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

F. Serinaldi (Referee)

francesco.serinaldi@ncl.ac.uk

Received and published: 26 April 2015

General comments

In this paper, the Authors provide a regime-shift analysis of rainfall annual maxima at multiple time scales (30 min to 72 h) recorded in Australia and spanning several decades. The aim is to show the effect of possible regime-shifts on the computation of intensity-frequency-duration curves (IFD; aka IDF curves). The paper is clear and

well organized. As discussed below, my main concerns are about the concept of stationarity, the use of the statistical tests, and the interpretation of the results. I say in advance that my point of view is slightly different from that reported in the largest part of this type of studies; so, it will be easy for the Authors to provide a possible rebuttal based on a very extensive literature if they do not agree with me. Nonetheless, I hope they will consider my arguments. I also apologize in advance for mentioning some of my own papers in the following.

Specific comments

As mentioned in a comment to another HESSD paper that I posted recently, after viewing/reading/reviewing quite a large number of papers dealing with the application of tests to check for trends and change points (such as Mann-Kendall, Pettitt, and similar), I think that we should have a break and spend some time to reflect about the rationale of these tools, and the meaning of their outputs, as well as their interpretation. First of all these tests does not check for stationarity/nonstationarity. As discussed by Dr. Koutsoyiannis in some of his papers (with more emphasis in the most recent ones), stationarity is a concept referring to models rather than to time series. This is not a semantic issue but a concept that has serious consequences on our ability to make inference starting from data (see e.g., Koutsoyiannis and Montanari, 2014; Serinaldi and Kilsby, 2015a).

this context, the tests mentioned before simply check some hypotheses (such as stochastic ordering, equality in mean/median, etc.) and provide results (e.g. evidence for abrupt changes) that require attribution. However, dealing with finite time series, these results (e.g., shifts) imply nonstationarity only if we can identify a deterministic dynamics which caused e.g. the shift. Such a driver cannot be a climate index but something like dam building, i.e. a mechanism that is known with certainty (or negligible uncertainty) and almost perfectly predictable in the future. If we attribute e.g. rainfall regime shifts to regime shifts of climate indices, we simply "shift" the problem,

because climate shifts recognized on finite time series do not imply that the underlying dynamics is nonstationary, if we do not know a predictable mechanism of evolution of the phenomenon under study. In this respect, the regime shifts recognized by the Authors are typical of many natural phenomena and can be described by processes characterized by long range dependence, or "regime-switching process, which can be described by models such as hidden Markov models and Markov switching models, self-exciting threshold autoregressive models or similar ..., resulting in a mixture of distributions that alternate stochastically according to the transition probability from one regime (state) to another one." (Serinaldi and Kilsby, 2015a). These models generally exhibit locally persistent fluctuations but are still globally stationary owing to the lack of a deterministic and predictable temporal evolution. Of course, nonstationarity can also characterize stochastic processes such as the Wiener process (or Brownian motion), which however unlikely fit rainfall dynamics.

Figure below shows a concrete example of the practical implications of the above remarks. Panel (a) illustrates 115 years of annual rainfall data showing a pattern similar to that reported in P3472Fig4 in the paper. The signal shows a weak autocorrelation, and Pettitt and CUSUM detect two change points (1932 and 1980)... Actually, the signal is just a (rescaled) subsample of the longer time series shown in panel (b). The latter is a realization of a fractional Gaussian noise (aka Hurst-Kolmogorov process), with Hurst parameter $H = 0.75$, which is stationary and characterized by a power-law decaying autocorrelation (ACF). Notice that the true nature of the ACF is very difficult to be recognized in the 115-size time series, which exhibits a very weak empirical ACF (whose estimator is known to be strongly biased; see e.g. Koutsoyiannis, 2003 and references in Serinaldi and Kilsby, 2015b). Note also that Pettitt and CUSUM generally do not suffer from the problem of "sequential split" mentioned by the Authors in P3457L1-2: these tests usually detect the main change point (1932 in the example above) and then possible further changes in the obtained subsequences (referring to the example above, we have no further change in the first subsample (1901-1932) and one change in 1980 in the second subsample (1932-2015)). So, the behavior illus-

C1259

trated in P3472Fig4 seems to me a bit strange because, looking at that time series and based on my experience with CUSUM, Pettitt and other similar tests, I'm quite confident that such tests should recognize two changes located around the points identified by Mann-Whitney.

Thus, in my opinion, what the Author did in this paper was the recognition of regime shift dynamics for rainfall, which leads to conclusions similar to those discussed e.g. by Serinaldi and Kilsby (2015a) concerning the evolution of regime shifts in annual peak flows. In particular, the strategy of defining IFDs for two (or more) different regimes (as done e.g. by US Authorities in one of the case studies discussed by Serinaldi and Kilsby (2015a)) can only partially solve the problem, as we do not know neither the benning nor the end of a specific regime (P3464L15-20), and this holds both for rainfall and possible climate drivers.

