“Is time a variable like the others in multivariate statistical downscaling and bias correction?”, reviewer 1

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

General comments

In this manuscript, a new method of incorporating the temporal variable into a multivariable bias correction is introduced with sufficient motivations and with a thorough and clear description. This new method is versatile in that it can work with any existing MBC's and this is demonstrated via applying it to dOTC and to a more naive method they call Random Bias Correction. The method is first tested on a synthetic dataset for an explorative tuning of the parameters, then to a real dataset. A few points from the analyses from the real data experiments are unconvincing (this will be touched in the specific comments), but most results are well-supported. A new generalizable metric is introduced for measuring bias reduction relative to some ground-truth dataset but its benefits and shortcomings could be discussed further.

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.

Specific comments

Comment 1

In section 2.2, the concept of reconstruction by rows is introduced. Reconstruction by rows certainly seem to perform better than reconstruction by columns. It is asserted here that many reconstructions are possible and that these are determined by the "starting row". Starting the n^th row for $1< n < l$ for some lag l omits the first $n-1$ values, which are clearly needed in the final reconstruction. It is possible that those $n-1$ values are repeated more than once in the lagged matrix and a more specific description of how to include these values is needed.

Response:

We thank the reviewer for this question that allows us to clarify the methodology.

Actually, based on our "reconstruction by rows", it is not possible to repeat values. In the case considering *l* lags, this method by rows jumps (concatenate) from one given row (e.g., nth row) to another row located *l*+1 rows after (i.e., (n+*l*+1)th row) avoiding to repeat values. However, as noted by the reviewer, starting at row n > 1 will omit the n - 1 first values. Generally, this has no impact on the corrections in terms of their statistical properties as n is usually very small compared to the length of the time series. This is validated by the results of the test performed in section 3.1 with a VAR process. To clarify the reconstruction method by rows and the fact that no repeated values are possible, we have added the following sentence in section 2.2:

"*However, even if the values are repeated in the lagged matrix, no repeated values can appear in the final reconstruction. [...] Moreover, the choice of a starting row $r>1$ omits the first $r-1$ values in the final reconstruction. This leads us to wonder about the influence of the choice of the starting row.*".

Nevertheless, as no values are omitted when starting at first row for the reconstruction, this is a logical and practical choice. This is now clarified in the "Conclusions and discussion" section 5 as follows:

"I*n the case of a starting row $n>1$ (i.e., with a lag $s>0$), the reconstruction will omit the first $n-1$ time steps. In order to have as many reconstructed time steps as in the model simulations to correct, it is possible to sample from the first $s-1$ row(s) of the fully corrected lagged matrix, allowing to complete first $n-1$ time steps of the reconstruction matrix. However, as no values are omitted when starting at first row ($r=1$) for the reconstruction, this is a logical and practical choice*".

Comment 2

Section 3.1 asserts that the starting row has little impact on the overall bias correction performance, and this is attributed to the high correlations of the results of the TSMBC method to the biased data matrix X, as well as the high correlations between the results of the TSMBC methods with varying starting rows (as shown in Figure 3). In figure 3, it is also shown that all TSMBC results have very low correlations with Y, the reference matrix. Shouldn't the results of TSMBC be "corrected" and therefore aspire to exhibit higher correlations with Y more than X?

Response:

Regarding the conclusion that the choice of the starting row has little impact on the BC performance: we do NOT conclude this because of the high correlations of the TSMBC results to the biased data matrix X. We conclude this from the high correlations between the results obtained from different starting rows. Indeed, this indicates that, whatever the chosen starting row, the results are very close to each other. The resulting high correlations with data matrix X is an effect of the dOTC method, which tries to preserve as much as possible the temporal properties of the model simulations to be corrected (Robin et al., 2019; François et al., 2020).

This is now clarified in section 3.1, as follows:

> "*We can see for TSMBC(dOTC) that all corrections are highly correlated between them -- with values close to 1 -- whereas for TSMBC(RBC) no significant correlation appears. This indicates that, whatever the chosen starting row for TSMBC, the results are very close to each other. Remark that for TSMBC(dOTC) the corrections stay highly correlated with $\mathbf{X}$. It is an effect of the dOTC method, which tries to preserve as much as possible the temporal properties of the model simulations to be corrected(Robin et al., 2019; François et al., 2020).*"

Regarding whether or not the TSMBC results should aspire to exhibit higher correlations with Y than with X: Actually, the answer is "no". The goal of bias correction, be it univariate or multivariate and with or without including a correction of the temporal properties, is to have corrected simulations that have statistical properties similar to those from the reference data. Hence, the goal is not to get corrected data correlated to observations. Raw climate simulations and references are usually uncorrelated (except via seasonal cycles and potential trends). As some bias correction methods preserve the rank chronology (i.e., the temporal properties) of the simulations --- as is the case, to some extent, for the dOTC method (see Robin et al., 2019) ---, there is no reason for the bias corrected data to be correlated to the reference. In other words, if the BC procedure is efficient, corrected and reference time series can be seen as generated based on the same statistical distributions and/or properties but independently. Hence, they are not correlated.

