Review of

‘An improved statistical bias correction method that also corrects dry climate models’

by F. Lehner, I. Nadem and H. Formayer

Recommendation: reject and resubmit new manuscript

This manuscript presents a bias correction method that preserves the simulated climate change signal (CCS) and allows to increase the number of dry days. The method is called ‘Empirical percentile-percentile mapping’ (EPPM). As stated by the authors, it uses elements of published methods to achieve the CSS preservation and the adding of dry days, and in this respect there is no novelty.

However, EPPM deviates strongly from common practice with respect to how the corrections are applied to future climate. Usually in quantile mapping (QM) the percentile for a given simulated value in simulations for future climate is determined with respect to the cumulative distribution function (CDF) of the simulated values in the present climate and then mapped onto the observed values for the same percentile. This means a simulated value is always mapped onto the same corrected value regardless of whether it occurs in the present or in the future climate. This assumption of constant bias for a given value does not necessarily always hold, but in lack of knowledge how the bias might change from the present to the future climate, it is a reasonable assumption. In contrast EPPM determines for future climates the percentile for a given simulated value from the future simulated CDF and applies the correction derived in the present climate for this percentile (the difference or ratio between the observed and simulated values for this percentile). This means the corrections applied to a given simulated value can be very different in the present and the future climate. I cannot see any physical justification for doing this. Unfortunately the manuscript neither explains this key property in a transparent way nor does it give any justification for this approach. Even the name of the method is misleading, because for the future climate there is no mapping of percentile values for one distribution onto those for another.

Moreover the text is written as if it was obvious that the CCS should be conserved by postprocessing methods, but there a reasonable arguments for and against this. Again, there is no justification at all for this in the manuscript and related literature is not sufficiently discussed.

In its current form the manuscript is conceptually unclear, not systematically written, contains only superficial explanations of the approach, and if published it would add confusion to the discussion on bias correction rather than help to address open issues and contribute to methodological progress. There might be aspects in this work that are publishable, but getting the manuscript in a publishable form requires a clear identification of what is novel, and a systematic and sound explanation and justification of the approach. This would go substantially beyond a revision of the manuscript and constitute a new paper. I thus recommend rejecting the manuscript, but encourage submission of a new manuscript after addressing the problems.

More detailed comments on the issues mentioned above and some additional points are listed below.

Specific comments

1)
The introduction cites a good number of relevant publications, but it is not well structured and many details are unclear. The attempt at grouping the methods in lines 67-81 is not clearly linked to the discussion of individual methods at the beginning of the introduction.

The introduction should be organised from the beginning by explaining and giving examples for groups of methods with common structural elements and properties, including

- Additive and multiplicative scaling, and linear regression
- Empirical QM, parametric QM using one function, multi-segment parametric QM
- Modification or preservation of the simulated CCS, specifying what aspect of the CCS is preserved, e.g. mean or CCS for specific quantiles
- Treatment of biases in wet-day frequency
- Treatment of values outside observed range
- Which methods assume stationarity of the bias? For those who don't, how do they specify potential non-stationarities?

It should also include a paragraph on comparison studies. This overview should include Maraun and Widmann (2018), which is a standard reference for downscaling and bias correction, and the VALUE comparison of downscaling and bias correction methods (Maraun et al. 2018, Gutierrez et al. 2019, Widmann et al. 2019).

Gutierrez, J.M., D. Maraun, M. Widmann et al., 2019: An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *Int. J. Climatol*, **39**(9), 3750-3785.

Maraun, D., R. Benestad, S. Kotlarski, E. Hertig, M. Widmann, J. Wibig, J.M. Gutierrez, R. Huth, R.E. Chandler and R. Wilcke, 2019: Validation of temporal variability in the VALUE perfect predictor experiment. *Int. J. Climatol*., **39**(9), 3786—3818.

Maraun and Widmann, 2018: Statistical downscaling and bias correction in climate research. *Cambridge University Press*, ISBN 1107066050

Widmann, M, J. Bedia, J.M. Gutierrez, T. et al., 2019: Validation of spatial variability in downscaling results from the VALUE perfect predictor experiment. *Int. J. Climatol*., **39**(9), 3819-3845.

2)
The list of properties that the new method should satisfy (lines 97-100) is not clear. Property 1: What are 'long-term trends'? Differences between the simulated future and reference periods? If so, differences in the mean or in individual quantiles? Property 2: 'the model data' should be 'the corrected model data'. 'should match the observational data' with respect to what? Means, quantiles, trends?

3)

The outline of the method (lines 107-112) is unclear. Up to here the text has mentioned as key issues the correction of an underestimation of wet days, and the preservation of the CCS. Now the question of whether biases are stationary is the key point, but this and the various ways of how to define QM have not been properly introduced. Most QM methods assume stationarity, not because it is necessarily a correct assumption, but because deviations from this have to have specific forms, for instance the one given in this manuscript, which are difficult to justify. As mentioned at the beginning of the review any specific approach needs of course to be justified.

4)

Lines 102/103: the underestimation of the precipitation sums needs to be explained in more detail. If the full PDFs were matched the sums and thus the means would be identical by construction because $\text{mean}(x) = \int x \, \text{pdf}(x) \, dx$. If I understand correctly the problem is that the PDFs are not matched for the full range of values including zero, but only for the simulated wet days, which leads to the mean over the corrected simulated wet days matching the observed sum over these days, but the simulated mean over the whole period includes more dry days than the observed mean over the whole period.

5)

Whether the CCS of the original simulations should be preserved by bias correction is an important questions, but the correct answer depends on whether the raw CCS can be expected to be realistic. Discussions of this can be found for instance in Maraun and Widmann (2018), section 12.9.1, and in Maraun et al. (2017). The paper contains no justification for preserving the CCS.

Maraun, D, T. Shepherd, M. Widmann, et al., 2017: Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, **7**(11), 764-773.

6)

The validation of the OEKS15 data in section 4 is not convincing with the information available in the manuscript. In addition to structural elements in SDM that can lead to a difference in the means of the postprocessed models and the observations there are two more potential reasons for this difference. The first is the use of different observation datasets used for fitting SDM and for validation SDM. The second is a different reference period for fitting SDM and for the validation. There is no information in the manuscript on these two points, and therefore it cannot be concluded that the differences that are found are due to the structure of SDM.