

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

Łukasz Gruss et al.

lukasz.gruss@upwr.edu.pl

Received and published: 18 September 2020

Dear Referee, Thank you for your comment. We do appreciate your constructive suggestions. Below I present explanations, additions and corrections.

Reply to general comments:

Referee: 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.

Reply: We agree with the Referee. Below is a justification that we propose to add to the introduction: Madsen et al. (2013) created a report of flood frequency analysis where they state that both Poland and the Czech Republic have plans for further research

C1

activities. The important information is that in the case of Poland: the annual peak cycles of Polish rivers are a mixture of summer and winter flows. A flood regime can be affected by many factors such as land cover change, canal modifications, drainage works and presumably climate change. As further reported by Madsen et al. (2013) in the Czech Republic, in the FFA analysis, a flood regime can be affected by regional precipitation of a longer duration, occurring in catchments with an area of more than 100 km². For water bodies with a winter flood regime, snow melting should be taken into account. Additionally, it was noted in the report that, in addition to catastrophic floods, there are flash floods in mountainous areas. In both countries, an important role is attached to data compilation because in the Czech Republic and in Poland flood frequency estimation is necessary for the design of hydraulic structures, dams, urban, hydrology, flood-hazard mapping. In the report, these countries did not indicate the application of a specific likelihood distribution for these countries, hence our proposal is a new GGEV distribution.

References: Madsen, H., Lawrence, D., Lang, M., Martinkova, M., Kjeldsen, T.R.: A review of applied methods in Europe for flood-frequency analysis in a changing environment: Floodfreq COST action ES0901: European procedures for flood frequency estimation, Department of Architecture & Civil Engineering, Wallingford, U. K., 180, 2013.

Referee: 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.

Reply: Thank you for pointing this. We propose to extend the results with the selection

C2

of the best-fitted distribution using information criteria - Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Referee: 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?

Reply: Thank you for paying attention to this. After reviewing the documentation for the test, we find that we were unable to apply this test to determine the change point detection, and we want to withdraw from it. We propose a change to line 221: "Long time series of the six profiles were checked for trend and randomness". We propose to delete the entire paragraph on SNHT on the line 239-243 and the results and discussion on the line 329-334.

Referee: 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.

Reply: Thank you very much for this suggestion. We wanted to supplement our answer and propose to add the justification at the end of the sentence on line 75 (is below): As reported by Bezak et al. (2014) in the POT method the Exponential and generalized Pareto distributions can be used. Instead of these distributions, one can also use the LN distribution (Adamson and Zucchini, 1984, Rosbjerg, 1987), and the Weibull (Bačová-Mitková and Onderka 2010, Dimitrov 2016,) distribution functions. Also, Wong and Li (2010) use the Weibull and gamma distributions in the POT method. Likewise, Xu et al. (2019) applied 3W and GEV in POT method. The 3W distribution provides a very good estimation of short-term extreme value. They applied two assumptions: the selected peaks are Poisson distributed, and the exceedances should be approximately independent. In their study, the dispersion index is applied to select clusters and check

C3

the Poisson character. In turn, Dimitrov (2016) used 3W in the POT method. He points out that in the POT method, all independent response peaks, which exceed a certain high threshold level, are included in the analysis. Addition, make sure that each peak corresponds to an independent event.

References: Dimitrov N.: Comparative analysis of methods for modelling the short-term probability distribution of extreme wind turbine loads, *Wind Energy*, 19, 717–737 10.1002/we.1861, 2016.

Xu S., Ji C.Y., Guedes Soares C.: Estimation of short-term extreme responses of a semi-submersible moored by two hybrid mooring systems, *Ocean Eng*, p. 190106388, <https://doi.org/10.1016/j.oceaneng.2019.106388>, 2019.

Wong T.S.T., Li W.K.: A threshold approach for peaks-over-threshold modeling using maximum product of spacings. *Stat Sini* 20(3),1257–1272, www.jstor.org/stable/24309490, 2010.

Adamson, P.T. and Zucchini, W.: On the application of a censored log-normal distribution to partial duration series of storms. *Water SA*, 10 (3), 136–146, 1984.

Rosbjerg, D.: On the annual maximum distribution in dependent partial duration series. *Stochastic Hydrology and Hydraulics*, 1 (1), 3–16. doi:10.1007/BF01543906, 1987,

Referee: 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.

Reply: We propose to supplement the text with the following entry at the end of line 305: The smaller the χ^2 , the better the expected fit of the model to the sample being tested (Haktanir, 1991). In calculating the statistics, the R package 'stats' was used - 'chisq.test' function. with continuity correction performed.

References: R Core Team and contributors worldwide: The R Stats Package, 'stats', <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>, 2020.

