Hydrol. Earth Syst. Sci. Discuss., 12, C2673–C2684, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C2673/2015/

Hydrology and Earth System Sciences Discussions

HESSD
12, C2673–C2684, 2015

Interactive Comment

Interactive comment on "Non–stationarity in annual maxima rainfall across Australia – implications for Intensity–Frequency–Duration (IFD) relationships" *by* D. C. Verdon-Kidd and A. S. Kiem

D. C. Verdon-Kidd and A. S. Kiem

Danielle.Verdon@newcastle.edu.au

Received and published: 21 July 2015

We would like to thank Dr Serinaldi for his review of the paper "Non-stationarity in annual maxima rainfall across Australia-implications for Intensity-Frequency-Duration (IFD) relationships". We have considered the Reviewers' comments and provided detailed descriptions of how each comment will be addressed in the revised manuscript below:

Specific comments Dr Serinaldi's specific comments center around three main themes.

The first is the use of change point analysis to test for non-stationarity in a data series, the second is the application of the CUSUM test (and identifying multiple change points) and the third is the use of the term “regime shift” in preference to “non-stationarity” when discussing variability in the annual maxima rainfall timeseries. Each of these issues were further built on in the technical remarks provided by Dr Serinaldi and therefore are addressed in detail below.

Technical remarks 1. Please, consider to check the significance of (bias-corrected) serial correlation (if this was not done) because it can affect the results of change point analyses (see e.g. Serinaldi and Kilsby (2015a) and references therein for a discussion on Mann-Kendall and Pettitt, which however holds true also for e.g. CUSUM and similar). As shown above, apparent regime shifts can be artifacts resulting from hidden persistence.

Response: In response to the Dr Serinaldi’s suggestion, the Durbin-Watson (DW) statistic was used to test for autocorrelation (serial correlation) in the annual maxima timeseries (Durbin and Watson (1950, 1951)). The Durbin-Watson statistic tests the null hypothesis that the residuals from an ordinary least-squares regression are not autocorrelated against the alternative that the residuals follow an AR1 process. The Durbin-Watson statistic ranges in value from 0 to 4. A value near 2 indicates non-autocorrelation; a value toward 0 indicates positive autocorrelation; a value toward 4 indicates negative autocorrelation. Typically, tabulated bounds are used to test the hypothesis of zero autocorrelation against the alternative of positive first-order autocorrelation. For the sample size in our case ($\sim$100) and a linear trend model with intercept the dlower = 1.522 and dupper=1.562 for 1% significance.

All DW statistic values were found to be greater than the 1.562 (the upper bound for 1% significance) providing no evidence to reject the null hypothesis (see figure 1 showing the distribution of all DW statistic values for the 1-day annual maxima timeseries at each site).

HESSD
12, C2673–C2684, 2015

The DW test results will be included in the revised paper to demonstrate that the annual maxima data does not suffer from serial correlation and therefore the statistical tests used in the change point analysis is appropriate.

2. P3453L20-25: In my opinion, such lines reflect some confusion on this topic. Trends or change points in finite time series do not imply nonstationarity. Nonstationarity cannot be in principle significant or not significant, because it is an assumption made on the underlying process that can be introduced only if we know the underlying nonstationary dynamics (physical equations, well-defined changes with a clear cause such as flow regime changes due to dams operation, etc.). Please consider to reword this type of sentences throughout the text in light of the discussion and references above.

Response: Dr Serinaldi's review has highlighted some important points regarding the use of the term "stationarity" and if/where it is applicable in our study. In particular he questioned whether the term "regime shifts" was more fitting in describing our findings of change points in the annual maxima rainfall timeseries that are possibly attributable to climate shifts. Dr Serinaldi stated that (following Koutsoyiannis and Montanari, 2014; Serinaldi and Kilsby, 2015) stationarity is a concept referring to models rather than to timeseries. In our case the model is the IFD curve. Thus the text that describes the assumptions in the IFD development of stationarity in the underlying processes (i.e. the statistical properties of the rainfall do not change over time and that the chance of an extreme event occurring is the same at any point in time (past or future)) is relevant, however we agree that the text discussing change points in the rainfall data may be misleading where the term non-stationarity has also been used. However it is interesting to note that many studies have also used the word stationarity and non-stationarity when describing similar timeseries (e.g. Ishak et al. 2013, Westra and Sisson 2011, Wagesho et al 2013, Wilby 1998, etc), therefore there appears to be widespread disagreement on the use of this term. Despite this, on further review of the journal papers provided by the Reviewer we agree that in our case the term "regime shift" is more suitable. Given the above we will revise the text in the paper to reflect this. In particular, we will use the term non-stationarity only when referring to the IFD development, however the sections of the paper that are focused on identifying change points in the rainfall timeseries will be edited and the term “regime shift” will be used in preference. Further, as per the reviewer’s suggestion (in his specific comments) we will also change the title of the paper to “Regime-shifts in annual maxima rainfall across Australia – implications for Intensity–Frequency–Duration (IFD) relationships”. The associated text will include the references provided by the reviewer.

