall data at different spatial scales, Hydrol. Earth Syst. Sc., 4, 653–667, https://doi.org/10.5194/hess-4-653-2000, 2000.

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, https://doi.org/10.1002/joc.3830, 2014.

Beven, J. K.: Rainfall-runoff modelling: The primer, Wiley, Chichester, England, 2001.

Boé, 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, https://doi.org/10.1002/joc.1602, 2007.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?, J. Climate, 28(17), 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.

Charles, S. P., Chiew, F. H. S., Potter, N. J., Zheng, H., Fu, G., and Zhang, L.: Impact of downscaled rainfall biases on projected runoff changes. Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-375, in review, 2019.

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, https://doi.org/10.1016/j.jhydrol.2011.02.020, 2011.

Chiew, F.H.S.: Estimation of rainfall elasticity of streamflow in Australia. Hydrol. Sci. J., 51, 613–625, 2006.

Chiew, F. H. S., Kirono, D. G. C., Kent, D. M., Frost, A. J., Charles, S. P., Timbal, B., Nguyen, K. C., and Fu, G.: Comparison of runoff modelled using rainfall from different downscaling methods for historical and future climates, J. Hydrol. 387, 10–23, https://doi.org/10.1016/j.jhydrol.2010.03.025, 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, https://doi.org/10.1029/2008GL035694, 2008.

Cox, D. R., and Miller, H. D.: The theory of stochastic processes, Methuen, London, 1965.

CSIRO: Climate and water availability in south-eastern Australia: A synthesis of findings from Phase 2 of the South Eastern Australian Climate Initiative (SEACI), CSIRO, Australia, 41 pp., 2012.

CSIRO and Bureau of Meteorology: Climate Change in Australia. Information for Australia's Natural Resource Management Regions: Technical Report, CSIRO and Bureau of Meteorology, Australia. Available at: http://www.climatechangeinaustralia.gov.au/en/publications-library/technicalreport/, 2015.

Di Virgilio, G., Evans, J. P., Di Luca, A., Olson, R., Argüeso, D., Kala, J., Andrys, J., Hoffmann, P., Katzfey, J. J., and Rockel, B.: Evaluating reanalysis-driven CORDEX regional climate models over Australia: model performance and errors, Clim. Dynam., 53, 2985–3005, https://doi.org/10.1007/s00382-019-04672-w, 2019.

Dosio, A.: Projections of climate change indices of temperature and precipitation from an ensemble of bias-adjusted high-resolution EURO-CORDEX regional climate models, J. Geophys. Res. Atmos., 121, 5488–5511, doi:10.1002/2015JD024411, 2016.

Downes, S., Beyer, K., Ji, F., Evans, J., Di Virgilio, G., and Herold, N.: The next generations of NSW/ACT Regional Climate Modelling (NARCliM), AMOS-ICTMO 2019, Darwin, Australia, 11–14 June 2019, #102, 2019.

Ekström, M.: Metrics to identify meaningful downscaling skill in WRF simulations of intense rainfall events. Environ. Modell. Softw., 79, 267–284, https://dx.doi.org/10.1016/j.envsoft.2016.01.012, 2016.

Evans, J. P., and Argüeso, D.: Guidance on the use of bias corrected data. NARCliM Technical Note 3, 7pp., NARCliM Consortium, Sydney, Australia, available at http://www.ccrc.unsw.edu.au/sites/default/files/NARCliM/publications/TechNote3.pdf, 2014.

Evans, J. P., and Ji, F.: Choosing GCMs. NARCliM Technical Note 1, 7pp., NARCliM Consortium, Sydney, Australia, available at http://www.ccrc.unsw.edu.au/sites/default/files/NARCliM/publications/TechNote1.pdf, 2012a.

Evans, J. P., and Ji, F.: Choosing the RCMs to perform the downscaling. NARCliM Technical Note 2, 8pp., NARCliM Consortium, Sydney, Australia, available at http://www.ccrc.unsw.edu.au/sites/default/files/NARCliM/publications/TechNote2.pdf, 2012b.

Evans, J. P., and McCabe, M. F.: Regional climate simulation over Australia's Murray-Darling basin: A multitemporal assessment. J. Geophys. Res.-Atmos., 115, D14114, https://doi.org/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, https://doi.org/10.1007/s00382-011-1244-5, 2012.

Evans, J. P., Ji, F., Abramowitz, G., and Ekström, M.: Optimally choosing small ensemble members to produce robust climate simulations, Environ. Res. Lett., 8, https://doi.org/10.1088/1748-9326/8/4/044050, 2013.

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

García-Díez, M., Fernández, J., and Vautard, R.: An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation, Clim. Dynam., 45, 3141–3156, https://doi.org/10.1007/s00382-015-2529-x, 2015.

Gilmore, J. B., Evans, J. P., Sherwood, S. C., Ekström, M., and Ji, F.: Extreme precipitation in WRF during the Newcastle East Coast Low of 2007. Theor. Appl. Climatol., 125, 809–827, https://doi.org/10.1007/s00704-015-1551-6, 2016.

Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bull., 58, 175–183, 2009.

Gobiet, A., Suklitsch, M., and Heinrich, G.: The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal, Hydrol. Earth Syst. Sc., 19, 4055-4066, https://doi.org/10.5194/hess-19-4055-2015, 2015.

Goodrich, D. C., and Woolisher, D. A.: Catchment hydrology, Rev. Geophys. (Suppl.), 29, 202–209, 1991.

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. Sc., 16, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012, 2012.

13

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, https://doi.org/10.1007/s00704-013-0834-z, 2013.

Haerter, J. O., Hagemann, S., Moseley, C., and Piani, C.: Climate model bias correction and the role of timescales. Hydrol. Earth Syst. Sc., 15, 1065–1079. https://doi.org/10.5194/hess-15-1065-2011, 2011.

5 Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., and Piani, C.: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. J. Hydrometeor., 12, 556–578, https://doi.org/10.1175/2011JHM1336.1, 2011.

Hnilica, J., Hanel, M., and Puš, V.: Multisite bias correction of precipitation data from regional climate models, Intl. J. Climatol., 37, 2934-2946, https://doi.org/10.1002/joc.4890, 2017.

