Authors' Response and Marked-up Manuscript

Reviewer #2

The revised version of the paper now includes some of the proposed changes from all reviewers. It has definitely improved, also in terms of reducing the number of figures, so that a clearer structure emerged.
5 Nevertheless, in my opinion there are still some open issues – in particular concerning the expectations on bias correction (BC) / hypothesis of the analysis, model/data description, and method description. These issues are either still unclear or insufficiently presented, but are directly relevant to the research question of the paper: what is the effect of BC on rainfall characteristics relevant for runoff modelling? Therefore, I still suggest that major revisions are required.

10 *Thank you for your insightful review and we have revised the manuscript accordingly as described in our response that follows.*

In the following, I will outline these issues.
Major comments
- The Introduction still lacks some relevant aspects. Most of all, the authors should state what a BC is
15 meant to correct (amount of precipitation), what cannot be corrected (e.g. temporal errors, such as persistence of transition probabilities), and what is known to be affected by BC (change signals, persistence in some studies) and whether this effect is intended. Based on this literature review (which is still scattered over the whole text), hypotheses/expectations should be formulated that guide the analysis. For instance, it could be expected that a BC would remove the systematic errors in rainfall
20 amounts, but would not correct the transition probabilities or 3-day precipitation totals. Since other studies found that BC affects persistence and transition probabilities, this should be a motivation for the study and for introducing the new diagnostic presented by the authors.

We add "Most bias-correction methods alter daily amounts with the application of distributional mappings. The temporal structure of the occurrences is most often unaltered.
25 *Further, bias correction can and does affect the magnitude of change signals (Hagemann et al., 2011; Gutjahr and Heinemann, 2013; Dosio, 2016)." Before "The underlying assumption…" on p.2, also adding "Given that QQM bias-correction approaches are specifically tailored for correcting daily amounts while retaining the existing sequencing of occurrences, we seek to understand the effect of bias correction on rainfall sequencing and*
30 *transition probabilities." in the final paragraph of the introduction.*

- As another motivation, previous results with the here used WRF R1-R3 simulations should be summarized already in the Introduction. Obvious would be the results of Jin et al. (2016), who evaluated rainfall (and temperature) of the same WRF simulations on climate time scale. Jin et al. even concluded that a bias correction is required, which renders a natural motivation for this study. I further
35 noticed that the WRF simulations presented here actually stem from a double nesting (50km CORDEX Australasia domain and the inner 10km NARCliM domain, Fig.1 in Jin et al.). It would be good to write this in the model description – maybe also mention that it is the same simulations as in Jin et al.

We included Ji et al., 2016 in the discussion, but also including in the introductory sections is a good idea, this information has been included in section 2.1 and 2.2, which discusses the
40 *NARCliM data.*

- Other important evaluation results of the here used WRF simulations, such as presented on p. 9, L. 6ff of the Discussion, should be presented already in the Introduction or model description. Thereby, note that Gilmore et al (2016) only evaluated 15-day simulations focusing on a specific extreme precipitation event; thus, it is not guaranteed that WRF behaves similar at the daily scale for other synoptic situations. These evaluations should then constitute the basis from where you start your analysis. Of particular importance is the information that WRF R2 renders the best configuration for hydrological applications (p.9, L.13-15) (why?) with respect to Olson et al. (2016). This information has to go into the Introduction, or into the model/data description. A natural question arises then whether your analysis support their conclusion? Finally, make clear how the models have been selected (e.g. as it is done in the Introduction of Jin et al., 2016) and present this information in the model/data description. Only buried in the text (p.6, L.31), you state that WRF selection was based on credible simulations of 2-week heavy-precipitation periods. Do you then expect that these configurations faithfully represent other rainfall metrics, e.g. transition probabilities? If for instance transition probability/persistence of rainfall was not a selection criteria, but is important for hydrological applications, then the a-priori expectation should be that BC would not help and a possible conclusion from your results could be e.g. that hydrologically relevant metrics should be added to a GCM/RCM selection process (based on CMIP5 etc.).

- In the discussion (p.8,L.28ff) you state that WRF has relatively large rainfall biases. You then conclude that there are model errors whose origin needs to be analysed and whether they render the physics implausible. In contrast, on p.8, L.37, you write these WRF simulations have been extensively tested,

2

and on p.9, L.6f that there is reasonable confidence in these simulations, although they have a general cold and wet bias. So, I guess the origin of these errors is understood to some degree and was already attributed in other studies?

The rainfall amounts are acknowledged to be significantly biased, but the spatial occurrences, seasonal and interannual dynamics are recognised to be reasonably well modelled for Australia (as per the provided citations). The precise origin of the biases is suggested to include problems with subgrid cloud cover parameterisation (García-Díez et al., 2015; Di Virgilio et al., 2019), and is the subject of ongoing research. For our purposes we contend that the WRF modelling ensemble provides climatically and physically realistic representations of the study area's climate, albeit with a bias that needs to be corrected.

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, 2015, https://doi.org/10.1007/s00382-015-2529-x

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

We add "It is suggested that the wet 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)." to the second paragraph of the discussion (p.9)

- On p. 8, L.2f you raise the need for bias correction methods that correct the occurrences of rainfall events – in other words, the temporal structure. Once again, temporal errors indicate fundamental model errors that cannot be corrected by BC methods (e.g. Maraun et al., 2016; Maraun and Widmann, 2018). They are directly linked to errors in the dynamics of the climate model, which could result from many different sources, e.g. from coarse topography that affects the dynamics in a wrong way or missing relevant regional processes. BC is not meant to correct these errors. In contrast, on p.2, L.24f you state: "The underlying assumption of bias correction is that the RCM output faithfully represents climate processes relevant for rainfall, although the amounts themselves may not be accurate"; and on p.3, L. 15ff: "QQM bias correction cannot remove biases in rainfall sequencing …". This implicitly implies that the temporal aspect (persistence, transition probabilities, etc.) is assumed to be realistically simulated by the RCM and should thus be part of the model evaluation procedure prior to any BC. Any attempt to correct temporal errors of the model, however, could introduce unwanted major artefacts.

We agree with these sentiments. In an ideal world the models would be faithfully reproducing the correlation structure of the rainfall fields. However, this appears to not necessarily be the case, and we seek to mitigate this. We approach the problem from a practical point of view, in terms of analysing and hopefully correcting data that is available for hydrological projections applications. Our research questions are: i) what climate data is available for hydrological applications? ii) what are the benefits and limitations of using this data? and iii) how to properly bias correct if necessary.
We are not aware of other studies that have analysed and evaluated temporal structure of rainfall (at least in the study area), and indeed all the bias correction and projection research we have reviewed have (implicitly) assumed that the correlations are correct. Our research findings will hopefully lead to better appreciation of the necessity of evaluating correlation

structure and/or methods (and their limitations) for dealing with this problem for future hydroclimate projections.

- A motivation is missing why the authors repeat the bias correction with an empirical QQM, when there already is a corrected rainfall set using a double-gamma QQM (p.4, L.14f) that is provided by Evans and Argüeso (2014). Is there any reason not to use this data set? At least, provide a statement why this already existing data set was not used.

