

Interactive comment on “Frequency Analysis of Extreme Sub-Daily Precipitation under Stationary and Non-Stationary Conditions across Two Contrasting Hydroclimatic Environments” by Eleonora M. C. Demaria et al.

F. Serinaldi (Referee)

francesco.serinaldi@ncl.ac.uk

Received and published: 3 August 2017

General comments

This paper compares Intensity-Duration-Frequency (IDF) curves resulting from stationary and nonstationary frequency analysis. This exercise is not new in the literature, even though the Authors emphasize the quality of the analysed (fine scale) rainfall data set. The statistical methodology devised for such a type of analysis is already available both in frequentist and in Bayesian flavour. In this respect, the paper does not intro-

C1

duce any improvement; on the other hand, I found the application quite confused and questionable. To summarize, in my opinion, this paper is in line with a growing literature adopting some misconceptions and misinterpretations about these topics. In the following I'll provide some remarks, which however are secondary compared with the fact that (non)stationarity is a property of models and cannot be inferred from data only (as done in this paper), thus making these studies (based on trend tests and model selection) no very informative, if not meaningless and misleading. I regret to provide such a negative opinion. I hope that the other reviewers will provide a positive feedback.

Specific comments

P2L14-19: the Authors are right when they talk about 'stationarity of the model', but then they confuse stationarity (of the model) with low-frequency fluctuations of ENSO, AMO or whatever climate index as possible sources of nonstationarity. Low-frequency fluctuations resulting in local upward or downward trends (in observed time series) can be due to persistence or other causes. If we do not know the cause of such local trends, they can result from both nonstationary processes and stationary processes (with persistence, regime switch, etc.). The class of stationarity models goes far beyond the *i*/*id* case, and comprise (non)*i*/*id* processes whose realizations show local trends that can be confused with the effect of nonstationarity. Unfortunately, without identifying the generating mechanism, no definite conclusion can be made about the stationarity or nonstationarity of the underlying process. Concerning the positive increases in observed precipitation linked to a warmer atmosphere, it is worth recalling P1L28-29: the increase of water vapor is a model projection (we can also project vapor decrease, if we want), and the change of the magnitude and frequency of intense precipitation events is a hypothesis (among many others equally reasonable based on the available data).

P2L20: see also Rootzén and Katz (2013) and Serinaldi (2015) for a wider discussion.

C2

P2L22-24: 'A few climate studies have shown that design storm estimates from stationary models are significantly lower than estimates from nonstationary models'. If we assume an increasing trend in the parameters of a distribution (as done in those works and in the present paper), an increase of return level for a given return period or an increase of the risk of failure for a given design value are unavoidable, and trivial, as they are a direct consequence of the model structure. On the other hand, if we incorporate decreasing trends, we should conclude that the estimates from nonstationary models are significantly lower than estimates from stationary models (which is also a trivial conclusion, given the model structure). The point is that we do not know the actual direction without identifying the mechanism driving the evolution of the system. Only if we know such a mechanism and we identify a deterministic component, then the use of nonstationary modelling is legitimate; otherwise, it reduces to a simple numerical exercise of parameter fitting on a bunch of data (which is always possible, but uninformative). P2L1-6 and the outcome of the trend tests confirm that we actually have both positive and negative trends: if we fit stationary and nonstationary distributions, we will find that stationary estimates are lower (higher) than nonstationary estimates in the former (latter) case.

P3L21-27: Sincerely, I cannot see any large inter-decadal variability in Fig. 1b; on the contrary, the 11-yr moving average denotes a very regular (flat) behaviour. Please, do not confuse isolated minima and maxima with decadal patterns. Data and resolution are not enough to draw such conclusions, and the diagrams do not support them in any case. The same holds for Fig. 1c: please stop fitting untenable, unjustified, and evidently unsuitable straight lines to hydro-meteorological time series. I hope not to be the only one who sees that they are ill-posed and uninformative...maybe not (Poppick et al. 2017). If you do not agree with me, please try this experiment: fit straight lines to data for the periods 1961-1980, 1981-2002, and 1961-2002, then assess the significance of the trend slope, extrapolate to 2015, and finally draw conclusions about the nonstationarity of the process and the design values that we should have adopted. Actually, I think there is no steady increasing pattern, but only a very short time series

C3

whose length is not enough to draw conclusions about the nature of the observed variability, and to support nonstationary models. Concerning sample size requirements and the unreliability of nonstationary design values in future time windows, you can refer to Prosdocimi et al. (2014), Serinaldi and Kilsby (2015), and Luke et al. (2017).