10 Hoffmann, H., and Rath, T.: Meteorologically consistent bias correction of climate time series for agricultural models. Theor. Appl. Climatol., 110, 129–141, https://doi.org/10.1007/s00704-012-0618-x, 2012.

Hope, P., Timbal, B., Hendon, H., Ekström, M., and Day, K.: Victorian Climate Initiative annual report 2015-16. Bureau Research Report BRR-015, 2016.

Hope, P., Timbal, B., Hendon, H., Ekström, M., and Potter, N.: A synthesis of findings from the Victorian Climate Initiative
15 (VicCI), Bureau of Meteorology, Australia, 56pp., 2017.

Ines, A. V. M., and Hansen, J. W.: Bias correction of daily GCM rainfall for crop simulation studies. Agr. Forest Meteorol., 138, 44–53, https://doi.org/10.1016/j.agrformet.2006.03.009, 2006.

Jaun, S., Ahrens, B., Walser, A., Ewen, T., and Schär, C.: A probabilistic view on the August 2005 floods in the upper Rhine catchment, Nat. Hazard. Earth Sys., 8, 281–291, https://doi.org/10.5194/nhess-8-281-2008, 2008.

20 Ji, F., Evans, J. P., Teng, J., Scorgie, Y., Argüeso, D., and Di Luca, A.: Evaluation of long-term precipitation and temperature Weather Research and Forecasting simulations for southeast Australia, Clim. Res., 67, 99-115, 2016.

Jones, D. A., Wang, W., and Fawcett, R.: High-quality spatial climate data-sets for Australia. Aust. Meteorol. Ocean., 58, 233–248, 2009.

Kent, D. M., Kirono, D. G. C., Timbal, B., and Chiew, F. H. S.: Representation of the Australian sub-tropical ridge in the
25 CMIP3 models. Intl. J. Climatol., 33, 48–57, https://dx.doi.org/10.1002/joc.3406, 2013.

Kirono, D. G. C., and Kent, D. M.: Assessment of rainfall and potential evaporation from global climate models and its implications for Australian regional drought projection. Intl. J. Climatol., 31, 1295–1308, https://dx.doi.org/10.1002/joc.2165, 2011.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in Combining Projections from Multiple Climate
30 Models. J. Climate, 23, 2739–2758, https://dx.doi.org/10.1175/2009JCLI3361.1, 2011.

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. Sc., 11, 1145–1159, https://doi.org/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
35 Panel on Climate Change AR4 models using equidistant quantile matching. J. Geophys. Res., 115, D10101, https://doi.org/10.1029/2009JD012882, 2010.

Li, X., Pijcke, G., and Babovic. V.: Analysis of Capabilities of Bias-corrected Precipitation Simulation from Ensemble of Downscaled GCMs in Reconstruction of Historical Wet and Dry Spell Characteristics, Procedia Engineering, 154, 631–638, https://doi.org/10.1016/j.proeng.2016.07.562, 2016.

40 Lockart, N., Willgoose, G., Kuczera, G., Kiem, A. S., Chowdhury, A., Manage, N. P., Zhang, L. Y., and Twomey, C.: Case study on the use of dynamically downscaled climate model data for assessing water security in the Lower Hunter region of the eastern seaboard of Australia, Journal of Southern Hemisphere Earth Systems Science, 66, 177–202, 2016.

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, J. Climate, 26, 2137–2143, https://doi.org/10.1175/JCLI-D-12-00821.1, 2013.

Maraun, D.: Bias Correcting Climate Change Simulations - a Critical Review, Current Climate Change Reports, 2, 211–220, https://doi.org/10.1007/s40641-016-0050-x, 2016.

5   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, https://doi.org/10.1029/2009RG000314, 2010.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutierrez, J. M., Hagemann, S., Richter, I., Soares, P. M.
10   M., Hall, A., and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nature Clim. Change, 7, 664–773, 10.1038/nclimate3418, 2017.

Matsoukas, C., Islam, S., and Rodriguez-Iturbe, I.: Detrended fluctuation analysis of rainfall and streamflow time series, J. Geophys. Res.-Atmos., 105, D23, 29165–29172, https://doi.org/10.1029/2000JD900419, 2000.

McMahon, T. A., Peel, M. C., and Karoly, D. J.: Assessment of precipitation and temperature data from CMIP3 global climate
15   models for hydrologic simulation. Hydrol. Earth Syst. Sc., 19, 361–377, https://dx.doi.org/10.5194/hess-19-361-2015, 2015.

Meehl, G. A., Covey, C., Delworth, T., Latif, M., Mcavaney, B., Mitchell, J. F. B., Stouffer, R. J., and Taylor, K.E.: The WCRP CMIP3 multimodel dataset - A new era in climate change research, B. Am. Meteorol. Soc., 88, 1383–1394, https://doi.org/10.1175/BAMS-88-9-1383, 2007.

Mehrotra, R., Johnson, F., and Sharma, A.: A software toolkit for correcting systematic biases in climate model simulations.
20   Environ. Modell. Softw., 104, 130–152, https://doi.org/10.1016/j.envsoft.2018.02.010, 2018.

Milly, P. C. D., and Wetherald, R. D.: Macroscale water fluxes 3. Effects of land processes on variability of monthly river discharge, Water Resour. Res., 38, 1235, https://doi.org/10.1029/2001WR000761, 2002.

Nakićenović, N., Alcamo, J., Davis, G., De Vries, H. J. M., Fenhann, J., Gaffin, S., Gregory, K., Grubler, A., Jung, T. Y., Kram, T., La Rovere, E. L., Michaelis, L., Mori, S., Morita, T., Pepper, W. J., Pitcher, H., Price, L., Riahi, K., Roehrl, A.,
25   Rogner, H. H., Sankovski, A., Schlesinger, M., Shukla, P., Smith, S. J., Swart, R. J., Van Rooijen, S., Victor, N., and Dadi, Z.: Special Report on Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change, Cambridge, Cambridge University Press, 2000.

Olson, R., Evans, J. P., Di Luca, A., and Argüeso, D.: The NARCliM project: model agreement and significance of climate projections, Clim. Res., 69, 209–227, https://doi.org/10.3354/cr01403, 2016.

