Dear prof. Sharma,

Thank you for your constructive comments to our manuscript. The point of the original paper was not against the multivariate bias correction methods correcting the auto- and cross dependence as such, rather we were demonstrating how few data points may alter the estimated dependence structure and providing a method to minimize these effects. We call these points dependence outliers and although we understand that they are important e.g. for hydrological impact assessment, we show that their presence may distort the results of bias correction.

In the main point of the review a demonstration of the effect of outliers on the multivariate bias-correction methods was requested. Furthermore, several additional comments were included in the review, related to various methodological points. All points are addressed bellow. Following the comments, our reply is divided into two parts - the main point and detailed comments.

MAIN POINT

Comment: ***My request to the authors is to use the multivariate bias correction software now publically available and described in [Mehrotra, R., F. Johnson, and A. Sharma (2018), A software toolkit for correcting systematic biases in climate model simulations, Environmental Modelling and Software, 104, 130-152, doi:10.1016/j.envsoft.2018.02.010.] to show the impact these robust correlation metrics have on results.***

As requested, we analysed the effect of outliers using the recommended software. Since observed data for our study area that are available to authors differ significantly in their spatial resolution from the used RCM data, which might raise additional discussion on the effects on the correction results, we decided to use the data attached as an example to the recommended software toolkit, in particular the data from the example 3. The provided observed data consists of monthly rainfall values for 15 locations covering the period from 1921 to 1990. The model data consists of two datasets originating from a univariate rainfall generator, which does not reproduce the spatiotemporal correlation structure of the observations. Both model datasets cover the same period as observations, the first dataset is used for the calibration of the bias correction, the second for its validation.

To compare the results of the bias correction with and without outliers, two different model calibration datasets were used. The first was an original dataset from the example 3. The second was the same dataset, in which one value in the time series of Variable1 (i.e. rainfall for the station Oberon) was changed:

ORIGINAL DATA

Year	Month	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5
1975	6	48.5	115.1	24.5	4.9	1525.3

DATA WITH OUTLIER

Year	Month	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5
1975	6	$\boxed{900}$	115.1	24.5	4.9	1525.3

The time 6 / 1975 was chosen because there is another large value in this month – the value 1525.3 mm in the time series for the Variable 5, as seen from the figure above. The following table shows the lag-0 cross-correlations between variables 1 and 5 (denoted hereafter as $r_{1,5}$) for the observed and model validation data (which was not modified) and for the model calibration data (with and without outlier). Only the values for the month 6 were included, since the effect of the outlier is more evident in such case. As seen from the table, the outlier strongly affects the value of $r_{1,5}$:

Data	$r_{1,5}$
Observed	0.84
Model calibration - original	-0.05
Model calibration - with outlier	0.76
Model validation	-0.15

Four different bias corrections (denoted hereafter as BC1 … BC4) were performed using the model calibration data with and without outlier. The settings of particular corrections are in the following table:

Correction	Method	Nesting
BC1	MRNBC	no (monthly data only)
BC2	MRQNBC	no (monthly data only)
BC3	MRNBC	monthly, quarterly, annual (3 iterations)
BC4	MRQNBC	monthly, quarterly, annual (3 iterations)

In all cases (BC1 – BC4) the following statistics were corrected for all time steps included in the nesting scheme:

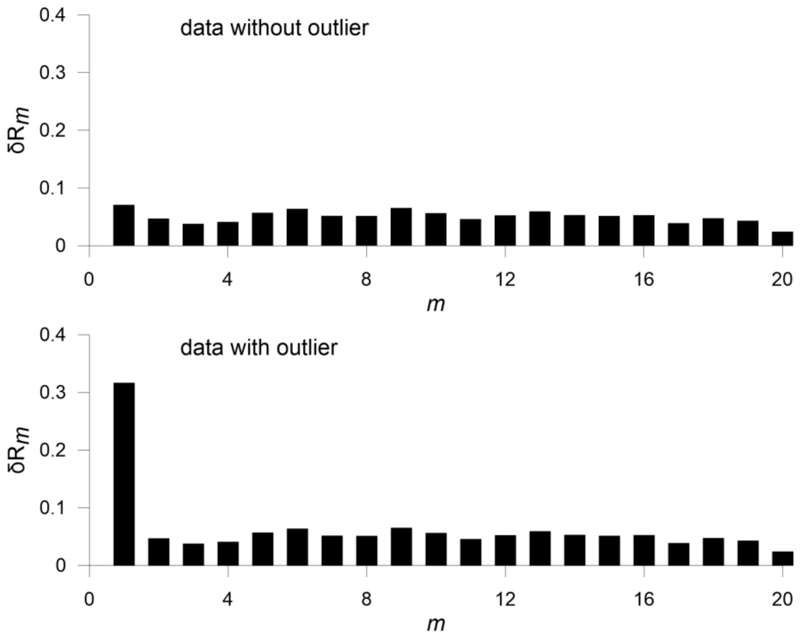
- mean
- sd/distribution
- lag-1 auto-correlation
- lag-0 cross-correlation.

The results of the corrections ($r_{1,5}$) are presented in the following table. We note that all bias corrections were calculated with complete datasets; nevertheless the results are presented using the data for the month 6 only, since the effect of outlier is more evident in such sub-datasets:

Correction	Outlier in the model calibration data?	$r_{1,5}$	
		Corrected data - calibration	Corrected data - validation
BC 1	no	0.803	0.745
	yes	0.830	0.283
BC 2	no	0.794	0.660
	yes	0.820	0.521
BC 3	no	0.833	0.792
	yes	0.835	0.290
BC 4	no	0.829	0.670
	yes	0.835	0.557

The table shows that the outlier strongly affects the results, the correction schemes with MRNBC method are affected more distinctly. Although the model data from the calibration and validation period originate from the same population, their correlation coefficients differ strongly *due to one outlying point*. This strong difference consequently affects the correction procedures.

The outlier presented above can be simply detected by the method proposed in our manuscript. The following figure compares the plots of $\delta\mathbf{R}_m$ (calculated for the complete 15-dimensional data) for the model calibration data without and with outlier:



The procedure should be performed in two steps. In the first step the plot of $\delta\mathbf{R}_m$ for several outliers (say 20 as in the figure above) is obtained from the complete data. In the second step the plot is assessed and the most significant outliers should be removed from the data. The source codes attached with the submitted paper can simply perform such operations. Certainly, our procedure is not "omnipotent". The effect of outliers can be more subtle than in the example presented above, nevertheless it can detect at least the most significant outliers.

Finally let us present another short example. Two synthetic outliers were now introduced into the time series of Variable 1 in the *observed* data:

ORIGINAL DATA

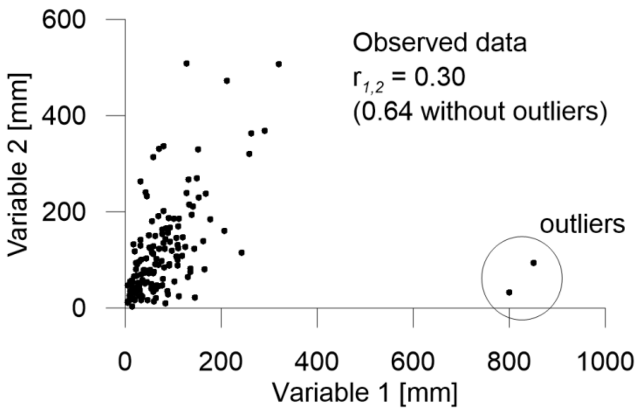
Year	Month	Variable 1
1922	4	47.5
1922	5	35.5