C4

Referee: Please indicate how the K-S test statistic was converted to a p-value. Was the limit distribution used?

Reply: In the one-sample two-sided case, exact p-values are obtained as described in Marsaglia, Tsang & Wang (2003). In order to obtain p-value, the `gofTest` function uses an algorithm written in C. The p-value is an indication of how likely it is to get a specific test statistic value for a random sample from a given distribution. Using the `gofTest` functions, there was no limit to the use of the distribution. R Package 'EnvStats' was used - `gofTest` function that calculates statistics and p-value. We propose to supplement the text with the following entry at the end of line 308: In calculating the statistics, the R package 'stats' was used - '`chisq.test`' function. with continuity correction performed.

Referee: 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.

Reply: The K-S statistic (D_{max}) was calculated using the '`gofTest`' function and a p-value was obtained for each tested distribution. The p-value was calculated only if the data follow a specified distribution.

Referee: 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 overfitting 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.

Reply: In the methods, we showed which distributions have as many parameters. Afterwards, in the results in Figure 4 (a) we showed the sample sizes, and in Figure 4 (b) the threshold sizes in the POT method. We have a lot of results that we did not include in the article. Please specify more precisely what we could include?

C5

Referee: 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.

Reply: We agree with the Referee. Therefore, we propose to delete the second conclusion from line 429.

STYLE

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

Reply: We agree with the Referee. We propose to correct the sentence: However, in order to use distributions, the data must be random, without a change-point and should not have a trend.

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

Reply: We agree with the Referee. We propose to correct the sentence: This distribution turned out to be the best fit especially for the sample whose independence is affected by the presence of a water reservoir.

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

Reply: We agree with the Referee.

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

Reply: We agree with the Referee. We propose to correct the sentence: Many floods of different intensity and extent took place on the Oder and its tributaries in the 20th century and in the beginning of the 21st century (Dubicki et al., 2005). The flood that

C6

occurred in Poland in the Oder and the Vistula basins in the summer 1997 caused 54 fatalities and material losses estimated at billions of USD (Kundzewicz et al., 1999). Afterwards extreme fluvial flooding took place in many parts of the Czech Republic in August 2002. This flood overwhelmed most of existing flood protection systems and caused damage exceeding EUR 3 billion (Holick and Skora, 2010).

We propose to delete some of the text that begins with the sentence "During the catastrophic flood ..." on line 38 up to the end of line 45.

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

Reply: We agree with the Referee. We propose to correct the sentence: Therefore, we decided that our research would be carried out on rivers, whose proper use has a significant impact on water management and which play an important role in designing hydrotechnical structures of the Upper Oder basin.

Referee: Line 70 "analyzes" should be "analyses".

Reply: We agree with the Referee. We propose to correct the sentence: Long time observation series were processed using the Flood Frequency Analysis (FFA), so that the distribution analyses could be carried out later.

Referee: 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.

Reply: The authors meant: FFA is also used to fit a probability distribution to a given

C7

maximum flow series dataset in order to estimate the annual exceedance probability for a given flood flow. We propose to correct the sentence: Long time observation series were processed using the Flood Frequency Analysis (FFA), so that the distribution analyzes could be carried out later. FFA is often adopted to investigate the relationships between flood magnitude and the corresponding frequency of occurrence (Gharib et al., 2017). FFA is also used to fit a probability distribution to a given maximum flow series dataset in order to estimate the annual exceedance probability for a given flood discharge (Rahman et al., 2015; Haktanir, 1991; Lang et al., 1999; Silva et al., 2012; Yadav and Pande, 1998).

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

Reply: We agree with the Referee. We would like to give it from a new paragraph.

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

Reply: We agree with the Referee. We suggest adding a sentence: In turn, Baretts (1982) reports that testing of sample randomness is of fundamental importance in statistics. In the estimation of the distribution of AM and POT method's, it is a generally accepted assumption that the sequence of observations is the independent and identically distributed (Gharib et al., 2017, Szulczewski and Jakubowski, 2018). Over the past 20 years, research has been conducted on testing various distributions and methods for estimating their parameters have been developed.

Referee: 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?

Reply: On lines from 106 to 109 is the information about 3-parameter distributions. Their disadvantage is the difficulty in estimating parameters, as reported by Kidson

C8

and Richards (2005). We also mention this topic in the results and discussions. We suggest leaving it like this.

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

Reply: We agree with the Referee. We suggest adding a sentence: The parameters of the probability distributions should be estimated. However, the estimators may not be unique in a given dataset, and thus can provide multiple solutions (Langat et al. 2019). Various methods of estimating distribution parameters have been studied. Different scientists came to different conclusions. For the LN, P3 and GEV the Maximum Likelihood Estimator (MLE) is recommended (Szulczewski and Jakubowski, 2018), whereas the L-moments method was used for the GEV, LN3, P3, GLO, KAP and WAK (Cassalho et al., 2018).