30   Piani, C., and Haerter, J.O.: Two dimensional bias correction of temperature and precipitation copulas in climate models. Geophys. Res. Lett., 39, L20401, https://doi.org/10.1029/2012GL053839, 2012.

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, https://doi.org/10.1016/j.jhydrol.2010.10.024, 2010.

35   Post, D. A., Chiew, F. H. S., Teng, J., Wang, B., and Marvanek, S.: Projected changes in climate and runoff for south-eastern Australia under 1°C and 2°C of global warming. A SEACI Phase 2 special report, Commonwealth Scientific and Industrial Research Organisation, Australia, 2012.

Post, D. A., Timbal, B., Chiew, F. H. S., Hendon, H. H., Nguyen, H., and Moran, R.: Decrease in southeastern Australian water availability linked to ongoing Hadley cell expansion. Earth's Future, 2, 231–238,
40   https://doi.org/10.1002/2013EF000194, 2014.

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, https://doi.org/10.1029/2010WR010333, 2011.

Potter, N.J., Petheram, C., and Zhang, L.: Sensitivity of streamflow to rainfall and temperature in south-eastern Australia during the Millennium drought. In Chan, F., Marinova, D. and Anderssen, R.S. (eds) MODSIM2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2011, pp. 3636–3642. http://www.mssanz.org.au/modsim2011/I6/potter.pdf, 2011.

5 Potter, N. J., Chiew, F. H. S., Zheng, H., Ekström, M., and Zhang, L.: Hydroclimate projections for Victoria at 2040 and 2065. Commonwealth Scientific and Industrial Research Organisation, Australia, 2016.

Potter, N. J., Ekström, M., Chiew, F. H. S., Zhang, L., and Fu, G.: Change-signal impacts in downscaled data and its influence on hydroclimate projections, J. Hydrol., 564, 12–28, https://doi.org/10.1016/j.jhydrol.2018.06.018, 2018.

Power, S., Casey, T., Folland, C., Colman, A., and Mehta, V.: Inter-decadal modulation of the impact of ENSO on Australia. 10 Clim. Dynam., 15, 319–324, 1999.

Racherla, P. N., Shindell, D. T., and Faluvegi, G. S.: The added value to global model projections of climate change by dynamical downscaling: A case study over the continental U.S. using the GISS-ModelE2 and WRF models, J. Geophys. Res.-Atmos., 117, D20118, https://dx.doi/org/10.1029/2012JD018091, 2012.

Rajczak, J., Kotlarski, S., and Schär, C.: Does Quantile Mapping of Simulated Precipitation Correct for Biases in Transition 15 Probabilities and Spell Lengths? J. Climate, 29, 1605–1615, https://doi.org/10.1175/JCLI-D-15-0162.1, 2016.

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

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3. NCAR Technical Note NCAR/TN-475+STR, 20 https://dx.doi.org/10.5065/D68S4MVH, 2008.

Skamarock, W. C., and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, J. Comput. Phys., 227, 3465–3485. https://doi.org/10.1016/j.jcp.2007.01.037, 2008.

Smith, I., and Chandler, E.: Refining rainfall projections for the Murray Darling Basin of south-east Australia—the effect of sampling model results based on performance. Climatic Change, 102, 377–393, https://dx.doi.org/10.1007/s10584-009-9757-25 1, 2010.

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., and Evans, J. P.: How does bias correction of regional climate model precipitation affect modelled runoff? Hydrol. Earth Syst. Sc., 19, 711–728. https://doi.org/10.5194/hess-19-711-2015, 2015.

Terink, W., Hurkmans, R. T. W. L., Torfs, P. J. J. F., and Uijlenhoet, R.: Evaluation of a bias correction method applied to 30 downscaled precipitation and temperature reanalysis data for the Rhine basin, Hydrol. Earth Syst. Sc., 14, 687–703, https://doi.org/10.5194/hess-14-687-2010, 2010.

Tessier, Y., Lovejoy, S., Hubert, P., Schertzer, D., and Pecknold, S.: Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions, J. Geophys. Res.-Atmos., 101, 26427–26440, https://doi.org/10.1029/96JD01799, 1996.

35 Teutschbein, C., and Seibert, J.: Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? Hydrol. Earth Syst. Sc., 17, 5061–5077, https://doi.org/10.5194/hess-17-5061-2013, 2012.

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, https://doi.org/10.1002/joc.2168, 2011.

Themeßl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models 40 and its impact on the climate change signal, Climatic Change, 112, 449–468, https://doi.org/10.1007/s10584-011-0224-4, 2012.

von Storch, H.: On the Use of ''Inflation'' in Statistical Downscaling. J. Climate, 12, 3505–3506, https://doi.org/10.1175/1520-0442(1999)012<3505:OTUOII>2.0.CO;2, 1999

16

Wang, L., and Chen, W.: Equiratio cumulative distribution function matching as an improvement to the equidistant approach in bias correction of precipitation. Atmos. Sci. Lett., 15, 1–6, https://doi.org/10.1002/asl2.454, 2014.

Wilcke, R.A.I., Mendlik, T., Gobiet, A.: Multi-variable error correction of regional climate models. Clim. Change, 120, 871–887, https://doi.org/10.1007/s10584-013-0845-x, 2013.

5   Wolock, D. M., and McCabe, G. J.: Explaining spatial variability in mean annual runoff in the conterminous United States, Clim. Res., 11, 149–159, https://doi.org/10.3354/cr011149, 1999.

Yang, W., Andréasson, 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, https://doi.org/10.2166/nh.2010.004, 2010.

