Referee Report of
« A stochastic event-based approach for flood estimation in catchments with mixed rainfall/snowmelt flood regimes»
(NHESS-2018-174, version 2)

General comments:

This paper presents a significant enhancement of a Norwegian method for the estimation of extreme floods, based on an event-based rainfall-runoff simulation. It introduces a stochastic process for the assignment of the initial hydrological conditions before the simulated events, as well as for the intensity and the temporal dynamic of the simulated precipitation events. This method is compared to the initial method (which considers only a reference precipitation on given condition), and to a classical FFA.

The presented method is interesting, both in terms of methodology and statistical results. It is well explored, with a detailed sensitivity analysis.

Most of the issues raised during the review of the initial manuscript have been addressed by the authors, with new and/or complementary information provided, which now allows to better understand the method and the sensitivity analysis.

However, in some sections the writing remains perfectible, and some figures would deserve to be improved to ease their reading and their interpretation.

I would then recommend a minor revision of this paper, based on the detailed comments/suggestions provided to the authors in that follows.

Detailed comments/questions:

Quoted sentences are written in italic.

Page 4, line 4: *Australian Rainfall and Runoff 2016 (?)*
Reference is missing.

Page 4, lines 4-5: *this method has not yet been established in Norway*
Should be "this kind of methods has not yet…"

Page 4, line 24 to page 5, line 7:
This summary is not easy to read. A block diagram (or an enhanced version of the Figure 1) would better illustrate the process, especially the seasonal and stochastic loops. The "calibration" steps (1 to 4) should be better distinguished from the "simulation" steps.
In step 10, I think that the steps 5 to 9 (instead of step 6 to 9) are repeated 100 000 times.

Page 5, line 24:
Reference/link to NEVINA tool should be put into the "Reference" section.

Page 8, line 14: *we have preferred to use the GP due to the inclusion of the shape parameter*
Two of the seasonal GP distribution for Hørte (SON and JJA) show a bounded asymptotic behaviour. According to the authors, is this an "acceptable" hypothesis for extreme flood estimation? What about coercing the shape parameter to be positive or null only?

Page 9, equation 2:
This equation should rewritten properly: x is a vector of parameters (and consequently to be written with uppercase), with 3 components for G1, and 2 for G2.

Page 9, lines 14-19:
The solution based on copulas could be evoked at the beginning of the section (line 2?) as it introduces the use of the multivariate normal distribution.

Page9, line 29: *was calibrated for the 45 highest flood events*
Is it a fixed number for each catchment, whatever may be the length of discharge record?

Page 11, line 6: *More detailed information on the set up is given in Table 3*
This detailed information would better fit into the section §2.7. See specific comments on Table 3.

Page 11, line 11: *A high sensitivity to the shape of the hyetograph was also found by Alfieri et al. (2008).*
More details about this reference (context, what kind of catchment, connection to the present work) would be welcome to increase its relevance.

Page 11, line 15: *A high sensitivity to the parameters of the rainfall model was also described by Svensson et al. (2013)*
Same remark as above.

Page 11, line 18: *using a higher quantile for the threshold leads to selecting fewer events, leading to a higher degree of uncertainty in the GP fit.*
It's the same for all the catchments (higher quantile -> less selected events). This doesn't explain why Øvrevatn is more sensible to this.

Page 11, line 20 to 22:
Krinsvatn seems more sensible to the parameters of the PQRUT model. Any comment on that?

Page 11, line 26 to 28: *A reason for this […] and for Horte, it is 16.7mm.*
In this analysis, I guess that when *soil moisture conditions* are evoked, it is in fact the *soil moisture deficit*. Please check the writing, and also the captions of Table 2.
I don't clearly understand what is explained in these sentences. A possibly better explanation for the sensitivity of Øvrevatn to initial soil conditions can be provided by considering the ratio between the maximum soil moisture deficit and the mean precipitation, both provided in the Table 2: for Krinsvatn, Hørte and Øvrevatn, these ratios are 33%, 26% and 65% respectively, meaning that for Øvrevatn, the maximum potential interception of precipitation is relatively two to three times higher than for the other catchments.

