Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

A stochastic event-based approach for flood estimation in catchments with mixed rainfall/snowmelt flood regimes

Valeriya Filipova[1], Deborah Lawrence[2], and Thomas Skaugen[2]

[1]University of Southeast Norway,INHM, Gullbringvegen 36, 3800 Bø, Norway, e-mail: valeriya.filipova@usn.no
[2]Norwegian Water Resources and Energy Directorate,P.O. Box 5091 Maj., N-0301 Oslo, Norway, e-mail: dela@nve.no

Correspondence: Valeriya Filipova (valeriya.filipova@usn.no)

Abstract. The estimation of extreme floods is associated with high uncertainty, in part due to the limited length of streamflow records. Traditionally, flood frequency analysis or event-based model using a single design storm have been applied. We propose here an alternative, stochastic event-based modelling approach. The stochastic PQRUT method involves Monte Carlo procedure to simulate different combinations of initial conditions, rainfall and snowmelt, from which a distribution of flood peaks can be constructed. The stochastic PQRUT was applied for 20 small and medium-sized catchments in Norway and the results show good fit to the observations. A sensitivity analysis of the method indicates that the soil saturation level is less important than the rainfall input and the parameters of the PQRUT model for flood peaks with return periods higher than 100 years, and that excluding the snow routine can change the seasonality of the flood peaks. Estimates for the 100- and 1000-year return level based on the stochastic PQRUT model are compared with results for a) statistical frequency analysis, and b) a standard implementation of the event-based PQRUT method. The differences between the estimates can be up to 200% for some catchments, which highlights the uncertainty in these methods.

1 Introduction

The estimation of low-probability floods is required for the design of high-risk structures such as dams, bridges, levees, etc. For example, floods with a 500-year return period are sometimes used to evaluate the risk of scour to bridges (Ries, 2007), and the design and safety evaluation of high-risk dams requires that, in some cases, floods with magnitudes of up to the Probable Maximum Flood (PMF) are estimated. An overview of design flood standards for reservoir engineering in different countries is provided in Ren et al. (2017). Flood mapping also usually requires input hydrographs for flood events with return periods of up to 1000 years. Methods for estimating these floods can be generally classified into three groups: 1) statistical flood frequency analysis; 2) the single design event simulation approach; and 3) derived flood frequency simulation methods.

At gauged sites, statistical flood frequency analysis involves fitting a distribution function to the annual maxima or peak over threshold flood events and calculating the quantile of interest. When longer return periods are needed, the process requires extrapolation of the fitted statistical distribution, which introduces a high degree of uncertainty due to the number of limited observations relative to the estimated quantile (e.g. Katz et al., 2002). Significant progress has been made in methods for reducing this uncertainty by incorporating historic or paleo-floods data (Parkes and Demeritt, 2016), where available. Another

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

way to "extend" the hydrological record in order to reduce the uncertainty is to combine data series from several different gauges by identifying pooling groups or hydrologically similar regions, where this is possible. It has been found, however, that the identification of such hydrological regions can be difficult in practice (Nyeko-Ogiramoi et al., 2012). The application of statistical flood frequency analysis in ungauged basins is also problematic. As the physical processes in the catchments are not directly considered in the analysis, estimating the flood quantiles in ungauged basins using regression or geostatistical methods have been shown to produce average errors between 27 and 70% (Salinas et al., 2013) or even higher. In addition, the complete hydrograph is often needed in practice. Although multivariate analysis of flood events (e.g. flood peaks, volumes and durations) can be used to generate hydrographs for specific return periods, the methods are not easily applied (Gräler et al., 2013).

The second method for extreme flood estimation is the design event approach in which single realizations of initial conditions and precipitation are used as input in an event-based hydrological model. Another feature of the approach is that when event-based models are used, a critical duration defined as the storm that results in the highest peak flow, needs to be set.Advantages of this method over statistical flood frequency analysis is that rainfall records are often widely available (e.g in the form of gridded datasets) and that the event hydrograph is generated in addition to the flood peak magnitude. This approach has been traditionally used due to its simplicity (e.g. Kjeldsen, 2007; Wilson et al., 2011). However, its application often involves the assumption that the simulated flood event has the same return period as the rainfall used as input in the hydrological model. This assumption is not realistic and, depending on the initial conditions, the return period of the rainfall and the corresponding runoff can differ by orders of magnitude (e.g. Salazar et al., 2017). A reason for this is that flood events are often caused by a combination of a high level of saturation, rainfall and snowmelt, and a joint probability distribution needs to be considered if one is to fully describe the relationship between the return period of rainfall and of runoff.

The third possible approach is the derived flood frequency method in which the distribution function of peak flows is derived from the distribution of other random variables such as rainfall depth and duration, and different soil moisture states. Although a statistical distribution or plotting positions are then used to calculate the required quantiles, as in conventional flood frequency analysis, a hydrological model can be used to derive discharge values and thus extend and complement the observed discharge record. The derived distribution can be solved analytically by using a simple rainfall-runoff model (e.g. a unit hydrograph), assuming independence between rainfall intensity and duration, and considering only a few initial soil moisture states. However, because of these simplifying assumptions, the method can produce poor results (Loukas, 2002). For this reason, methods based on simulation techniques are most often used, and these range from continuous simulations to event-based simulations with Monte Carlo methods.

