Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-325-RC1, 2017

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Does Nonstationarity in Rainfall Requires Nonstationary Intensity-Duration-Frequency Curves?" *by* Poulomi Ganguli and Paulin Coulibaly

Anonymous Referee #1

Received and published: 23 June 2017

The manuscript presents an interesting topic, and discuss the crucial question of whether there is enough evidence of changes in hydrometric series to warrant a change in the IDF curves used for the design and maintenance of hydraulic structures.

Although the topic discussed is interesting and worthy, the paper is quite inconclusive and does not manage, in my opinion, to provide a clear point of view on the matter. The authors have definitely done a lot of work and have looked very carefully at the data, but they fail to summarise their finding in any useful way and simply provide a lot (too much maybe) of information. The presentation of the methods and results is quite unclear and it has several opaque points. The statistical methods are often presented with some imperfections and in general the paper could greatly benefit from some proof-reading and re-organisation. In particular the authors should make more of an attempt to summarise their findings from all the non-parametric tests in a way that is more informative.

Some more specific comments

The title of the manuscript indicate that IDF curves are the main topic, although the authors limit themselves to the (hard) task of fitting different frequency curves to the each series with different duration separately. This could result in non-consistent estimates eventually. The type of studies the authors perform is laudable and would be the first step to take to assess whether new IDF curves would need to be derived.

The authors do a lot (a lot!) of tests to the data series of each duration - definitely the issue of multiple comparisons arise and it is to be expected that some tests will turn out to be significant just by randomness. I have to say it is difficult to follow the authors in all their testing, there is very little effort made to summarise the finding in any useful way and the results are simply presented/dumped as they are in the SI (except I am not sure whether the results are reliable given the authors have p-values larger than 1).

Some further remarks given in more order:

1. The beginning of Section 3.3 is very messy and should be rewritten. Distributions do not contain parameters, they are characterised by parameters. Line 25, " a value of the shape parameter equal to zero". Line 28: "In the case of a negative shape parameter, the distribution is a Weibull". Note that the Frechet is also a bounded distribution, except it has a lower bound. Overall I would write down the whole thing in a a formula, specifying the limits of the distribution for the different values of the shape.

2. Page 2 line 13. It is often the case though that IDF curves are derived not only

from at-site data but using a pooled set of stations see for Svensson and Jones (2010, doi:10.1111/j.1753-318X.2010.01079.x) for a review of methods used in several countries

3. Page 3 - line 8-9: the authors seem to imply that the Gumbel distribution is symmetric - which is not the case, as it is easy to see by plotting the pdf of a Gumbel distribution.

4. Section 3.1: I think the information of the percentage of missing values of each station/duration should be given somewhere - ideally in the main text and not in SI. I can not judge whether the MCR technique is the most appropriate one, as this is too far away from my area of expertise.

5. Page 8 - line 8: if the 5% and 95% quantile of the posterior samples are taken then a 90% credibility interval is constructed. A 95% interval is taken to be one that contains 95% of the distribution.

6. Section 3.3: it is not clear to me why the authors go through the trouble of fitting both an ML and Bayesian fit for the stationary model if they only use a Bayesian model for the non-stationary models. Just use the Bayesian methods and embrace Bayesian Inference. Also, seeing in Table SI16-S24 that the more complex non-stationary model GEVII is often selected I wander whether the authors have tried to only fit models with the scale taken as the only varying function? Lastly, why not to formally test stationary/non-stationary model is better by using a Bayesian factor or some pre-set rule on the 95% credibility interval not-containing zero?

7. Section 3.3: what do you do with the results of the Pettitt test? One could use it to build a model with a step-change rather than a continuous function of time. In general, why doing all the non-parametric test AND the parametric models? What is the use of the non-parametric tests exactly?

Printer-friendly version

Discussion paper

8. Page 8 - line 14: it is very good that the authors verify the goodness of fit by using PP, but it is unclear to me how they "select the model with fewer parameters as the best model when two models have comparable performances.". This is exactly what the AIC should do, so even if the AIC does not indicate that a simpler model should be used the authors might cull a non-stationary model out if the stationary model give a better fit in the PP plot?

9. Page 8 - line 25-26: a positive skewness is just an indication of an asymmetric/skewed distribution, it doesn't necessary indicate a change in the distribution. I mean "extreme values are more frequent in the time series" compared to what?

10. Page 9, line 29: Bayesian measures of uncertainty are normally called credibility and not confidence intervals. Also as I mentioned above - unclear if the 95% or the 90% intervals are derived.

11. Page 10/Figure 4: how are the DSI calculated for the non-stationary models? Is the last value of the parameters used to compute the quantiles? Why do you show boxplots of the posterior sample and not a 95% credibility interval? As I said I would drop the estimation using ML completely, but if you do use it, you could show confidence intervals based on the delta-method (see Coles, 2001).