Page 11, line 34: *and also a step change in the frequency curve*
I don't see to what this "step change" refers in the frequency curves of the Figure 8.

Page 12, line 33: *the duration of the storm event was assumed to be the same as that used for the stochastic PQRUT model*
I suppose it is the critical duration? If yes, it is worth mentioning it.

Page 13, Equation 5:
The quantile score should be noted QS, as it is in the text above the equation.
The index I, going from 1 to the number of years of record should be mentioned under/above the Σ.

Page 13, line 29: *the results of test 2 indicate that both the GEV distribution and the stochastic PQRUT provide similar fits to observed quantiles*
This not obvious considering the Figure 11, where the stochastic PQRUT show more variant but also higher QS scores than GEV. A simple boxplot of the QS scores would be more informative than the barplot of the Figure 11.

Page 15, line 2: *Due to the considerable seasonal variation in the initial conditions…*
And also (and mostly) in the extreme rainfall distribution.

Pages 17 to 19:
Some references are incomplete (journal name is missing): Blanchet, Gräler, Nathan, Wilson.

Figure 1:
As suggested in the comments above, a simple, vertical block diagram would better illustrate the process, especially the seasonal and stochastic loops. The "calibration" steps (1 to 4) should be better distinguished from the "simulation" steps.

Figure 2:
The “zoom maps” for the three catchments will likely be unreadable in the final document if the size of the figure is reduced. Furthermore, I am not sure they provide an interesting information.

Figure 4:
I assume the duration of these precipitations is the critical duration of each catchment. If yes, it deserves to be mentioned, whether in the axis name or in the caption of the figure.
The catchments should be ordered in the same way they are in the rest of the manuscript, i.e. Hørte, Krinsvatn and Øvrevatn.

Figure 5:
Units are missing.

Figure 7:
The “peak discharge” mention should appear whether in the axis name or in the caption of the figure.

Figure 8:
In order to “spare space”, the name of the catchments could only be mentioned once in the top of each “column”.
Bolder lines could be used, colours are difficult to distinguish on very fine lines.
Legend which are specific to each “setup” should appear within the corresponding box. To this end, a “portrait” setup could be chosen for this page.

Figure 9:
I am unable to interpret the left plot. I am not sure that violin plots used in this context are useful. Ten different stochastic hyetographs could be presented instead of it, to illustrate their potential diversity.

Figure 10:
Once again, this plot looks odd to me. Why not presenting a simple matrix [catchment x models] with check crosses where the Q100 falls within the confidence interval of the GP?
The reference to the three example catchments should be written in bold, and identified in the caption.

Figure 11:
As mentioned above, simple boxplot of the QS scores for both methods would be more informative, and also a scatter plot QS(flood frequency) v/s QS(Stoch. PQRUT). The three example catchments should also be outlined.

Table 1:
In the caption of the columns, the units could be written below the caption for a better readability.

Table 2:
Column and row caption should be written in bold characters.
The width of the columns could be uniformized.
I am not sure that it is worth providing both mean and median values. I would chose the median.

The P100 fitted by GPD (for the season at risk) would be worth to be added in the table.
The critical duration should also be mentioned, at least in the Pmean caption (i.e. Pmean 48h).
“Soil moisture” should be replaced by “soil moisture deficit”.

Table 3:
Column and row caption should be written in bold characters.
The different setups presented in the table should be organized in the same order than they are in the text and in the Figure 8.

Table 4:
As for Table 3, the different setups presented in the table should be organized in the same order than they are in the text and in the Figure 8, and “blocks” of setups should be named a) b) c) etc. as they are in Figure 8.
The setups should be described in the text, not summarized in the captions here.
The percentage values could be rounded to the next integer.
The 10/100/1000 quantiles could be separated more clearly (bold line) in the Table.