Response to Referee Comment 2 on **"Multivariate bias corrections of climate simulations: Which benefits for which losses?"** ***by*** **Bastien François et al.**

Anonymous Referee #2

General comments:

Comment:

1) A comparison of methods is especially helpful in an emerging area of research such as multivariate bias correction (MBC), where few guidelines are available. Overall, I like the article but have some suggestions for improvement.

Response:

We would like to thank the anonymous referee for her/his very positive comments and the detailed questions. All the comments and our point-by-point responses are given below.

Comment:

2) I was surprised that the authors re-gridded the 0.5 degree precipitation product to the coarser climate model grid using nearest neighbor interpolation. The authors don't say what method is used for the 8km precipitation product, but presumably that also is nearest neighbor? Overall, this approach seems to ignore a lot of spatial information in the "observed" data and effectively makes this a quasi-regular resampling of the observed data and not an interpolation. Is the goal to get area-averaged precipitation or a gridded product of point precipitation? Some discussion of this choice and the tradeoffs is warranted, unless the authors decide to use a different approach.

Response:

We think that there is a misunderstanding concerning the re-gridding step. We did not re-grid the 0.5 degree precipitation and temperature products (WFDEI) to the coarser climate model grid (IPSL) as understood in your comment, but the opposite, as explained in Section 2, L94. However, with the aim of clarifying better this point, we propose to rewrite L94 as follows :

*"Note that, as spatial resolution between WFDEI and IPSL-CM5 are different, IPSL **model** data are regridded by a nearest neighbour technique to associate each IPSL grid cell to its nearest WFDEI grid cell center. **Hence, in the following, the IPSL data will be used at the 0.5° spatial resolution corresponding to that of the WFDEI reference dataset.**"*

Concerning the method used for the 8km precipitation product, nearest neighbour technique is indeed used. To better clarify this, we propose to rewrite L99 in Section 2 as follows:

"*IPSL data are regridded to **the 8 km x 8 km** SAFRAN resolution **using nearest neighbour technique***".

Comment:

3) A lagged version of a variable, whether spatial or temporal, can be thought of as just another variable. Adding a lagged variable to MBC, therefore, is just MBC. As a result, I think the results that show the impacts of poorly conditioned matrices is the key takeaway here. One should parsimoniously add variables that are important to preserving the kind of variability that one is most interested in. For instance, if heatwaves are of interest, then one should emphasize temporal correlation. I'd like to see a bit more on the tradeoffs between emphasizing temporal vs spatial

correlation, in terms of choosing the dimensionality of the bias correction. The authors touch on this in the discussion, but some more specific guidance for making these choices would be helpful.

Response:

We agree on the fact that adding variables for MBC must be done wisely by end-users, and propose to add the following sentence in the discussion (in blue), on current L586 in the discussion (Section 6):

"More generally, for most MBCs, for a given number of statistical dimensions (e.g., number of grid cells), as going from a large (e.g., France) to a smaller (e.g., Brittany) area reduces the "effective dimension", it facilitates the multivariate corrections and therefore improves the results (e.g. compare Figs. 1, S1, 4, S4, 5 and S5). This raises the question of whether applying MBC on climate simulations over large geographical areas is justified, i.e. if it is worth striving for the correction of correlation structures between distant sites presenting weak statistical relationships, and, by doing so, taking the risk of losing global effectiveness of the BC methods. ***It also highlights the importance of choosing parsimoniously the variables to correct, in order to adjust dependence structures that are relevant without potential quality loss induced by additional (and unneeded) variables."***

Concerning the tradeoffs between emphasizing temporal vs spatial correlation, this is a relevant remark. However, providing guidance on the compromise between correcting temporal and spatial correlation would require additional evaluations by implementing another dimensional configuration in the study (for-example a 14d-version would be needed to correct autocorrelations of 2 physical variables until lag 7 at one given location). Although useful for end-users interested in correcting temporal correlations, it goes beyond the scope of this paper, and is left for further work.

Comment:

4) Given that the methods use covariance (which is the basis for Pearson's correlation) to constrain the bias correction, it seems somewhat inconsistent that Spearman's correlation is used to evaluate how well the various methods preserve the inter-variable"correlation". I understand the reasoning for using nonparametric correlation, but it also raises questions about what the goal of the bias correction should be. That is, should bias correction preserve covariance in instances where covariance can't reliably be estimated?

