“Is time a variable like the others in multivariate statistical downscaling and bias correction?”, reviewer 2

Yoann Robin & Mathieu Vrac, https://doi.org/10.5194/esd-2021-12

General comments

This manuscript deals with a new approach (TSMBC) of how to incorporate the time as additional variable into a multivariable bias correction. The approach can be conducted with existing multivariate BC methods such as MBCn, R2D2, or MRrec. Here, the dOTC approach is followed and the results are compared to a “naive” method, the “Random Bias Correction” (RBC).

The method is first tested on a synthetic dataset, following a VAR process, before applying it to “real” climate data, based on a pseudo reality approach, i.e. treating the RCM results as observations.

The approach could potentially be interesting and innovative. It seems that this is the first time that the time is treated as separate variable in the bias correction. However, I have some doubts that the results are reliable for the application with the real data case (see detailed comments below). Moreover, I think that the evaluation of the TSMBC using synthetic data based on the VAR process is of limited value. It did not convince me technically and scientifically, nor did it help me to better understand the proposed procedure.

On the other hand, more information is required to understand the potential value of the TSMBC. Authors did not convincingly present the methodological background. Critical questions remain unanswered, e.g. what is a VAR process? How is the sampling from the VAR process done? How does the dOTC works?

The Wasserstein metric is also not well introduced in the method section.

General response:

First, we would like to thank this anonymous reviewer for her/his thorough reading and interesting comments. We tried to take them into account and we provide point-by-point responses below in blue.
More specifically, questions/remarks regarding synthetic data generated with the VAR process, the dOTC methodology, the Wasserstein metric and the naive RBC method are treated in the responses to comment 6 of the reviewer.

Specific comments

Comment 1

It remains spurious how and why the increase of the numbers of dimensions (could be time lags or other "variables") affects the stability of the approach. It is just mentioned that the dimensionality should not exceed 10.

Response:

This is related to the well-known problem of "curse of dimensionality": having ~2500 values in 4576 dimensions for TSMBC10 / S2V indicates that we may not have enough data to explore such a high-dimensional space and, thus, that the MBC inference/procedure performed by dOTC may not be robust. However, even in this TSMBC-10/S2V configuration, the "shape" of the DCP set appears improved (first normalization, Fig. 8) whereas a bias appears in the DCP set when the intensity of the correlations are also accounted for (second normalization, Fig. 9).

Another potential explanation is also the linear dependencies arising from our datasets. Two kinds of linear dependency might appear:

- The linear dependence between two "close" grid points (especially for temperatures). However, this effect seems limited, as dOTC works correctly at lag 0.
- The linear dependence in the lagged matrix when duplicating and shifting the columns. However, this is difficult to distinguish from the "curse of dimensionality" problem.

At the end of the section 4.2 we have added the following text:

"*One potential explanation for this is the well-known problem of "curse of dimensionality" (e.g.,Wilcox, 1961; Finney, 1977): having 2500 values in 4576 dimensions for TSMBC10 / S2V indicates that we may not have enough data to explore such a high-dimensional space and, thus, that the MBC inference/procedure performed by dOTC may not be robust. In addition, an increased number of dimensions could potentially lead to two types of linear dependencies that could interfere with the underlying MBC method being used (dOTC): (i) a linear dependence between two "close" grid points (especially for temperature), although this effect seems limited as dOTC performed correctly at lag0; and (ii) a linear dependence in the lagged matrix by duplicating and shifting the columns. However, the latter is difficult to distinguish from the curse of dimensionality problem.*"

Added references:

Finney, D. J.: Dimensions of Statistics, Journal of the Royal Statistical Society: Series C (Applied Statistics), 26, 285–289, https://doi.org/https://doi.org/10.2307/2346969, https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2346969, 1977.

Wilcox, R. H.: Adaptive control processes—A guided tour, by Richard Bellman, Princeton University Press, Princeton, New Jersey, 1961, 255pp., Naval Research Logistics Quarterly, 8, 315–316, https://doi.org/https://doi.org/10.1002/nav.3800080314, https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800080314, 1961.