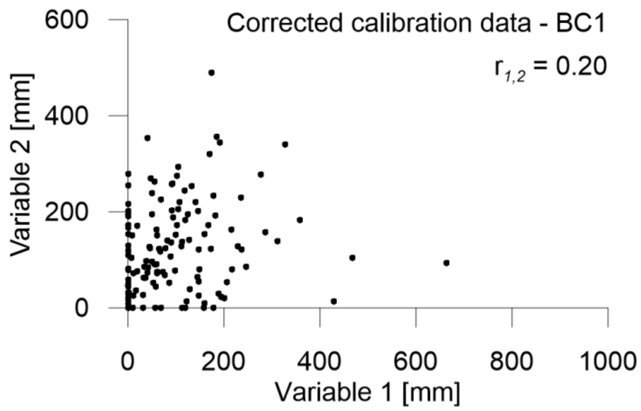
OUTLIERS

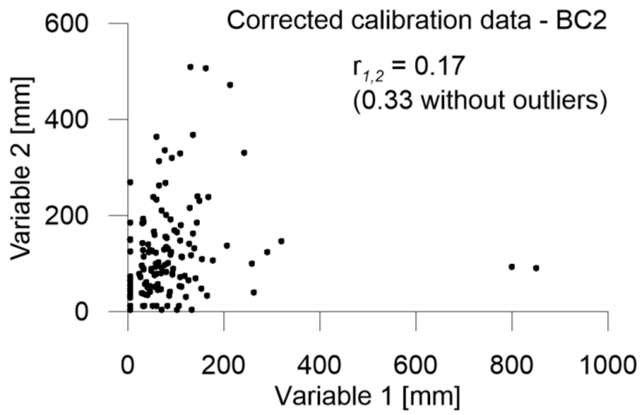
Year	Month	Variable 1
1922	4	800
1922	5	850

The values 800 and 850 mm are indisputably large, but they are not over physical limits (even values over 1000 mm can be found in the observed dataset). When only the values from months 4 and 5 are considered, the outliers change the cross-correlation between variables 1 and 2 (denoted hereafter as $r_{1,2}$) from 0.65 to 0.30 in the observed dataset.

The situation slightly differs from the previous case, when the outlier was in the model data. Now the corrections seemingly perform well, because the observed and corrected data show similar correlations. The problem emerges when the data from the Variable 1 are plotted against the data from the Variable 2. The following figure depicts such plots for the observed data and for the corrected calibration data (for BC1 and BC2):

