Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174-RC1, 2018 © Author(s) 2018. This work is distributed under the Creative Commons Attribution 4.0 License.





Interactive comment

Interactive comment on "A stochastic event-based approach for flood estimation in catchments with mixed rainfall/snowmelt flood regimes" by Valeriya Filipova et al.

Anonymous Referee #1

Received and published: 9 July 2018

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.





However, the paper could be greatly improved by a better writing and more illustrations, particularly about the stochastic PQRUT which deserves a detailed step by step explanation of the simulation procedure (text and diagram), and also the probabilistic models for precipitation.

Regarding the sensitivity analysis, which is very important to understand the key factors and options of this new method, its writing also should be better organized and illustrated. It lacks a basic but important study of the impact of the random drawing (e.g. by performing 100 different simulations) and of the number of the simulated events on the extreme quantiles estimation. The later seems to be an issue here for high return periods.

I would recommend a significant revision of this paper, mostly to improve its structure, its writing and the illustrations provided. Detailed comments/suggestions are provided to the authors in that follows.

NHESSD

Interactive comment

Printer-friendly version



Detailed comments/questions:

Quoted sentences are written in italic.

Abstract

For those not familiar with PQRUT, it could be added in the abstract that the stochastic PQRUT is an extension/evolution of the "standard" PQRUT routine, applied since many years in Norway (dates and references to be provided).

The differences between the estimates can be up to 200% for some catchments, which highlights the uncertainty in these methods

This is not a good message for hydrological engineering, a less pessimistic phrasing could be " $[\ldots]$ 200% for some catchments where the uncertainties of the compared methods are high and combine unfavourably".

§1 - Introduction

Page 1, line 14: For example, floods with a 500-year return period are sometimes used to $[\dots]$

As most of the estimates evoked in the paper are 100 or 1000-yr. floods, and example of the use of such quantiles in Norway could useful.

Page 1, line 17: Flood mapping also usually requires input hydrographs

This is also the case for dam safety assessment.

Page 1, line 21: When longer return periods are needed

I guess the author means "longer than the record length", i.e. return period of 100 yr. or above.

Page 2, line 6: have been shown to produce average errors between 27 and 70%

Please mention on what estimation this error is computed (observed quantiles or esti-

NHESSD

Interactive comment

Printer-friendly version



mated ones, of which return period).

Page 2, line 32: they are computationally inefficient...

Another writing could be " [...] they are computationally demanding, as long continuous periods have to be simulated to estimate extreme quantiles".

Page 3, line 6: millions of rainfall events can be sampled from the MEWP model...

More exactly, millions of synthetic rainfall events can be generated, assigned to a probability estimated from the MEWP model, and inserted...

Page 3, line 21: it requires the generation of a temperature sequence for the event

I would add " a temperature sequence for the event, coherent with the simulated rainfall, and a snow water..."

Page 3, line 22: The assumption of a fixed rate of snowmelt [...] and a joint probability model needs to be considered

Does it mean that a fixed snowmelt is usually added without consideration to the rest of the variables which will characterize the simulated event? What kind of joint probability model should be added?

Page 24, line 24: SEFM which has been applied in several USGS studies

To my knowledge, I am not sure it is USGS (although SEFM is evoked in the USGS Bulletin 17C "Guidelines for Determining Flood Flow Frequency" of 2018), but several application of SEFM for dam safety studies have been delivered to USBR (US Bureau of Reclamation).

Page 3, line 34: due to the large uncertainty in both the event-based model and the statistical flood frequency analysis

I am not comfortable with this writing. With two identical methods (say a classical FFA), but with two distributions fitted on two different samples, estimations would be different,

Interactive comment

Printer-friendly version



and in that case this difference is completely linked to the uncertainties of the FFA (and mainly the sample uncertainty). But with different methods, these differences can also be produced by discrepancies between methods which should be treated *per se*, in order to assess the method themselves. So the interpretation of the difference should, in my opinion, not only rely on uncertainties.

Page 3, line 35: To better understand the differences between these methods, a sensitivity analysis of the stochastic PQRUT is performed

Here I have somehow the opposite comment from above: this sensitivity analysis of stochastic PQRUT is more about dealing with the uncertainties of stochastic PQRUT, not the differences between methods.

§2 - Stochastic event-based model

Page 4, line 18: The study area consists of a set of 20 catchments

A more logical phrasing could be "The study area in Norway, with a dataset of 20 catchments located throughout the whole country"

Page 5, line 6: for Krinsvatn, Lk and the area covered by marsh, M, is more than 10%

In the Table 1, Lk and M values are 9 and 1.1 %, respectively.