In the continuous simulation approach, a stochastic weather generator is used to simulate long synthetic series of rainfall and temperature, which serve as input in a continuous rainfall-runoff model. The resulting long series of simulated discharge are then used to estimate the required return periods, usually using plotting positions (e.g. Calver and Lamb, 1995; Camici et al., 2011; Haberlandt and Radtke, 2014). A disadvantage of these methods is that they are computationally inefficient in that long periods between extreme events are also simulated. Several newer methods, therefore, use a continuous weather generator coupled with an event-based hydrological model. For example, the hybrid-CE (causative event) method uses a continuous rainfall –runoff simulation to determine the inputs to an event-based model (Li et al., 2014). Another disadvantage of continuous simulation models is that stochastic weather generators require the estimation of a large number of parameters (e.g. Onof et al., 2000; Beven, Keith & Hall, 2014). In addition, models such as the modified Barlett-Lewis rectangular pulse model have limited capacity to simulate extreme rainfall depths, which can lead to an underestimation of runoff (Kim et al., 2017). In order to avoid the limitations of the continuous weather generators, the semi-continuous method SCHADEX (Paquet et al., 2013) uses a probabilistic model for centered rainfall events (MEWP;Garavaglia et al. (2010)), identified as over threshold values that are larger than the adjacent rainfall values. Using this approach, millions of rainfall events can be sampled from the MEWP model and inserted directly into the historic precipitation series to replace observed rainfall events. In this manner, the SCHADEX method is similar to the hybrid-CE methods because a continuous hydrological model is used to characterize observed hydrological conditions, and synthetic events are only inserted into the precipitation record for periods selected from the observed record. Despite the many advantages of the hybrid-CE and SCHADEX methods over continuous simulation methods, they still require sufficient data for the calibration of the hydrological model, modelling of the extreme precipitation distribution and for ensuring that an exhaustive range of initial hydrological conditions are sampled during the simulations.

Another method for derived flood frequency analysis is the joint probability approach (e.g. Muzik, 1993; Loukas, 2002; Svensson et al., 2013; Rahman et al., 2002). In this approach, Monte Carlo simulation is used to generate a large set of initial conditions and meteorological variables, which serve as input to an event-based hydrological model. This approach requires that the important variables are first identified and any correlations between the variables are quantified. Most often, the random variables that are considered are related to properties of the rainfall (intensity, duration, frequency) and to the soil moisture deficit. Some of these methods, such as the Stochastic Event Flood Model (SEFM, (Schaefer and Barker, 2002)), similarly to SCHADEX, require the use of a simulation based on a historical period to generate data series of state variables from which the random variables are sampled. Although the contribution of snowmelt can be important in some areas, it is rarely incorporated as it requires the generation of a temperature sequence for the event and a snow water equivalent as an initial condition. The assumption of a fixed rate of snowmelt, as is often used in the single event-based design method, can introduce a bias in the estimates and a joint probability model needs to be considered to obtain a probability neutral value (Nathan and Bowles, 1997). One of the few methods that incorporates snowmelt is the SEFM which has been applied in several USGS studies and uses the semi-distributed HEC-1 hydrological model (Schaefer and Barker, 2002). Considering that, most often, simple event-based hydrological models are used (e.g. unit hydrograph), the joint probability approach is particularly advantageous in ungauged catchments or data-poor catchments, where the use of parsimonious models is preferred.

The purpose of this study is to develop a derived flood frequency method using a stochastic event-based approach to estimate design floods, including those with a significant contribution from snowmelt. In this way, the results for any return period can be derived, taking into account the probability of a range of possible initial conditions. The results are then compared with results from an event-based modelling method based on a single design precipitation sequence and assumed initial conditions and with statistical flood frequency analysis of the observed annual maximum series for a set of catchments in Norway. The methods give different results in many of the catchments due to the large uncertainty in both the event-based model and the statistical flood frequency analysis. To better understand the differences between these methods, a sensitivity analysis of the stochastic PQRUT is performed by considering the effect of the initial conditions, model parameters and rainfall intensity on the flood frequency curve.

2 Stochastic event-based flood model

The stochastic event-based model proposed here involves the generation of several hydrometeorological variables: precipitation depth and sequence, temperature during the precipitation event, antecedent discharge, soil moisture conditions and antecedent snow water equivalent. A simple 3-parameter flood model PQRUT (Andersen et al., 1983) is used to simulate the streamflow hydrograph for a set of randomly generated conditions based on the hydrometeorological variables and a flood frequency curve is constructed from all of the simulations using plotting positions. As the method requires initial values for soil moisture and snow water equivalent, i.e variables which cannot be sampled directly from climatological data and which depend on the sequence of precipitation and temperature over longer periods, the Distance Distribution Dynamics (DDD) hydrological model (Skaugen and Onof, 2014) was calibrated and run for a historical period to produce a distribution of possible values for testing the approach. The study area and data requirements are described in section 2.1, while section 2.2 describes the method for determining the critical duration, and section 2.3 and 2.4 describe the generation of antecedent conditions and meteorological data series. The hydrological model is presented in section 2.5 and the method for constructing the flood frequency curve is outlined in section 2.6.

2.1 Study Area and Data requirements

2.1.1 Catchment selection and available streamflow data

The study area consists of a set of 20 catchments located throughout Norway (fig 1). All catchments have at least 10 years of hourly discharge data, and in all cases the length of the daily flow record is considerably longer than 10 years. All selected catchments are members of the Norwegian Bench Mark dataset (Fleig, 2013), which ensures that the data series are unaffected by significant streamflow regulation and have discharge data of sufficiently high quality suitable for the analyses of flood statistics. The catchment size was restricted to small and medium-sized catchments (maximum area is 854 km^2), as the structure of the 3-parameter PQRUT model does not take into account all of the storage processes within the catchment which possibly contribute to delaying runoff during storm events. Previous applications of PQRUT in Norway indicate that this shortcoming is most problematic for larger catchments. Discharge datasets with both daily and hourly time steps were obtained from the national archive of streamflow data held by NVE (https://www.nve.no/). In order to illustrate the application of the method, we have selected three catchments which can be considered representative for different flood regimes in Norway: Krinsvatn in western Norway, Øvrevatn in northern Norway and Hørte in southern Norway (fig 1).