In assessing the effect of bias correction, we wished to apply a controllable method in order to understand fully the method and its effects, rather than relying on previous methods. In this way we have better understanding of the way bias correction is applied and its flow-on effects. Since Teng et al. showed there is little difference between methods, we have no reason to believe that this has any impact on the findings of the paper.

- The description of the models is still somewhat confusing. Please clearly refer to WRF simulations instead of either "bias-corrected reanalysis" or "bias-corrected GCM rainfall" (e.g. p.7, L.10f). Otherwise, it is very confusing (also in the figure labels), and it is just wrong because reanalysis/GCMs were downscaled with WRF and not directly bias corrected.

This has been corrected in the line indicated and throughout, as well as the headings in figure 8 and 9.

- I further recommend to revise section 2. First, all models that are used in the paper should be named here, including the runoff model GR4J that is first introduced in the results section (section 3.2 on p. 6, L.35). Then provide a short description of the WRF model with exact model version, model resolution, number of vertical levels, etc., and in particular briefly mention the differences of the WRF R1 – R3 configurations. It is still unclear to the reader whether R1 - R3 are different realisations based on slightly different initial fields or on different physics. Part of that information (e.g. model resolution, time period, etc.) is provided in section "2.2 Daily data", but that is not the correct place (maybe rename 2.2 to "Observations"). Furthermore, on p.4, L.6f, be more precise in describing that you use these three (R1-R3) WRF configurations for each of the four chosen GCMs, resulting in 12 simulations in total. One could misread p.4, L.6f that you downscale only three GCMs and that these downscaled WRF simulations are labelled R1-R3. Mention also that these result from double-nesting into the GCM/NCEP: 50km CORDEX-Australasia domain and 10km NARCliM domain. In section 2.2 only NCEP is stated as the reanalysis data used for this study, but in Fig. 8 & 9 you show results from WRF nested into ERA-Interim?

As mentioned elsewhere, we prefer not to describe GR4J in too much detail given that it is not the main focus of the paper and is described elsewhere (citation given).

Increased detail of the WRF model version, resolution, and citation to specifics of R1-R3 physics configurations is provided in section 2.1

Included "The full NARCliM model ensemble thus consists of twelve members (four selected GCMs, each downscaled by the three RCM configurations)." in order to precisely indicate the size of the model ensemble.

4

Double-nesting of the modelling domains has been included (see response above).

Finally, ERA-I was also downscaled by the NARCliM project at a later date; as mentioned below, we have updated figures 8 and 9 using NCEP/NCAR reanalysis to be consistent with the other figure. Thus there is no need to include information on ERA-I in the present manuscript.

- The description of the BC method still lacks some critical information. Although a reference to Teng et al. (2015) was added to the Introduction (but not to section 2.2?), it is still not clear how exactly the QQM was constructed for the here presented WRF simulations. Teng et al. (2015) use different time periods, different observations to construct the QQM, and different WRF simulations. Apart from the different distribution, is the QQM method identical to the QQM of Evans and Argüeso (2014), as suggested on p.4, L.12 ("… which is similar in many respects to the non-parametric procedure we apply …")? Did you use the full historical period (1990-2009) to construct QQM and then applied it to the complete period? Or did you split it into calibration (dependent) and validation (independent data) subsets and then apply the so constructed QQM to the complete period? If you did the first without validating QQM with independent data, you cannot judge how strong overfitting is. If the complete historical period was used to construct the transfer functions, then a BC always reduces the mean error to zero (neglecting linear interpolation from 'qmap'), but the overfitting only takes effect on independent data (future time period data). Thus, overfitting could be large depending on how the QQM was constructed. Without such information, it is not possible to repeat the method/analysis in combination with these WRF simulations.

Updated information is included in section 2.2 (citation to Teng) and section 2.3 (specifically that the full historical period was used to calibrate the transfer functions). We agree that with this setup it is not possible to judge overfitting, hence why bias in rainfall amount metrics (Figure 3 and 5) is effectively reduced to zero. We have added a parenthetical sentence acknowledging this when discussing figure 3 in section 3.2.

Even with zero calibration-validation error from applying BC to the same period, we see errors in transition probabilities (hence the motivation of this manuscript). Teng et al. looked specifically at the cross-validation error using two time different historical periods, and our manuscript effectively complements Teng's analysis by considering error/bias in secondary rainfall characteristics separately. We agree that in applying the WRF-BC method for end-user applications, proper calibration-validation methods should be applied to correctly analyse the error.

- Given these above comments, I think part of the conclusion is already known a-priori by considering existing literature. One conclusion, that the change signals are much smaller than the bias should include how statistically significant the change signals are in relation to annual variability. Assume, for instance, that annual variability is larger than the change signal of mean precipitation. Would such a small change then be relevant to be discussed at all?

We address this comment below.

Minor / Specific Comments

- Some of the figures lack units on the y-axis (e.g. Fig. 5) and the legend is not correct (e.g. Fig.8/9), for instance "B/C reanalysis (ERA-I, R2)" suggests that ERA-Interim was bias corrected. Correct the legends and mention WRF, e.g. "B/C WRF reanalysis (ERA-I, R2)".

---

y-axis in figure 5 is unitless as these are relative errors (percentages): include "as percentages" in caption to figure 5.

Titles on figures 8 and 9 have been updated as suggested

- Fig.2: (a) and (b) are exactly the same plots?

Thank you for noticing, this error crept in while preparing the figures for the revision, and has been redrawn as per the original version

---

- Fig.2: add "GCM (ECHAM5,R1)" to the caption (d). And mention which configuration (R1-R3) was used for (b).

Figure 2 has been updated as suggested.

- Fig.3 & 4: add "WRF" to the figure caption, or to the y-axis label. Otherwise, it could be misread as GCM corrected output and not corrected WRF output.

Captions updated

- It would be clearer to label every subplot with (a), (b), and so on.

Each subplot is not discussed separately and we prefer to keep the labelling as is due to the number of panels in figures 3 and 4.

- Fig.8 & 9: I think the bottom row can be deleted. The changes in transition probability are already shown in the middle row. The percentage change of the probability change does not add to it.

We agree, the bottom row has been deleted, and reanalysis has been updated to be consistent with figure 2.

- Fig.8 & 9: why show here WRF results from downscaling ERA-Interim, whereas in Fig.2 you show results from WRF downscaling NCEP?

Figures 8 and 9 have been updated to be consistent with figure 2, i.e. NCEP/NCAR reanalysis

---

6

- p.1, L.37: also mention statistical downscaling approach here?.

A small sentence is included mentioning statistical downscaling

- p.2, L.5: delete "generates rainfall sequences by simulating physical climatic processes" and replace "does generate" by "generates".

Done, thank you

- p.2, L.34: not only hindcast RCM rainfall, but also historical or in principle any data that is used for calibrating QQM.

"Modelled historical rainfall (in this case" inserted before "hindcast RCM rainfall"

- p.3, L.4: Teng et al. were not the first to use a double-gamma QQM, see for instance Yang et al. (2010, doi: 10.2166/nh.2010.004).

Citation added

- p.3, L.12: do other studies agree with that? Otherwise, I suggest to add "for Australia".