P4L21-26: $T = 1/P$ and $T = 1/rP$ are not definitions but expected values obtained under specific model assumptions (see, Fernández and Salas, 1999; Douglas et al. 2002; Serinaldi 2015; Volpi et al. 2015; Salvadori et al. 2016)

P5L5: for an updated discussion on threshold selection see Langousis et al. (2016) and references therein.

P5L11-14: MK test does not check for linear trends but for stochastic dominance (double check Mann's paper and the rationale Mann-Whitney U-statistic). Moreover, if you check for monotonic trends, why should the trend pattern be restricted to the linear one? Why not an S-shaped pattern or something else?

P5L20-23: Please, pay attention to cookbook recipes! In their Eq. 10 for equivalent sample size, Yue and Wang (2002) merged the results for lag-1 ACF corresponding to a Markov model (from Matalas and Langbein (1962)) with those for all lags (lag-1 included) from Bayley and Hammersley (1946), which in turn do not rely on Markov assumption. Such hybridization is not required and mixes results relying on different hypotheses. Moreover, ACF estimates are prone to strong bias and required corrections depend (once again) on the assumed model (see, e.g. Koutsoyiannis, 2003; Serinaldi and Kilsby 2016).

P5L25-33: I do not understand which estimation method you use. In these few lines you mention a Bayesian framework without giving any detail, but in the same paragraph, P5L8-10, and in the remainder, you talk about maximum-likelihood estimates, and perform a frequentist analysis based on classical GoF tests, likelihood ratio tests (P6L12-22), etc. Please clarify. My guess is that nothing is Bayesian in this study. If I am wrong, please use a unique approach throughout the paper. Both frequentist

C4

and Bayesian approaches come with a full set of tools for complete inference. Choose the method that you prefer, but avoid exotic hybridizations because there is already enough confusion (and misuse) regarding statistics in hydrology.

P6L1-6: I am sorry, but I think that this is not the correct way to perform this type of analysis. You pretend to compare results from a GEV model where the location parameter (controlling the shift of the distribution) varies linearly with a GPD where the linear behaviour is assigned to the scale parameter related to the variability (spread) of the distribution. This is simply nonsensical: you should choose if the trend is in mean or in variance! Note that you performed trend tests only on the average levels, not on the squared residuals. At most, you should use a GPD with time varying threshold. Moreover, linear trends in the GEV and GPD cannot be justified by the outcome of the trend tests because the distribution introduces a nonlinear transformation between quantiles and location and scale parameters. Again, what means that you multiplied the 40000 parameters by the linear trend? This is not how NS inference works. In the NS framework your distribution is parametrized as $F(x; \mu_0, \mu_1, \sigma, \xi)$; so, your set of MC(MC?) parameters already describes the linear trend and its uncertainty. You do not need to multiply for anything. I only hope that you did not fit the model in a stationary set up, then using the fitted straight lines as correction factors. If so, your approach is not only empirically and theoretical inconsistent, but it also strongly underestimates one of the main sources of uncertainty, i.e., the variability of shape and magnitude of the parameter's trend.

P7L1: Eq. 8 confirms the confusion about the meaning and derivation of the concept of return period: T_t simply does not exist. In both stationary and nonstationary framework, the return period and the corresponding return level are unique value. NS risk of failure can only be written in terms of time-varying probabilities of exceedance P_t because T results from integration (averaging) over a time windows; writing $T_t = 1/P_t$ is meaningless. As mentioned above, $T = 1/P$ is not a definition (axiom) but a relationship derived from calculations under specified model assumptions.

C5

P7L19-31: It seems that the results of trend tests are somewhat mixed with positive and negative trends. So, why do you select two sites with only positive trends, and, in this set, you discard the case showing significant negative trend (temperate, PDS, 60-min)? Leaving aside that the use of trend tests to justify NS models makes little sense, a fair picture should show results for both increasing and decreasing cases. Of course, this would contradict the main conclusion, resulting is a less catchy but more realistic and obvious result: we have both positive and negative trends; they are likely related to random (short-term and long term) fluctuations (or at least the information is not enough to identify a true deterministic evolution), and then NS design values are lower or higher than the stationary counter-part.

P8L9-10: Why are you sure that negative trends are related to natural variability and do not deserve an NS model telling us that the risk is decreasing, whereas positive trends are due to a deterministic mechanism (which is required to justify the NS models) and need NS modelling? Why are negative trends 'children of a lesser god'?

P8L15-24: What means that the theoretical distributions were found statistically significant? Did they pass GoF tests or not? Fig. 3 shows that the fitting is pretty good. Since you have 40000 MC parameter sets, please complement point estimates with confidence intervals (or Bayesian credible intervals).

