**Hydrology and
Earth System
Sciences**

Open Access

Discussions

HESSD

Interactive
comment

Interactive comment on "Technical note: Changes of cross- and auto-dependence structures in climate projections of daily precipitation and their sensitivity to outliers" *by* Jan Hnilica et al.

G. Pegram (Referee)

pegram@ukzn.ac.za

Received and published: 19 March 2018

https://doi.org/10.5194/hess-2018-7: Jan Hnilica et al.

General comments

The authors outline the purpose of this paper succinctly in the paragraph starting in line 44: "This paper evaluates the biases and temporal changes in cross-and auto-correlation structures in multivariate precipitation data simulated by climate model. We further investigate whether the magnitude of the changes exceeds considerably the natural variability. The attention is finally paid to the effect of outlying values, which can

significantly affect the correlations and can thus lead to artefacts in bias-corrected time series." Sadly, I found the paper difficult to follow on the first read through and had to reread it twice more to understand fully the import of the argument.

My biggest worry about the paper is that simple ideas are not explained well and that the mathematical treatment is relatively naïve and somewhat muddled. The result is that the intercomparisons between the GCM and RCM historical and future data are not treated straightforwardly. My biggest worry is the treatment of the correlations between the historical and projected daily rainfall as estimated by climate models and historically comparing correlation coefficient (cc) outputs with observations. The measured rainfall is averaged over 25 km square areas covering catchments, some of which are of smaller dimension. The other comparisons are with the cc modelling products of a GCM and RCM which have a foot-print of 0.11 degrees square (about 8 by 11 km at 50 degrees North). There is no figure showing this spatial mismatch, which makes it difficult for the reader to juggle the ideas mentally to understand the consequences. Furthermore, the confusion between correlations and covariance (the latter briefly touched on in section 4.6) is not sensible. Covariance depends not only on correlation, which is meaningfully comparable, but also variance, which will alter depending on scale. The statement in line 263: ' The overall result is that the covariance structures change substantially more than correlations.' therefore comes as no surprise and is indicative of the naivety of much of the paper's message. Spearman correlation coefficients are used for comparison, eliminating the effect of outliers in the various time series, which is clear from Figure 11, but rather briefly commented upon in the supporting text. However, the declared purpose of the paper is to identify outliers and remove them so they do not contain the comparisons between the time series, so rank correlations are not useful for this task and more usefully indicate bias in the intercomparisons.

Autocorrelations of the series were calculated, but included the zeros, which were pooled with the non-zero amounts in section 3 and illustrated in Figures 5 and 6. It is only in the discussion on Figure 11, in lines 251-2 where attention is drawn to the

problem: "The precipitation data contain a high number of zero values which must be treated otherwise the rank correlation is distorted by severe artefacts. A possible way is to consider the correlations only for non-zero values, however this approach results into a loss of information and furthermore it is not applicable to auto-correlations. Therefore the joint (average) rank was assigned to each zero value in this study (we note that the results can differ if e.g. random rank is applied instead)." In my opinion, the inclusion of the zeros by ranks masks the problem and artificially increases the ccs. Treatments of zero and non-zero data need to be done separately; even if there are mismatches in the wet periods in the time series - there should be enough common wet days to obtain a fair comparison between the wet day sets to determine the effect of serious outliers on the serial and cross-correlation coefficients. Incidentally, the serial correlation coefficients should be compared both forward and backward over a few intervals to eliminate 'drag' and find the highest cc.

After much thought, I recommend that this paper be rejected. The particular comments are noted below in two paragraphs. I am returning the paper with my annotations.

Geoff Pegram

19 March 2018

Specific comments, by line number

55: It's difficult to match table 1 and figure 1. Please consider re-labelling the links in Fig. 1 in a more orderly pattern, South to North, perhaps as I have suggested by the blue lines on the images?

62: Some of the areas in Figure 1 are less than 20 km across and are likely to be covered by a 25 km square Hydro grid element - please show the grid superimposed on the basins.

99-100': 'the numbering was done according to rows of correlation matrix'. Why not use the convention of (i,j) in each cell for row (i) and column (j)? It would save the reader

from having to do arithmetic deciphering 156 numbers! Please see my partitioning of figure 4 as an alternative.

113: In the caption of Figure 3, I do not understand the difference between bias and change in (a) & (b). Please Insert the phrase 'Horizontal axis labelling from table 2' at the end of the figure caption

116: after CCLM-4-8-17 and RCA4, please insert (A) then (B) to help the reader, as their definition is 2 pages back

124: 'range from -0.1 to 0.04.' what are the confidence limits for these statistics? 95% CLs for 78 independent data pairs are about +/- 0.19 so I reckon that these biases are relatively negligible...!

Figure 4: it would help the reader if you partitioned the elements from successive matrix rows as I have, then add the following to the caption: The blue dividers identify the successive rows below the diagonal in the matrix.

Figure 8: to compile this figure did you visually check through 78 of these bar charts or did you devise an algorithm to make the summary?

Figure 9: I do not understand these figures - it would help if you gave the axes labels and improve the description in the text.

253-4 In these lines we find the following: ' Therefore the joint (average) rank was assigned to each zero value in this study (we note that the results can differ if e.g. random rank is applied instead).' I do not like this procedure - there are too many zeros in daily rainfall records, unless you have a very humid geography. What is the proportion of dry-time? Why not calculate the binary cross correlation coefficients for wet and dry days over a few days' lags so you can determine whether the wet periods match? Then treat the overlapping wet periods.

261: the paragraph on 'Changes of covariance' - I do not understand the emphasis on covariance, because differences in variance mask the dependence - that's why we use

correlations! Why not compare variances and means separately from correlations?

285: ' covariance structures change considerably (even without outliers) and their changes are tightly associated with the changes in variances.' Exactly!! See my previous remark.

Technical corrections, by line number - text between apostrophes are quotes from the text - my remarks and insertions are not

63: 'with 0.11 degree resolution' at this latitude that is about 11 x 8 km?

78-82: did you exclude dry days? did you consider seasons?

107: 'the' correlations of the 'model'

110: 'individual models' in Figure 3 or 4?

122: 'independent and' does 'not'

143: 'the changes of the model 3B (not shown).' Reword - 3B is shown in the figure!

149: 'in many cases' in contrast to 3B

162: ' decrease' is 'caused'

174: 'thousands' of 'values'

195: remove 'either'

200: ' the most outlying (multi-variate) value is found in the data' - in what alternative coordinates?

215: 'evolution of dRm' , defined in equation (5),

220: instead of 'actual' insert remaining or surviving?

234: '4.6 per mille' - mille is rarely used - why not put '0.46%' so we don't have to look up what it means and divide by 10?

242: replace 'eliminated' by much reduced

246: 'significant changes' , especially in Figure 3B, where a large proportion of pairs of data bars do not overlap at all.

253: 'auto-correlations' , unless these are calculated only for blocks of overlapping wet days.

258: 'correlations' in Figure 11

259: ' the precipitation data often show a more or less tight linear dependence' NOT to my mind in Fig 6b!!

290: ' covariance is affected by a' lack of (!!) 'linear transformation. The high changes of' variance ..

292-4: 'From this point of view a reasonable scenario of future precipitation can be obtained by the corresponding linear transformation of observations, i.e. by the multiplicative delta method (Déqué, 2007).' In my opinion, this is a rather weak conclusion . . .

Please also note the supplement to this comment:
https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-7/hess-2018-7-RC1-supplement.pdf

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-7, 2018.