Added ", at least in the Australian context studied by Teng…"

- p.4, L.8-10: This sentence could be deleted as it is not of direct relevance to the study. If you want to keep this information, then move it to the end of the discussion section (or conclusion depending on what you hope to gain from it).

We believe the fact that ongoing research is providing better and updated products is still relevant to the underlying issues of bias correction and future projections, and as such we have moved the indicated sentence to the 2nd paragraph of the discussion.

- p.5, L.37: Here and maybe on other locations you refer to "NARCliM rainfall" which is not correct, because NARCliM is no climate model. Better refer to e.g. "12-member WRF ensemble rainfall" or to WRF ("driving model", RX).

We agree that NARCliM is not a climate model, however we argue that "NARCliM rainfall" is meaningful since it refers to an ensemble from the NARCliM project. However, we agree that it may lead to confusion so we include the following: "Daily accumulated precipitation

7

from the NARCliM 12-member WRF-downscaled ensemble (referred to from here onwards as "NARCliM rainfall") is produced…" in section 2.2.

- p.6, L.9: I think you refer to Figure 5 here (not 6).

Corrected, thank you

5 • p.6, L18: This also depends on the calibration period that was used (see main comments above).

We agree with this comment as noted in our response to the major comment above. We acknowledge this by adding "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)…" in section 3.2.

10

- p.6, L.28ff: Is there any explanation for that and would it be sensitive if another BC method is used?

There is no particular explanation suggested here, and it would depend presumably on the correlation structure of the RCM rainfall. We suggest that it is unlikely to be different if another (daily) BC method is used, since the corrections applied from different daily BC methods is similar as noted in the introduction (see also Teng et al., 2015).

- p.6, L.31: how the WRF model/configuration was selected belongs to the data & methods section.

We agree, however the point mentioned in the manuscript is relevant in the results section as calibration to the largest historical rain events would tend to reduce the bias at very large (historical) rainfall accumulations

- p.6, L.34ff: Information on GR4J output belongs into the data & methods section.

We agree on this as well; runoff modelling is described in the companion paper that we cite, and so we reduce this to "runoff was modelled… using GR4J (see Charles et al., 2019 for more details)." Since runoff is not the key focus of the current manuscript, we decided not to describe the runoff modelling in greater detail here or in the introduction.

- p.7, L.2-4: This information and discussion belongs into the Introduction (see main comments).

Agreed, and this has been moved to the relevant paragraph in the introduction

- p.7, L.6: add the information that "more bias" means in this case shorter wet spell durations. Then the next sentence is more meaningful.

Added "(i.e. they are closer to zero)"

- p.7, L.6: you might want to refer to Fig.5c here again.

Cross reference added

- p.7, L.14: Do you refer to the middle row of Fig. 8 here?

Yes, and this reference has been included

- p.7, L.21: you mean WRF rainfall transition probabilities?

9

- p.7, L.21: somewhat unclear to what subplot of Fig. 9 you refer --> insert (a), (b), (c),... to the subplots?. From the sublots in the middle row I read a difference of about -0.02 to -0.04 for WRF (ERA-I,R2) and -0.04 to -0.1 for WRF (ECHAM5, R2). That is -2% to -4% and -4% to -10%. I do not see "over 10% over almost all of Victoria" from the plots. Or do you refer to the bottom right subplot? But, as I understand this figure, what is shown here is %-change of the difference. That means if the observed wet-wet prob. is e.g. 0.8, the difference in WRF is -0.08, then the %-change of this difference is -10%. But the error in wet-wet prob. is -8% and not -10%, isn't it? I think the bottom row plots are misleading, and I recommend to remove them. The information you want is already in the subplots of the middle row, isn't?

  We agree, and the bottom rows of fig. 8 and fig. 9 have been removed; on reflection percentage changes of probabilities are unclear and misleading.

- p.7, L.41: I suggest to add this line to the legend in Fig.10 and also to explain in the caption that a position right to it means higher serial correlation.

  Fig. 10 caption updated to include "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."
  The relevant sentence(s) on p.7 have been emended to reflect the revised caption.

- p.8, L.2f: a bias correction is not intended to correct rainfall occurrences (see my major comment).

  We definitely agree with the reviewer on this. However, bias correction is ultimately intended to provide physically plausible rainfall, of which rainfall occurrences are then indirectly assumed to be correct, which as we have demonstrated is not necessarily true. We alter "We surmise that, as bias correction is ultimately intended to produce physically plausible rainfall, bias correction methods…"

- p.8, L.8f: Yes, you can compare the percentiles, as they are statistical measures of a distribution.

  Yes, they can be compared directly, but we argue that it is not meaningful to do so since the underlying amount (which is the real quantity of interest) is different. "compared directly" changed to "meaningfully compared"

- p.8, L.16f: Is that a result of the seasonally applied BC? In what way is that problematic? Does not need to be a long discussion here, but briefly outline the problem associated to changed mean precipitation after BC. Would this change in mean be sensitive to a different BC method (e.g. monthly or annual)?

  The problem comes from deciding which is the "true" change signal in relevant quantities. Are raw change signals more realistic than bias corrected change signals? How can we tell?

10

A "small" change in future projections can in theory have large real-world consequences in terms of water planning and decision making. There is considerable debate on this in the literature, and we provide a cross reference to the discussion section where this is alluded to:

"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)."

- p.8, L.24ff: are the found changes signals statistically significant in relation to annual variability?

Ordinarily we would report significance for any changes discussed in the manuscript. However, since the focus of this manuscript is change signals in relation to bias and bias correction, rather than in relation to annual variability, we prefer not to discuss this directly, however, Olson et al. (2016) analysed the statistical significance specifically, and we include a citation in section 3.4: "... SON rain which has a decrease projected by the NARCliM ensemble of around 20% (found to be statistically significant by Olson et al., 2016)."

- p.9, L.12-13: Does this affect the errors in the raw transition probabilities? (see major comment)

STR is a major determinant of the amount and timing of the seasonal rainfall progression through the year in SE Australia, and is important for correctly modelling climate in SE Australia. We do not expect that modelling this well will have any particular impact on transition probabilities.

- p.9, L.14: unclear what is meant by "land and atmospheric circulation schemes". There is no atmospheric circulation scheme, you can either change the dynamical core or the physics, both of them then affect the circulation. Be more precise what is meant here.

This section has been moved to the introduction in response to other comments, and "land and atmospheric circulation schemes" has been deleted. Since it is now mentioned directly in relation to the discussion on R1-R3 this should now be more clear.

- p.9, L16: First, the information that WRF R2 is the most credible configuration in terms of runoff modelling belongs to the data & methods section.

These sentences have been moved to section 2.1 and corrected

- p. 10, L. 26: "multi-".

Corrected to "multi-decadal"

- p. 10, L. 34f: again, BC cannot correct temporal errors.

11

- p. 10, L. 39: what is completely unmentioned is how statistically significant the change signals are in relation to annual variability.

> *Kindly see our response to the similar comment above.*

- p. 10, L. 40: replace "dynamical downscaling process" with "RCM".