Table 1 summarises the climatological and geomorphological properties of these three catchments, including: area, mean annual runoff (Q in mm/year), mean annual precipitation (P in mm/year), mean elevation ($Hm50$), percent of forest-covered area (For), percent of marsh-covered area (M), percent area with sparse vegetation above three line (B), 'effective' lake percent

(*Lk*) and mean annual temperature (*Temp*). The effective lake percent (*Lk*) is used to describe the ability of water bodies to attenuate peak flows such that lake areas which are closer to the catchment outlet have a higher weight than those near the catchment divide. It is calculated as $\frac{\sum A_i \times a_i}{A^2} \times 100$, where *ai* is the area of lake i, *A*i is the catchment area upstream of lake i and *A* is the total catchment area. The dominant land cover for Krinsvatn and Øvrevatn is sparse vegetation over tree line *B*, while the land cover for Hørte is mainly forest *For*. The effective lake percent *Lk* is insignificant for Hørte and Øvrevatn but for Krinsvatn, *Lk* and the area covered by marsh, *M*, is more than 10%. The catchment Krinsvatn, being located near the western coast of Norway, has much higher mean annual precipitation (*P*), i.e. an average of 2354 mm/year, compared to Hørte (1261mm/year) and Øvrevatn (1558 mm/year). The dominant flood regime for Krinsvatn is primarily rainfall-driven high flows, as the catchment is located in a coastal area and is characterised by high precipitation values and an average annual temperature of around 4^0 C. The highest observed floods, however, also have a contribution from snowmelt. The season of the AMAX (annual maxima flood) is the winter period, i.e. December – February, although high flows can occur throughout the year. Hørte has a mixed flood regime with most of the AMAX flood events in the period September–November, but in some years annual flood events occur in the period March–May and are associated with rainfall events during the snowmelt season. Øvrevatn has a predominantly snowmelt flood regime with most AMAX flood events occurring in the period June – August, due to the lower temperatures in the region such that precipitation falls as snow during much of the year. The catchments were delineated and their geomorphological properties were extracted using the NEVINA tool:http://nevina.nve.no, except for *Q*, which was calculated using the available streamflow data and *P*, which was calculated using available gridded data (further details are given in 2.1.2 below).

2.1.2 Available meteorological data

Data for temperature and precipitation with daily time resolution were obtained from seNorge.no. This dataset is derived by interpolating station data on a 1 km^2 grid and is corrected for wind losses and elevation (Mohr, 2008). In addition, meteorological data with a sub-daily time step is needed for calibrating the PQRUT model, as many of the catchments have fast response times. For this, precipitation and temperature data with a three-hour resolution, representing a disaggregation of the 24-hour gridded seNorge.no data using the HIRLAM hindcast series (Vormoor and Skaugen, 2013), were used. The HIRLAM atmospheric model for northern Europe has 0.1 degree resolution (around 10 km^2) and was first downscaled to match the spatial resolution of the seNorge data (Vormoor and Skaugen, 2013). The precipitation of the HIRLAM data was rescaled to match the 24-hour seNorge data, and these rescaled values were used to disaggregate the seNorge data to a 3-hour time resolution. The method was validated against observations, and the correlation of the method was found to be higher than that obtained by simply dividing the seNorge data into eight equal parts. These datasets were further 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 be used.

2.1.3 Initial conditions

The stochastic PQRUT method requires time series of soil moisture deficit, *SWE* and initial discharge. These data series are used to generate initial conditions, which serve as input in the event-based PQRUT model. Sources of these data can be e.g.

remotely sensed data or gridded hydrological models, which can be used for modelling in ungauged basins. In this study, the DDD hydrological model was used to simulate these data series. The DDD model is a conceptual model that includes snow, soil moisture and runoff response routines and is calibrated for individual catchments using a parsimonious set of model parameters. The snowmelt routine of DDD model uses a temperature-index method and accounts for snow storage and melting for each of 10 equal area elevation zones. The soil moisture routine is based on one dynamic storage reservoir, in which we find both the saturated- and the unsaturated zone, having capacities which vary in time. The flow percolates to the saturated zone if the water content in the unsaturated zone exceeds 0.3 of its (dynamic) capacity. The response routine includes routing of the water in the saturated zone using a convolution of unit hydrographs which are based on the distribution of distances to the nearest river channel within the catchment.

2.2 Critical Duration

When simulating flood response with an event-based model, it is important to specify the so-called critical duration to ensure that the complete flood hydrograph is modelled. In order to determine the length of the precipitation input series producing the most extreme flows, a critical duration for storm events, defined as the duration that results in the highest observed peak value, had to be determined for each catchment. To determine the critical duration, flood events over a certain quantile threshold (0.9) were extracted. The POT (peak over threshold) flood events were considered to be independent if they were separated by at least seven days of lower values than the threshold. The day with the maximum value (peak) of the streamflow was then identified for each event. The peak values were tested for correlations with the precipitation on the day of the peak flow and on days -1, -2 and -3 before the peak. The critical duration was determined as the number of days in which the correlation between the precipitation and the streamflow was higher than 0.25. This threshold value was selected because it gave realistic durations for the catchments in the study area. At first, the critical duration was set to equal the number of days for which the correlation was significant at p=0.01, which however resulted in very long durations in some cases. A possible reason is that if there are only a few observations, even relatively low Pearson correlation coefficients can produce statistically significant p-values. In some catchments (mostly those having snowmelt flood regime), no significant correlation was found between discharge and precipitation, and in this case the critical duration was fixed to 24 hours. If the critical duration was more than one day, the precipitation was aggregated to the critical duration by applying a moving window to the data series. For Hørte and Øvrevatn, the critical duration was set to 24 hours and for Krinsvatn to 48 hours (fig 2).