10
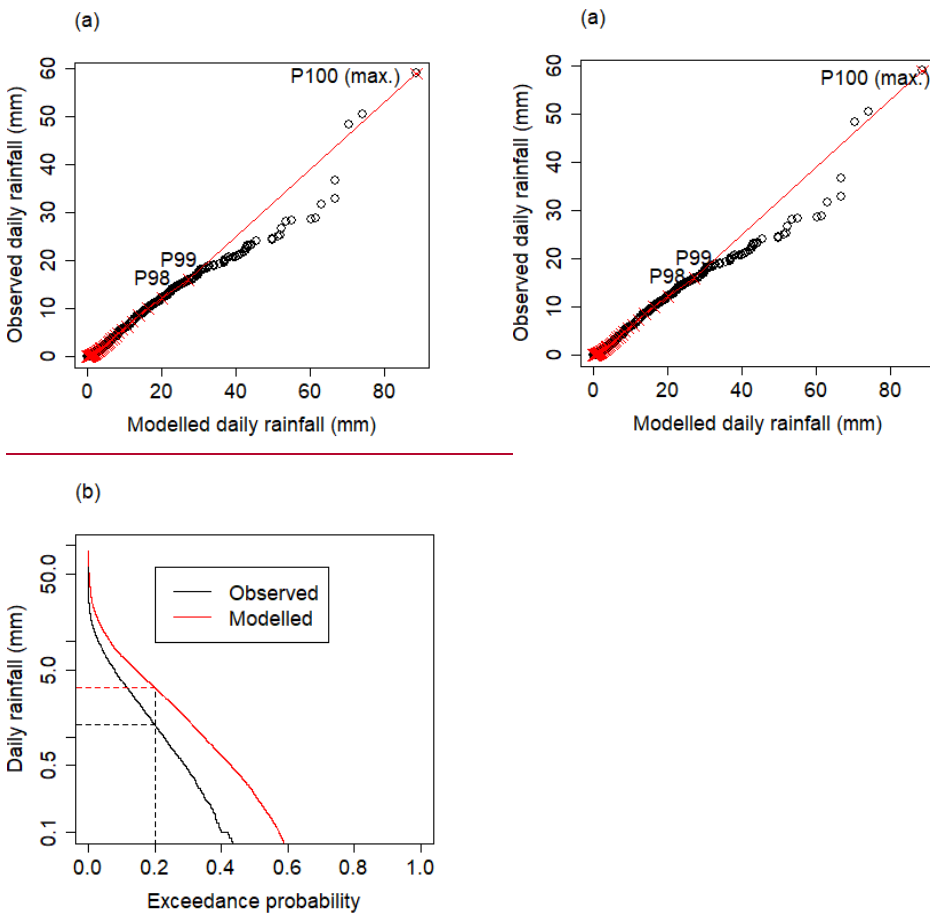
Figure 1: Schematic of QQM bias correction. (a) Empirical cumulative density functions for both observed and modelled rainfall in a given gridcell. Percentiles are estimated from both distributions, which are equated in bias correction to generate a mapping function (b). Values lying between percentiles or outside the modelled maximum value are interpolated or extrapolated linearly.

17

(a) Observed **(b) WRF downscaled from reanalysis, R1** **(c) WRF downscaled from GCM (ECHAM5, R1)** **(d) WRF downscaled from ECHAM5, R1 (future)**

(e) **(f) Bias/Error** **(g) Bias/Error** **(h) Change (future relative to historical)**

-600 -400 -300 -200 -100 -50 0 50 100 200 300 400 600 1000 1500 2000

(a) Observed **(b) WRF downscaled from reanalysis, R1** **(c) WRF downscaled from GCM (ECHAM5, R1)** **(d) WRF downscaled from ECHAM5, R1 (future)**

(e) **(f) Bias/Error** **(g) Bias/Error** **(h) Change (future relative to historical)**