> *Replaced*

Reviewer #3

The authors missed the point regarding bias with respect to spatial correlation or persistence. They discuss the "inflation" issue as raised by Maruan (2013). This is regarding the inflation of the subgrid aggregated precipitation amount. This stems from applying the same quantity from a lower-resolution RCM/GCM to all of the points or grids that fall within each low-res grid cell. So, for example, when the model simulates 100 mm for a day, this amount is similarly bias corrected for all subgrid cells contained within the larger model grid cell. This often will lead to inflating the area-average, because the subgrid cell variability is not well represented.

Entirely separate, but equally important, is the issue of modeled versus observed spatial correlation (or what the authors refer to as spatial structure) when they are on the same grid (Bardossy and Pegram, 2012). Consider an example where both the model and observations are on the same 10x10 grid. Also, consider the model to have, on average, larger spatial correlation (correlation in time between each grid cell, this would be a 100x100 correlation matrix) than observations (i.e., when it rains at one grid cell, it is much more likely to rain at an adjacent grid cell than what observations show). This effect will translate to inflating the area-averaged precipitation amounts of the bias-corrected modeled data. This is because the model has larger covariance or spatial correlation or coherence or spatial structure than the observations, so the extent of individual model storms are, on average, greater than those seen in observations. The impact of the model over- or under-estimating the spatial correlation matrix, with respect the observations, is not effectively dealt with by bias correction. And again, this is an entirely different issue than the "inflation" issue discussed by Maraun (2013). The authors end the paragraph that begins page 10 with, "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." I would agree that a multivariate method that effectively deals with temporal and spatial biases simultaneously is a direction for further research. However, a researcher can correct for spatial structure via recorrelation (Bardossy and Pegram, 2012) and temporal structure via a commonly used univariate bias correction method. The result of correcting the biases of time and space independently will not perfectly correct all possible types of bias within the modeled data, but it will be superior to doing one or the other alone. At the very least, the impact that model bias (in spatial structure) has on area-averaged precipitation needs to be further acknowledged and discussed.

We thank the reviewer for their comments, and acknowledge that we didn't treat the issue of underestimation of modelled spatial correlation sufficiently in the previous version. We add the following section to the discussion:

"Equally, Bárdossy and Pegram (2012) noted that RCM rainfall was considerably less spatially correlated than observations in the Rhine 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."

Bias in downscaled rainfall characteristics

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
[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, dynamical downscaling can provide higher resolution hydroclimate projections, as well as independent (from historical timeseries) and 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 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 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. This 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.

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 need more spatially detailed information as well as information on water metrics for low- and high-flow events, as well as 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 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
5   (Charles et al., 2019) focuses attention on the effect of these biases on runoff.

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 extracted from the larger scale GCM
10 change information (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 historical record of rainfall, future changes in
15 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 by simulating physical climatic processes, does 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;
20 Teutschbein and Seibert, 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
25 rainfall occurrences. Themeßl et al. (2012) bias corrected RCM output over Europe and found residual biases in rainfall extremes and a modification of the climate change signal.

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-
30 season rainfall in Victoria since 2000 has averaged 15% less than the long-term average during the 20th century (Hope et al., 2017). This 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 as well as different metrics of daily rainfall amounts and
35 occurrences.

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
40 accurate. Water resource projection modelling is concerned with future changes, and so an argument could be made that although the rainfall amounts are biased for hindcast (historical) simulations, they will presumably be equally biased for future simulations, so that 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 biases in high and low runoff amounts. This makes inferences on the changes to runoff characteristics highly uncertain when biased rainfall inputs are used.

5   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 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

10   (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
- Parametiric (distributional) QQM.

15   Scaling methods simply consider the change in mean, and apply a constant factor to correct bias in RCM rainfall. Quantile-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

20   results so long as the distribution is sufficiently flexible. Arguably, the choice between non-parametric (empirical) or parametric (distributional) mapping is a representation of the bias-variance tradeofftrade-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

25   mapping for each season across integral percentiles as described below. Overall there is a small but relatively unimportant difference between 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

30   remove biases in rainfall sequencing and multi-day accumulations that might not be readily apparent when considering only the daily distribution of rainfall amounts (Addor and Seibert, 2014). A number of studies have highlighted the residual bias in rainfall sequencing after QQM bias correction (e.g. Addor and Seibert, 2014; Li et al., 2016; Ines and Hansen, 2006), however whereas Terink et al. (2010) and Rajczak et al. (2016) contend that bias correction produces reasonable transition probabilities and spell durations.

35   UnfortunatelyUnfortunately, 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 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

40   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 (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 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 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 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. 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 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, as well as future changes, was 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 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 al., 2016). The three WRF physics configurations are labelled R1, R2 and R3 (see Evans et al., 2014, Table 1), detailed further 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 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. ~~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).~~

2.2 Daily Rainfall Data

5 Daily accumulated precipitation from the NARCliM 12-member WRF-downscaled ensemble (referred to from here onwards 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
10 to the non-parametric procedure we apply to the raw NARCliM data in this paper (Teng et al., 2015).

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
15 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

20 Quantile-quantile mapping bias correction works by estimating the cumulative density function for observed and modelled historical 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
25 timeseries. We use the R package 'qmap' (Gudmundsson et al., 2012), to estimate the cumulative density functions for the mapping function. This 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
30 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~~2~~), 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.

35 **2.4 Defining Transition Probabilites**

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
40 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

17

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 $w = \Pr(W_n|W_{n-1})$ and the corresponding dry-to-dry transition probability as $d = \Pr(D_n|D_{n-1})$. 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. 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 ~~here~~ 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

Figure 2 shows mean annual rainfall across Victoria from (a) AWAP observations, (b) downscaled (NCEP/NCAR) reanalysis and downscaled GCM ~~downscaled~~ (c) historical and (d) future. 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 ~~. There is~~ less rainfall to the more arid north-west region ~~as well~~. 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~~ 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~~ 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, are 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

18

increasing prediction uncertainty. ~~This residual bias in QQM bias corrected rainfall is more pronounced at P99 (Fig. 4a) where interpolation between P99 and P100 (maximum) can have a larger effect.~~

Figure 5<u>6</u> 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 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 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 reduces at around the 99th percentile but is moderated somewhat for the 3-day maximum (i.e. 100th percentile). As noted <u>in section 2.1,</u> ~~by Olson et al. (2016),~~ 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 ~~(see Charles et al., 2019)~~ using GR4J <u>(see Charles et al., 2019 for more details).</u> ~~calibrated on observed rainfall and potential evapotranspiration data using 90 catchments in and around Victoria. Ungauged areas use parameters donated from the nearest neighbour calibration catchment. Further details are provided by Charles et al. (2019).~~ Figure 7 plots the percentage difference in <u>bias-corrected</u> 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 suggesting 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 <u>(Fig. 7c)</u>, this region has the highest absolute biases with bias in runoff of more than −20 mm <u>(Fig. 7b)</u>. Characteristics and biases of runoff from bias corrected NARCliM rainfall is explored in more detail by Charles et al. (2019).