3. P3457L25: Please consider to reword, e.g. “LP3 was not rejected at x% significance level for all series (or n series out of N)”.

Response: The sentence in question will be reworded as suggested to read “Here the null hypothesis is that the data fits the Log-Pearson III distribution (the alternate is that the data does not follow the Log Pearson III distribution). All p-values were greater than 0.05 (average p-value was 0.75), for all series (30min to 72hr durations at Brisbane, Sydney and Melbourne), therefore we accept the null hypothesis at the 5% significance level.

4. P3458L12-15: I do not know AR&R, but it is not clear to me why return periods defined on annual maxima should be adjusted for PDS. Usually we do the opposite when we start from PDS and we need the actual AMAX return periods (under suitable conditions such as Poisson arrival dynamics, etc.). Please clarify.

Response: The methodology adopted in this paper to calculate return periods of annual maxima specifically follows that outlined in Australian Rainfall and Runoff (1987), Engineers Australia’s guide to estimating and utilising IFD information. Published IFD currently used by industry in Australia are based on this method. The updated IFD (which are NOT currently used in operation) are based on a revised statistical methods (for example, a Generalised Extreme Value (GEV) frequency distribution was fitted to the annual maxima rather than Log Pearson III and extension of sub-daily rainfall statistics to daily read stations is conducted with Bayesian Generalised Least Squares

Regression rather than PCA). The purpose of adopting the AR&R 1987 method was to assess the implications of varying data lengths and climatic variability on the resulting IFD (which have been historically used and are currently still in use) and to highlight the issue of underlying variability in the annual maxima that should be appropriately considered and addressed in the current (and future) revision of the IFD estimates.

5. P3459L9: As mentioned above, step changes and nonstationarity are very different concepts and surely not synonyms.

Response: This will be revised and clarified as per discussion above (Comment 2).

6. P3459L17-24: Leaving aside the use of the term nonstationarity, CUSUM identifies automatically the change point location and does not split the time series in two halves. If the Authors mean that the test proceeds based on subsequent dyadic partitions, this is right, but for such short time series it is actually quite difficult (and not meaningful) to go beyond 2-4 changes. Please note that many other refined techniques are available for segmentation... of course, a question rises about the (physical) meaning of such refined segmentations...

Response: This will be clarified in the text. The data is not split in equal halves for the CUSUM test, it is split into two portions, which may or may not be equal. However, unless a moving window is used (say 20 years as we did for the Mann- Whitney, multiple regime shifts could still be missed using this method.

7. P3459L25-P3460L10: Following the previous remark, my interpretation of P3473Fig5 is a bit different. The almost uniform spread of changes across the decades denotes that such changes occur quite randomly, and sincerely I cannot see a tendency to cluster in the east coast. We may see something in panel (b), but the spatial distribution of the stations is not uniform and we cannot exclude that such stations are spatially correlated, as they are subject to similar climate forcings (thus reducing the evidence for changes). Note that spatial correlation is another factor that can strongly affect the outcome of such a type of tests (see e.g. Douglas et al. (2000), Yue et al.

HESSD
12, C2673–C2684, 2015

(2003), Guerreiro et al. (2014), among others)

Response: We agree with the Reviewer that there is almost a uniform spread of changes across the decades based on panel (a). We state in the paper that ", the large-scale climate phenomena impact various regions of Australia at different times of the year and to varying degrees, therefore it is not surprising that the timing of shifts in the annual maxima timeseries varies spatially and temporally." However this will be further clarified in the revised paper.

The clustering along the east coast can only be clearly seen in panel B. The text will be revised to clarify this.