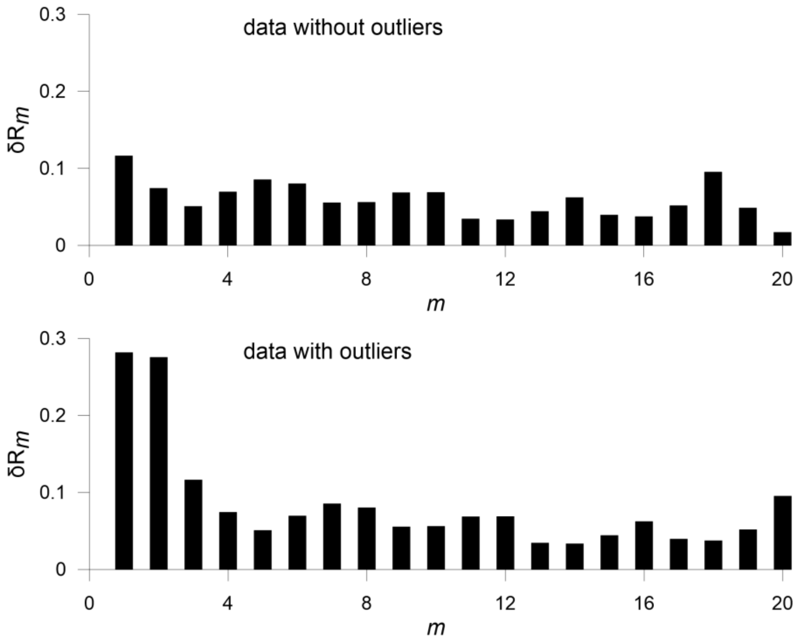






The figure shows that although the corrected and observed data correspond in their correlation coefficients (more or less precisely), the internal configuration of the corrected data differ significantly from the observations (note e.g. the concentration of points with Variable 1 = 0 and Variable 2 >> 0 and vice-versa, or the spread of the point cloud), which can affect hydrological behavior of the area. The reason is the same as in the previous case – the correlation coefficient 0.30 does not correspond to the configuration of the observed data, because it is strongly affected by two outliers. Nevertheless the value 0.30 is used in the bias correction procedure, which leads to misleading results. We note that the cross-correlations between the variable 1 and other variables are affected similarly as $r_{1,2}$.

The plots of $\delta\mathbf{R}_m$ (calculated for the complete 15-dimensional data) for the observed data without and with outliers indicates the presence of outliers:



DETAILED COMMENTS:

Comment: ***l79 - Using the block approach will alter the lag-one correlation at the end of year boundaries. I presume the impact will not be much but should be stated by the authors. On the same point, I would expect the cross dependence to remain unchanged, and the lag 1 correlation to only slightly be changed.***

Reply: We are aware of this problem; the calculations were designed such that the lag-1 autocorrelations were not affected in this manner. This point is briefly mentioned in the methodological part of the submitted paper (line 87), where the following sentence can be found "*The joints of the adjacent blocks were not included in the calculation*". Nevertheless, prof. Pegram in his review mentioned that this sentence is not intelligible. As a reply to his comment we suggested it´s removing, but we can reformulate the sentence, for example as follows:

"*Due to random selection of the blocks the beginning part of blocks is independent on the end of the previous blocks. To minimize bias introduced by block resampling, data that are potentially influenced (joints of the adjacent blocks) were not considered for the calculation of the serial correlation*".

Comment: ***And I am unable to figure out how these confidence intervals are finally used? Were all the correlations from the raw data and the resampled opnes pooled in deriving the results in Figs 3 and 4? Usually one does bootstrap tests to assess the significance of correlation from zero - here it seems the idea is to assess the significance of correlation from what it would be if the year to year dependence is made null. Some clarification is needed.***

Reply: The procedure for estimation of confidence intervals is standard [see e.g. Davison and Hinkley 1997] and relies on sampling with replacement of annual blocks of data. The resampling of blocks is done to preserve seasonal variation of rainfall. It is true that year-to-year dependence of rainfall is ignored, but we do not expect that this would significantly affect the estimates of confidence intervals of dependence indicators. In the original manuscript we did not perform real test on the significance of changes in auto- and cross dependence, instead we only visually assessed the overlap of estimated confidence intervals for correlations and autocorrelations in figs. 4 a 5 (boxplots in fig. 3 represent distribution of indicators between grid boxes). If the confidence intervals overlap, then the changes are not significant. Standard test would be easily performed e.g. by subtracting the estimated correlations for control period from those for the future period. If the confidence interval of this difference contains zero, then the change is not statistically significant. We will modify the description if we are invited to revise the manuscript. We can also quantify the significance explicitly.

Comment*: l98 - it would be nice to know what is the fraction of zeroes and non-zeroes in the data used, and how that might be impacting the binary cross-correlation results. here. From my experience, storms in warmer climates are getting smaller in size, hence the fraction of zeroes is increasing.*

Reply: We calculated the fractions of zeroes in all time series. Two different thresholds were applied to determine dry days: 0 mm and 0.1 mm. The results for the model 1A are presented in the following table:

grid-box	threshold 0.0 mm			threshold 0.1 mm		
	historical	Future	difference	historical	future	Difference
1	0.267	0.300	0.033	0.379	0.406	0.027
2	0.259	0.293	0.034	0.367	0.396	0.029
3	0.252	0.287	0.036	0.359	0.387	0.028
4	0.248	0.287	0.039	0.358	0.387	0.028
5	0.267	0.302	0.035	0.388	0.413	0.025
6	0.261	0.298	0.037	0.378	0.407	0.028
7	0.255	0.292	0.037	0.371	0.397	0.026
8	0.250	0.290	0.040	0.368	0.396	0.028
9	0.271	0.310	0.039	0.398	0.425	0.027
10	0.268	0.306	0.038	0.390	0.416	0.026
11	0.261	0.299	0.037	0.377	0.407	0.030
12	0.257	0.296	0.039	0.376	0.405	0.029