12. Page 11/Figure 6: has any assessment been done on whether the stationary version of the fitted curves has a good overlap to the EC-curves? Surely if these two curves are very different, any mis-match between the non-stationary results and the EC-curves could be due to the fact that the EC curve doesn't fully fit the data of a site. This links to a comment on the statements in page 13 between line 20-25: you are saying that from the comparison of stationary to non-stationary models there seems to be no indication of a need to update DSI, but when comparing the outputs of a non-stationary model to the EC-curves (obtained assuming stationarity) then the evidence is that we should update the DSI. This points in the direction of the EC-curves being different from the at-site stationary curves.

13. Page 14: I don't understand what the last sentence of the paper means.
14. SI3: I would give the lower and upper bound of the GEV in a formula to give a simpler indication of the effect of the value of the shape parameter
15. SI3.1: why using ML in one case and Bayesian methods for another?
16. SI3.1, paragraph after equation 3.8: $p(y|\lambda, x)$ does not give information on the parameters. The formulation of the sentence seem to imply that the likelihood $p(y|\lambda, x)$ gives information on the parameters under non-stationarity, which is not the case.
17. SI4.1 - the definition in eq 4.1 for the Akaike information criterion is not correct (or better it is correct for a normal model, but not for a GEV). AIC is generally defined as AIC $= -2log(L(\omega, x)) + 2m$. That's how the two references cited by the authors define the AIC as well. From what I understand from the explanation of the observed/expected values the authors are doing a model selection using AIC based on the quantiles, which is not made explicit in section 3.3. If that's the case, which quantiles are used?
18. Equation 5.1 and 5.2, what happens if $\zeta = 0$?
19. Table S7, 24-hours, the p-value for the pettitt test is larger than 1 - this can not be right. (see also S9 30min, S11 15min to 2hr, S14 15min to 2hr, S15 12hr)
20. Table SI16 - not sure if the red and blue are right in all stations
21. Pg 14 Supplement : the definition of return level has the word expected in the wrong place. *... often referred as return level in the literature is the expected value to be exceeded on an average once in every...* should be *... often referred as return level in the literature, is the value which is expected to be exceeded on an average once in every...* - see Coles, 2001 - end of section 3.1.3 (pg 49 in my edition)

I also find some of the Figures - and in particular their captions - could be improved

1. Figure 3 caption
   - durations higher than an hour are also shown I would say "Spatial distribution of trends, change points and non-stationarities in rainfall extremes of several durations in nine urbanized locations, Southern Ontario"
   - Drop the information on the population - it's in Figure 2 and in the text (several times)
   - Drop the information on the tests performed or at least reduce it since it's given in the text (for example drop the references)
   - Include information on the color coding in the legend.
   - If tests are performed at 5% and 10% - what is considered statistically significant? p-values < 0. 05 or p-value < 0.1?
2. Figure 4 caption: drop the list of the name of the station - it is given in the plot.
3. Figure 5 caption: add the information on the cyan shading representing the site with significant autocorrelation in the legend and drop from the legend. The second last sentence grammar is not correct.
4. Figure 7: I would include the information on solid/dotted lines in the legend

The paper has several grammar mistakes, with articles missing or appearing in the wrong place and several sentences which have non-concordant subject and verb. I list here a minuscule sample of the typos/mistakes I found

- Page 3, line 16 slowly or varying are not antonyms. Line 18-19 does should have a singular subject (not signal). Same in line 25-26.

- Page 3 line 23-24: The structure of the sentence is confusing. It is not the signatures that necessitate IDF. Maybe use "...make necessary the use..."
- Page 5: line 4-5 more repeated twice.
- Page 8: Line 27-28: the sentence is not complete
- Page 10 - line 16: less uncertainty (not lesser)
- Page 11 - line 17: More genralLY - and the sentence has a singular subject so line 19 should be is not are
- Page 12 - line 2: smaller, not lesser
- Page 12 - line 17: It? I think you need a "We"?
- Page 13 - line 6: does/is?
- Page 13 - line 12: several studies HAVE

Further inconsistencies I identified

- Page 4 Line 10: the ref to Jien and Gough is missing in the reference list and I think is not needed since it states a basic fact about the geography of Canada
- Page 9, Line 28 - ξ, instead of ζ used in he SI, for the shape parameter of the GEV
- Reference list: Cheng, L. and AghaKouchak, A. 2014 - just give the doi, not the ncbi link
- Supplement references: Coles and Tawn (1996) cited in text missing in the ref. Anyway, for that formula Coles, 2001 is probably enough as a citation.

- The citation to *Coles 2001, An introduction to statistical modelling of extreme values, Springer* in the supplementary material is wrong, as it has additional authors other than Coles.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-325, 2017.