Page 5, line 28: the correlation of the method was found to be higher

To which values the results of this disaggregation method have been correlated? Hourly or 3-hourly rainfall observations?

Page 5, line 29: *simply dividing the seNorge data into eight equal parts*

To be clearer, I suggest "simply dividing the seNorge daily data into eight equal 3-hourly values".

Page 5, line 29: *disaggregated to a 1-hour time step using a uniform distribution to match the time resolution of the discharge data, although a 3-hour time step could also*

NHESSD

Interactive comment

Printer-friendly version



be used

I don't fully understand this. Was the 3-hour value affected randomnly to one hour or divided into three? Why is it possible to use the 3-hour value with an hourly model?

Page 6, line 1: remotely sensed data

Assimilating remotely sensed data in a hydrological model is not an easy task, especially soil moisture. Any reference to provide that would apply to the context of this paper?

Page 6, line 2: the DDD hydrological model was used

Please provide a reference which introduces this model.

Page 6, line 7: exceeds 0.3 of its (dynamic) capacity

Please define what is a dynamic capacity here.

Page 6, line 11: *the so-called critical duration*

A reference can be provided here to define this concept:

Meynink, W. J., & Cordery, I. (1976). Critical duration of rainfall for flood estimation. Water Resources Research, 12(6), 1209-1214.

Page 6, lines 12-14: In order to determine [...] had to be determined for each catchment

I don't find this sentence useful, considering what is written before and after it.

Page 6, lines 14: flood events over a certain quantile threshold (0.9) were extracted.

On what data this POT sample was extracted: daily or hourly discharges?

Page 6, lines 18-25:

An alternative to this could be to study the correlation between the peak daily value and

NHESSD

Interactive comment

Printer-friendly version



the precipitation of that day (which we could call P0), the sum P0 to P-1, P0 to P-2 and so on... When the correlation coefficient stops to increase significantly, it means that the correct length of the "precipitation window" is reached, thus the critical duration is estimated. This is likely to be more robust than studying the correlation between the peak daily discharge and the individual precipitations the days before.

Page 6, lines 22-24: In some catchments (mostly those having snowmelt flood regime), no significant correlation was found between discharge and precipitation

In that case, some processing of the flood is needed, e.g. only considering the "snowfree" seasons, or adding a threshold on the precipitation over the preceding days in the POT selection of floods. This could prevent using an arbitrary duration.

Page 6, line 28: the sequence of the input data must be prescribed for the stochastic simulation

What means "prescribed"? Is it generated? Is it randomly drawn from the observed sequences?

Page 7, line 1: a Generalized Pareto distribution was fitted to the series of selected events

A figure with the corresponding fits and observations for the example catchments would be welcome.

Page 7, line 6: *introduced in section ??*

A paragraph number is lacking.

Page 7, line 7: Using the fitted Generalized Pareto (GP) distribution, precipitation depths were simulated

Does it mean that probabilities where randomly drawn then the corresponding precipitation depths deduced from the fitted GPD? How many events are drawn?

NHESSD

Interactive comment

Printer-friendly version



Page 7, line 8: a storm hyetograph was first sampled

How is it sampled? I guess it comes from the hyetographs collection corresponding to the POT selection of precipitation events, but with what consideration to season, intensity, etc.?

Page 7, line 10:

Pi and P are not defined in the disaggregation formula.

Page 7, line 15: Output from DDD model runs

Have the DDD models been calibrated on local data? If yes, some words about the calibration method are welcome. I guess the DDD models are used at daily time-step, is it true?

Page 7, lines 21-25:

The writing is not clear, and neither is the equation of the mixed distribution (what is x?). As far as I understood, it is about randomly switching between a trivariate (discharge, moisture, snow) and a bivariate (discharge, moisture) distribution depending on the probability of having snow on a given season. Is *p* also drawn for the simulation?

Page 7, line 26: The correlation between the observed and simulated variables is shown in Figure 4

Apparently, *sl* is the soil moisture deficit. Contrary to SWE and Qobs which are "observable" variables, *sl* is linked to a model (here DDD). So it should be introduced, in relation with the DDD model structure.

Page 8, line 5: for estimating design floods and safety check floods for dams in Norway

This type of application is perhaps documented in (Andersen, 1983), but this reference is not easily accessible on line, and is written in Norwegian, so a accessible reference documenting this type of application would be welcome.



Interactive comment

Printer-friendly version



Page 8, line 14: The general procedures used for the PQRUT calibration are described in Filipova et al.

Some details about this calibration would be welcome, e.g. which flood events sample is considered (is it the same as the one used in §2.2 for critical duration)?

