

## ***Interactive comment on “The application of new distribution in determining extreme hydrologic events such as floods” by Łukasz Gruss et al.***

### **Anonymous Referee #2**

Received and published: 3 August 2020

Dear authors,

The paper represents a large amount of work, but I feel the presentation does not do your work justice. There are also some methodological questions that need to be clarified. The review is split into two parts: questions and comment concerning the content, and questions and comments on the style (formulations used, possible typing errors).

### SUMMARY

The paper studies the applicability of several distributions for six rivers in the Upper  
C1

Oder basin. It concludes that further study is needed to determine the choice of distribution for the region.

### CONTENT

General remark: it would be nice to have more justification for the use of GGEV. For instance, theoretical reasons or practical considerations such as use by one or more governments. A better fit to the data on its own is not a very strong argument. In this context the paper of Vogel and McMartin (1991) is interesting: “Probability plots for the P3 and LP3 distribution based on an estimate of the sample skew will, in general, appear more linear than they should. Essentially, the estimated sample skew acts to adjust the probability scale to make the sample, when plotted, appear more linear than it would if the the skew had been used to construct the plot.” This suggests that great care must be taken to avoid overfitting and misleading fits, specially when comparing distributions with different numbers of parameters.

Line 211. Alexandersson (1986) originally intended his test to be used on series of ratios or differences with respect to a series of, possibly weighted, means of the measurements of a group of surrounding stations. Could you elaborate on how it was applied here? Given that Alexandersson (1986) assumed the ratios to be normally distributed, can you indicate why it should be suitable for series of extremes?

Line 224. To the best of my knowledge, the POT method is closely linked to extreme value theory, and the corresponding distribution to be used in fitting the data is the Generalized Pareto distribution. Please justify its use with other distributions.

Line 303. Please specify the details of the Chi-square test such as class boundaries and degrees of freedom after correction for number of fitted parameters. Please indicate how the K-S test statistic was converted to a p-value. Was the limit distribution

used? Please explicitly state how a correction was made for the number of parameters being fitted, because the standard KS test statistic distribution does not apply when comparing an empirical distribution for given data to a distribution fitted to the same data.

Line 425-438. It is customary to look not only at goodness of fit but also at the number of parameters when selecting a distribution. This is done to avoid rewarding the over-fitting of data. I feel this should be added to your analysis. Especially because in a combination of POT and GGEV there are actually five parameters being chosen.

Line 429. The purpose of both tests in your paper is not to simply reject the null hypothesis, but to reject the null hypothesis when the alternate hypothesis is true. In that case the power of the test should be examined, not the number of combinations of distribution and fitting method it rejects. The number of rejected combinations of distribution and fitting method includes type one errors. Please clarify your meaning.

## STYLE

Abstract line 3: I think "with a change-point" should be "without a change-point".

Abstract line 28: "a GGEV water reservoir". What is a GGEV water reservoir?

Line 34. Is a new paragraph here necessary? It seems a continuation of the previous lines.

Line 38-45. "During ... (Pollert, 45 2006)." This seems a series of disconnected sentences, please consider rewriting.

Line 59, 60. "Therefore ...". The preceding part of this paragraph states the importance of time series analysis and the study of extremes. But in this sentence you decide to investigate rivers that are important to the water management of the Upper Oder

C3

basin, seemingly unconnected to the preceding part of the paragraph. So why use "therefore"?

Line 70 "analyzes" should be "analyses".

Line 72. "FFA is also used to fit a probability distribution to an empirical distribution function ... ." As far as I know, flood frequency analysis is the process of studying past floods. Fitting a distribution to an empirical distribution function can be part of that process, but I do not see how a generic process can be used to do distribution fitting. Please clarify what you mean by FFA.

Line 75-80. "In time series modeling ...". Jump to a new topic (independence, trends, etc.); please improve coherence.

Line 81-109. New topic (choice of distribution); please link it to preceding material.

Line 108. It would be nice if a clear motivation for both your choice (three or more parameters) and that of many others (to parameters) was presented. Are there specific disadvantages to three parameter distributions?

Line 110-124. New topic (choice of fitting method); please add introduction linking it to this paper.

Line 125-144. New topic; please link it to preceding material.

Line 145-156. New topic; please link it to preceding material.

Line 160. "Additionally the GGEV distribution is the best suited empirical distribution irrespective of sample independence". The GGEV is not an empirical distribution. The empirical distribution is a clearly defined concept in statistics. Do you mean the GGEV fits the data best? Are you drawing a conclusion in the introduction?

C4

Line 165. There is a part of a sentence missing between “The catchments of these last two rivers are” and “The Budkowiczanka River is 56.5 km long.”

Line 170. “MM” in “80.04 MM m<sup>3</sup>” should be “M”, but even then it is not correct as ISO prefixes bind closely to the unit, so 1000000m<sup>3</sup> = 1hm<sup>3</sup>.