2.3 Precipitation and temperature sequence generation

In addition to the critical duration of the event, the sequence of the input data must be prescribed for the stochastic simulation. Snowmelt can be important in the catchments considered in this study, so both the sequence of precipitation and temperature must be considered. In order to account for seasonality, the meteorological data series were first split into standard seasons: DJF, MAM, JJA and SON. In this way, we ensure that more homogeneous samples are used to fit the statistical distributions. Although the season at risk could have been defined for each catchment individually (e.g. Paquet et al., 2013), the standard season definition was used for all catchments. Precipitation events over a threshold (POT events) were identified in the 24h precipitation data and a Generalized Pareto distribution was fitted to the series of selected events. In order to select a threshold value for event selection, two criteria were used: 1) the threshold must be higher than the 0.93 quantile, and 2) the number of selected events must be between two to three per season. Although other methods for threshold selection exist, such as the use of mean life residual plots, the described method gives adequate results (e.g. Coles, 2001). The selected threshold varied between the 0.93 to 0.99 quantiles, depending on the season and catchment. In addition, storm hyetographs and temperature sequence with a 1-hour time resolution were identified from the disaggregated seNorge data, introduced in section **??**, and extracted for each POT events. Using the fitted Generalized Pareto (GP) distribution, precipitation depths were simulated and the storm hyetographs were used to disaggregate the precipitation values as follows: a storm hyetograph was first sampled (fig 3) and the ratios between the 1-hour and the total precipitation for the event were calculated according to:

$$P_h sim = \frac{P_i}{sum(P)} P_d sim \tag{1}$$

where Phsim is the 1-hour precipitation intensity and Pdsim is the daily intensity. The calculated ratios were then used to rescale the simulated values (fig 3).

2.4 Antecedent snow water equivalent, streamflow and soil moisture deficit conditions

In order to determine the underlying distribution for various antecedent conditions, the relevant quantities were extracted from simulations using the DDD hydrological model of Skaugen and Onof (2014). Output from DDD model runs were used to extract values for initial streamflow, snow water equivalent (*SWE*) and soil moisture deficit, prior to the seasonal POT events. The POT event series used for this is the same as that used for identifying the critical duration (described in section 2.2).

After extracting the initial conditions, the correlation between the variables was tested for each season for each catchment. As the correlation between the variables is in most cases significant, the variables were jointly simulated using a truncated mul- tivariate normal distribution. In order to achieve a normality for the marginals, the *SWE* and the discharge were log-transformed. In the spring and summer, the *SWE* is often very low or 0 in some catchments. If the proportion of non-zero values,*p*, was greater than 0.3 (around 15 observations), the values were simulated using a mixed distribution as:

$$F(x) = pG_1(x) + (1-p)G_2(x) \tag{2}$$

where G_1 and G_2 represent the multivariate normal distribution with discharge, soil moisture deficit and *SWE* as variables and the bivariate normal distribution for the discharge and soil moisture, respectively. In addition, because the initial conditions are not expected to include extreme values, the values of the initial conditions were truncated to be between the minimum and maximum of the observed ranges. The correlation between the observed and simulated variables is shown in Figure 4 for the Krinsvatn catchment, and although the distribution of simulated values exhibits a very good resemblance to that of the observed, there is not a perfect correspondence between the two. A reason for this may be that the variables (even after log transformation) do not exactly follow a normal distribution. We considered using copulas for the correlation structure of the initial conditions Hao and Singh (2016) however, as the data were limited (around 50 observations per season), these were much more difficult to fit. Similarly, nonparametric methods such as kernel density estimation were not deemed to be feasible

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

due to the limited number of observations. Therefore, the multivariate normal distribution was chosen as the best alternative for modelling the joint dependency between the variables comprising the initial conditions for the stochastic modelling.

2.5 PQRUT model

The PQRUT model was used to simulate the streamflow for the selected storm events. The PQRUT model is a simple, event-based, 3-parameter model (fig 5) which is used, amongst other things, for estimating design floods and safety check floods for dams in Norway. In practical applications, a hypothetical precipitation design sequence of a given return period is routed through the PQRUT model, usually under the assumption of full catchment saturation. For this reason, only the hydrograph response is simulated, and there is no simulation of subsurface and other storage components, such as are found in more complex conceptual hydrological models. Of the three model parameters, K_1 corresponds to the fast hydrograph response of the catchment, and the parameter K_2 is the slower or 'delayed' hydrograph response. The parameter Trt is the threshold above which K_1 becomes active.

The PQRUT model was calibrated for flood events for each catchment by using the DDS (Dynamically Dimensioned Search) optimization (Tolson and Shoemaker, 2007) , and the Kling Gupta efficiency (KGE) criterion (Gupta et al., 2009) was used as the objective function. The general procedures used for the PQRUT calibration are described in Filipova et al. (2016). As described in that work, an additional parameter, lp, was introduced to account for initial losses to the soil zone and is necessary if one is to achieve calibration of the model to actual events (rather than hypothetical events in which the catchment is fully saturated).

For the work presented here, the value of this parameter was set to the initial soil moisture deficit, estimated using DDD. This parameter functions as an initial loss to the system, such that the input precipitation to the reservoir model is 0 until the value of lp is exceeded by the cumulative input rainfall. In order to model flood events involving snowmelt, a simple temperature index snow melting rate was used:

$$S = C_s(T - T_L) \tag{3}$$

where S is the snow melting rate in mm/hour, Cs is a coefficient accounting for the relation between temperature and snowmelt properties and T_L is the temperature threshold for snowmelt (here fixed at 0^0 C) . The model used regional values for the Cs parameters related to the catchment properties, based on the ranges given in Midtømme and Pettersson (2011). In addition, the temperature threshold between rain and snow was set to T_X =0.5 ^{0}C which is typically used in Norway.