-600 -400 -300 -200 -100 -50 0 50 100 200 300 400 600 1000 1500 2000 mm/yr

5   **Figure 2: Regional mean annual rainfall (mm yr^{-1}). Inset map (e) shows the location of the State of Victoria in Australia. Other panels show: (a) observed (AWAP) rainfall; (b) rainfall downscaled from reanalysis (NCEP/NCAR); (c) historical rainfall downscaled from median GCM (ECHAM5/R1); (d) future rainfall downscaled from GCM; (f) bias in reanalysis downscaling (compared to observed); (g) bias in GCM downscaling; (h) change factor of GCM downscaled rainfall.**
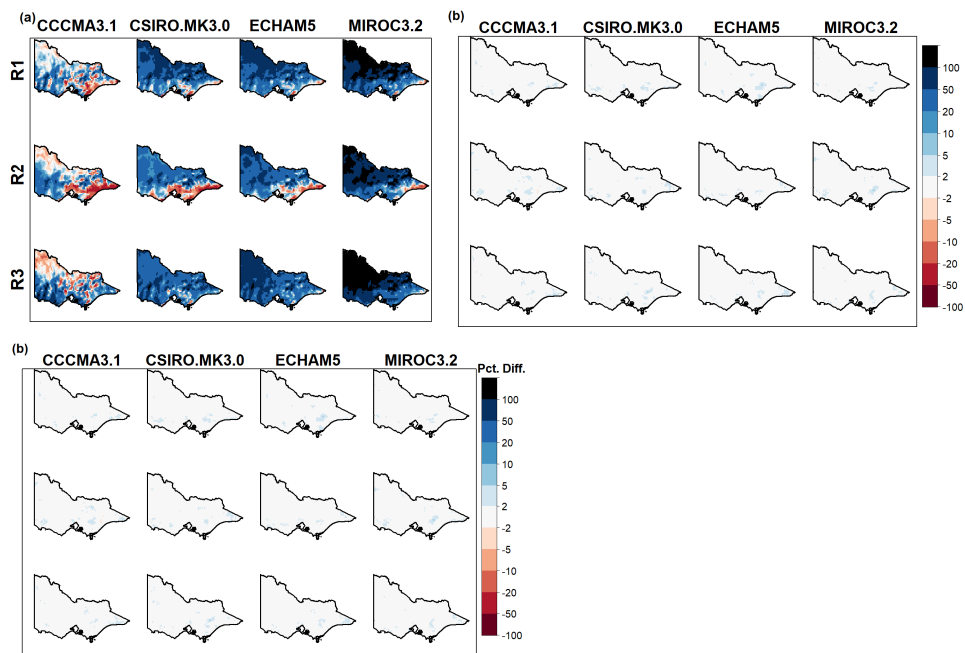
Figure 3: Percentage bias in WRF-downscaled mean annual rainfall from the GCM/RCMs indicated: (a) raw data; (b) residual bias (after bias correction).
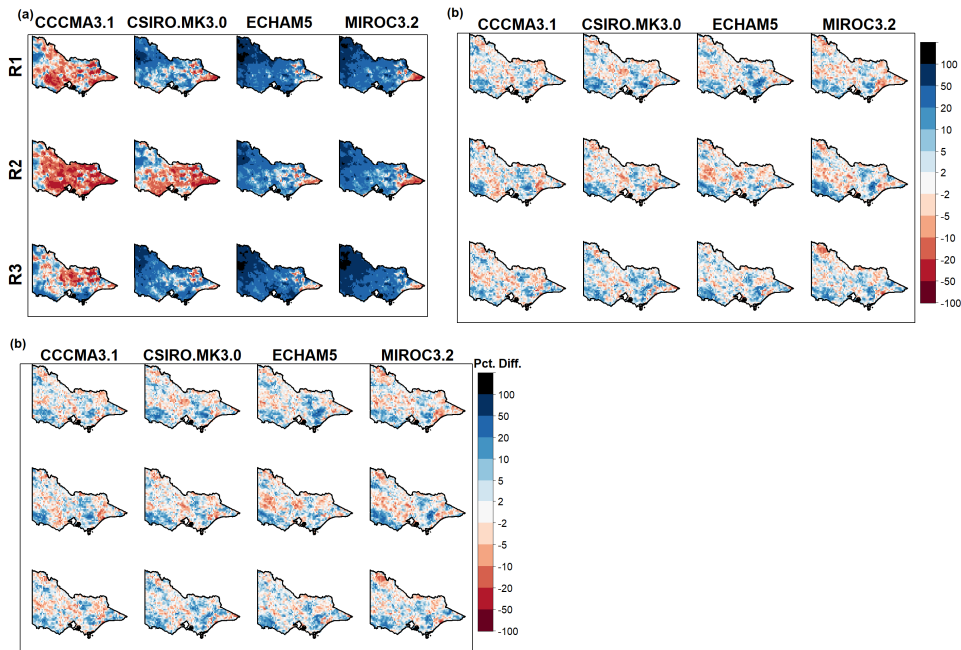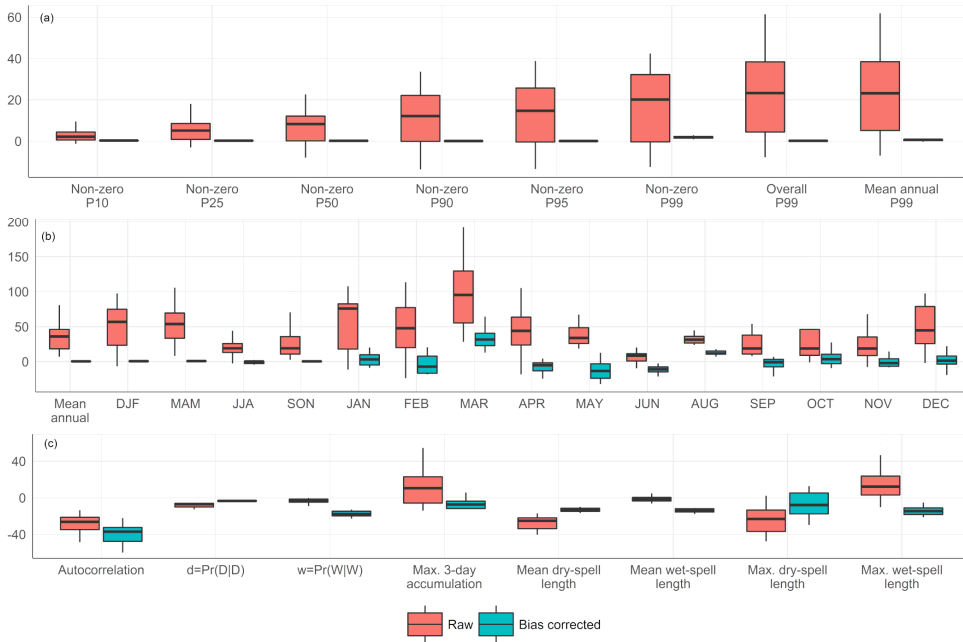
Figure 4: Percentage bias in WRF-downscaled 99th percentile of rainfall (P99) from the GCM/RCMs indicated: (a) raw data; (b) residual bias (after bias correction).
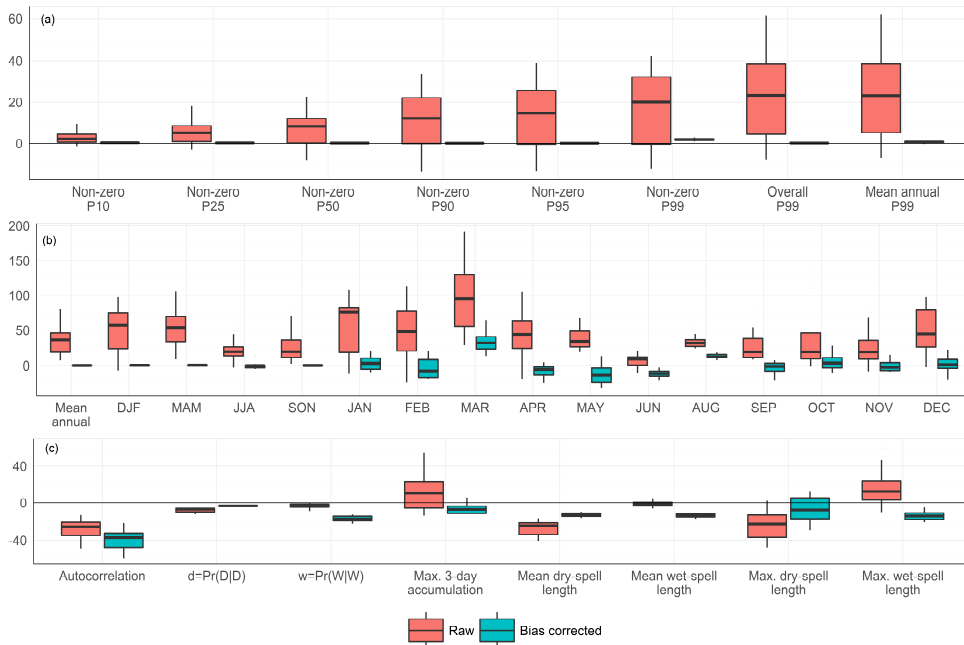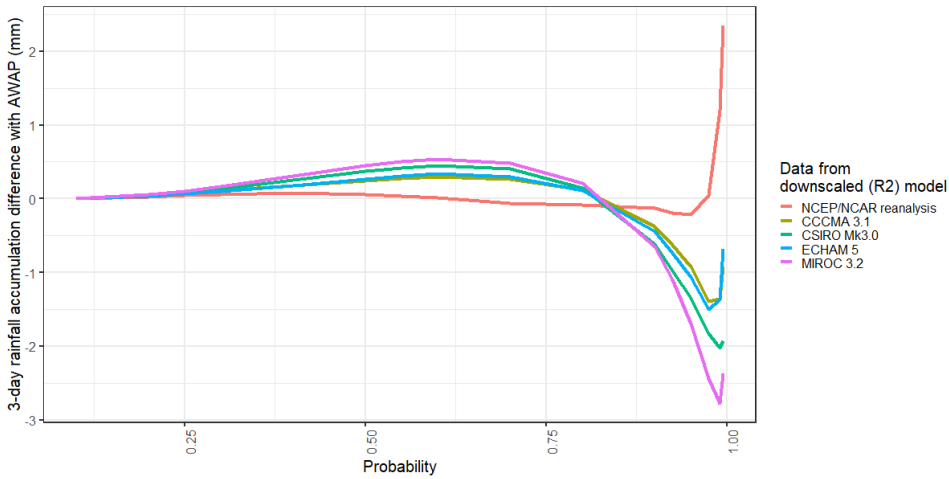
Figure 5: Relative bias (modelled compared to AWAP, as percentages) of: (a) rainfall percentiles; (b) mean annual, seasonal and monthly rainfall; (c) rainfall-sequencing metrics both before and after bias correction. The range of results represents the spread of downscaled GCM hindcast spatial averages over Victoria from the NARCliM 12-model ensemble