Line 175. Gruss et al (2019) place the source of the Widawa at 109.02 km of the river’s course. How does that relate to the length of 114.6 km mentioned here?

Line 179. Sentence ends with “a Normal Pool Capacity of 1 MM cm<sup>3</sup>”; I expect this should be 1 hm<sup>3</sup>.

Line 187-193. Should most of this not be in the introduction?

Line 201. “and change point detection” should be “and the presence of change points”.

Line 211. “used to analyze the change-point”. Phrasing seems to assume there is a change point; do you mean: “used to check for the presence of a change point” ?

Line 247. What is meant here by “verified”?

Line 255. The term "empirical input moments" is not in use as far as I know; please write "empirical moments" instead.

Line 257. “The probability of this sample must be maximal, because the sample observed comes from many other possible samples (Haktanir, 2009).” Please either remove this sentence or replace it by a longer explanation. As it stands, it does not help the reader to understand the method.

Line 260. “In the gamma distribution developed by Becker and Klößner (2017), ...”. Becker and Klößner (2017) did not develop the Gamma distribution but a package

C5

for the Pearson distribution system. Moreover, the Pearson III distribution has three parameters and is therefore not usually referred to as “the” Gamma distribution which traditionally has two parameters.

Line 302. “The Chi-squared Test (q2), Kolmogorov-Smirnov (K-S), and the Mean absolute relative error (MARE) tests were widely used to indicate the adequacy of the distribution functions being tested”. Meaning of “widely used” in this sentence is unclear. Do you mean in the literature, in practice, in this paper?

Line 316. “The MK test showed no trends neither in the AM method (except for the O sample) nor in POT (except for samples BB and O).” This means the MK test showed trends in both methods. I assume you meant: “The MK test showed trends neither for the AM values (except for the O sample) nor for POT (except for samples BB and O).”

Line 320. “Also, based on the test result, which was not statistically significant (5%) Cassalho et al. (2018) rejected 7 out of 113 series for the Rio Grande do Sul in Brazil.” Too brief, please rewrite to make meaning clearer because at the moment it can be misunderstood. Cassalho et al. (2018) state: “Based on the non-parametric Mann-Kendall test, at a significance level of 5%, only 7 out of 113 series (Fig. 2) presented significant monotonic trend, thus, they were not used for the sequence of this study.” Thus, 7 series are rejected because for those series the result was statistically significant at a significance level of 5%.

Line 322. “They also relied on a significance level of 5%. Most samples did not meet this criterion.” What is the criterion you refer to? In the reference 3 out of 9 series have p-values below 5%. In your sentence the criterion is: the null hypothesis of no trend is rejected at the 5% significance level. In the present context where the aim is to select series without trend, the term “criterion” might be misinterpreted. Please rewrite this line.

C6

Line 324. "Test B showed that for two samples: MPT and O analyzed in the AM method, the series are not random. Thus, in these cases the H0 hypothesis was rejected." Please make clear what H0 is. Given the context of this paper there are 7 candidates:

a: "There is no trend"

b: "The series is random"

c: "There is no change point"

and their combinations: a and b; a and c; b and c; a,b, and c.

Line 327. "Bezak et al. (2014) used the von Neumann's ratio test whose test statistics were compared with a critical value. This test is based on a rank version proposed Bartels (1982) for testing a series for randomness." Why is this sentence here? Should it not be in Section 2.2.1 or in the introduction?

Line 348. Typo: "He" should be "he".

Line 349. "He obtained a p-value close to 1 in the K-S goodness-of-fit test for each of the four estimates in the analyzed distribution, which indicates that GPD distribution fits very well with empirical data." The p-value is not a measure of fit; it is an indication of how likely it is to get a specific test statistic value for a random sample from a given distribution. Please emphasize this somewhere in the paper.

Line 367. "In the case when the value of  $p > 0.05$  for the analyzed distribution, then it showed the lack of the best fit of the empirical distribution with the theoretical distribution." If I read Table 2 in Szulczewski and Jakubowski (2018) correctly, then  $p < 0.05$  leads to rejection of the hypothesis that the sample is from the given distribution; here you state the opposite. Please clarify.

Line 372. "in the case of the GGEV distribution it is more difficult to work with four parameters trying to adjust this distribution". This is a highly unusual finding; normally,

C7

more parameters result in a better fit. Please discuss this some more.

Line 379. chi square symbol is not displayed correctly.

Line 425. "Out of the many methods used for estimating the 3-parameter distributions in accordance with ... the best-fitted parameters were obtained by the MMM and by the MLE". MM, MMM, and MLE are the only methods mentioned in the paper; the sentence mentions two of out of three, thus the phrase "Out of the many methods" seems out of place.

Table 2. What is meant by "rH0 - H0 hypothesis was rejected."? It does not seem related to the p-values in the same column.

Table 3, footnote. The K-S statistic itself is a measure of the distance between two cumulative distribution functions, but the associated p-value is not.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2020-173>, 2020.