Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-139-RC2, 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Bias in downscaled rainfall characteristics" *by* N. J. Potter et al.

Anonymous Referee #2

Received and published: 20 May 2019

The main objective of the article is the analysis of how bias correction (BC) affects rainfall characteristics in a 12-member WRF-ensemble (NARCLiM simulations) forced by reanalysis and pre-selected GCMs. In particular, the authors investigate whether BC modifies the temporal persistence of dry/wet spells, transition probabilities and simulated climate change signals. Precipitation fields with realistic rainfall amount and persistence are crucial for runoff modelling and BC is thus a commonly applied technique. The paper nicely shows that the chosen BC method (empirical quantile mapping) corrects for the biases in the mean and variability, but it does not alter the temporal structure of the precipitation field. The introduced diagnostic for transition probabilities is interesting and valuable in judging the quality of simulated precipitation fields. The authors further show that the climate change signals of precipitation characteristics are mostly unaffected (with some exceptions) by the BC. Although the paper is well written, it has some major deficiencies that have to be revised in order to improve the paper.

In summary, the paper has to point to a knowledge gap the authors want to address and frame the current state and consensus of BC research, which has advanced compared to what the authors describe, in that it became much more critical (Maraun et. al, 2017). Thus, the construction and evaluation of the BC has to be better described, already beginning with the credibility of the climate models. Finally, the models (WRF and the rainfall-runoff model GR4J) need to be described in more detail. Overall, I encourage the authors to take up the points I will elaborate below.

Major comments

- The study must be better embedded in the current state of bias correction research. Therefore, the Introduction needs to be revised in order to better frame the current edges of bias correction, which has advanced compared to what is written by the authors. Another issue in this context is the somewhat unclear research question of the authors. The research question should be stated much more clearly, immediately after the opening, to define the overall thrust of the study. At present, the research question is weak (post-processing WRF output by a well known method and see what happens) and only appears at the end of the Introduction, after the reader went through long descriptions. Sharpening the research question is important because it is unclear where the authors see a knowledge gap they would like to close/target with their study. Recent discussion on bias correction research became much more critical on using BC as a one-tool-does-it-all or Swiss-Army Knife (Maraun et al., 2017; Maraun 2016), also the very basics of applying BC are questioned (Ehret et al., 2012). Therefore, I think the conclusion on P11L16 "..., a different approach to bias correction is needed ..." maybe points into the wrong direction, because a BC should not modify the transition probabilities, trends or climate change signals when the simulated regional climate (change) is credible. It does not mean that BC should not be used at all, but that the whole procedure has to be viewed more critical and applied with great care. That said, the authors should take care on what their expectation and hypothesis are; it already begins with the credibility of the climate models. When it is known that BC affects climate change signal / persistence, what is the expectation when applying BC to the WRF ensemble? So instead of stating “the aim of paper is to investigate” (i.e. collecting information), it should be “our question was” (i.e. knowledge hope to gain). For instance, one interesting question to follow might be: “How to exploit/subsample an RCM-ensemble to produce the most credible bias-corrected precipitation fields for runoff modelling?”.

- It is known that BC modifies the climate change signals or trends (e.g. Hagemann et al. (2011); Gutjahr and Heinemann (2013); Dosio et al. (2016)), which is why trend preserving BC methods have been developed (e.g. Switanek et al., 2017), so that the user can choose how to modify trends or change signals. In particular, if the simulated regional climate change is credible, then Maraun and Widmann, (2018, p.199) recommend to use a trend preserving BC instead of the standard quantile mapping.

- Persistence and BC is a more difficult field, but most BC methods do not affect the temporal structure of the data (persistence), so that the corrected fields retain the inherit persistence of the driving model (Maraun et al., 2017; Maraun and Widmann, 2018). In contrast, however, Rajczak et al. (2016) demonstrate the opposite, although artefacts might be introduced (Maraun and Widmann, 2018, p. 182). This discussion should be taken up by the authors.

