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. 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. 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 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 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.
- 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 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.
- 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, 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?

- 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.
- 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.
- 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.
- 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?
- 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.

- 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?

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)".
- Fig.2: (a) and (b) are exactly the same plots?
- Fig.2: add "GCM (ECHAM5,R1)" to the caption (d). And mention which configuration (R1-R3) was used for (b).
- 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.
- It would be clearer to label every subplot with (a), (b), and so on.
- 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.
- Fig.8 & 9: why show here WRF results from downscaling ERA-Interim, whereas in Fig.2 you show results from WRF downscaling NCEP?

- p.1, L.37: also mention statistical downscaling approach here?.
- p.2, L.5: delete "generates rainfall sequences by simulating physical climatic processes" and replace "does generate" by "generates".
- p.2, L.34: not only hindcast RCM rainfall, but also historical or in principle any data that is used for calibrating QQM.
- 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).
- p.3, L.12: do other studies agree with that? Otherwise, I suggest to add "for Australia".
- 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).
- 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).
- p.6, L.9: I think you refer to Figure 5 here (not 6).
- p.6, L18: This also depends on the calibration period that was used (see main comments above).

- p.6, L.28ff: Is there any explanation for that and would it be sensitive if another BC method is used?
- p.6, L.31: how the WRF model/configuration was selected belongs to the data & methods section.
- p.6, L.34ff: Information on GR4J output belongs into the data & methods section.
- p.7, L.2-4: This information and discussion belongs into the Introduction (see main comments).
- 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.
- p.7, L.6: you might want to refer to Fig.5c here again.
- p.7, L.14: Do you refer to the middle row of Fig. 8 here?
- p.7, L.21: you mean WRF rainfall transition probabilities?
- 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?
- 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.
- p.8, L.2f: a bias correction is not intended to correct rainfall occurrences (see my major comment).
- p.8, L.8f: Yes, you can compare the percentiles, as they are statistical measures of a distribution.
- 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)?
- p.8, L.24ff: are the found changes signals statistically significant in relation to annual variability?
- p.9, L.12-13: Does this affect the errors in the raw transition probabilities? (see major comment)
- 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.
- 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.
- p. 10, L. 26: “multi-“.
- p. 10, L. 34f: again, BC cannot correct temporal errors.
- p. 10, L. 39: what is completely unmentioned is how statistically significant the change signals are in relation to annual variability.
- p. 10, L. 40: replace “dynamical downscaling process” with “RCM”.