Supplement of Hydrol. Earth Syst. Sci., 23, 1741–1749, 2019
https://doi.org/10.5194/hess-23-1741-2019-supplement

Hydrology and
Earth System
Sciences
Open Access

Supplement of

Technical note: Changes in cross- and auto-dependence structures in climate projections of daily precipitation and their sensitivity to outliers

Jan Hnilica et al.

Correspondence to: Jan Hnilica (hnilica@ih.cas.cz)

This supplement briefly presents the effect of outliers on the multivariate bias correction methods described in Mehrotra and Sharma (2015) and Mehrotra and Sharma (2016). A software toolkit employed to perform bias corrections in this study is available and described in Mehrotra et al. (2018).

The data used in this study are attached as an example 3 to the above mentioned software toolkit. The dataset 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	**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 following table, the outlier strongly affects the value of $r_{1,5}$:

Data	$r_{1,5}$ (without any correction)
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)

Note that the abbreviation "MRNBC" denotes the Multivariate Recursive Nesting Bias Correction method, the abbreviation "MRQNBC" denotes the Multivariate Recursive Quantile-matching Nested Bias Correction method. In all cases (BC1 – BC4) the following statistics were corrected for all time steps included in the nesting scheme:

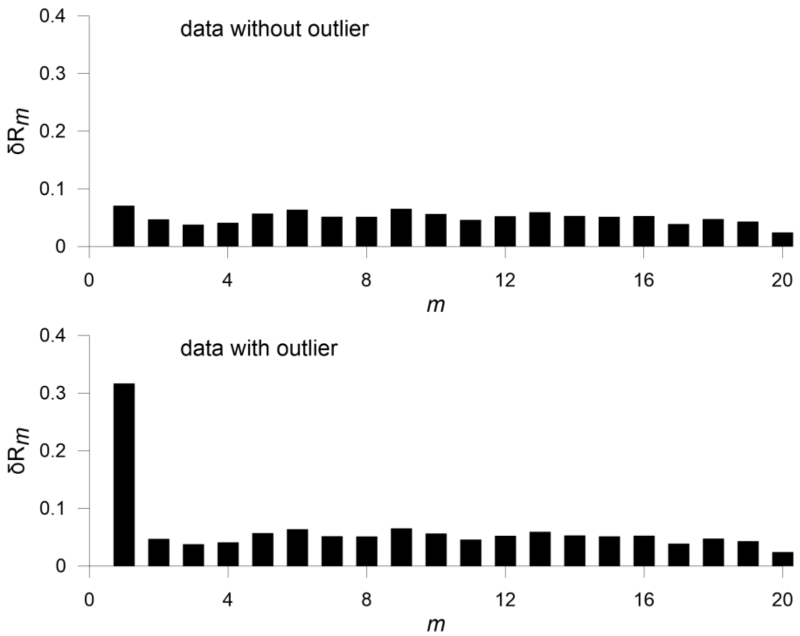
- mean
- sd/distribution
- lag-1 auto-correlation
- lag-0 cross-correlation.

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 Hnilica et al. 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 available through https://doi.org/10.5281/zenodo.1407992 (attached to Hnilica et al.) simply perform such operations. Certainly, this exploratory 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