3.3 Residual Bias in Rainfall State Transition Probabilities

~~A number of studies have highlighted the residual bias in rainfall sequencing after QQM bias correction (e.g. Addor and Seibert, 2014; Li et al., 2016; Ines and Hansen, 2006), whereas Terink et al. (2010) and Rajczak (2016) contend that bias correction produces reasonable transition probabilities and spell durations.~~ 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. This results in the 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
5 (Fig. 5c). Figure 8 shows the dry-dry transition probabilities calculated from observed rainfall (top left), bias corrected WRF-downscaled reanalysis (middle column), bias corrected historical WRF-downscaled GCM (right column). Both the bias -corrected WRF-downscaled reanalysis and bias -corrected WRF-downscaled GCM results show a spatial pattern very similar to observations with higher dry-dry transition probabilities to the northwest of the State, and at 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
10 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 probabilities are largest over the high-runoff producing region (Fig. 9), which corresponds to the high-relief, high-altitude part
15 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 almost all of Victoria. This results in underestimation of wet spell occurrences and durations and multiday accumulations of rainfall (Fig. 5c). The bias in wet-wet transition probabilities is more problematic for modelling runoff than the bias in dry-
20 dry transition probabilities, not only because it is of larger magnitude, but because:

- runoff is sensitive to multiday wet spells
- the larger wet-wet probabilities occur in high runoff producing areas, which we would like to model correctly for regional water availability projections
- QQM bias correction reduces the bias in dry-dry transition probabilities, but increases the magnitude of the bias in
25 wet-wet transition probabilities (Fig. 5c).

Figure 10 shows the observed (green), raw (blue) and bias-corrected (red) historical downscaled GCM rainfall amounts 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). Quantile-quantile mapping bias correction equates the quantiles for each rainfall amount such that equal rainfall amounts for observations and bias-corrected rainfall
30 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 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 black line) is 0.774. The
35 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 from Sect. 2.4 that the slope =1 line in Fig. 10 where p = d corresponds to an independent sequence of events and to the right of this line in the transition probability space represents more (positive) serial correlation. This implies that the observed rainfall timeseries contains more
40 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, a bias correction methods that adjusts occurrences areis needed to properly correct biases for hydrological modelling.

20

3.4 Change signals

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 directly. NeverthelessNevertheless, a similar pattern is found with the 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 which 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 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 somewhat 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 effect in more detail).

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 in Fig. 11a).

4 Discussion

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 the 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 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, as well as 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 selection of GCMs representing important climate processes has been an on-going research strand both for the CMIP3 climate model ensemble (Smith and Chandler, 20100⁹; 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 the RCM component of NARCliM, WRF, has been tested extensively, including the general cold and wet bias over south-eastern Australia (Evans et al., 2012; Andrys et al., 2016; Ekström, 2016; Olson et al., 2016; Ji et al., 2016; Gilmore et al., 2016). It is suggested that the wet 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 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 changes are correctly modelled (e.g. Knutti et al., 2010; Racherla et al, 2012).

As mentioned in the introductionOverall, 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). 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 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). For hydrological applications, a specific combination of land and atmospheric circulation schemes (R2) is recommended for hydrological applications (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. 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 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, 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 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 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, andgeneration and neglecting this can lead to 'inflation'. Inflation refers to a phenomenon in bias correction (Maruaun, 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

22

of a timeseries 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
5   amounts become overestimated and low amounts underestimated. Preserving the correct spatial correlation between gauges or gridpoints is an 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 ~~Bardossy and Pegram, 2012;~~ 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
10 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 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
15 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 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
20 should be taken into consideration when applying bias correction for projection applications.

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
25 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, and care must be taken as to which metrics are of interest, and which are the most appropriate
30 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
35 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, as well as spatial correlations between rainfall gauges, remains a challenge for stochastic methods, however.

5 Conclusions

40 Projections of future changes to rainfall and runoff from dynamically downscaled climate models often necessitates a form of bias correction to rainfall fields to obtain sufficiently realistic rainfall inputs for hydrological models. Dynamical downscaling

23

offers potential benefits to regional hydroclimate projections, such as the ability to better model daily rainfall metrics, low flow metrics (after modelling runoff with a hydrological model), and finer spatial scale information, but comes with challenges related to bias. Whereas bias in rainfall amounts can be corrected using quantile-quantile mapping (QQM) methods, biases in rainfall occurrences (such as rainfall autocorrelation, dry-dry and wet-wet transition probabilities) are retained and in some cases increased are not properly corrected with QQM.

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 RCMdynamical downscaling process. 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. Some interpolation and extrapolation occurs in the approach used here, so there is some random residual bias in higher percentiles (i.e. high rainfall amounts). This 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. 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 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 as well as rainfall sequencing metrics are largely preserved after bias correction, with the exception of maximum 3-day rainfall accumulation. HoweverHowever, there is a slight tendency for DJF and MAM rainfall change signals to increase after bias correction and this leads to a tangible reduction in the magnitude of the projected decrease in mean annual rainfall. This is problematic for applications since mean annual change is the most commonly used metric for hydroclimate projections. One possible solution is to rescale the bias corrected rainfall according to raw changes signals but this depends on whether we believe the raw or the bias corrected change signal is correct. However, the fact that 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.

(https://climatechange.environment.nsw.gov.au/Climate-projections-for-NSW/About-NARCliM) for freely providing the climate data. We thank three anonymous referees for their helpful and constructive comments and ~~feeback~~feedback.

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). Int~ernational~. J~ournal~. ~of~ Climatol.~ogy~, https://dx.doi.org/10.1002/joc.4641, 2016.

B~á~árdossy, 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/10.1029/2011WR01152~410.1175/JCLI-D-14-00754.1~, 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. Clim~ate~, 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~.~, ~2015.~ 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. Environmental. Modelling l.& Software., 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. Sci., 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.

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

26

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, International. J.ournal of Climatology., 37, 2934-2946, https://doi.org/10.1002/joc.4890, 2017.

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 (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.

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.mate Res.earch, 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 CMIP3 models. International. Journal. Cof Climatology., 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. of 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 Models. J. Clim.ate, 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 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.

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.

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.

5   Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutierrez, J. M., Hagemann, S., Richter, I., Soares, P. M. 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.

10 McMahon, T. A., Peel, M. C., and Karoly, D. J.: Assessment of precipitation and temperature data from CMIP3 global climate models for hydrologic simulation. Hydrol. Earth Syst. Sci., 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.

15 Mehrotra, R., Johnson, F., and Sharma, A.: A software toolkit for correcting systematic biases in climate model simulations. Environ. Modell. & Software, 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.,
20 Kram, T., La Rovere, E. L., Michaelis, L., Mori, S., Morita, T., Pepper, W. J., Pitcher, H., Price, L., Riahi, K., Roehrl, A., 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
25 projections, Clim. Res., 69, 209–227, https://doi.org/10.3354/cr01403, 2016.

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

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℃ and 2℃ 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
35 water availability linked to ongoing Hadley cell expansion. Earth's Future, 2, 231–238, 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.

40 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.

28

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. Clim. Dynam., 15Climate Dynamics, 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 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, 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/s10584-009-9757-1, 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 downscaled precipitation and temperature reanalysis data for the Rhine basin, Hydrol. Earth Syst. Sci., 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.

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 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

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.

