Responses to Referee #1

Dear Referee #1,

We really appreciate your rapid and constructive comments on our manuscript entitled "Comparative study of flood projections under the climate scenarios: links with sampling schemes, probability distribution models, and return level concepts" (Number: hess-2016-619) that are very helpful to improve our study and paper. We have carefully followed these comments and accordingly made the revisions. Please see our point-by-point reply below.

General comment 1:

The paper discusses the estimation of flood return levels in the nonstationary context and applies the ENE (Expected Number of Exceedances) concept for both block maxima and Peak Over Thresholds approaches. It is an interesting, well presented and documented study using pertinent methodologies. The flow data used for the application extend over the period 1960-2009, which leads to 50 annual maxima, which is good but not perfect for a robust statistical fitting.

Response:

We thank the referee very much for the positive evaluation of our paper and the valuable comments. Indeed, a long-period dataset is essential for a robust statistical fitting as the longer series are considered to better represent the population. Nevertheless, a compromise has to be made for a real-world application since long-historical records are often sparse due to accidental (e.g., equipment malfunction) or human-induced reasons (e.g., management failure). This is why even shorter data series are used in many other studies, e.g., Villarini et al. (2009) used 26 annual flood peak of the Little Sugar Creek watershed in Charlotte, North Carolina to fit a gamma distribution model under nonstationarity. Seckin et al. (2011) conducted flood frequency analysis with the annual flood peak records varying in lengths from 15 to 57 years.

Following the referee's suggestion and in order to further improve the statistical practice, we have attempted to collect more available data for our study, which is now extended to 2012, thus the study period is from 1951 to 2012 (totally 62 years). Due to the augmentation of data for the study, the related statements, figures and tables in the data description and result analysis sections have been changed accordingly in the revised manuscript.

<u>Reference</u>

Villarini, G., Smith, J.A., Serinaldi, F., Bales, J., Bates, P.D., and Krajewski, W.F.: Flood frequency analysis for nonstationary annual peak records in an urban drainage basin, Adv. Water Resour., 32, 1255-1266, doi:10.1016/j.advwatres.2009.05.003, 2009.

Seckin, N., Haktanir, T., and Yurtal, R.: Flood frequency analysis of Turkey using L-moments method, Hydrol. Process., 25, 3499-3505, doi:10.1002/hyp.8077, 2011.

General comment 2:

Could you argue why other distributions than the asymptotic limit GEV distributions are considered for fitting annual maxima? LP3 is pointed as the best choice, but it is the one with the largest standard errors on the parameter estimations in the stationarity case.

Response:

Thank you for the valuable comments. In many available literatures, flood frequency analysis for annual maxima (AM) data has been frequently carried out with several theoretical distributions selected from the Normal family (e.g., normal, lognormal), the Gamma family (e.g., gamma, Pearson type 3, Log-Pearson type3), and the Extreme Values family (e.g., Weibull, Gumbel, Generalized Pareto) since there is no conclusive standard to define the adoption of a single type of optimum distribution for AM series. All the distributions presented above have been widely used, and their respective advantages found at different gauging sites and/or in different river basins have been proven (Strupczewski et al., 2001; Kidson and Richards, 2005; Villarini et al., 2009). The asymptotic limit GEV distribution is a theoretical result based on Extreme Values Theory and may not be always applicable under the complex hydrological circumstances and/or with the finite observation data. Additionally, distributions recommended for flood frequency analysis by national standards in many countries are often different, e.g., LP3 distribution is officially specified in both USA and Australia (Vogel et al., 1993), while P3/LP3 is recommended in China. Therefore, except for the GEV distribution, we have employed other distributions (i.e., LNO3, LP3) from two different distribution families in this paper. These three distributions all have the parameters of location, scale, and shape, which we consider can allow a flexible fit for annual maxima data. To better account for the choice of distributions for AM series, we have clarified these points in the newly revised manuscript.

The model selection follows a generalized Akaike information criterion, i.e., AIC and BIC, to make a tradeoff between model structure complexity and goodness of fit. The model with minimum generalized Akaike information criterion value is preferentially considered, and it will be finally chosen as the best once if the significance test of parameter estimations and goodness-of-fit test can pass at the 5% significance level. From this, we can see that the selected best model does not necessarily mean a lowest standard error for parameter estimations (due to the uncertainty problem incurred by limited observation data). As the referee pointed out, this study indeed has found that the LP3 model yields a larger standard error (but the difference is minor) in the estimated statistical parameters than other used probability distribution models, including GEV, in the stationary context. However, this LP3 model does have the minimum AIC/BIC value when compared to other used probability distribution models such as GEV, and has passed the 5% significance test for statistical parameters.