As seen from the table, the fraction of zeroes slightly increases regardless the applied threshold, which is in accordance with your experience. The results for the other models were similar (except the model 2B, where the fraction of zeroes remains almost constant). We can add this information (in a reduced form) to the paper.

Comment: ***What I think the authors are doing is to estimate sample correlations of the current and the future independently (i.e. taking their respective sample means and standard deviations). As a result of which they may be finding the change is insignificant, whereas the change with respect to a faxed reference (sat the historical climate) may be more. At the very least, some clarification on how the correlations are estimated as well as the change in the first order statistics that are used in its estimation is needed.***

Reply: Indeed, the sample correlations are estimated independently, and we agree that the results may change slightly if a fixed reference is considered. We will extend the whole description of the bootstrap procedure and the details on estimation of cross- and auto- correlation will be given in detail in the revised manuscript, if we are invited to submit revised version.

Comment: ***l110 - the negative change in autocorrelations is consistent with my experience. If one were to consider changes in the associated means and standard deviations this becomes even greater. Additionally, these changes manifest themselves at longer time scales as well, as a AR1 model structure is not a great characterisation of the system. This has formed the argument for the range of "nesting" approaches in the literature for rainfall generation and bias correction. This needs to be discussed somewhere in the paper at the very least.***

Reply: We agree that the AR1 model is not sufficient characterization of the system. We will discuss this explicitly in the revised version of the manuscript if we are invited.

Comment: ***l115 - The figure title states 95% confidence of correlations. Does this mean 95% of the 66 correlations, or all the resampled correlation estimates as well?***

Reply: The confidence intervals presented in Fig. 4 are derived for each correlation coefficient separately. The title of the figure should state this more clearly, therefore we suggest changing the first sentence of the title as follows: "*The 95% confidence intervals of the individual cross-correlation coefficients for overlapping wet periods for all models…*". The first sentence of the Figure 5's title will be changed in the same way.

Comment: ***l180 - I believe the authors need to write a simple equation to show how they will ascertain their dependance outlier, and give us results of some tests that help argue these are genuine outliers and not examples of real extremes that would be of interest in hydrology. This is kind of important as this seems to be the key contribution the paper is making.***

Reply: We do not propose any new dependence measure; rather we are offering a procedure how to obtain robust estimates of correlation and autocorrelation by removing few data points that have large influence on the estimates. Therefore, there is no straightforward formula. We agree that real extremes are of special interest in hydrology, but our point is that dependence outliers (no matter if genuine or real extremes) affect data transformation within some bias correction methods and may possibly distort the dependence structure of the corrected data. In our concept, the dependence outlier is <u>any</u> value deviating from the correlation structure (as demonstrated in Fig. 7), regardless of the origin of this value. The distinguishing between real extremes and "true" outliers (say for example measurement errors) can be only hardly based clearly on statistics. This would involve an expert assessment of particular event, considering local conditions and the data from surrounding locations. Therefore we cannot design a simple equation for such purposes. Nevertheless, the real extremes as well as genuine outliers affect the correlation structures in the same way, which subsequently affects the bias corrections (or stochastic generators). Therefore the dependence outliers, regardless of their origin, can be detected and should be removed from the calibration data. Note that it is possible to insert these data back later.

Reference: Davison, A. C., & Hinkley, D. V. (1997). Bootstrap methods and their application (Vol. 1). Cambridge university press.