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.
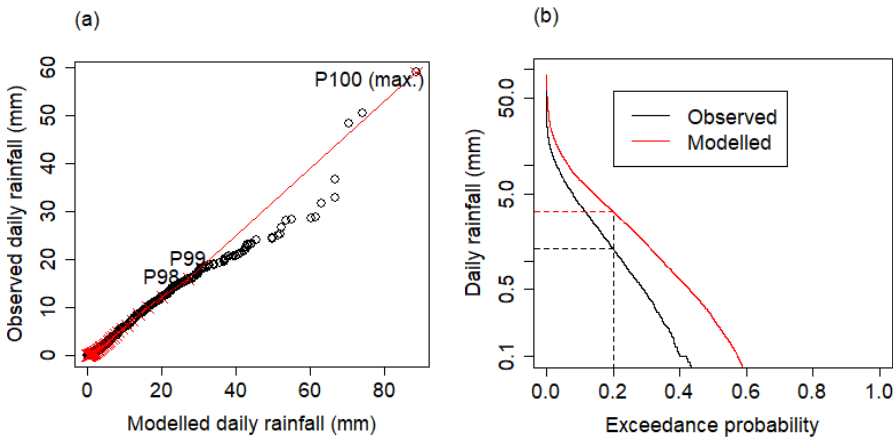
5

Field Code Changed

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.

10
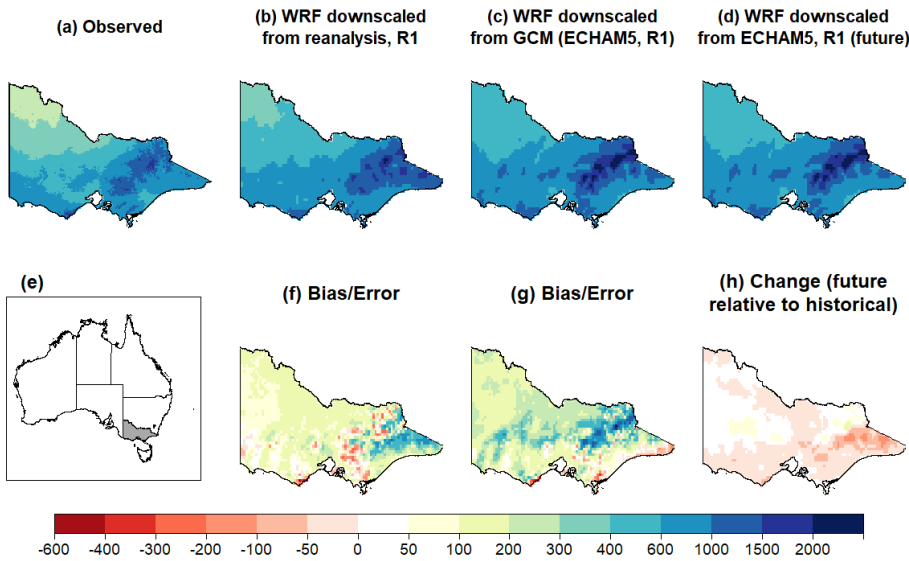
Figure 2: Regional mean annual rainfall (mm yr^{-1}). Inset map (e) shows the location of the State of Victoria in Australia. Other panels show: (a) observed (AWAP) rainfall; (b) rainfall downscaled from reanalysis (NCEP/NCAR); (c) historical rainfall downscaled from median GCM (ECHAM5/R1); (d) future rainfall downscaled from GCM; (f) bias in reanalysis downscaling (compared to observed); (g) bias in GCM downscaling; (h) change factor of GCM downscaled rainfall.

30
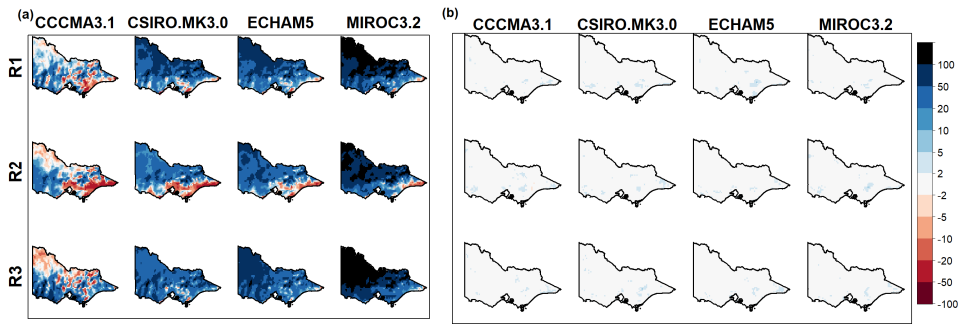
Figure 3: Percentage bias in WRF-downscaled mean annual rainfall from the GCM/RCMs indicated: (a) raw data; (b) residual bias (after bias correction).
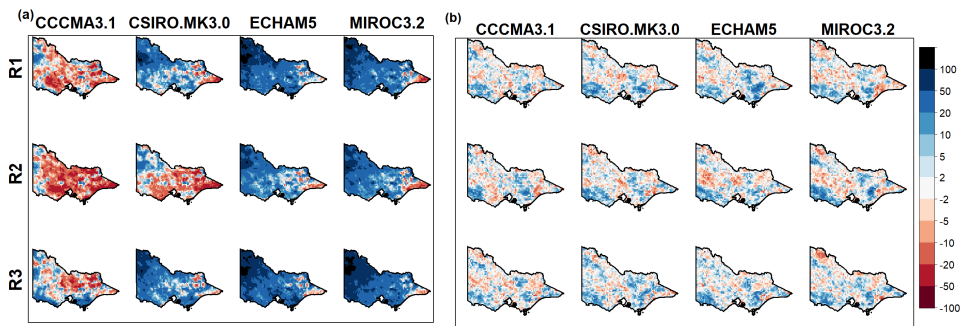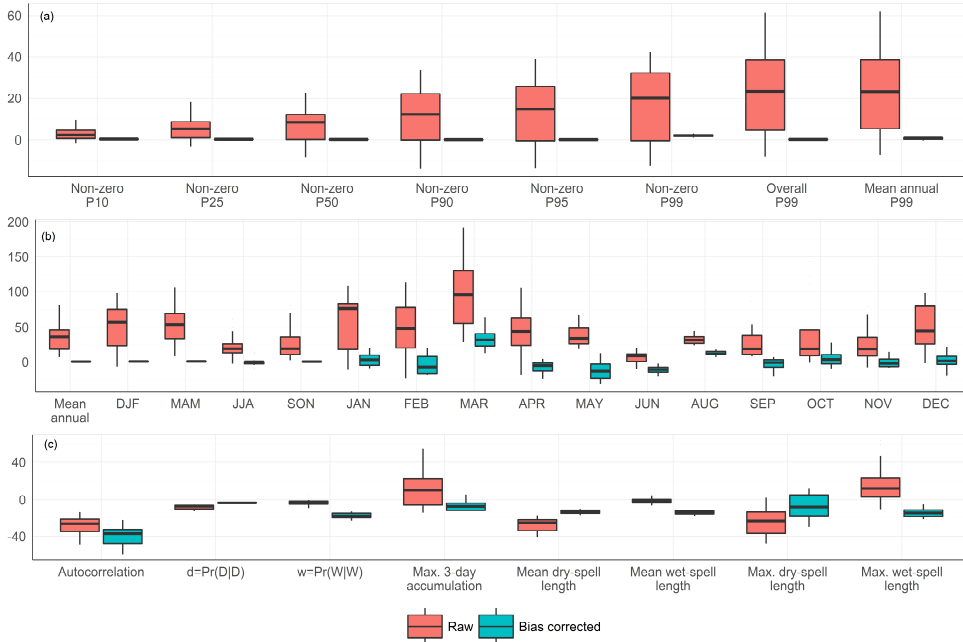
5