P8L26-31: The QQ plots in Fig. 4 should correspond to the results in Fig. 3. However, panels in Fig. 4 seem to contain much more data points, and the bad agreement on the upper and lower tails does not match with the good fit shown in Fig. 3. It seems that the QQ plots show results for the stations altogether. I hope to be wrong. Moreover, the temperate site is excluded from QQ plots because it does not show significant trends. Leaving aside that this site shows two significant trends for PDS at 60 and 1440 minutes (see Fig. 2), in P6L10-11 you write that 'This method [QQ plots] was only implemented for the stationary models since the parameters of the non-stationary models change with time'. Thus, you apply QQ plots for stationary cases, but you do not show QQ plots for the stationary case! Is this a three-card Monte

C6

game? :-) By the way, QQ plots can be used in the NS case after suitable rescaling (see, Furrer and Katz, 2008).

P9L1-9: I think that a fairer assessment should include all sites.

P9L15-19: As mentioned above, there is no evidence for any linear trend, and in any case, this does not translate into linear trends in GEV location, and even less in GPD scale parameter. Extrapolating the observed trend is not what NS models do. They extrapolate the law of variation of their parameters with the corresponding uncertainty.

P9L21-23: Please, use transparency to show the overlap of the confidence bands. As mentioned above, obviously NS design values are systematically higher than the stationary values according to the magnitude of the increasing trend introduced in the GEV location parameter. This is a trivial result. Please show also the NS cases with negative trends, where design values decrease compared with the stationary ones.

P9L25-30: 'These results indicate that a specific precipitation intensity will be more likely to occur, i.e., shorter return period, in the future if the observed trends are incorporated in the IDF design.'... For sure, and it will be less likely in the cases of negative trends... and will be unchanged if the trend is null... and water is wet :-) I hope you will understand that the above sentence is pleonastic, actually tautological.

P10L1-12: In text and Fig.6, you refer to absolute differences, but the caption reports the expression of the relative percentage difference. So, what is shown in Fig.6? Which values are discussed in the text? Absolute or relative?

P10L1-12: 'This demonstrates that the stationary framework currently used for structural design systematically underestimate short-duration precipitation extremes which might lead to more frequent infrastructure damage'. I think that this demonstrates that confidence intervals are missing around the point estimates! Add them, and probably the story will change a little bit.

Fig 7: Please add confidence intervals, show cases with negative trends and perform
C7

a fair comparison.

Fig. 8: I cannot see where the variability of the project life appears in these diagrams. Please, clarify.

Fig. 9: Please, add confidence intervals.

Editing remarks

Please, check a few typos throughout the text.

Sincerely,

Francesco Serinaldi

References

Douglas, E. M., R. M. Vogel, and C. N. Kroll (2002), Impact of streamflow persistence on hydrologic design, *J. Hydrol. Eng.*, 7(3), 220–227
Fernández, B., and J. D. Salas (1999), Return period and risk of hydrologic events. I: mathematical formulation, *J. Hydrol. Eng.*, 4(4), 297–307

Furrer, E.M., and R.W. Katz, 2008: Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, 44, W12439

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

Langousis, A., A. Mamalakis, M. Puliga, and R. Deidda (2016), Threshold detection for the generalized Pareto distribution: Review of representative methods and application

to the NOAA NCDC daily rainfall database, *Water Resour. Res.*, 52, 2659–2681

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., Sanders, B. F., (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water Resources Research* 53.

Poppick, A., Moyer, E. J., Stein, M. L., 2017. Estimating trends in the global mean temperature record. *Advances in Statistical Climatology, Meteorology and Oceanography* 3 (1), 33-53.

Prosdocimi, I., Kjeldsen, T. R., and Svensson, C. (2014) Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK, *Nat. Hazards Earth Syst. Sci.*, 14, 1125-1144.

Rootzén H, Katz RW (2013) Design life level: quantifying risk in a changing climate. *Water Resour Res* 49(9):5964–5972

Salvadori, G., F. Durante, C. De Michele, M. Bernardi, and L. Petrella (2016), A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities, *Water Resour. Res.*, 52, 3701–3721

Serinaldi F. Dismissing return periods! (2015). *Stochastic Environmental Research and Risk Assessment*, 29(4), 1179-1189

Serinaldi F, Kilsby CG. The importance of prewhitening in change point analysis under persistence. *Stochastic Environmental Research and Risk Assessment* 2016, 30(2), 763-777

Volpi, E., A. Fiori, S. Grimaldi, F. Lombardo, and D. Koutsoyiannis (2015), One hundred years of return period: Strengths and limitations, *Water Resour. Res.*, 51, 8570–8585

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2017-247>, 2017.