The spatial correlation of the annual maxima timeseries was investigated as per the reviewer's suggestion. We found that less than 9% of all possible pairings of rainfall data sets display a significant yet weak correlation at the 5% level (r >0.2, significance based on n=100). Only 8 pairings (out of 4465) were correlated at 0.5 or higher. It was also found that stations located more than 500km apart were unlikely to be correlated and that the strength of the correlation reduced as distance increased between the pairs (see Figure 2 and 3).

8. Section 3.2: Again, my interpretation of P3474Fig6 and P3475Fig7 is a bit different. If I'm right, box plots for IPO(-) summarize the distribution of 41 AMAX (1913-1920 and 1945-1977), while we have 67 AMAX for IPO(+) box plots. For such sample sizes, inferring difference in distribution based on box plots is a bit hard (at least). My suggestion is to use some formal two-sample goodness-of-fit tests such as the two-sample Kolmogorov-Smirnov or similar, thus accounting for sampling uncertainty and different sample sizes. In any case, comparing box plots (overlooking the large uncertainty of the quantile estimates) is not informative and does not provide a quantitative assessment, especially in this case where differences between IPO(-) and IPO(+) regimes are really hard to recognize.

Response: The results of the two-sample KS test will be included in Figure 6 as suggested.

9. The same holds for P3475Fig7: if I'm right, this diagram shows the differences Δ (in %) between the point estimates of rainfall return levels obtained by LP3 distributions fitted on 41 and 67 AMAX. It is almost superfluous to highlight how large the uncertainty of such a point estimates can be. I suggest a fairer check based on a simple bootstrap procedure. For each duration:

1. resample with replacement IPO(-) and IPO(+) time series to obtain two new B-samples;

2. for each B-sample refit LP3, compute the required LP3 return levels and calculate the difference Δ(B) as for the observed data;

3. repeat previous steps B times (e.g. 1000) and store the obtained B differences (for each ARI). These values can be used to build the empirical distribution of the differences Δ(B) i , i = 1, ...,B. This distribution describes the effects of sampling and parameter estimation uncertainties under the hypothesis of existence of two different regimes;

4. Use the B Δ(B) i values to build confidence intervals (CIs) at a given confidence level (e.g.95%). If these CIs include $\Delta = 0$, then there is not evidence for a significant difference, otherwise we can conclude the opposite.

I think this is a better way to provide a quantitative assessment. Of course, conclusions concern the effects of possible regime shifts and not of nonstationarity. Section 3.2 should be reworded according to the results of the analyses suggested above.

Response: We agree this would be a more robust test for regime shifts and thank the Reviewer for his suggestion. We have not yet completed the new analysis, however this bootstrap method will be carried out and included in the revised paper. We would like to thank Dr Serinaldi for this suggestion.

10. Section 4: as for Section 3.2, this section should be reworded according to the

updated results.

Response: Noted, this will be revised.

11. Please avoid sentences such as that in P3464L27-29 and P3465L1-3: even after more accurate analyses, there is not way to make unquestionable conclusions about nonstationarity if we do not identify a well-defined mechanism of evolution which is almost perfectly predictable (at least, at the time scales of interest).

Response: The sentence "Based on the results of this study, and literature cited within this paper, we emphasise that there undoubtedly is non-stationarity in historical short duration rainfall extremes but the characteristics and causes of this non-stationarity vary from location to location and decade to decade – something which must be considered and accounted for when attempting to estimate IFD design rainfalls and prior to quantifying how those IFD estimates might change in the future." will be revised and expanded to read "This study has highlighted the existence of regime shifts in annual maxima rainfall data in Australia. The driving mechanisms of these regime shifts are likely to vary from location to location and decade to decade. However, these shifts are typical of many natural phenomena and can be described by processes characterized by long range dependence (or regime-switching processes) and captured by hidden Markov models (or similar), resulting in a mixture of distributions that alternate stochastically according to the transition probability from one regime to the next (e.g. Serinaldi and Kilsby, 2015a). While the strategy of defining IFDs for two (or more) different regimes (e.g Serinaldi andKilsby (2015a)) currently only partially solves the problem, as we often do not know the beginning or the end of a specific regime (be it rainfall or climate driver), recent work has focused on optimizing designs and planning strategies based on the range of what is plausible rather than a reliance on knowing the current and future climate state (e.g. Mortazavi-Naeini et al., 2015). At the same time, work is also underway on seamless prediction at a range of timescales and if/when this eventuates the results discussed here become even more important/useful. Nevertheless, the immediate usefulness of the insights presented here occurs when first establishing the IFD, as an approach similar to that employed here can be used to determine if the underlying data are biased to a mostly wet or mostly dry regime (or a mix of both) which then provides an indication as to whether the IFD is likely to be an over- or underestimate of the true risk. Importantly, this issue needs to be considered and accounted for when attempting to estimate IFD design rainfalls and prior to quantifying how those IFD estimates might change in both the near and long-term future."

