Interactive comment on "Impact of bias nonstationarity on the performance of uni- and multivariate bias-adjusting methods" *by* Jorn Van de Velde et al.

Bastien FRANCOIS (Referee)

bastien.francois@lsce.ipsl.fr

Received and published: 11 January 2021

The authors of "Impact of bias nonstationarity on the performance of uni- and multivariate bias-adjusting methods" apply different methods of bias correction (BC) for the adjustment of temperature, precipitation and evaporation time series. Then, adjusted precipitation and evaporation time series are given to a hydrological model for impact assessment. In particular, the performance of 2 univariate and 3 multivariate BC methods are assessed under climate change conditions in order to determine the influence of bias nonstationarity on results. The authors find that the multivariate BC (MBC) methods often perform worse than the univariate BC (1d-BC) methods concerning the

adjustment of the marginal and inter-variable properties of the variables for the projection period. These underperformances from MBCs for the correction of the selected simulated variables are linked with their changes under climate change conditions, suggesting that MBCs are failing to cope with nonstationary climate conditions.

While I see the interest in assessing the influence of bias nonstationarity on results from BC methods, I am not sure that the present study is comprehensive enough to do so precisely and to provide appropriate nuances. This however can be improved by several means.

General comments:

1. Results are often not well explained when bad performances for (M)BCs are obtained. The validation metrics used in this study (residual biases relative to the observations (Rbo) and to the model bias (Rbmb)) tend to make, in my opinion, the analyses of results difficult for readers. In that respect, I would find helpful to provide raw values of the indices for, respectively, the calibration and projection periods (i.e. without computing directly their biases as in Table A1) and their changes between the calibration and projection periods for: the reference, the simulations and the 5 corrections from the (M)BC methods. In a general way, it would permit to better explain the results obtained from the (M)BC methods, i.e. whether good or bad results for a specific method come from the nonstationary problem (i.e. from the fact that simulated changes are "wrong", which alter the quality of results from MBC methods) or from the characteristics of the method (e.g., the statistical technique used, its stochastic nature, etc.). Identifying the reasons for good or bad results from (M)BC methods is of key importance for this study, to be able to conclude with more certainty that bias nonstationarity is the main source of problem.

2. The authors conclude that "the univariate bias-adjusting methods, computationally less complex and not taking (potentially changing) correlations into account, seem to be more robust." (L787). Concerning correlation, I am afraid that this result of robust-

ness is only specific to the present application, and hence the generalization of this conclusion cannot be done as the authors do (e.g., L825 "we advise to use univariate bias-adjusting methods, until it becomes more clear how it can be ensured that multivariate methods certainly perform well in a climate change context."). Indeed, one of the advantages of considering MBC methods instead of univariate ones is that MBCs are able to adjust correlations between variables. Univariate BC methods are not designed to do so. For example, 1D-BC methods such as QDM globally conserve the rank structures of the climate model to correct, and hence the simulated dependence structure between variables is preserved. According to Table A1, simulated correlations often present little bias compared to observations (for example, for corr(P,E), corr(E,T), crosscorr(P,E,0), crosscorr(E,T,0), crosscorr(E,T,1)). Then, by preserving the simulated rank structures, 1d-BC methods like QDM mechanically present correlations with little bias as well. This result is really specific to this study, and would not be obtained if the raw climate simulations would have presented strong biases in correlations. Hence, concerning correlations, results from 1d-BC as QDM depend on how well the models simulate relevant dependencies between the climate variables. This is something already suggested by Zscheischler et al. (2019). This point is not explained in the present study, while it is one of the key points of discussion, and the principal reason why QDM often performs well in this study for correlations. This point should be mentioned and discussed to provide the appropriate nuances to initial conclusions.

3. Linked with the comment 1. above, I would find interesting to provide (at least in Appendix) the results for the multivariate bias correction and the hydrological model for the calibration period. In my opinion, verifying that MBC methods perform well for the calibration period (which has to be verified) would validate the global methodology and would better support the conclusions on the effect of bias nonstationarity on the results from MBCs for future periods.

4. As explained by the authors (L79), climate models can present statistical biases compared to observations that are nonstationary, that is, that the differences of bias

between the calibration and future period are not the same. What is often not clarified clearly through the article is that, in a changing climate context, another way to see bias nonstationarity is that it results from the fact that observed and simulated variables do not present the same changes between the calibration and projection periods. This is of key importance, as some of the (M)BC methods included in this study can take into account the simulated changes in their correction procedures (such as dOTC or MBCn). It must be better highlighted through the article, and has to be used to provide a better analysis of the results. 5. I think the reasons why the three MBCs (MRQNBC, dOTC and MBCn) are selected in this study must be better specified. In particular, their differences of assumptions and attributes could be better indicated, for example with a table (as recommended in 6.).