2.6 Flood frequency curves

Seasonal and annual flood frequency curves were constructed by extracting the peak discharge for each event and estimating the plotting positions of the points using the Gringorten plotting position formula:

$$P_e = \frac{(m - 0.44)}{(N + 0.12)k} \tag{4}$$

whereP_e is the exceedance probability of the peak, m is the rank of the peak value, N is the number of years, k is the number of events per year. The number of events per year, k, was set to be equal to the average number of extracted POT storm (precipitation) events per year. These simulated events were compared with the POT flood events extracted from the observations (fig 6). After calculating the probability of the simulated events using Eq. 4, the initial conditions and seasonality for a return period of interest can be extracted. For example, the events with return period between 90 and 110 years were extracted (representing around 80 events), and the hydrological conditions for those events were identified (table 2). The results show that there is a large variation in precipitation values and initial conditions that can produce flood events of a given magnitude and this is the reason why it is difficult to assign initial conditions in event-based models. However, it is still useful to extract the distribution of these values in order to ensure that the ranges are reasonable and the catchment processes are properly simulated. For example, the average snowmelt is negative (there is snow accumulation) for Krinsvatn, which means that in most cases snowmelt does not contribute to the extreme floods. This is reasonable as the catchment is located in western Norway, where the climate is warmer (the mean temperature is around 4 ^{0}C) and the mean elevation is low. The average snowmelt contribution for Øvrevatn is much higher as this catchment has a predominantly snowmelt flood regime. The soil moisture deficit for the three catchments is larger than 0, even though fully saturated conditions are used in the event-based PQRUT model. The seasonality of the simulated values is consistent with the seasonality of the observed annual maxima (table 1).

2.7 Sensitivity analysis

A sensitivity analysis was performed for the three test catchments, Hørte, Øvrevatn and Krinsvatn, in order to determine the relative importance of the initial conditions, precipitation and the parameters of PQRUT on the flood frequency curve. As these catchments are located in different regions and exhibit different climatic and geomorphic characteristics, we hypothesize that the flood frequency curve will be sensitive to different parameters, hydrological states, precipitation, snow and catchment charateristics. The results are presented in figure 7 and summarised in table 3.

Considering the effect of the initial conditions, using fully saturated conditions results in the slight overestimation for all catchments of flood values, as expected, and the impact is higher at lower return periods. In addition, Øvrevatn shows higher sensitivity (around 30% for *Q1000*) to the initial soil moisture conditions than the other two catchments. A possible explanation is that the baseflow index (*BFI*) is higher (*BFI*=0.6) for Øvrevatn, than the *BFI* for Krinsvatn (*BFI*=0.4) and Hørte (*BFI*=0.5). This indicates that the runoff at Øvrevatn is less responsive to the rainfall input. Similarly, a sensitivity analysis by Svensson et al. (2013) shows that high sensitivity of floods to soil moisture deficit is more present for permeable catchments, where the *BFI* is also high. The initial discharge value does not seem to have a large impact for any of the catchments. This means that in ungauged catchments the median value can be used. If no snow component (no snowmelt and no snow accumulation) is used, there is not much difference in the results for Øvrevatn and Hørte, but the seasonality of the flood events is changed. For example, the season when the *Q1000* is simulated for Øvrevatn is SON instead of JJA when most of the AMAX values are observed. Due to the change of seasonality, the precipitation values that produce *Q1000* are higher (the median is around 30% higher). The soil moisture deficit, as expected, is also somewhat higher and shows much more spread with values up to 60 mm. In addition, Krinsvatn shows high sensitivity to snowmelt (29% higher) and also a step change in the frequency curve, even though the soil moisture deficit is higher. This can also be explained by the fact that the snowmelt contribution is negative, as can also be seen in table 2.

The results show that the temporal patterns of the rainfall input have high impact (up to 50 %) on the flood frequency curve for Hørte and Krinsvatn, as these catchments have a predominantly rainfall-dominated flood regime, but the impact is very little for Øvrevatn. High sensitivity to the shape of the hyetograph was also found in Alfieri et al. (2008). They found that using rectangular hyetograph results in a significant underestimation of the flood peak while the Chicago hyetograph (e.g. Chow et al., 1988) resulted in overestimation. In addition, Øvrevatn and Hørte showed sensitivity (28.9%) to the choice of the statistical distribution for modelling precipitation. This means that the uncertainty in fitting the rainfall model can propagate to the final results of the stochastic PQRUT, and therefore, it is important to ensure that the choice of distribution and parameters should be carefully considered. A high sensitivity to the parameters of the rainfall model was also described by Svensson et al. (2013), who suggest that this is a main source of uncertainty. Both Hørte and Krinsvatn showed relatively lower sensitivity to the threshold value for the GP distribution, compared to Øvrevatn. A reason for the high sensitivity to the threshold value for Øvrevatn is that using a higher quantile for the threshold leads to selecting fewer events, which are less representative and the fit of the GP becomes more uncertain.

In general, all catchments are very sensitive to the parameters of the PQRUT model and there is a high uncertainty in these values. Because of the higher sensitivity to the calibration of the rainfall-runoff model, a conclusion can be made that, in practice, if streamflow data is available it is important that this is used for calibrating the PQRUT model. Other studies have also shown that the soil saturation level is not as important in comparison with the parameters of the hydrological model. For example, Brigode et al. (2014) tested the sensitivity of the SCHADEX model using a set of block bootstrap method. In each of these experiments, different sub-periods selected from the observation record were used in turn to calibrate the rainfall model, the hydrological model and to determine the sensitivity to the soil saturation level. The results showed that for extreme floods (1000–year return period), the model is sensitive to the calibration of the rainfall and the hydrological model, but not so much to the initial conditions.