5

Figure 6: Biases in spatial average 3-day rainfall accumulation percentiles. Here a 3-day moving average filter was applied to each hindcast timeseries and equivalent quantiles taken at increasing probability values (x-axis).

10

(a) Runoff from AWAP (mm) **(b) Bias in median runoff (mm)** **(c) Bias in median runoff (%)**
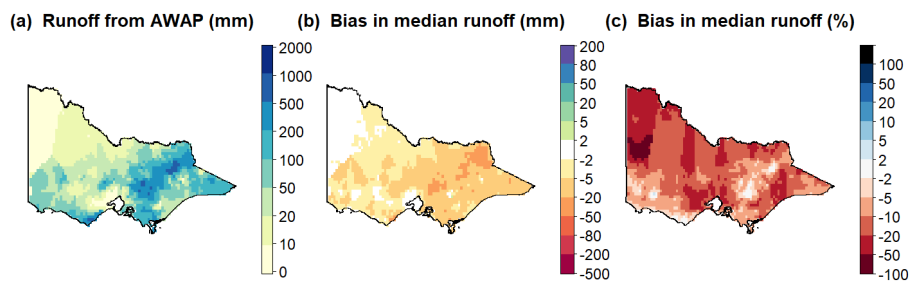
Figure 7: Ensemble (downscaled GCM hindcasts from the NARCliM 12-model ensemble) median modelled runoff over Victoria: (a) from AWAP historical observations; (b) absolute bias (mm); (c) percentage bias (%).
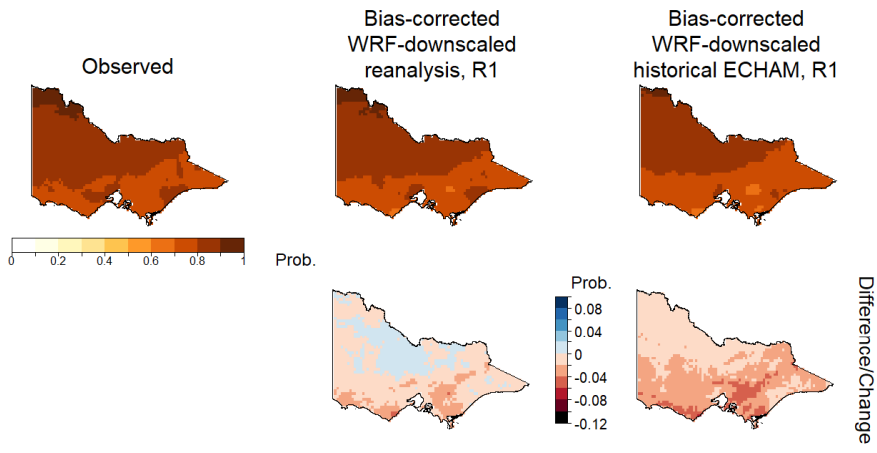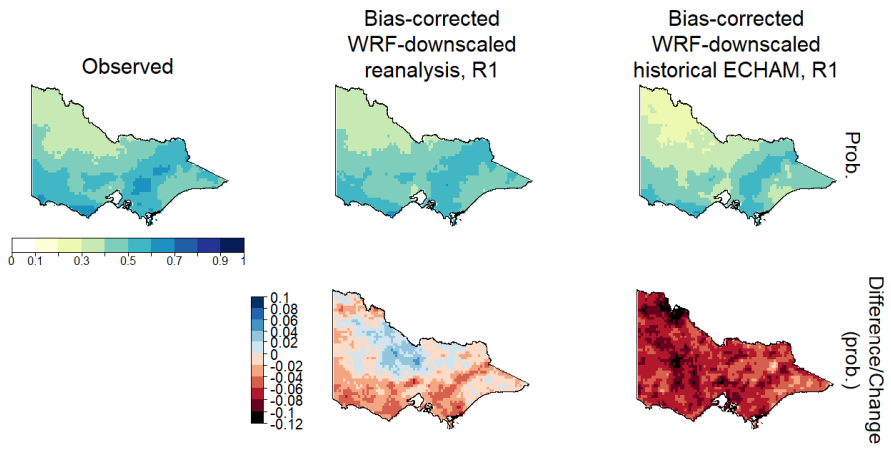
Observed

Bias-corrected
WRF-downscaled
reanalysis, R1

Bias-corrected
WRF-downscaled
historical ECHAM, R1

Prob.