6. If I understood well, all the three MBC methods are able to take into account (some of) the potential simulated changes in their correction procedures. It also exists in the literature MBC methods that assume dependence structure to be stable in time (R2D2, Vrac, 2018, Vrac and Thao, 2020). It would have been interesting to include in the study such MBC methods as benchmarks for a better assessment of the potential losses (or benefits) of considering stability of dependence structure, even in a changing climate context.

7. The quality of presentation should be improved.

- The "Bias-adjusting methods" section is too long. Please consider to summarize each of the BC methods and their main properties in a short text (and with the help of a table, as already explained) to keep from overwhelming the reader with technical details that are not necessarily useful to get the main points. Technical details and algorithms can eventually be placed in Appendix.

- Please consider to indicate the letters for each figure when describing the results (for instance, for QDM: Figure 3a, mQDM: Figure 3b, etc). This would better guide the reader through the different plots.

Specific comments:

L55: I would replace the word "uncertainty" as it can be misleading here. I would rather say "error (or bias) that can propagate in the impact models".

L72: in François et al. (2020), it is explained that all multivariate methods that are under study fail in adjusting the temporal structure of simulated times series. Of course, it already exists methods that can adjust (part of) the temporal properties: MRQNBC is one of them.

L75: what do you mean by "more recent validation period". I think that just mentioning "validation period" is enough.

L225: Did you verify that assuming the stationarity of the frequency of dry days holds for your application? Also, did you apply a thresholding after bias correction? I know that, for example, dOTC can generate negative values for precipitation. How did you deal with this problem?

L282: The successive conditional approach performs successive corrections conditionally on the variables already corrected. This can be applied to more than two variables. It should be specified. It does not "adjust a second variable conditionally on the second variable", as written.

L286: This is not well explained. The problem of robustness of the successive conditional approach is specific for a high number of dimensions to correct. Indeed, in a successive conditional approach, as the number of variables already corrected increases at each successive step, it progressively reduces the number of data available for the correction, making it less and less robust.

L396: Do you know why the results are exacerbated?

L429-440: I am not sure that the equations to explain the Schaake Shuffle are necessary, as it can simply be explained with words.

L442: Do you have a reference to account for ties by introducing small random values? There exists also other ways to compute ranks to handle problems of equal values.

L445: This result is really "case-specific". I would precise it, as it cannot be generalized for each application (depending on the bias-nonstationarity, for example).

L454: Did you check if, even after early stopping, overfitting was not a problemĂă? In particular, how did you choose the tolerance of 0.0001?

L463: I would precise that dOTC extends the CDFt method to the multivariate case.

L465: dOTC is indeed designed to preserve the trend of the model for marginal properties but also for dependence structure (or copula). It should be specified, as it is a major difference with univariate methods.

L520: I don't think that the 10 calculations made for dOTC are necessary. What are the bin sizes chosen to implement dOTC? If the bin sizes are small, the influence of the stochastic components in dOTC is rather small, and hence introducing 10 calculations does not seem necessary to me.

L593: I do not understand the sentence "Yet, QDM has the best RBMB values and might thus be preferable", that seems in opposition with the previous sentence.

L644: "an influence" on what? Correlations? Please be more precise.

L677: Actually, it surprises me as it has been found in François et al. (2020) that, for a small number of dimensions to correct, MBCn and dOTC preserve roughly the rank sequences of the model to correct. It normally does not have such "important shuffling" as specified. Do you know why?

L698: A difference between QDM and MBCn is also the adjustment of dependence structure. It should be specified.

L705-706: I do not understand. Please rewrite.

L790: Did you mean "possibly"?

Bibliography

Zscheischler, J., Fischer, E. M., and Lange, S.: The effect of univariate bias adjustment on multivariate hazard estimates, Earth System Dynamics, 10, 31–43, https://doi.org/10.5194/esd-10-31-2019, 2019.

Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R2D2) bias correction, Hydrology and Earth System Sciences, 22, 3175, https://doi.org/10.5194/hess-22-3175-2018, 2018.

Vrac, M. and Thao, S.: R2D2 v2.0: Accounting for temporal dependences in multivariate bias correction via analogue ranks resampling, Geosci. Model Dev., 2020, 1–29, https://doi.org/10.5194/gmd-2020-132, 2020.

François, B., Vrac, M., Cannon, A. J., Robin, Y., and Allard, D.: Multivariate bias corrections of climate simulations: Which benefits for which losses?, Earth Syst. Dyn., 2020, 1–41, https://doi.org/10.5194/esd-2020-10, 2020.