<u>Reference</u>

Kidson, R., and Richards, K.S.: Flood frequency analysis: assumptions and alternatives, Prog. Phys. Geog., 29, 392-410, doi:10.1191/0309133305pp454ra, 2005.

Strupczewski, W.G., Singh, V.P., and Mitosek, H.T.: Non-stationary approach to at-site flood frequency modelling. III. Flood analysis of Polish rivers, J. Hydrol., 248, 152-167, doi:10.1016/S0022-1694(01)00399-7, 2001.

Villarini, G., Smith, J.A., Serinaldi, F., Bales, J., Bates, P.D., and Krajewski, W.F.: Flood frequency analysis for nonstationary annual peak records in an urban drainage basin, Adv. Water Resour., 32, 1255-1266, doi:10.1016/j.advwatres.2009.05.003, 2009.

Vogel, R.M., McMahon, T.A., and Chiew, F.H.S.: Flood flow frequency model selection in Australia, J. Hydrol., 146, 421-449, doi:10.1016/0022-1694(93)90288-K, 1993.

General comment 3:

Concerning the POT approach, in theory, under stationary conditions, the return levels should be the same whatever the approach used to estimate them, which is the case here if we compare POT2 and GEV with their confidence intervals (which are larger for GEV because the fitting is made with less values). I have questions concerning the POT approach.

3.1 Threshold choice: there are some rules to choose a convenient threshold for the POT estimation, based on the mean excess plot and/or the constancy of the shape and modified scale parameters. Considering the very different results obtained with POT4 compared to POT 2 or 3 (and GEV), it is doubtful that the threshold used for POT4 is a convenient threshold.

Response:

Thanks for this comment and suggestion. In the original paper, we followed the criteria in USWRC (1982) to define time span between successive extreme events, which guarantees the independence of the selected peak flow values. We then choose the threshold value *u* for POT flood samples according to a widely-used method proposed in Lang et al. (1999), i.e., the preselected annual number of peaks per year on average. For example, POT2 threshold is determined based on that the number of selected peaks is twice the number of years of the study record. This procedure for threshold choice includes somewhat a level of subjectivity since there is no absolutely the best approach to determine the threshold, and no unique value that must be selected but rather a range of appropriate values. Considering this lack of standard methods in determining the threshold, we apply three different threshold values to define the POT samples for our study objectives and want to see if the results are dependent on the threshold choice.

In accordance with the referee's comment and suggestion, we have added in the newly revised manuscript the tests, namely, the mean excess plot and the plot for estimated shape and scale parameters, to evaluate the reasonability of the selected POT samples. In fact, the three used threshold values are all tested to be acceptable, but the use of POT4 as the referee pointed out may not be the most appropriate choice for flood design in the study basin of the Weihe. For example, the results of the mean excess plot indicate that the POT4 threshold almost approaches (but is within) the lower bound of domain where the mean excess should be an approximately linear function of *u* for a valid choice of the generalized Pareto (GP) distribution.

Reference

Lang, M., Ouarda, T.B.M.J., and Bobée, B.: Towards operational guidelines for over-threshold modeling, J. Hydrol., 225, 103-117, doi:10.1016/S0022-1694(99)00167-5, 1999.

USWRC: Guidelines for Determining Flood Flow Frequency, United States Water Resources Committee, Washington, DC, USA, 1982.

3.2 Seasonality: the threshold exceedances have to be independent (which is correctly dealt with in the excesses selection) and identically distributed. This second condition is not considered here, but may not be straightforward for environmental variables, because of seasonality and possibly inter-annual variability. Regardless of inter-annual variability, is there a preferred season for the occurrence of floods in this basin? If so, then it may be necessary to restrict the analysis to this season. This has an impact on the estimated Poisson intensity. The Poisson process however does not need to be homogeneous, when nonstationarity is introduced, it is a non-homogeneous Poisson process.

Response:

Thanks so much for this constructive comment. The seasonal (intra-annual) variability of POT flood, as the referee stated, does exert an important effect on the assumption of Poisson process, which however was not taken into account before. In the new version, we have supplemented the analysis of the distribution and strength of the seasonality based on the longer flow record (as explained in the response to General comment 1) in the Weihe basin, China. The result shown as circular data (Pewsey et al., 2013) in the Figure 1 indicates that most flood events tend to occur during the July-October period with an approximately unimodal distribution. The above information indicates that the intra-annual variability of POT arrival rate is not prominent in this study. Choosing a whole year or a restricted season will lead to the same results.