- I am not sure if I correctly understood where the 12-member WRF ensemble originates from. I suppose Evans et al. (2014) chose this subsample from the 36-member WRF ensemble, described by Evans et al. (2013), based on their skill and independence. Thereby the skill of the WRF simulations was based on how they simulate two-week periods centred on eight storm events. My con-

cern is in line with the remarks by Evans et al. (2014) in that it is unknown how skillful the simulations are at climate time scales, when the long-term memory becomes important, e.g. for dry spells. As stated by Maraun et al. (2017), bias correction is only partly a statistical issue, but already begins with the selection of credible climate models. If these pre-selected 12 WRF simulations already have fundamental errors in precipitation temporal structure (e.g. because of misplaced circulation or misrepresented weather-types), then it cannot be expected that BC corrects the persistence (and probably should not). Maybe a different set of GCMs and subsequent WRF simulations are better suited as input for runoff modelling. I recommend discussing how suited the 12 WRF simulations are for runoff modelling in Victoria on the climate scale. In this sense, maybe the number of WRF simulations considered for bias correction and eventually for runoff modelling could be reduced to only the most credible ones. Although I understand that the authors want to show the range of uncertainty from the WRF ensemble, in practice one does not want to use RCM simulations that have fundamental errors (e.g. in circulation, precipitation persistence, etc.) for runoff modelling, so these could be discarded from the analysis in advance.

- It should be discussed how bias-corrected precipitation fields might be inconsistent with other fields of the WRF model (e.g. temperature). BC reduces the precipitation volume error but might create other problems for impact models due to inconsistencies with other atmospheric fields. The problem here is that BC is not intended to modify the persistence and it mostly inherits the temporal structure of the driving model. It does, however, implicitly modify the transition probabilities as it adjusts the number of wet days (the wet day threshold is a percentile of the distribution). Maraun et al. (2017) and Maraun and Widmann (2018) argue that if the persistence of the driving GCM or RCM is all too wrong, then this model should not be used for downscaling and it is not recommended to apply a bias correction (which the authors also state on P2 L27). Errors in persistence originate mainly from wrong dynamics in the climate models, so that specific circulation features are misplaced or misrepresented. These errors in dynamics then manifest themselves in errors of transition probabilities or as seasonal biases. For instance, the WRF models used here simulate too much precipitation throughout the year, but mostly so in January – May. Could this indicate fundamental errors in the circulation of the models in these months? As outlined above, a BC cannot overcome this problem and is thus not recommended (Maraun et al., 2017). If a BC is still applied, it causes inconsistencies with the other fields. Suppose for instance, if the precipitation fields were corrected to match longer dry spells, then the large-scale synoptics would still simulate wet conditions, which causes inconsistencies and might produce errors/artefacts in runoff models. The authors should discuss these issues and whether these inconsistencies affect the runoff model.

- The evaluation strategy of the bias correction method is unclear. BC evaluation is a difficult task (Maraun et al., 2017) and has to be done very carefully. The important part here is how the BC performs for data outside the calibration period. As QM may be prone to overfitting, in particular for the tails, such an evaluation is important to understand for instance why the 'Non-zero P99' distribution in Fig. 14 is corrected to higher values, which might also be relevant for why the DJF and MAM change signals of annual mean precipitation increase after BC (Fig. 15). Maybe a Q-Q plot of a representative time series of the raw and corrected precipitation might help to understand how the BC performs (and whether it is too flexible/stiff) for data not used to construct the transfer functions. This might also help to understand why the rainfall is corrected to too low values in Feb, Apr, May, and June (P6, L31f) (issue of overfitting?). Was it tested whether a seasonal correction is the best strategy? There are papers that analyse time scales issues for bias correction (e.g. Haerter at al., 2011; Reiter et al., 2018), concluding that the biases are not independent of the timescale and might introduce errors if assumed so. Has something similar been tested, or what is the argument for a

seasonal correction?

- Section 2 lacks a detailed model and data description. Mention that a WRF-ensemble consisting of 12 members was used (except you decide to reduce number of members). By which data set was the WRF hindcast simulation forced? I suppose from NCEP (Phase 1 described by Evans et al. (2014)), but why does ERA-I later appear in the legends of Fig. 11 and 12?. Define here also what R1, R2, R3 means, and what the main differences between these three WRF configurations are. It should be mentioned whether these 12 WRF simulations (plus the hindcast) were evaluated on the climate scale for the Victoria state, as mentioned before. If not, it should be at least included in the Discussion. Furthermore, there is no description of the runoff model GR4J, it is suddenly mentioned in section 3.2. A description should be included in section 2 (also mention which atmospheric input fields are used to force GR4J, inconsistency issue).