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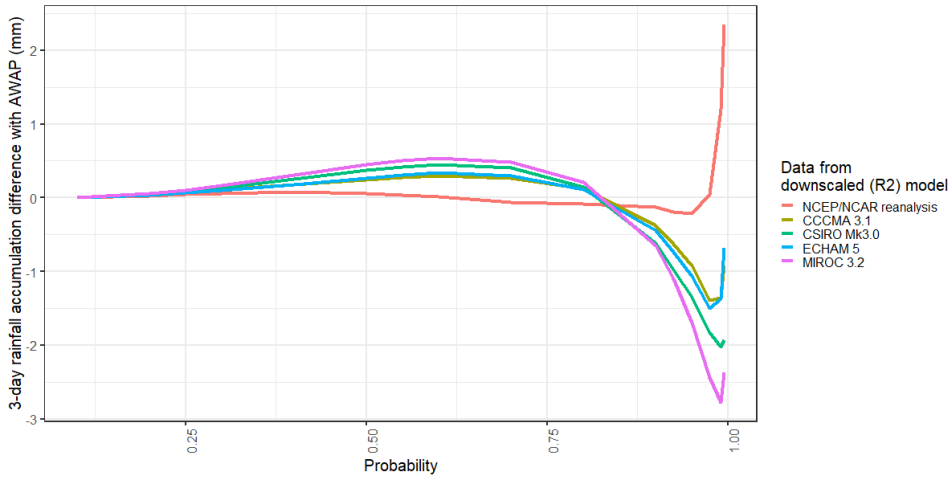
Figure 6: Biases in spatial average 3-day rainfall accumulation percentiles. Here a 3-day moving average filter was applied to each hindcast timeseries and equivalent quantiles taken at increasing probability values (x-axis).

(a) Runoff from AWAP (mm)   **(b) Bias in median runoff (mm)**   **(c) Bias in median runoff (%)**
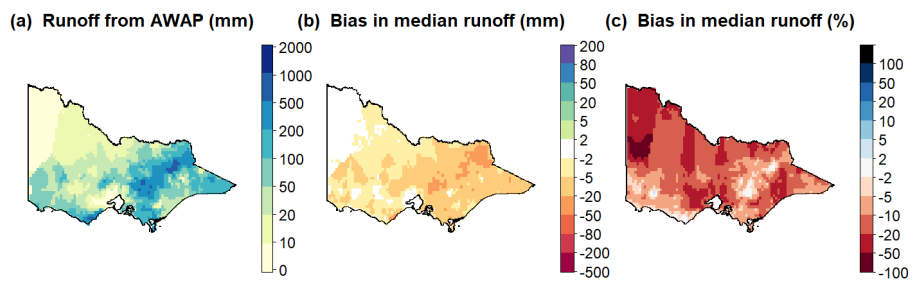
Figure 7: Ensemble median modelled runoff over Victoria: (a) from AWAP historical observations; (b) absolute bias (mm); (c) percentage bias (%).
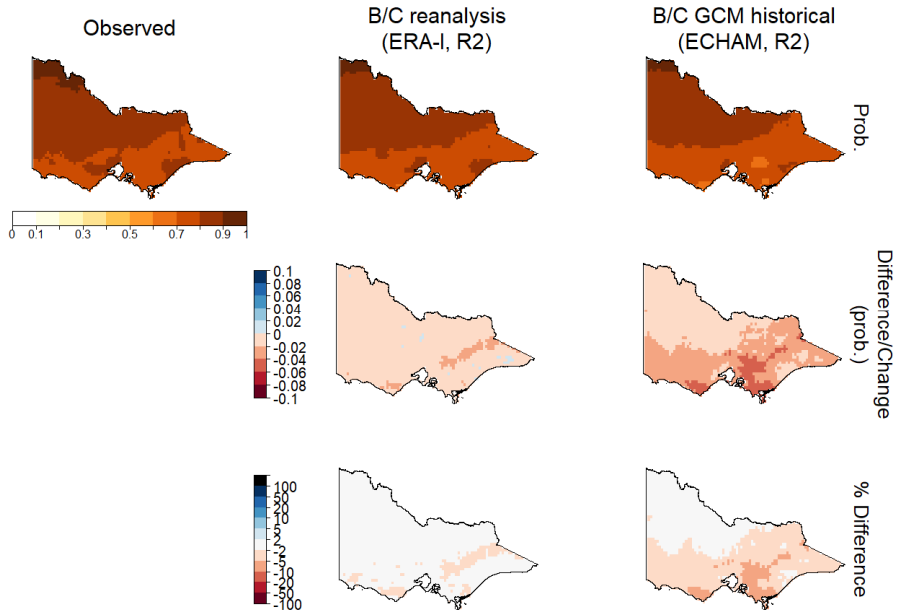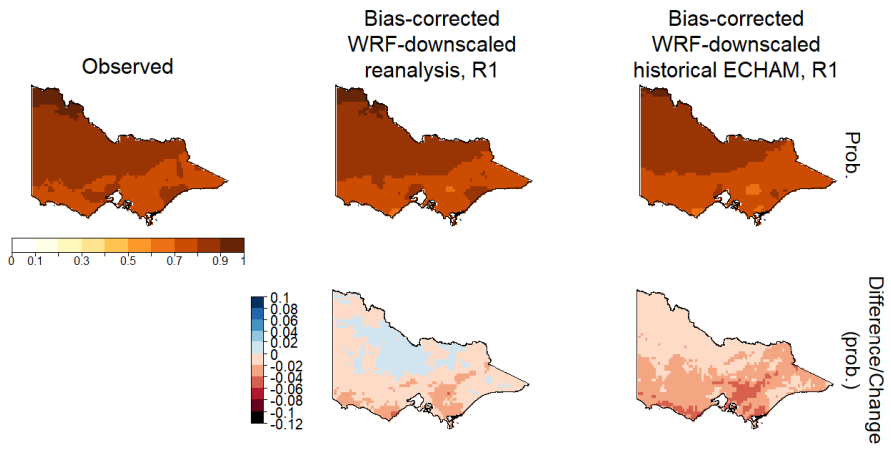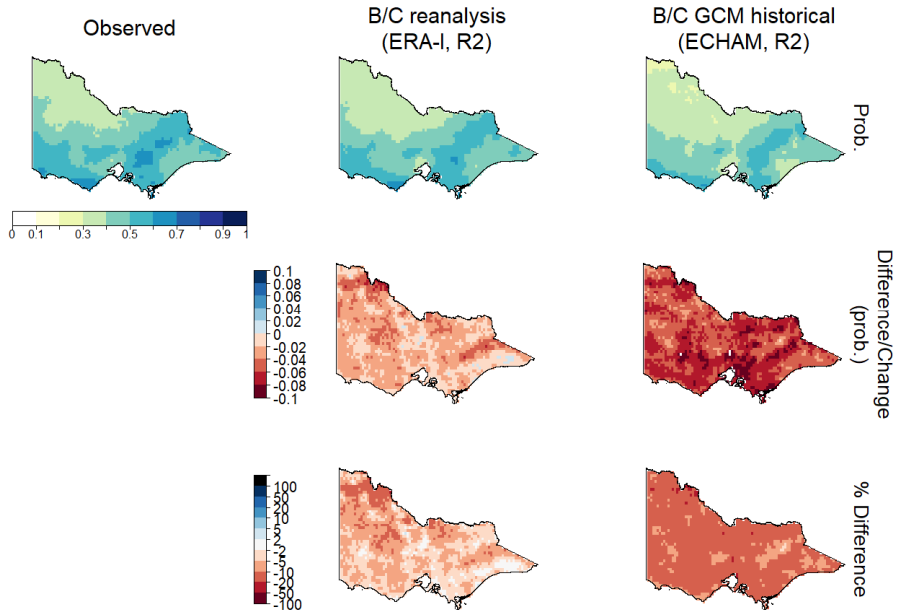
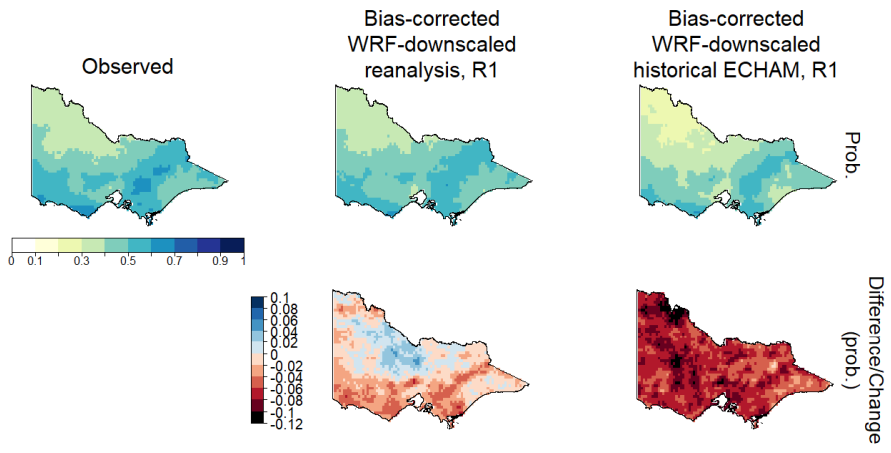**Figure 8: Dry-dry transition probabilities (1mm threshold)**