0 0.1 0.3 0.5 0.7 0.9 1

0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12

Difference/Change
(prob.)
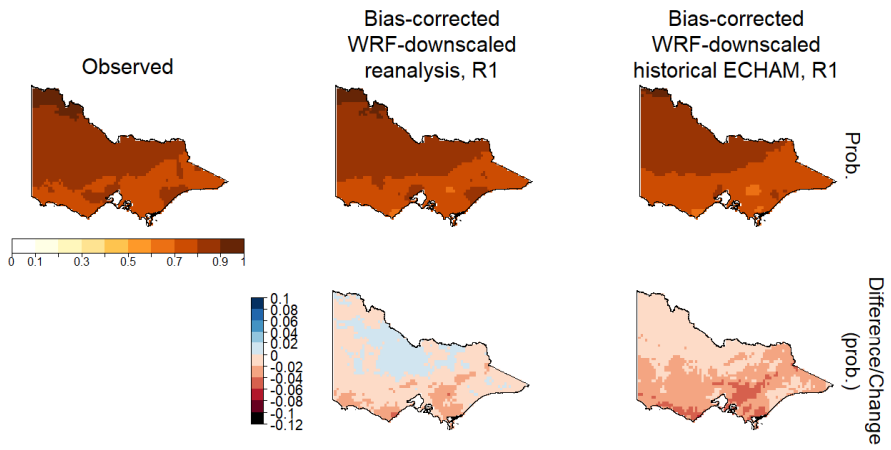
Figure 8: Dry-dry transition probabilities (1mm threshold)

Observed

Bias-corrected
WRF-downscaled
reanalysis, R1

Bias-corrected
WRF-downscaled
historical ECHAM, R1

Prob.