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 3449, 2015.

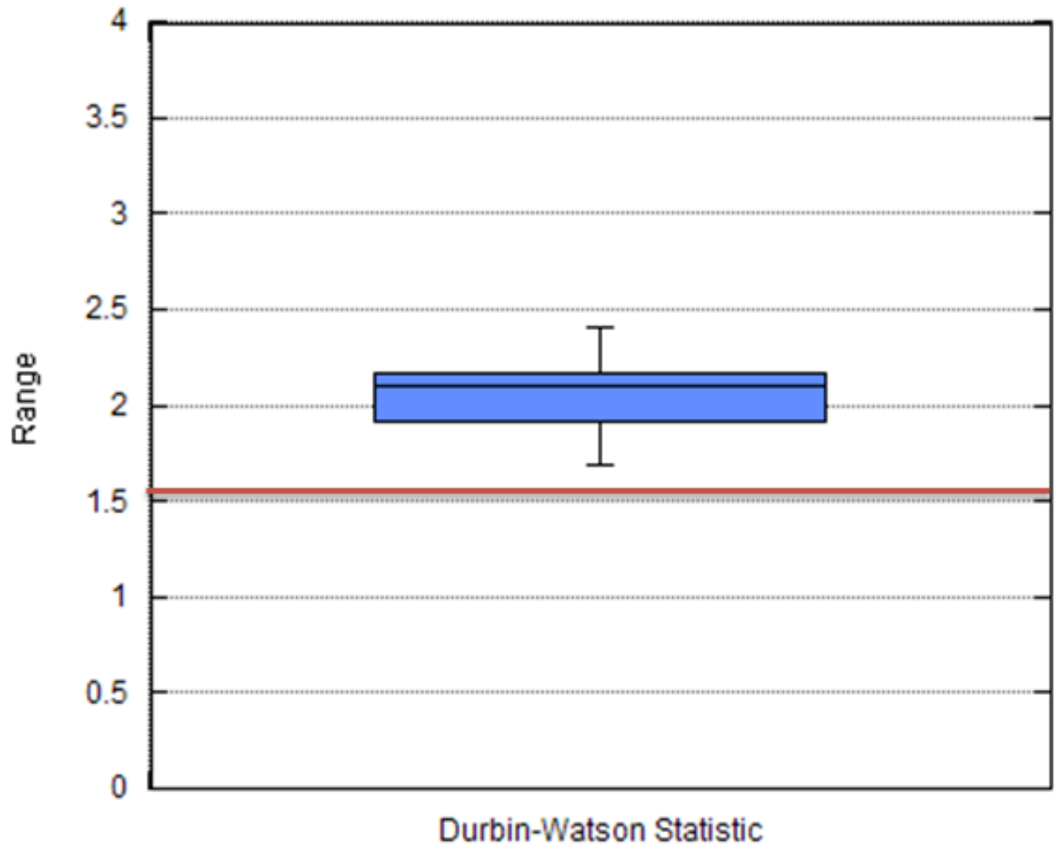

Fig. 1. Box plot of Durbin-Watson statistic for the 1 day-annul maxima timeseries at 96 stations (red line indicates dupper=1.562 for 1% significance)