Page 8, lines 15-17: This additional parameter *lp* should be documented in the structure of the PQRUT model presented in the Figure 5. Furthermore, I am not sure that it can be considered as a parameter, more likely it is an internal state variable which vary from event to event.

Page 8, line 18: the value of this parameter was set to the initial soil moisture deficit, estimated using DDD

This is an important assumption: it means that some internal variables of DDD (which ones, this is not documented) are used to estimate another one in PQRUT. This is far from obvious to accept for two very different models, running at different time steps: what has be done to check this "compatibility"?

Page 8, line 23: *Cs is a coefficient accounting for the relation between temperature and snowmelt*

Properties

It is usually called a "degree-day" coefficient (although used at a hourly time step here).

Page 8, line 30: The term under the bar should be "power to k" not be multiplied by k.

Page 9, line3: These simulated events were compared with the POT flood events extracted from the observations

At this point, I don't clearly understand the simulation process. Some lines detailing the simulation process (sequence of random drawings, number of simulation, processing of events, etc.), as well as a diagram, are really necessary to the reader before entering into the analysis of the simulations.

NHESSD

Interactive comment

Printer-friendly version



The simulated CDFs look affected by under-sampling above the 500 yr. return period (i.e. not enough simulations of this range), which interrogates the robustness of the 1000 yr. estimations which are assessed in the paper.

Page 9, line 7: large variation in precipitation values

Which duration is considered here? Daily?

The comparison to the 100 yr. precipitation depths estimated thanks to the GP fit evoked in §2.3 would be useful.

Page 9, line 14: even though fully saturated conditions are used in the event-based PQRUT model

I don't understand this: the *lp* variable (variable initial loss) has been introduced in §2.5 to depart from this fully saturated hypothesis.

Page 9, line 16: A sensitivity analysis was performed for the three test catchments

Once again, the detailed protocol of this analysis deserves to be presented for a better understanding of the results. Some information is given in Table 3 but would deserve to be detailed in the text. A more logical "progression" of the different setups could be: 2, 3 (statistical hypothesis on precipitation), then 4 (temporal disaggregation), then 5,6,7 (simple hydrological assumptions) and finally 1 (PQRUT parameters). This would apply for the Table 3, as well as for the writing of §2.7.

Page 9, lines 24-28:

I am not fully convinced by this explanation based on BFI. The sensitivity to initial loss should be linked to the possible values of initial loss in relation with the high quantiles of precipitation. I would be interested by looking at those values (maximum initial loss and 10, 100 and 1000 yr. precipitation) for the three catchments.

Page 9, line 31: In addition, Krinsvatn shows high sensitivity to snowmelt

NHESSD

Interactive comment

Printer-friendly version



This is in contradiction with Page 9, line 10 (*for Krinsvatn* [...] *in most cases snowmelt does not contribute to the extreme floods*). Any comment?

Page 10, line 7: Ovrevatn and Horte showed sensitivity (28.9%) to the choice of the

statistical distribution for modelling precipitation

A figure showing the precipitation distribution for each catchment (both observed and modelled by GP, and EXP) would be welcome to illustrate lines 7 to 10.

§3- Comparison with standard methods

Page 10, line 27: the standard implementation of the event-based PQRUT method

This is the first mention of such a "standard" implementation. I think this would deserve to be presented at the very beginning of the paper, which proposes a "stochastic PQRUT" being a significant enhancement from the "standard PQRUT". The context of this study would thus be better understood.

Page 10, line 29: the annual maximum series were extracted from the observed daily mean streamflow series

Why not using a GPD with the POT sample of floods extracted for the study of the critical duration?

Page 10, line 31: to obtain instantaneous peak values, the return values were multiplied by empirical ratios, obtained from regression equations

Here I don't understand why the POT flood events extracted from observations (shown in the plots of the Figure 6) has not been used to fit either a GPD, or a GEV after extraction of annual maxima. More comments about this would be welcome.

Page 11, line 11: obtained from growth curves based on the 5-year return period value

If I understand properly, the shape of the design hyetograph is based on the growth curves considered at the 5 yr. return period. Are the ratios between the different

Interactive comment

Printer-friendly version



duration values at this return period deduced from empirical distribution, or inferred from a fitted distribution? Later on, this must be scaled to define a 100 or 1000 yr. hyetograph. What precipitation distribution (duration and model) are these extreme values deduced from?

Page 11, line 15: The performance of the three models was validated by using two different tests

In that case, dealing with 100 or 1000 yr. flood estimations, it's more about "comparing different approaches".