3 Comparison with standard methods

3.1 Implementation of the methods and results

The results of the stochastic PQRUT method for the 100- and 1000-year return level were compared with the results for statistical flood frequency analysis and with the standard implementation of the event-based PQRUT method (in which full saturation and snow melting rates are assumed a priori) for the twenty test catchments, described in section 2.1. For the statistical flood frequency analysis, the annual maximum series were extracted from the observed daily mean streamflow series. The GEV distribution was fitted to the extracted values using the L-moments method and the return levels were estimated. In order to obtain instantaneous peak values, the return values were multiplied by empirical ratios, obtained from regression equations, as given in (Midtømme and Pettersson, 2011). The ratios can vary substantially from catchment to catchment, and in this study, the values are from 1.02 to 1.82, depending on the area and the flood generation process. Although much more sophisticated methods could be used to obtain statistically-based return levels, the procedure used here is equivalent to that currently used in standard practice in design flood analysis in Norway. In addition, a study by Kobierska et al. (2017) showed that the GEV along with the GL(Generalised logistic) distribution gives the most reliable results based on a sample of 280 catchments in Norway. The standard implementation of PQRUT involves using a precipitation sequence that combines different intensities, obtained from growth curves based on the 5-year return period value (Førland, 1992). The precipitation intensities were combined to form a single symmetrical storm profile with the highest intensity in the middle of the storm event. Here, the duration of the storm event was assumed to be the same as that used for the stochastic PQRUT model. The initial discharge values were similarly fixed to the seasonal mean values. The snowmelt contribution for the 1000-year return period was, in this case, assumed to be 30 mm/day for all catchments which corresponds to 70% of the maximum snowmelt, estimated as 45mm/day by using temperature-index factor of 4.5 mm//^{0}C day and 10^{0}C. The snowmelt contribution for the 100-year return period was assumed to be 21 mm/day for all catchments. In addition, fully saturated conditions were assumed for both the estimation of the 100- and 1000-year return periods. A similar implementation of PQRUT for the purpose of comparing different methods has also been described in (Lawrence et al., 2014). The parameters of the PQRUT were estimated by using the regional equations derived in Andersen et al. (1983), as these are still used in standard practice.

The performance of the three models was validated by using two different tests. Test 1 assessed whether the estimated values for the flood frequency curve are within the confidence intervals of a GP distribution fitted to the streamflow data with a 1-hour time step. The stochastic PQRUT shows good agreement with the observations, and for 15 of the 20 catchments, all the points of the derived flood frequency curve were inside the confidence intervals. As expected, for most of the catchments (14 out of 20) the return levels calculated using statistical flood frequency analysis based on the GEV distributions using daily values were within the confidence intervals. As discussed, due to the difficulty in assigning initial conditions for the event-based PQRUT model, it was only possible to estimate values for the 100- and 1000-year return periods. For this model, the values of the 100-year return level were within the confidence interval for only six of the catchments when the regional equations were used and for only eight of the catchments when calibrated parameters are used. In addition, the results of the flood frequency analysis and the Stochastic PQRUT methods were compared, based on quantile score (QS), this is given in Eq 5:

$$Qscore = 1 - \sum (abs(Qmod_i - Qobs_i)(Qobs_i - Qobs_{i-1})) \tag{5}$$

In Eq. 5 the observed probabilities ($Qobs_i$) are calculated using Gringorten positions for the POT series. The modelled probabilities that correspond to the observed events are calculated by using the statistical flood frequency analysis and the Stochastic PQRUT model as described previously. The standard implementation of the event-based PQRUT model was not evaluated based on QS as initial conditions could not be assigned for low return periods. The results for the quantile score show similar performance, the median is around 0.83 for both methods. However, the results vary between catchments as shown in fig 8. Although it is difficult to evaluate the performance of the models when the dataseries are relatively short, based on the results of test 1, 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 for the selected catchments.

3.2 Discussion

A comparison of the three methods shows that there is a large difference between the results of the different methods (fig 9 and 10). The violin plots (fig. 9) show that the stochastic PQRUT method gives slightly lower results on average than the standard PQRUT model for *Q100*. This is probably due to assuming fully saturated conditions when applying the standard PQRUT for *Q100*, which might not be realistic for some catchments. For example, the results for the initial conditions for the three catchments, presented in section 2.6, show that the soil moisture deficit is larger than 0. However, when the results are compared for *Q1000*, the stochastic PQRUT gives slightly higher results. Reasons for this may be that higher precipitation intensity or snowmelt is used. Furthermore, the absolute differences between the two methods are larger in catchments with lower temperature (fig. 9). This indicates that the performance of the standard PQRUT model is worse in catchments with a snowmelt flood regime, which might be due to the difficulty in determining snowmelt contribution or the poorer performance of the regional parameters in catchments with a snowmelt flow regime. Although showing the same pattern, the standard PQRUT model, implemented using calibrated parameters results in much less spread than the implementation using the regionalised parameters, when compared to both the GEV distribution and the stochastic PQRUT model. This means that the hydrological model can introduce a large amount of uncertainty, as also indicated by the sensitivity analysis described in section 2.7.

The difference between the stochastic PQRUT model and GEV is much smaller than the difference between the standard PQRUT and GEV, even when calibrated parameters are used. In general, the stochastic PQRUT model gives higher values than the GEV distribution, which might be due to the uncertainty in estimating the parameters for the GEV distribution. For example, the study by Rogger et al. (2012) shows that the flood frequency analysis based on fitting a Gumbel distribution to AMAX series underestimates high flows in catchments with a high storage capacity, where a step change in the flood frequency curve occurs. The results of the study by Rogger et al. (2012) can be explained by the fact that the Gumbel distribution, which has a shape parameter of 0 and so is not as flexible as the GEV distribution. In this study we find very low correlation between the difference between the stochastic PQRUT and the GEV distribution and catchments with high values of the effective lake index or the percentage of the catchment covered by marsh (fig 10), where we would expect that the storage capacity is higher. In addition, the differences are larger (the stochastic PQRUT results are lower as shown in fig 10) in western Norway where *P* and *Q* are higher and for catchments with higher catchment steepness *Hl* (defined as (Hm75-Hm25)/L, where L is the catchment length and Hm25 and Hm75 are the 25 and 75 quantiles of the catchment elevation). A reason might be that the empirical ratios that are used to convert daily to peak flows in these catchments are inaccurate and possibly too high.