0   0.1   0.3   0.5   0.7   0.9   1

0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12

Difference/Change
(prob.)
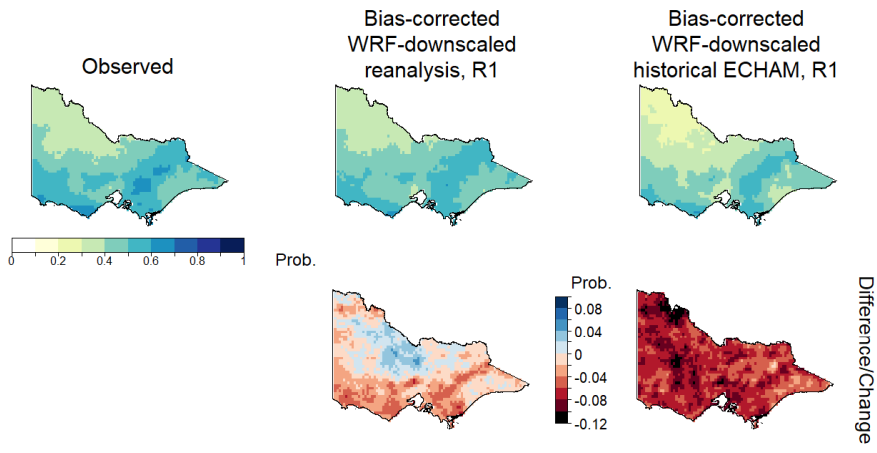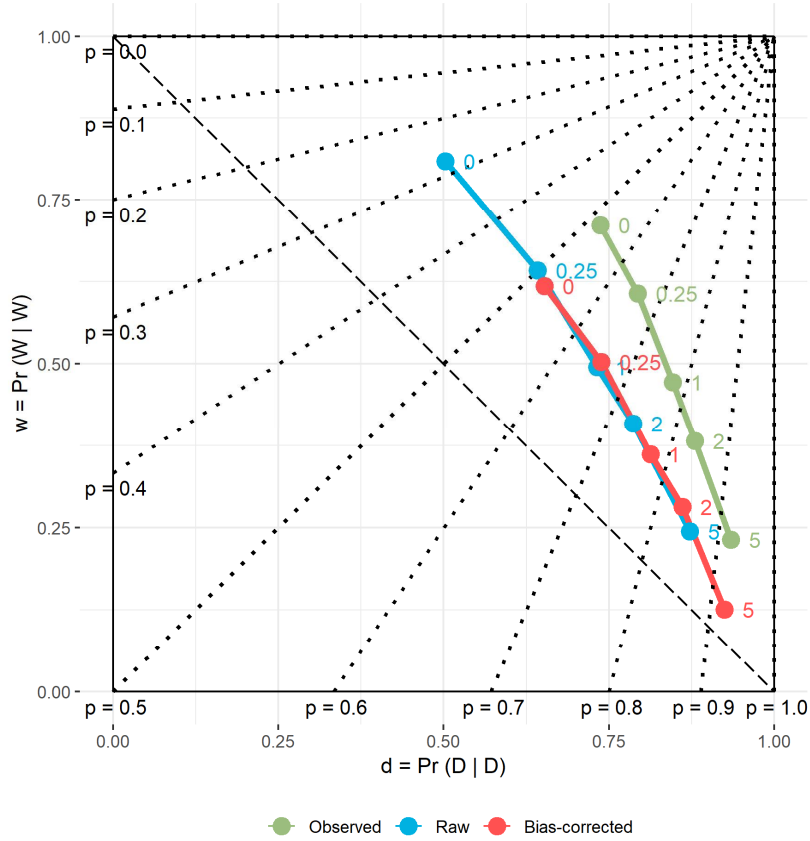
Figure 9: Wet-wet transition probabilities (1mm threshold)
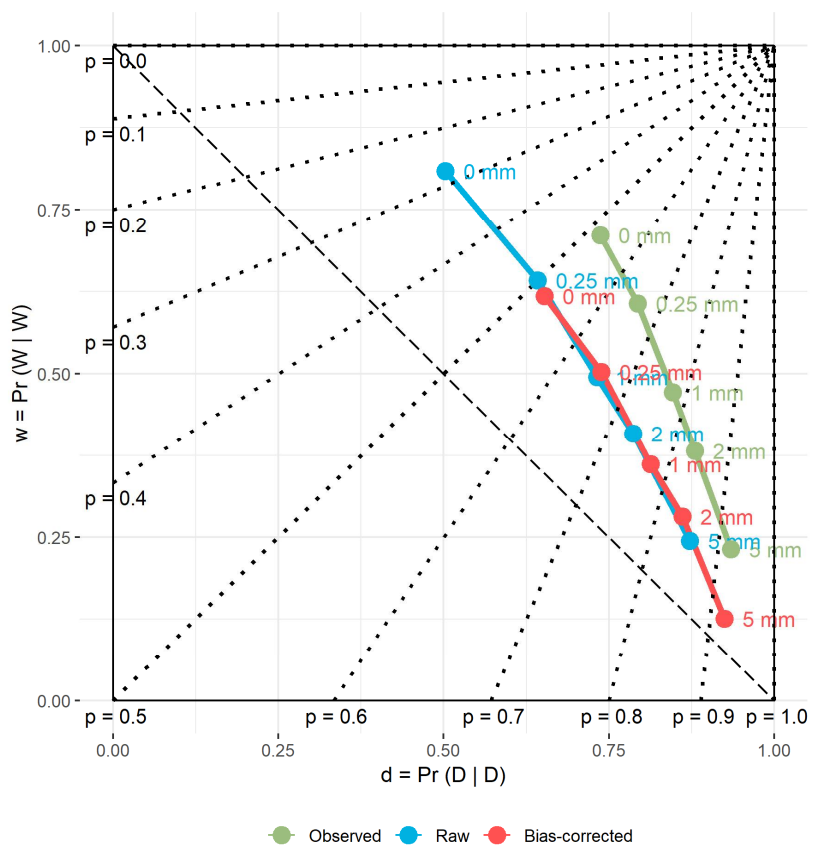
Figure 10: An alternative perspective on quantile-quantile mapping: daily rainfall amounts and associated probabilities plotted in *d-w* space. Quantile-quantile mapping bias correction (red) maps daily rainfall amounts from the raw data (blue) to the probability contours (dotted lines) corresponding to the appropriate observed daily amount (green). The dashed diagonal line represents *p=d* and hence an independent series of events (see sect. 2.4). Points lying above or to the right represent quantiles with greater (more persistent) autocorrelation.
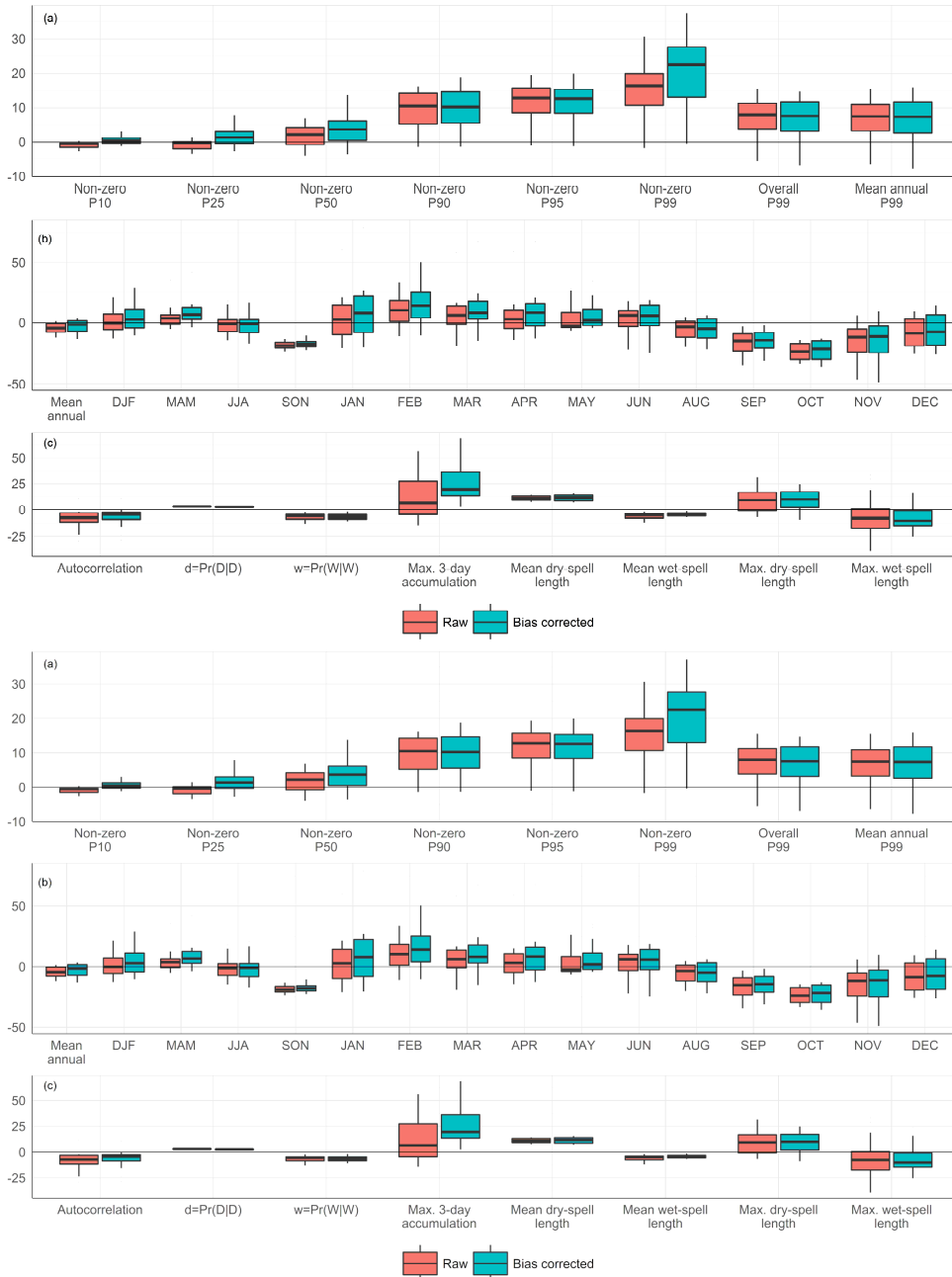
Figure 11: Change signal (percentage difference of RCM future relative to RCM historical) in (a) rainfall percentiles; (b) mean annual, seasonal and monthly rainfall; (c) rainfall-sequencing metrics both before and after bias correction.

29