Page 11, line 20: As discussed, due to the difficulty in assigning initial conditions for the event-based

PQRUT model

I don't understand this sentence, and to which discussion it refers.

Page 11, line 22: the regional equations were used

Which regional equations? For PQRUT parameters?

Page 11, line 25: equation of QS + *observed probabilities (Qobsi) are calculated using Gringorten positions for the POT series*

The POT series are used here, contrary to the daily (transposed to peak) annual maximum values that have been fitted in the statistical approach. Another option (already mentioned in my remark for page 10, line 29) could be to fit the statistical method on the POT sample, which would have allowed to keep it as a "benchmark" method, given more sense to the comparison presented (or conversely using the "peak-from-daily" observations for the QS calculation).

Page 11, line 30: the results vary between catchments as shown in fig 8

I don't find this figure very useful, the reader is unable to interpret the coloured dots.

Interactive comment

Printer-friendly version



An alternative, aside the boxplots, could be some scatter plots (statistical Q100 and Q1000 v/s standard and stochastic PQRUT, statistical QS v/s standard and stochastic PQRUT, etc.).

Page 11, line 32: we can conclude that the performance of the standard PQRUT model is poorer than the performance of the statistical flood frequency analysis and the stochastic PQRUT model

The results which ground this conclusion are not explicitly presented. The only clue given to the reader is the Figure 8 which only presents the distribution of QS scores for FFA and stochastic PQRUT. The results, in terms of QS score as well as confidence interval, should be presented in a table and in an adequate figure.

Page 12, line 3: The violin plots (fig. 9)

See remarks on Figure 9 below.

Page 12, line 7: *Reasons for this may be that higher precipitation intensity or snowmelt is used*

To assess this, the values of the reference hyetographs used in standard PQRUT deserve to be presented and compared to the simulated precipitation values of stochastic PQRUT (like the values of the Table 2 for Q100).

Page 12, line 8: the absolute differences between the two methods are larger in catchments with lower temperature (fig. 9)

I wonder how this can be deduced of illustrated by Figure 9, it is more likely somehow in Figure 10.

Page 12, line 17: which might be due to the uncertainty in estimating the parameters for the GEV distribution

I don't understand this interpretation which appears rather quick and subjective to me.

Interactive comment

Printer-friendly version



Page 12, lines 18-21:

This using of the study of Rogger et al. (2012) is off topic for me here, as it is based on Gumbel, whereas the FFA is done here with GEV, which is more flexible.

§4- Conclusions

Page 13, line 10-15:

Another modelling option could be to run the event-based simulation with the DDD model, already used for the initial condition. In that case, an hourly version of DDD should also be calibrated (with local observations or regionally), in compatibility with the daily version used for initial conditions. I am not fully aware of the potential difficulties of this, but it would be a more homogeneous approach in terms of hydrological modelling. Any comment about this?

Page 13, line 28: easily incorporate the uncertainty associated with this choice

This is a very good remark: the stochastic process here adequately models a variable which, when represented in a deterministic way (i.e. fixed initial conditions), appears as highly uncertain.

Page 13, line 31: based on an assessment of the uncertainty characterizing the individual methods

This is an interesting suggestion, but it has to be added that a proper expression of uncertainty for a rather sophisticated method like stochastic PQRUT is far from trivial, and is still to be investigated...

Tables

Table 1

Units are missing, as well as legend of the columns in the caption.

Table 2:

Interactive comment

Printer-friendly version



Units are missing, precipitations could be rounded to the next mm.

For Krinsvatn, the probability to find the Q100 events in one season or the other could be provided.

Figures

Figure 3:

Not very informative, re-scaling storm hyetograph is not something difficult to understand. A set of different "typical" hyetographs could instead be presented for the three catchments, ideally illustrating the potential diversity of storm dynamic.

Figure 4:

"for Krinsvatn catchment" could be added in the caption, as well as the number of observed and simulated events.

Figure 6:

The remarkable return periods (10, 100 and 1000 yr.) should be distinguished in the plots (by a bolder vertical line for example).

Figure 7:

There are too many distributions in the plots, their interpretation is not easy. Two plots could be edited for each catchment, having for example only the "calibrated" simulation in common.

An uncertainty band around the "calibrated" simulation would be useful to assess the intrinsic uncertainty of the simulation process.

Figure 8:

See comment of page 11, line 32.

Figure 9:

Interactive comment

Printer-friendly version



I am not convinced by the usefulness of the violin plots here considering the limited number of values per scores (20 catchments). Box plots with outliers would have been sufficient and more readable.

Captions of the methods sometimes overlap.

NHESSD

Interactive comment

Printer-friendly version