Similarly to the violin plots, fig 10 also shows that the results of the stochastic PQRUT closely match the GEV distribution with differences within 50% for most locations. There is no clear spatial pattern in the differences between the GEV distribution and the standard PQRUT model, except for the catchments in Trøndelag (including catchment Krinsvatn) where the GEV distribution produces higher results. However, a much larger sample of catchments is needed to assess whether there is a spatial pattern in the performance of the methods.

4 Conclusions

In this article, we have presented a stochastic method for flood frequency analysis based on a Monte Carlo simulation to generate rainfall hyetographs and temperature series along with the corresponding initial conditions. A simple rainfall-runoff model is used to simulate discharge, and plotting positions are used to calculate the final probabilities. In order to apply the method, we assume that the precipitation and temperature series are not significantly correlated with the initial conditions, which allows us to simulate them as independent variables. Although we have not performed statistical analysis, the independence between the flood events and the initial conditions has been verified by e.g. Paquet et al. (2013). Due to the considerable seasonal variation in the initial conditions, seasonal distributions were used. In addition to obtaining more homogeneous samples, this allows one to check the seasonality of the flood events, which can be of interest in catchments with a mixed flood regime. A limitation of the method is that PQRUT can only be used for small and medium-sized catchments, since its three parameters cannot take into account the spatial variation of the snowmelt and soil saturation conditions within the catchment. However, for the catchments presented in this study (all with area under 850 km^2), the model produces relatively good fits to the observed peaks, even though it uses a very limited number of parameters. For example, a semi-distributed temperature-index snowmelt model, such as the one used in HBV (Sælthun, 1996) , may improve the results in some catchments, though this would also increase the amount of data required.

In this study, initial conditions based on simulations using a hydrological model (DDD) were used. However, in other applications, initial conditions may be based on remotely-sensed data, or on the output of gridded hydrological models. This is particularly important for the application of the method in ungauged basins. Considering the results of the sensitivity analysis, the quality of the initial conditions is not as important as that of the precipitation data for the estimation of extreme floods (with return periods higher than 100 years). This means that if no other data is available, the output of gridded hydrological model could be considered.

The stochastic PQRUT model was applied to 20 catchments, located in different regions of Norway and was compared with the results of the flood frequency analysis and the event based PQRUT model. This comparison shows that there are large differences between the methods. Major sources of uncertainty for the flood frequency analysis are the use of short data series and the empirical peak to volume ratios that were used to calculate the instantaneous flow. There is also uncertainty in the event-based rainfall-runoff simulation method because of difficulties in assigning the initial conditions and in calibrating the rainfall-runoff model. This first of these is a reason why the use of a stochastic model is important, as it can simulate multiple initial conditions and easily incorporate the uncertainty associated with this choice. Due to the high uncertainty in estimating extreme floods, the application of different methods most often produces differing results. A possible way forward is to consider estimates based on different methods by calculating a weighted average of the various estimates, in which the weighting is based on an assessment of the uncertainty characterizing the individual methods.

Code and data availability. The R package StochasticPQRUT (https://github.com/valeriyafilipova/StochasticPQRUT) can be installed from github and contains sample data. A shiny app can be accessed at https://valeriyafilipova.shinyapps.io/PQRUT/

Competing interests. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgements. This work has been supported by a PhD fellowship to Valeria Filipova from USN-Bø. Additional funds from the Energix FlomQ project supported by the Norwegian Research Council and EnergiNorge have partially supported the contributions of the co-authors to this work.

References

Alfieri, L., Laio, F., and Claps, P.: A simulation experiment for optimal design hyetograph selection, Hydrological Processes, 22, 813–820, https://doi.org/10.1002/hyp.6646, 2008.

Andersen, J., Sælthun, N., Hjukse, T., and Roald, L.: Hydrologisk modell for flomberegning (Hydrological for flood estimation), Tech. rep., NVE, Oslo, 1983.

Beven, Keith & Hall, J.: Applied Uncertainty Analysis for Flood Risk Management, https://doi.org/10.1142/p588, http://www.worldscientific.com/doi/pdf/10.1142/9781848162716{_}fmatter, 2014.

Brigode, P., Bernardara, P., Paquet, E., Gailhard, J., Garavaglia, F., Merz, R., Micovic, Z., Lawrence, D., and Ribstein, P.: Sensitivity analysis of SCHADEX extreme flood estimations to observed hydrometeorological variability, Water Resources Research, 50, 353–370, https://doi.org/10.1002/2013WR013687, http://dx.doi.org/10.1002/2013WR013687, 2014.

Calver, A. and Lamb, R.: Flood frequency estimation using continuous rainfall-runoff modelling, Physics and Chemistry of the Earth, 20, 479–483, https://doi.org/10.1016/S0079-1946(96)00010-9, 1995.

Camici, S., Tarpanelli, A., Brocca, L., Melone, F., and Moramarco, T.: Design soil moisture estimation by comparing continuous and storm-based rainfall-runoff modeling, Water Resources Research, 47, https://doi.org/10.1029/2010WR009298, 2011.

Chow, V. T., Maidment, D. R., and Mays, L. W.: APPLIED HYDROLOGY_Chow_V T_Maidment, DR_Mays, L. W.pdf, McGraw-Hill International Editions, 2 edn., 1988.

Coles, S. G.: An introduction to Statistical Modeling of Extreme Values, https://doi.org/10.1007/978-1-4471-3675-0, 2001.

Filipova, V., Lawrence, D., and Klempe, H.: Regionalisation of the parameters of the rainfall–runoff model PQRUT, Hydrology Research, http://hr.iwaponline.com/content/early/2016/01/28/nh.2016.060.abstract, 2016.