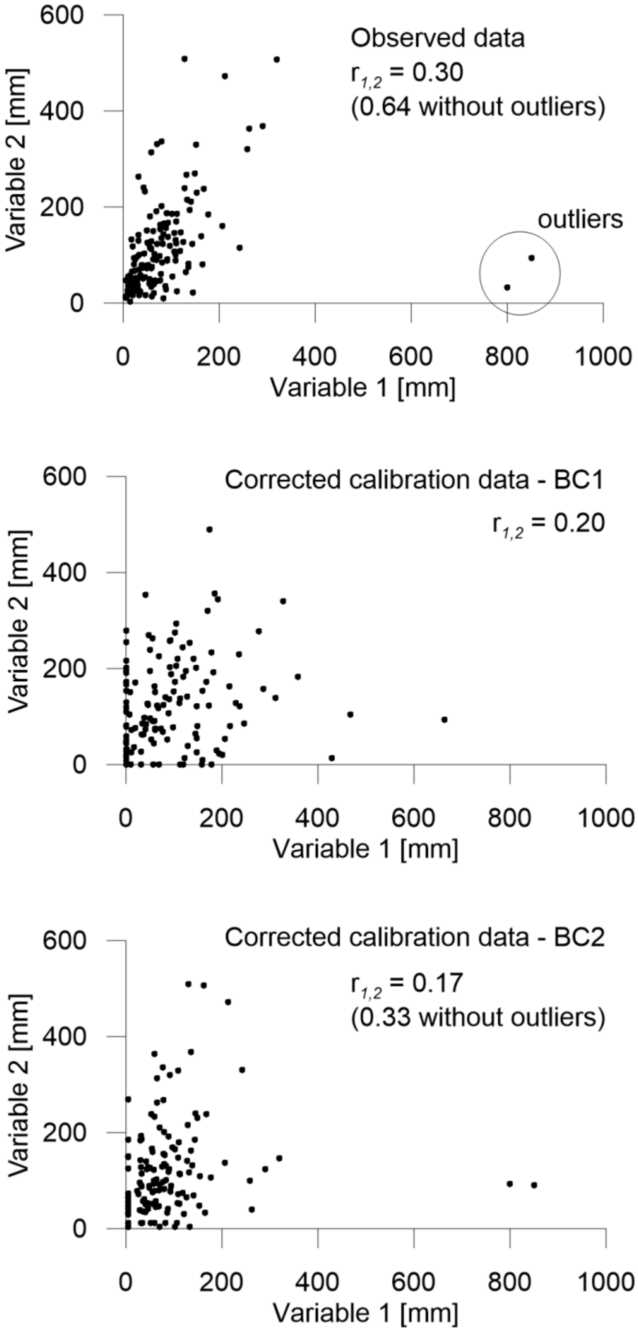
Year	Month	Variable 1
1922	4	47.5
1922	5	35.5

Outliers

Year	Month	Variable 1
1922	4	800
1922	5	850

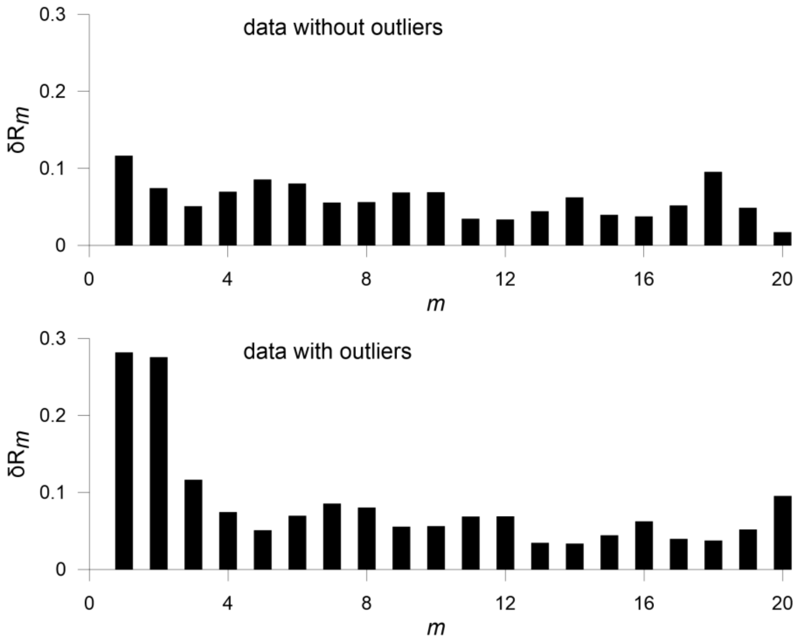
The values 800 and 850 mm are indisputably large, but they are not over physical limits (even values over 1000 mm can be found in the observed dataset). When only the values from months 4 and 5 are considered, the outliers change the cross-correlation between Variables 1 and 2 (denoted hereafter as $r_{1,2}$)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



References

Hnilica, J., Hanel, M. and Puš, V.: Technical note: Changes of cross- and auto-dependence structures in climate projections of daily precipitation and their sensitivity to outliers, Hydrol. Earth Syst. Sc.

Mehrotra, R., Johnson, F. and Sharma, A.: A software toolkit for correcting systematic biases in climate model simulations, Environ. Modell. Softw., 104, 130-152, doi:10.1016/j.envsoft.2018.02.010., 2018.

Mehrotra, R. and Sharma, A.: Correcting for systematic biases in multiple raw GCM variables across a range of timescales, J. Hydrol., 520, 214–223, doi:10.1016/j.jhydrol.2014.11.037, 2015.

Mehrotra, R. and Sharma, A.: A multivariate quantile-matching bias correction approach with auto-and cross-dependence across multiple time scales: Implications for downscaling. J. Climate, 29, 3519–3539, doi:10.1175/JCLI-D-15-0356.1, 2016.