Comment 2

I have some concerns about applying a BC using climate simulations (based on GCMs and not on reanalysis data) if the temporal sequence of variables is addressed, however, in this case I think it would be acceptable, since the reference is not observation data but downscaled results of the same forcing GCM.

Response:

Indeed, references are regional climate simulations forced by the same GCM to be downscaled/bias corrected. This kind of "perfect model experiment", considering simulations as "pseudo-observations", is a common approach to assess downscaling / bias correction methods (see, e.g. Charles et al. 2004[1]; Vrac et al., 2007[2], Frost et al. 2011[3]; Bürger et al. 2012[4]; Grouillet et al. 2016[5]).

A clarification has been added in Section 2.1:

"*This kind of "perfect model experiment", i.e., considering simulations as "pseudo-observations", is now a common approach to assess a downscaling / bias correction methods, (see, e.g. Charles et al. 2004; Vrac et al., 2007, Frost et al. 2011; Bürger et al. 2012; Grouillet et al. 2016).*"

Comment 3

My main concern stems from Figure 1 (right, top line). It seems that the mean precipitation and temperature fields do not correspond to the coast line, as I would strongly assume. Due to the coarse resolution, you would expect some distortions in the overlay, but this looks really erroneous. It seems that the projection of GCM and RCM is wrong, it could be reversed left to right.

Response:

To make sure that we did not make any mistake, we have plotted the equivalent of figure 1 but at the scale of France (see figure at the end of this file). When looking at the maps for France, it is clear that the main geographical patterns are correctly located (e.g., the Alps and the Pyrenees). The patterns visible in figure 1 of the article are also visible here. This shows that Figure 1 is correct.

[1] https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.1418
[2] https://doi.org/10.1029/2007GL030295
[3] https://doi.org/10.1016/j.jhydrol.2011.06.021
[4] https://doi.org/10.1175/JCLI-D-11-00408.1
[5] https://doi.org/10.5194/hess-20-1031-2016

Comment 4

Unfortunately, this would have tremendous impacts on the results and interpretations in the following (e.g. the spatial dependencies given in Figure 6). For instance, please explain the statement in lines 300-302. Why is the evolution of GCM variables so different from that of the RCM? Indeed, the RCM includes more spatially-detailed "processes", but is still driven by the GCM. Since the domain of the RCM is rather small, the impact of the forcing is expected to dominate the RCM simulations.

Response:

The area we have extracted is small, but the boundary of the RCM (i.e., the domain over which the RCM simulations are performed) is Europe, and more precisely the EURO-CORDEX domain, which is much larger than the south-east France domain shown in our article. Consequently, in our domain shown in Fig. 1, the impact of the GCM forcing is not dominating. Also, our domain is framed by 3 mountain ranges (Alps, Pyrenees, Massif Central) badly represented by GCMs while RCMs tend to improve their influences on climate. Hence,the RCM internal dynamics is certainly stronger than that of the GCM.

The following sentence has been added to section 2.1 for clarification:

"*Note that the extracted region of interest is small in comparison to the initial EURO-CORDEX domain (Jacob et al., 2014) over which the RCM simulations were performed.*"

Comment 5

Moreover, I cannot understand the differences the different performances of the calibration and the projection period (Figure 4 & 5). I would expect very similar performances. What is leading to the big discrepancies between the different periods?

Response:

In a cross-validation setting (i.e., calibration done on a dataset and evaluation/projection performed on a different dataset), the references are not used to fit the model (here, the correction) over the projection period. Therefore, it is expected to have lower quality results over the projection period.

Generally, in a cross-validation within a bias correction context, the quality of the results depends on two elements:

- The ability of the method itself to perform a relevant bias correction;
- The difference between the evolution of the model to be corrected and the evolution of the references. In other words, if the climate change (between calibration and projection) from the model simulations to correct is in disagreement with that from the references, the bias correction method will mostly preserve (i.e., not correct) this "bias of evolution".

For these reasons, in many studies, only the results over the projection period are shown. Here, we also wanted to incorporate results over calibration for comparison.