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

Reply: We agree with the Referee. We propose to correct the sentence: The distribution parameters are estimated based on the maximum annual series (Cassalho et al., 2018). However, factors such as anthropogenic impact, climate change or spatial distribution of precipitation generate changes in the frequency of observed floods.

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

Reply: Proponujemy dodanie krótkiego wstępu i rozpoczęcie zdania od nowego akapitu: We suggest adding a short introduction and starting the sentence with a new paragraph: In turn, according to Otiniano et al. (2019), new extensions of two - and three-parameter distributions were created, which may constitute a new class of distributions.

Referee: 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

C9

mean the GGEV fits the data best? Are you drawing a conclusion in the introduction?

Reply: Thank you for this suggestion. This is not a conclusion, but it is a second hypothesis. The sentence on line 160 should read: Additionally the GGEV distribution is the best suited to the empirical distribution irrespective of sample independence.

Referee: 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."

Reply: We agree with the Referee. The sentence "The catchments of these last two rivers are" will be deleted.

Referee: Line 170. "MM" in "80.04 MM m³" should be "M", but even then it is not correct as ISO prefixes bind closely to the unit, so 1000000m³ = 1hm³. Reply: We agree with the Referee. We propose to correct the units in this sentence.

Referee: 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?

Reply: Kilometres are correct. The total length of the watercourse is 114.6 km. The Widawa river is of the second order, therefore kilometre 0 is in the mouth of the river and grows in the opposite direction to the river. It results from the hydrographic division of Poland.

Referee: Line 179. Sentence ends with "a Normal Pool Capacity of 1 MM cm³ "; I expect this should be 1 hm³ .

Reply: We agree with the Referee. We propose to correct the units in this sentence.

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

Reply: We agree with the Referee. We suggest moving the sentence starting on line 189 to the end of the sentence from line 75. According to Bačová-Mitková and Onderka (2010), Bezak et al. (2014), Gharib et al. (2017), Langbein (1949), Lang et al. (1999), Kundzewicz et al. (2005), Svensson et al. (2005) the AM method is the most common

C10

because it samples only one extreme event per year. The POT includes all peaks above a certain flow value (the threshold) (Bezak et al., 2014; Gharib et al., 2017; Kundzewicz et al., 2005; Svensson et al., 2005).

Referee: Line 201. "and change point detection" should be "and the presence of change points". Reply: We would like to remove this phrase. We have included this in response to general comments.

Referee: 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" ?

Reply: We also suggest removing this phrase. We have included this in response to general comments.

Referee: Line 247. What is meant here by "verified"?

Reply: We agree with the Referee. We wanted to write that we used the GGEV distribution. We propose to change this word: Moreover, the authors used a four-parameter distribution called Dual Gamma Generalized Extreme Value Distribution (GGEV) described by Nascimento et al. (2016).

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

Reply: We agree with the Referee. We propose to correct the sentence: The Method of Moments is based on the empirical moments such as: mean, variance, skewness and kurtosis of the sample data.

Referee: 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.

C11

Reply: We agree with the Referee. We propose to delete this sentence:

Referee: 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 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.

Reply: We agree with the Referee. We propose to correct the sentence: In the DS Packages developed by Becker and Klößner (2017), the 3P3 distribution allows negative scale parameters to allow for negative skewness.

Referee: Line 302. "The Chi-squared Test (χ^2), 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?

Reply: The authors meant literature. We propose to complete the sentence:

The Kolmogorov-Smirnov (K-S) (Haktanir, 1991), the Chi-squared Test (χ^2) (Haktanir, 1991, Langat et al., 2019, Mamman et al, 2017, Zhang, 2007), and the Mean absolute relative error (MARE) tests were widely used to indicate the adequacy of the distribution functions being tested (Szulczewski and Jakubowski, 2018).

Referee: 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). "

Reply: We agree with the Referee. We propose to correct the sentence: The MK test showed trends neither for the AM values (except for the O sample) nor for POT (except for samples BB and O).

C12

Referee: 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%.

Reply: We agree with the Referee. We propose to correct the sentence: Cassalho et al. (2018) report that based on the MK test, at a significance level of 5%, seven samples from 113 for the Rio Grande do Sul in Brazil presented significant monotonic trend. Thus, these seven series are rejected.

Referee: 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.