It should also be noted that under nonstationarity we need to know the evolution of the probabilty distribution for periods (in the future) as long as the computed return period (at least), whereas, under (globally stationary) regime switching dynamics, stochastic fluctuations do not imply unrealistic future projections (see e.g. examples and discussion nonstationary return periods in Cooley (2013), Salas and Obeysekera (2014), Serinaldi and Kilsby (2015a), and references therein).

Thus, I think that the overall discussion in the paper generally agrees with the discussion above if we replace the word "non-stationarity" with "regime-shift", leading e.g. to a title like "Regime-shifts in annual maxima rainfall across Australia – implications for Intensity–Frequency–Duration (IFD) relationships". Indeed, this describes more carefully the content, does not change the overall structure of the paper, and leaves the way open to stationary and nonstationary approaches, making results more general and, in my opinion, more theoretically sound.
Please, find below some further technical remarks.

Technical remarks

C1260

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.

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.

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

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.

P3458L20-25: I can agree, but the comparison should not be limited to point estimates, and should include uncertainty (see below).

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

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...

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

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.
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 superfluos to highligth 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 $\Delta^{(B)}$ 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 $\Delta_i^{(B)}$, $i = 1, ..., B$. This distribution describes the effects of sampling and parameter estimation uncertainties under the hypothesis of existence of two different regimes;

4. Use the B $\Delta_i^{(B)}$ 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.

Section 4: as for Section 3.2, this section should be reworded according to the updated results. Please note that some of these remarks are already discussed elsewhere (e.g. in one of my papers mentioned above) along with discussions about regime shifts. 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).

Sincerely,

Francesco Serinaldi

C1263

References

Cooley D. (2013) Return periods and return levels under climate change. In: AghaKouchak A, Easterling D, Hsu K, Schubert S, Sorooshian S, editors. Extremes in a changing climate. Water science and technology library, vol. 65. Netherlands: Springer. p. 97–114.

Douglas EM, Vogel RM, Kroll CN. (2000) Trends in floods and low flows in the United States: impact of spatial correlation. Journal of Hydrology 240: 90–105.

Guerreiro S.B., Kilsby C.G., Serinaldi F. (2014) Analysis of time variation of rainfall in transnational basins in Iberia: abrupt changes or trends? Int J Climatol 34(1):114–133

Koutsoyiannis D. (2003) Climate change, the Hurst phenomenon, and hydrological statistics. Hydrol Sci J 48(1):3–24

Koutsoyiannis D., Montanari A. (2014) Negligent killing of scientific concepts: the stationarity case Hydrol Sci J http://dx.doi.org/10.1080/02626667.2014.959959

Salas J.D., Obeysekera J. (2014) Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. J Hydrol Eng, 19(3):554–68.

Serinaldi F., Kilsby C.G. (2015a),Stationarity is undead: Uncertainty dominates the distribution of extremes. Advances in Water Resources, 77, 17-36

Serinaldi F., Kilsby C.G. (2015b) The importance of prewhitening in change point analysis under persistence. Stochastic Environmental Research and Risk Assessment

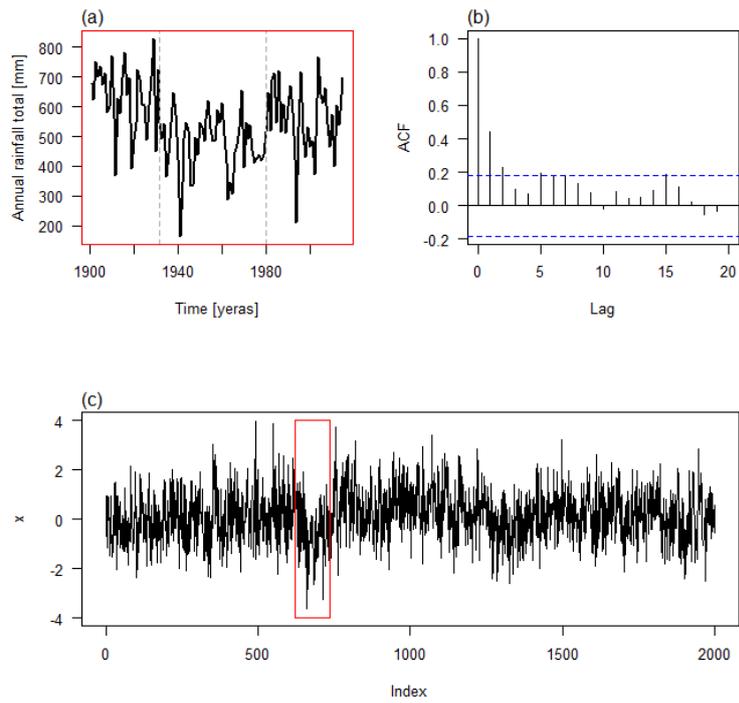Yue S, Pilon P, Phinney B. (2003) Canadian streamflow trend detection: impacts of serial and cross-correlation. Hydrological Sciences Journal 48(1): 51–64.

Fig. 1. Example time series

C1265