Clarifications have been brought to the manuscript by adding the following text into section 3.1:

> "*On the other hand, no correlation appears with Y. This was expected. Indeed, there is no reason for the bias corrected data to be correlated to the reference. If the BC procedure is efficient, corrected and reference time series can be seen as generated based on the same statistical distributions and/or properties but independently. Hence, they are not correlated.*"

Comment 3

The major aspect of TSMBC is that by adding lagged versions of the original time series data, the data is augmented to include the temporal variable as just another variable. This initial mapping from a dimension of size \$N_X\times d\$ to \$(N_X-s)\times d(s+1)\$ is injective but the inverse mapping is not. The authors chose to use a simple reconstruction that only relies on one extra parameter, the starting row, as a way to choose this inverse mapping, and assert in section 3.1 that the choice of the starting row does not have a big impact. Given that the analysis of figure 3 is unconvincing, it may be important to more carefully consider how to design the inverse mapping. For example, what is the variance of the repeated values? For TSMBC with lag \$s\$, there are some time indices that are repeated \$s+1\$ times total in the reconstruction. Are those \$s+1\$ values all very close to each other? If not, should some averaging scheme be used? If not, what does the variability in the reconstruction at some time index indicate about whether it should be trusted?

Response:

It is true that the initial mapping is injective and that, in general, the inverse mapping is not. However, with the suggested reconstruction (i.e., inverse mapping), when a starting row is given, the inverse mapping gives a unique time series and is thus injective.

This is now explained in section 2.2 of the updated article:

"*It is also worth noting the initial mapping (i.e.,going from $\bf{X}$ to $\bf{M_X}$ is injective, and that, in general, an inverse mapping is not. However, with the suggested reconstruction (i.e., inverse mapping), when a starting row is chosen, the inverse mapping gives a unique time series and is thus injective.*"

Moreover, as explained in response to comment 1, based on our suggested reconstruction method by row, there are no repeated values. Hence, statistics asked by the reviewer cannot be computed.

Comment 4

Regarding the analysis of figure 8 (pg 13, lines 372-389): The statement in line 374-375 "Generally speaking, for a specific configuration of the method (i.e., L1V, L2V, S1V or S2V), TSMBC (5 or 10) is better than dOTC that does not account for temporal properties. " is not well supported by figure 8. Apart from the plots for tas/tas (first column in figure 8), it is difficult to see that the TSMBC cells show darker (higher BR_w) values than the naive comparison dOTC. In addition, shouldn't the 3 methods (dOTC, TSMBC5, TSMBC10) all show the same value/color for lag 0 for each L1V, L2V, S1V, and S2V? What are some reasons they are not?

Response:

We thank this reviewer for this precious remark. Indeed, TSMBC (5 or 10) is not always better than dOTC. The text has then been modified to describe the results more precisely in section 4.2:

"*Generally speaking, for the local configurations (L1V and L2V), TSMBC (5 or 10) is better than dOTC that does not account for temporal properties. This is true for almost all lags $>$0 and any $BR_{\mathcal{W}}$ matrix (tas/tas, tas/pr, pr/tas, pr/pr). However, for the spatial configurations (S1V and S2V), TSMBC does not seem to provide better results than dOTC, except for the tas/tas matrix where TSMBC strongly improves dOTC.*"

Note also that we have changed the colormap used to better show the details.

Regarding whether or not the 3 methods (dOTC, TSMBC-5 and TSMBC-10) should give the same results for lag 0: The three methods dOTC, TSMBC-5 and TMSBC-10 works differently and with different variables due to the different configurations. For L1V, dOTC works in a univariate context, TSMBC-5 in a 6-dimensional context and TSMBC-10 in a 11-dim context. The number of variables managed by the 3 methods is different for each configuration and increases up to S2V where dOTC works with 416 variables, TSMBC-5 with 2496 variables and TSMBC-10 with 4576 variables (see Table 1). Thus a variability will necessarily appear in the various corrections. However, we can see in Fig. 8 that, for lag 0, the improvements are of the same order for all 3 methods.

Comment 5

One justification for why TSMBC10 performs worse than TSMBC5 is given by the fact that the inflated data size $(N_X-10)\times d(10+1)$ results in a higher complexity method. In line 412-413, it is stated "The increase in the complexity (i.e., the number of dimensions) of the method is made at the expense of the quality of the results." This is a vague statement and could be made stronger with more specific ideas. For example, the increased number of dimensions could potentially lead to linear dependence which then could interfere with the underlying MBC method being used. There could be some other ways that the increased complexity could have negative effects, and they should be discussed in more detail. Given the size of the problem, numerical instability should probably be ruled out.

Response:

We did not want to suggest that the problem was due to a numerical instability, but rather that 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).

Regarding the linear dependency, two kinds of dependency might indeed 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 6

Regarding the BR_{\Kappa} metric. One downside of this metric is explained well in the conclusions, in line 458-461: "However, biases in the intensities of the (intervariable, inter-site or temporal) correlations might remain. This is typically related to very small differences between two Wasserstein distances very close to zero: if the raw simulations already have a DCP set close to the reference, its Wasserstein distance will be near zero. Therefore, the relative reduction of bias BR can be strongly negative, even though the absolute difference is potentially very small."

Maybe this point should be suggested when the metric is first introduced in section 4.1.

Response:

This point has been added at the beginning of section 4.1, after Equation (6) :

"*Note that if the raw simulations already have a DCP set close to the reference, its Wasserstein distance will be near zero. If the correction gives also a Wasserstein distance very close to zero, then the relative reduction of bias BR_κ can have a very strong negative value if $\kappa(\bf{Z}) > \kappa(GCM)$, even if the absolute difference (i.e., $\kappa(\bf{Z}) - \kappa(GCM)$) is potentially very small.*"

Note also that the (previous) appendix A has been replaced by two appendices: 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.

Technical comments

1. *Should "corrected" in line 227 be "correlated" instead?*
2. *Line 289 should have [-\infty,1] instead of]-\infty, 1]*

Response:

These two technical comments have been corrected.