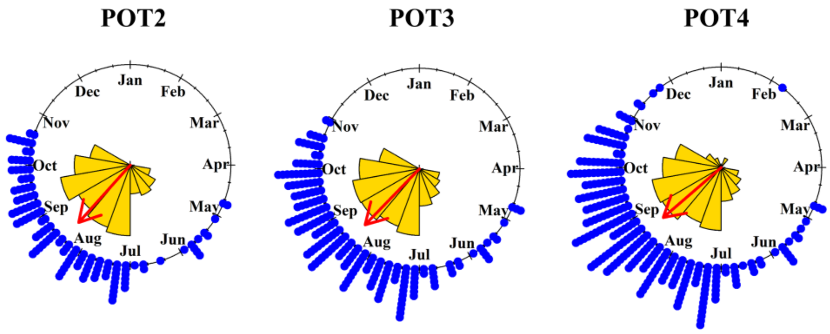

Figure 1 Flood events (blue points) of the POT2, POT3, and POT4 shown on the circular time axis. The red solid arrow is the mean resultant vector indicating the average occurrence time of the events.

Reference

Pewsey, A., Neuhäuser, M., and Ruxton, G. D.: Circular statistics in R. Oxford University Press, 2013.

3.3 Arrival rate: the use of a Negative Binomial is interesting, but it brings one more parameter to estimate with still quite few values. Could you discriminate the advantage brought by this approach compared to this necessity to estimate another parameter?

Response:

We thank the referee's approval on our work and the good suggestion. To address the question raised by the referee, we firstly recall the traditional use of the Poisson distribution in the POT analysis. The Poisson distribution has the single parameter termed Poisson process intensity. It is characterized by the identity of variance and mean of population, both of which equal to the Poisson process intensity. The POT arrival rate can be fitted by a stationary Poisson model (with constant Poisson process intensity) if it is independent and identically distributed and follows a homogeneous Poisson process, while a nonstationary Poisson model with time-varying Poisson process intensity can be assumed if the POT arrival rate is independent but not identically distributed and follows a nonhomogeneous Poisson process.

The attractiveness of the use of Negative Binomial (NB) distribution in contrast with the Poisson distribution is explained as follows:

(1) The derivation of the NB distribution is theoretically an extension of the Poisson distribution that mixes Poisson process intensity with a gamma distribution (Anscombe, 1950) (Please refer to Table 1 for the probability mass function in the original manuscript), i.e., the Poisson distribution is a special case of the NB distribution.

(2) The assumption of Poisson distribution is invalid when data is over-dispersed (variance-to-mean ratio greater than unity), while the NB distribution is theoretically justifiable for describing over-dispersed data and has been frequently used in literatures (e.g., Bhunya et al., 2013). Most studies have applied a stationary NB model (with constant parameters) to fit the over-dispersed data, but only a few ones have focused on evaluating whether the over-dispersed data have also shown a nonstationary behavior over a certain long time period. Therefore, this study has been partially aimed to compare the accuracy of stationary and nonstationary NB models for fitting the over-dispersed data.

(3) The requirement of independent data is relaxed when we fit a nonstationary NB model (which is strictly required by the Poisson model, whether with constant or time-varying Poisson process intensity). This advantage has made the NB distribution become increasingly popular especially when it is doubtful whether the observed arrival rates from a stochastic process satisfy the assumption of independence (Johnson et al., 1992).

Reference

Anscombe, F.J.: Sampling theory of the negative binomial and logarithmic series distributions, Biometrika, 37, 358-382, doi:10.2307/2332388, 1950.

Bhunya, P. K., Berndtsson, R., Jain, S. K., and Kumar, R.: Flood analysis using negative binomial and Generalized Pareto models in partial duration series (PDS), J. Hydrol., 497, 121-132, doi:10.1016/j.jhydrol.2013.05.047, 2013.

Johnson, N. L., Kemp, A. W., and Kotz, S.: Univariate Discrete Distributions, Second Edition, New York, John Wiley and Sons, 1992.

General comment 4:

Then the sensitivity analysis is very interesting, as well as the variations in return levels induced by the separate increases of Ptotal and Tmean. It could be much more informative if a choice had been done previously of the best model for the estimation.

Response:

We are pleased that our work can be appreciated by the referee and we would like to thank the referee for the valuable remark. We agree with the reviewer that it could be more informative and more instructive if the best model had been chosen in prior. But, in this study, both the choice of the best model and the sensitivity analysis of flood estimation to changing climate resulted from different models as measured by Ptotal and Tmean are the research topics.

Thanks again for your professional and valuable comments in reviewing our paper.

With best wishes

Yours sincerely

Lihua Xiong, PhD, Professor
State Key Laboratory of Water Resources and Hydropower Engineering Science
Wuhan University
Wuhan 430072, PR China
E-mail: xionglh@whu.edu.cn
Telephone: +86-13871078660
Fax: +86-27-68773568