Response:

We disagree with the statement that all the MBC methods presented in the paper use covariance to constraint the correction. Actually, only the "MRec" method, based on a matrix recorrelation technique, uses explicitly covariance/Pearson's correlation to constrain the bias correction. Pearson coefficient measures the strength of the linear relationship between normally distributed variables. Arguably, precipitation is not normally distributed and the relationship between temperature and precipitation can be non-linear. In that sense, to evaluate inter-variable correlations, we think that it is more appropriate to use the Spearman correlation that does not require the assumption of normal distribution of the variables or linear relationship.

Moreover, the Spearman's (rank) correlation is a measure of dependence between 2 variables rid of their marginals. Univariate BC methods are supposed to adjust marginal distributions. Most univariate BC methods (as quantile-quantile like methods) do correct the marginal distributions but leave the dependence (i.e., copula) structure of the model data unchanged. Hence, when applying a multivariate BC, the expectation is to adjust not only the marginals but also the dependence between the variables of interest, regardless of the marginal biases or correctness. The Spearman's correlation is thus appropriate for this specific aspect.

With the aim of justifying better the use of Spearman's correlation for the evaluation of inter-variable dependence structure, we propose to rewrite part of the paragraph starting at L294 in the sub-section 5.2 as follows:

"*To evaluate inter-variable dependence structure, Spearman correlations between temperature and precipitation are computed at each grid cell to measure the monotonic relationship between the two physical variables.* ***Using rank correlation presents the particularity of not being value-dependent, i.e. it measures the dependence between two variables rid of their univariate distributions. As the goal when applying MBC is to adjust not only the univariate distributions but also the dependence structure between the variables of interest, Spearman's correlation is appropriate for this latter aspect. Moreover, this measure does not require any assumption on the distribution of the variables or their statistical relationships. It is hence appropriate for temperature and precipitation studies presenting extreme values and/or lower bound (Vrac and Friederichs, 2015).***"

Specific comments:

Comment:

1) The article is mostly well written, but there's some awkward phrasing here and there including the Abstract (e.g., "climate variables evolutions", "not well apprehended") and elsewhere ("permits to relate").

Response:

Following this comment, we propose the following corrections (in blue) :

L1 (in the Abstract) ***"Climate models are the major tools to study the climate system and its evolutions in the future."***

L9 (in the Abstract) replacing "*not well apprehended yet*" by *"**not yet fully apprehended**"*

L117 (in Section 3) *"permits to relate"* by *"**permits to link**"*

Comment:

2) "methodology" refers to the study of methods. The authors should just use "method" wherever they have "methodology"

Response:

The word ***"method"*** instead of *"methodology"* will be used for L13 and L14 in the Abstract.

Comment:

3) The last paragraph of the Introduction can be deleted (this structure is obvious).

Response:

We prefer to keep the last paragraph of the Introduction. Although the structure of the paper is obvious, it permits to explicitly outline the sections of the paper.

Comment:

4) Section 2: "RPC" should be RCP.

Response:

Addressed, thanks.

Comment:

5) The framework for the CDF-t method described in text and in Appendix A seems to be a generic accounting of quantile mapping. What I didn't see was information on the transfer function itself (i.e., what criteria determine the degree to which the two distributions are required to match).

Response:

CDF-t and quantile mapping have indeed a similar philosophy. For the period of calibration, the two methods are theoretically equivalent. However, the difference between CDF-t and quantile mapping lies in the correction of variables during the period of projection. CDF-t takes into account the change in the modelled CDFs from the calibration to the projection period, while quantile mapping projects directly the modelled values onto the CDF of the calibration period to compute quantiles.

More specifically, Appendix A does not correspond at all to a generic quantile-mapping. Equations (A1) to (A4) describe the construction of a transfer function allowing to go from the reference CDF F_{RC} (i.e., over the calibration time period) to an estimate of a projected pseudo-reference CDF F_{RP} (i.e., over the projection time period). This projected pseudo-reference CDF F_{RP} is then used in a second step, involving quantile-mapping. While a traditional quantile-mapping approach performed to correct a dataset X_{Mp} of simulations over the projection period will use the formulation $\widehat{X}_{Mp} = F_{Rc}^{-1}(F_{Mc}(X_{Mp}))$ to get a corrected value $\widehat{X}_{Mp}$ (i.e., based on 2 distributions characterizing the calibration period), the CDF-t method relies on the following formulation: $\widehat{X}_{Mp} = F_{Rp}^{-1}(F_{Mp}(X_{Mp}))$, where the 2 involved distributions characterize projected distributions. Those points are already mentioned in Appendix A. However, with the aim of clarifying this point, we propose to rewrite the paragraph starting at L690 of Appendix A as following:

"*Once F_{Rp} has been estimated, a simple quantile-quantile method is performed between F_{Rp} and F_{Mp} to derive the bias corrected time series of CDFs $\widehat{X}_{Mp}$ for the projection period* ***as following:***

$$\widehat{X}^{d}_{M_p}(t) = F_{Rp}^{-1}(F_{Mp}(X^{d}_{M_p}(t))). \quad \textbf{(A5)}$$

While a traditional quantile-mapping approach performed to correct a dataset X_{Mp} of simulations over the projection period will use the formulation $\widehat{X}^{d}_{M_p}(t) = F_{Rc}^{-1}(F_{Mc}(X^{d}_{M_p}(t)))$, (i.e., based on two distributions characterizing the calibration period), the CDF-t method relies on Eq. (A5) where the two involved distributions characterize projected distributions. By proceeding this way, CDF-t takes into account the potential evolution of CDFs of the model between the calibration and projection periods to adjust the projection period. CDF-t is applied independently for each of the D statistical dimensions and for both calibration and projection period to derive the final bias corrected outputs $\widehat{X}_{Mc}$ and $\widehat{X}_{Mp}$.

Comment:

6) Table 1 is a helpful summary of the bias-correction methods, but it's hard to match it up with the several pages of text and the Appendices to try to figure out what makes them distinct. There seems to be a gap between describing the general characteristics of the methods (Table 1) and the lengthy text-based descriptions. Please consider adding another table (or other information to Table 1) that helps to determine the specific attributes that makes the methods distinct.

Response: Agree. Thanks for this suggestion. In order to provide a helpful summary of the methods, we propose to add another table in Section 3 (Multivariate bias correction methods):

Table 1. Summary of attributes of the different bias-correction methods.

Characteristics	CDF-t	R2D2	dOTC	MBCn	MRec
Type of BC	1d-BC	MBC	MBC	MBC	MBC
Category of MBC approach	n.a.	Marginal/dependence	All-in-one	Marginal/dependence	All-in-one
Statistical technique	Non-stationary quantile mapping	Conditionnal resampling	Optimal transport	Iterative partial matrix recorrelation	Matrix recorrelation
Dependence structure	~ same as the model	~ same as the reference	Allows changes in the dep. struct.	Allows changes in the dep. struct.	Allows changes in the *Gaussian* dep. struct.
Conceptual feature	Deterministic	Deterministic and stochastic	Stochastic	Deterministic and stochastic	Deterministic

We also propose to modify the paragraph starting at L109 (Section 3) as follows:

*“**This section presents a brief description of the univariate BC method and the four multivariate BC methods implemented in this study. As a reminder, results from the univariate CDF-t method serve as a benchmark to measure the benefits of considering multivariate aspects in the correction procedure instead of using univariate BC methods. For sake of clarity, Table 1 provides a concise summary of the different attributes that make the BC methods distinct.**”*

Numbering of tables will be changed accordingly.

Comment:

7) At the beginning of Section 4, it would be helpful for the authors to outline what using the three different designs (2d, spatial, and full) aims to accomplish. Which approach has been more commonly used in the literature? Some of the dimensionality tradeoffs could be introduced here as well.

Response:

The information concerning the aim of each dimensional design (2d, Spatial and Full) is already formulated at the beginning of Section 4 in the form of bullet points (starting at L210 in Section 4 - Design of Experiments). However, we propose the following clarifications (in blue) at L206 (Section 4 - Design of Experiments):

*« We tested and assessed this approach for each method, but also expanded the study to include high-dimensional configurations of MBC to adjust spatial and full (i.e. spatial and inter-variable jointly) dependence structures of climate simulations. **Depending on the dimensional configurations, the objectives of corrections for multivariate properties differ. Including different dimensional***

versions in the study will permit to better highlight the potential losses and benefits associated with them. *Therefore, in the following, each of the four MBC methods is applied according to the three following configurations: »*

Concerning the most commonly used approach, in our opinion, L204 of Section 4 is quite clear: It indicates that, in most cases, inter-variable configurations (i.e. in our study referred to as 2d-version) are applied in the literature and two papers are cited (Meyer et al., 2019 ; Guo et al., 2019) to illustrate this point. Of course, as shown by the results later in the article, this choice of inter-variable configuration can have important consequences on spatial and temporal dependencies.