Figure 9: Wet-wet transition probabilities (1mm threshold)
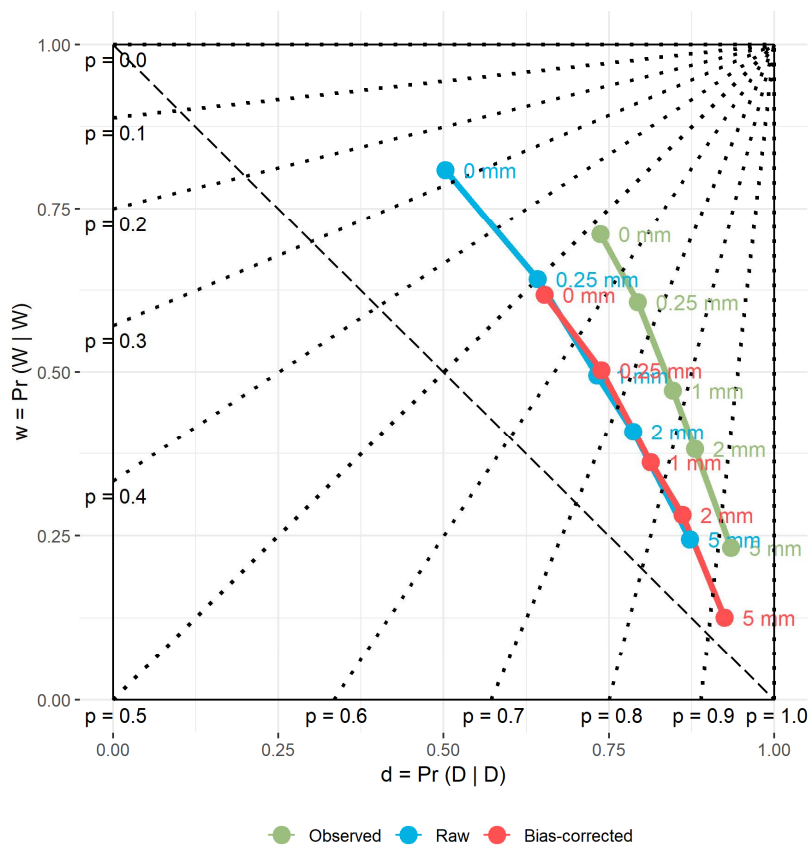
Figure 10: An alternative perspective on quantile-quantile mapping: daily rainfall amounts and associated probabilities plotted in
d-w space. Quantile-quantile mapping bias correction (red) maps daily rainfall amounts from the raw data (blue) to the probability
contours (dotted lines) corresponding to 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
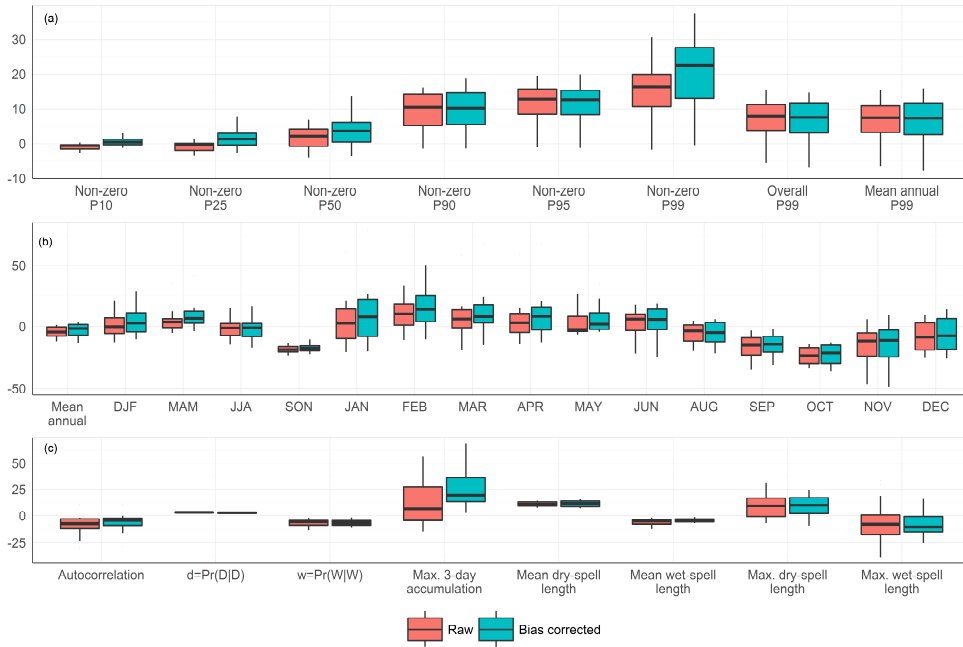**persistent) autocorrelation.**

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.

5