HESSD
12, C2673–C2684, 2015

Interactive Comment

HESSD

12, C2673–C2684, 2015

Interactive Comment

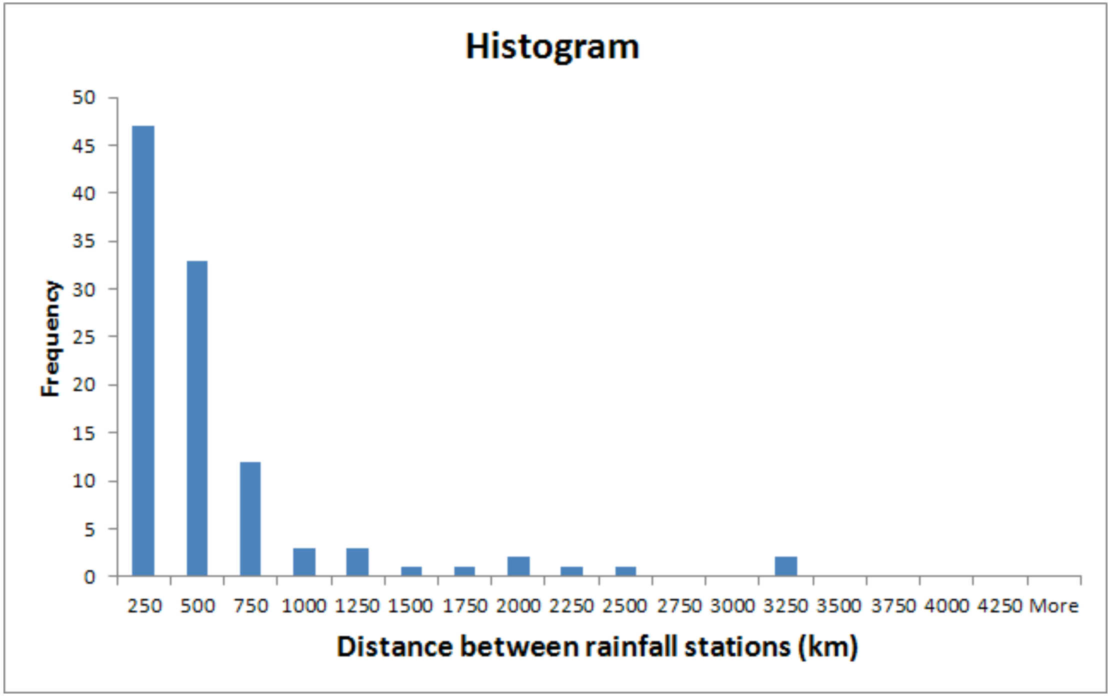

Fig. 2. Distance between stations with significant (r>0.2) correlations

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

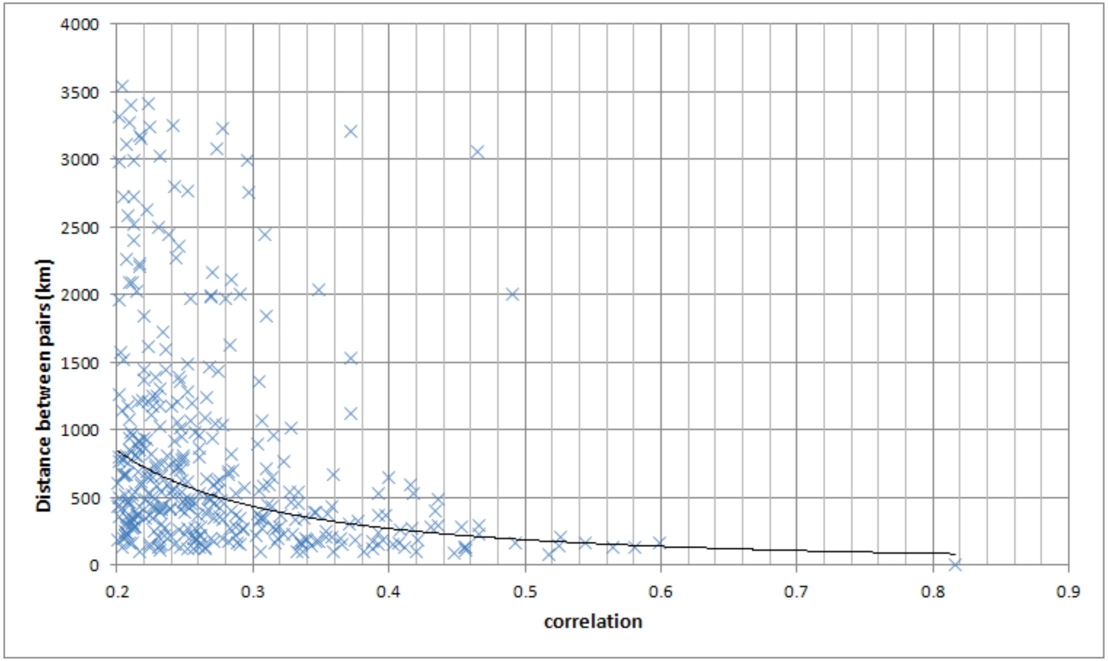

Fig. 3. Scatter plot of correlation vs distance between stations