Comment 6

The evaluation results of the TSMBC using synthetic data based on the VAR process are not convincing (whole section 3) and – at least for me – not fully understandable. For the revisions, I would suggest to leave out this synthetic exercise. Rather, I would focus on better explain the applied methods, i.e. the bias corrections approach applied here (dOTC), the Wetterstein-based metric, and how the naïve RBC (reference approach) works. I am also wondering if this naïve approach is really suitable for fair comparison.

Response:

The use of synthetic data from a VAR process was mainly (i) to test our TSMBC procedure on understandable data and (ii) to explore the influence of the TSMBC parameters on the results: namely, the choice of a starting row and the effect of the underlying bias correction method within the "row reconstruction method".

A "Vector AutoRegressIve" (VAR) process is a multivariate AutoRegressive (AR) process (i.e., allowing multivariate data) modelling the statistical link between the components of a vector (i.e., multivariate data) when they change in time.

The VAR process is very helpful in this case because the lag is fixed before the experiment, whereas with climate data only an estimation of the maximum lag must be done. So our results about the choice of the starting row and the importance of the dOTC method would not be well argued without this step. The definition of a VAR process is given by Equation (2), and the sampling from this kind of process is performed with this equation: In the case of an order-s VAR process, the first s vectors (i.e., from time 1 to s) are fixed and the VAR process allows generating new values for the components of the vector at time s+1.

The text now reads in section 2.1, before Equation (2):

"*A VAR process is a multivariate AutoRegressive (AR) process (i.e., allowing multivariate data) modelling the statistical link between the components of a vector (i.e., multivariate data) when they change in time. In the following, a VAR is used to generate multivariate time series…*"

and a few lines after Equation (2):

"*The sampling is performed based on Equation (2): the first s vectors (i.e., from time 1 to s) are initialised and the VAR process allows generating new values for the d components of the vector at time s+1.*"

Regarding the description of the bias correction methods used and the Wasserstein distance, two appendices have been written instead of the previous appendix A: the first one (new appendix A) describes how a bias correction method can be considered as a probability distribution and how dOTC works in this context; the second one (new appendix B) describes the Wasserstein metric.

These appendices are not cited here for sake of space.

Regarding RBC: As explained at the beginning of section 3, the RBC method is necessary to distinguish which part of the correction results comes from the reconstruction by row, and which part from the underlying bias correction method. Moreover, the “Random Bias Correction” (RBC) method has been more detailed at the beginning of section 3:

“*Because the reconstruction step preserves the dependence structure, we propose to test which part of the correction is due to the underlying method (here dOTC), and which part is due to the reconstruction. To do so, a second underlying bias correction method is then used as a benchmark. It corresponds to a very naive method: the correction is randomly drawn from the reference dataset, i.e., for any x ∈ X, the correction is given by a random value y generated according to the distribution of Y. In practice, values from Y are resampled.*”

Comment 7

The introduction should be improved, e.g. the statement given in line 28 (… (ii) from inherent biases in the model simulations.") is not very helpful. Potential reasons for the biases shall be mentioned. More and more recent references are required, e.g. for strong statements given in lines 39 & 40.

Response:

The main reason for biases is the inability of models (whether it is a GCM or a RCM) to exactly reproduce the observations, due to e.g. parameterizations, or processes not or poorly represented. The sentence regarding “inherent biases” has then be rewritten to mention reasons for the biases as:

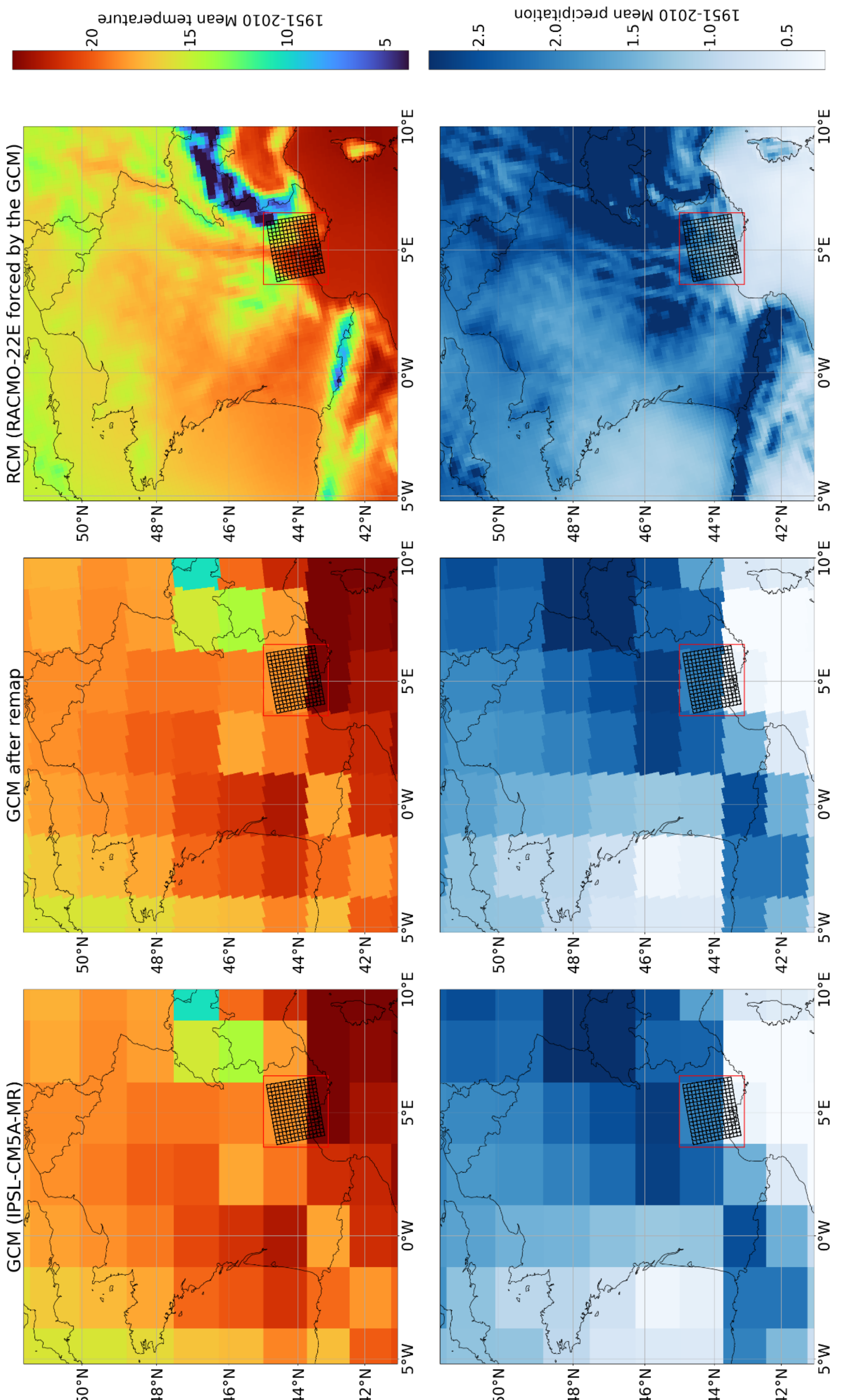
“*... (ii) from inherent biases in the model simulations, due to parameterizations, or processes not or poorly represented (e.g., McFarlane, 2011).*”

and the following reference has been added:

“McFarlane, N.: Parameterizations: representing key processes in climate models without resolving them, WIREs Climate Change, 2, 482–497, https://doi.org/https://doi.org/10.1002/wcc.122, https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.122, 2011.”

Regarding the fact that “*the obtained downscaled/bias corrected climate data can then serve as input into impact models*” (lines 39&40 of the initially submitted article), this is indeed a common assumption made in many impact studies. More references have been given to support this statement:

“*The obtained downscaled/bias corrected climate data can then serve as input into impact models (e.g., Teutschbein and Seibert, 2012; Galmarini et al., 2019; Bartók et al., 2019; Chen et al.,2021, among many others*).”



<u>Figure to answer comment 3 from reviewer 2</u>: Maps equivalent to those given in Figure 1 of the article but at the scale of France.