

Earth Syst. Dynam. Discuss., referee comment RC1  
<https://doi.org/10.5194/esd-2021-12-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on esd-2021-12

Anonymous Referee #1

---

Referee comment on "Is time a variable like the others in multivariate statistical downscaling and bias correction?" by Yoann Robin and Mathieu Vrac, Earth Syst. Dynam. Discuss., <https://doi.org/10.5194/esd-2021-12-RC1>, 2021

---

### 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.

### Specific Comments:

- 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  $n^{\text{th}}$  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.
- 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  $X$ ?

- 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?
- 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  $\tau_s/\tau_s$  (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?
- 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.
- 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.

## Technical Corrections

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