Reply: We agree with the Referee. We propose to correct the sentence: Also, The MK test was used by Młyński et al. (2018) to check the significance of the trend. The study was conducted for the significance level of $\alpha = 5\%$. The values received from MK test revealed that the trends of annual peak flow, for the investigated periods, in the catchments of the Grajcarek, Wołosaty and Hoczewka streams (the three from 9 investigated streams from the Upper Vistula River basin) were significant.

Referee: 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

C13

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.

Reply: In subsection 2.2.1. the null and alternative hypotheses were described for each test (MK and B) separately. I used the MK test to check the presence of a trend and the B test to check the randomness of the sample. We propose to merge the paragraphs on lines 202 and 208. The existing paragraphs may have confused the reader into thinking that the hypotheses on line 208 are for all tests in this section. In our opinion, the text should be legible after these changes. However, in the results, the MK and B tests are separated by paragraphs. we would like to remove the SNHT test as it cannot be used for this data. I wanted the tested samples to be verified in terms of trend and randomness, which should raise the level of work.

Referee: 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?

Reply: We agree with the Referee. We propose to remove the sentence from line 328: "This test is based on a rank version proposed Bartels (1982) for testing a series for randomness" and move it to line 217 as follows:

In the Bartels test for randomness (Bartels, 1982), (B) the null hypothesis that the sample is random is tested against the alternative hypothesis that the data is significantly different from random. This test is based on a rank version proposed Bartels (1982) for testing a series for randomness. A two-sided test was performed.

Additionally, we suggest extending the discussion of test B with the sentence on line 327:

Bezak et al. (2014) used von Neumann's ratio test whose to assess the homogeneity of data from the Litija hydrological station on the Sava River. The statistic values received

C14

from this test were compared with critical values. The test showed that the analysed data are homogeneous for the AM series for the periods 1895–2010, 1895–1952 and 1953–2010.

Referee: Line 348. Typo: “He” should be “he”.

Reply: We agree with the Referee. We propose to correct the sentence: Zhang (2007) reports that he studied the GPD distribution using MLE, MM, PWM, likelihood moment estimators (LMEs) estimators.

Referee: 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.

Reply: We agree with the reviewer that p-value is an indication of how likely it is to get a specific test statistic value for a random sample from a given distribution. We will insert this comment in the subsection 2.2.1. at the end of the paragraph about K-S test.

Referee: 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.

Reply: We agree with the Referee. We propose to correct the discussion from line 366: They stated that the goodness-of-fit hypothesis is rejected for two distributions, for the Oder River in the Trestno (LN, p-value = 0.026; GEV, p-value = 0.005) and Korzeńsk profile (P3, p-value = 0.016, LN, p-value = 0.005) and only the mixed distribution (MIX Gamma + GEV) ensures the best fit.

The sentence: "In the case when the value of $p > 0.05$ for the analyzed distribution,

C15

then it showed the lack of the best fit of the empirical distribution with the theoretical distribution" will be deleted.

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

Reply: The parameters of the GGEV distribution are more difficult to estimate than the two-parameter or three-parameter distribution. This is discussed in articles on mixed distributions. Szulczewski and Jakubowski (2018) and Vaidyanathan and Lakshmi (2016) write about it. The number of parameters is important in the chi-square test. Haktanir (2009) states that "the chi-squared value of a three-parameter model can be less than that of a two-parameter model, the probability level of acceptance of the former can be worse than the latter. There does not exist such an" effect of the number of model parameters in the KS GOF test however ". Whereas Wilks (2011) states that if "the parameters have been estimated from the data sample, then the estimating the parameters from the same batch of data used to test the goodness of fit results in the fitted distribution parameters being" tuned "to the data sample. In practice this provision can be a limitation to the use of the KS test, since it is often the correspondence between a fitted distribution and the particular batch of data used to fit it that is of interest ". In turn, for continuous distributions the K-S test usually will be more powerful than the χ^2 test and so usually will be preferred (Wilks, 2011).

References: Wilks D.S.: Statistical Methods in the Atmospheric Sciences, 3rd Edition, Academic Press, in International Geophysics, San Diego, Calif, 2011.

Referee: Line 379. chi square symbol is not displayed correctly.

Reply: We agree with the Referee.

Referee: 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

C16

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.

Reply: We agree with the Referee. We will correct this sentence.

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

Reply: For the MK, B, SNHT tests (the latter we would like to remove from the table) we compared the calculated statistic with the critical value. We made a note of it. If H0 was not rejected, we showed the p-value. In table 2 we show the p-value results.

Referee: 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.

Reply: We agree with the Referee. We propose to leave the title of Table 3 because, as suggested by Denis et al., (2018) three complementary methods are available for comparing models: p-value, by the difference between the theoretical and empirical survival functions, by the likelihood value.

References: Laurent Denis, ... David Delaux, in Reliability of High-Power Mechatronic Systems 2, 2017

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