Minor comments

- The authors may consider to revise the title of the paper to account for the runoff aspect.
- Also, consider to reduce the number of figures and correct the figure legends (be clear that WRF simulations are referred to, not the GCMs). Also add more explanations to the figure captions (abbreviations, before/after BC, add data set used to construct the figure, e.g. Fig.7 and similar). I further suggest to enlarge the axis annotations.
- P3, L12 and L18: Better refer to parametric and non-parametric distributions, because also empirical distributions are 'distributions'.

- P3, L16ff: Does this statement refer to corrections on the calibration data, or when applied to data outside the calibration period? Overly flexible methods might introduce artifacts at the tails outside the calibration period. Inappropriate (too stiff) parametric methods may introduce unrealistically high values, when much higher values than observed appear in a future scenario (Volosciuk et al., 2017) because of the extrapolation.

- P3, L23: How was this evaluated? Again, only on the calibration data or also on data outside the calibration period? For the calibration data, an almost perfect correction results by design by all methods.

- P4, L16ff: Evans et al. (2014) is not the correct reference for how WRF reads in lateral/lower boundary data from a GCM. Put this reference to the next sentence: "The NARCLiM projections (Evans et al., 2014)..."

- P5, L26 and L29: correct 'Figure' to 'Figure 2.' (you might want to drop this figure, as the information is already in Fig. 13).

- P6, L.16f: Only if it is assumed that the bias is time-independent, otherwise it was demonstrated that the chosen timescale for constructing the transfer functions impacts for instance the annual mean (Haerter et al, 2011; Reiter et al., 2018).

- P6, L27: remove "after bias correction".

- P6, L31: effect of overfitting? Would this also occur if a monthly BC is used instead of a seasonal?

- P7, L6: maybe better use "considerably" (or similar), "significantly" has a connotation related to statistical testing.

- P7, L8: Be more precise: Instead of "This results .." write e.g. "Due to good performance in correcting dry-dry transition probabilities, the bias in mean and maximum dry spells is well corrected, ...".

- P7, L10 and Fig.9: Does Figure 9 show the3-day rainfall bias of the WRF simulations (text) or of the GCMs (legend)? And then, before or after the bias correction?

- P7, L18: Explain what PET means; if it is a data set, then describe it in section 2.

- P7, L27: Give more than one reference for "a number of studies..".

- P8, L3: Be precise which models are used where. Sometimes reanalysis and GCM results are wrongly used, when WRF models forced by either NCEP or a GCM are meant. There are other places in the manuscript as well (e.g. legend of Fig. 9, or titles of Fig. 11 and 12).

- P8, L8: Add "... wet-wet transition probabilities are largest over the high-runoff producing region (Fig. 12),...".

- P8, L32: replace "northeast" with "to the right".

- P9, L18: Replace "Fig.16" by "Figure 16".

- P10, L4f: Replace "Charles et al. (2019, submitted). (Or insert reference if already published).

References:

Dosio et al. (2016), doi: 10.1002/2015JD024411.
Ehret et al. (2012), doi: 10.5194/hess-16-3391-2012.
Gutjahr and Heinemann (2013), doi: 10.1007/s00704-013-0834-z.
Haerter et al. (2011), doi: 10.5194/hess-15-1065-2011.
Hagemann et al. (2011), doi: 10.1175/2011JHM1336.1.
Maraun et al. (2017), doi: 10.1038/NCLIMATE3418.
Maraun and Widmann (2018): Statistical Downscaling and Bias Correction for Climate

Research. Cambridge University Press, ISBN: 978-1-107-06605-2.
Rajczak et al. (2016), doi: 10.1175/JCLI-D-15-0162.1.
Reiter et al. (2018), doi: 10.1002/joc.5283.
Switanek et al. (2017), doi: 10.5195/hess-21-2649-2017.
Volosciuk et al. (2017), doi: 10.5194/hess-21-1693-2017.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-139, 2019.