Fleig, A. K.: Norwegian Hydrological Reference Dataset for Climate Change Studies, Tech. rep., Oslo, http://webby.nve.no/publikasjoner/rapport/2013/rapport2013{_}02.pdf, 2013.

Førland, E.: Manuel for beregning av påregnelige ekstreme nedbørverdier (Manuel for estimating probable extreme precipitation values), Tech. rep., DNMI, Oslo, 1992.

Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garaon, R., and Bernardara, P.: Introducing a rainfall compound distribution model based on weather patterns sub-sampling, Hydrology and Earth System Sciences, 14, 951–964, https://doi.org/10.5194/hess-14-951-2010, 2010.

Gräler, B., Van Den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. E.: Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation, https://doi.org/10.5194/hess-17-1281-2013, 2013.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Haberlandt, U. and Radtke, I.: Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows, Hydrology and Earth System Sciences, 18, 353–365, https://doi.org/10.5194/hess-18-353-2014, 2014.

Hao, Z. and Singh, V. P.: Review of dependence modeling in hydrology and water resources, Progress in Physical Geography, 40, 549–578, https://doi.org/10.1177/0309133316632460, http://journals.sagepub.com/doi/10.1177/0309133316632460, 2016.

Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, Advances in Water Resources, 25, 1287–1304, https://doi.org/10.1016/S0309-1708(02)00056-8, 2002.

Kim, D., Cho, H., Onof, C., and Choi, M.: Let-It-Rain: a web application for stochastic point rainfall generation at ungaged basins and its applicability in runoff and flood modeling, Stochastic Environmental Research and Risk Assessment, 31, 1023–1043, https://doi.org/10.1007/s00477-016-1234-6, 2017.

Kjeldsen, T. R.: The revitalised FSR/FEH rainfall-runoff method, pp. 1–64, http://www.ceh.ac.uk/sections/hrr/RevitalisationofFSRFEHrainfall-runoffmodel.html, 2007.

Kobierska, F., Engeland, K., and Thorarinsdottir, T.: Evaluation of design flood estimates – a case study for Norway, Hydrology Research, p. nh2017068, https://doi.org/10.2166/nh.2017.068, http://hr.iwaponline.com/lookup/doi/10.2166/nh.2017.068, 2017.

Lawrence, D., Paquet, E., Gailhard, J., and Fleig, A. K.: Stochastic semi-continuous simulation for extreme flood estimation in catchments with combined rainfall-snowmelt flood regimes, Natural Hazards and Earth System Sciences, 14, 1283–1298, https://doi.org/10.5194/nhess-14-1283-2014, 2014.

Li, J., Thyer, M., Lambert, M., Kuczera, G., and Metcalfe, A.: An efficient causative event-based approach for deriving the annual flood frequency distribution, Journal of Hydrology, 510, 412–423, https://doi.org/10.1016/j.jhydrol.2013.12.035, 2014.

Loukas, A.: Flood frequency estimation by a derived distribution procedure, Journal of Hydrology, 255, 69–89, https://doi.org/10.1016/S0022-1694(01)00505-4, 2002.

Midtømme, G. and Pettersson, L.: Retningslinjer for flomberegninger 2011, Tech. Rep. 4/2011, NVE, Oslo, http://publikasjoner.nve.no/retningslinjer/2011/retningslinjer2011{_}04.pdf, 2011.

Mohr, M.: New Routines for Gridding of Temperature and Precipitation Observations for "seNorge. no", Met. no Report, 8, 2008, http://met.no/Forskning/Publikasjoner/Publikasjoner{_}2008/filestore/NewRoutinesforGriddingofTemperature.pdf, 2008.

Muzik, I.: Derived, physically based distribution of flood probabilities, Proceedings of the Yokohama Symposium, p. 183 to 188, 1993.

Nathan, R. J. and Bowles, D.: A Probability-Neutral Approach to the Estimation of Design Snowmelt Floods A Probability-Neutral Approach to the Estimation of Design Snowmelt Floods, pp. 125–130, 1997.

Nyeko-Ogiramoi, P., Willems, P., Mutua, F. M., and Moges, S. A.: An elusive search for regional flood frequency estimates in the River Nile basin, Hydrology and Earth System Sciences, 16, 3149–3163, https://doi.org/10.5194/hess-16-3149-2012, 2012.

Onof, C., Chandler, R., Kakou, A., Northrop, P., Wheater, H., and Isham, V.: Rainfall modelling using Poisson-cluster processes: a review of developments, Stochastic Environmental Research and Risk Assessment, 14, 384–411, https://doi.org/10.1007/s004770000043, http://www.springerlink.com/index/10.1007/s004770000043, 2000.

Paquet, E., Garavaglia, F., Garçon, R., and Gailhard, J.: The SCHADEX method: A semi-continuous rainfall-runoff simulation for extreme flood estimation, Journal of Hydrology, 495, 23–37, https://doi.org/10.1016/j.jhydrol.2013.04.045, http://dx.doi.org/10.1016/j.jhydrol.2013.04.045, 2013.

Parkes, B. and Demeritt, D.: Defining the hundred year flood: A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates, Journal of Hydrology, 540, 1189–1208, https://doi.org/10.1016/j.jhydrol.2016.07.025, 2016.

Rahman, A., Weinmann, P. E., Hoang, T. M. T., and Laurenson, E. M.: Monte Carlo simulation of flood frequency curves from rainfall, Journal of Hydrology, 256, 196–210, https://doi.org/10.1016/S0022-1694(01)00533-9, 2002.

Ren, M., He, X., Kan, G., Wang, F., Zhang, H., Li, H., Cao, D., Wang, H., Sun, D., Jiang, X., Wang, G., and Zhang, Z.: A comparison of flood control standards for reservoir engineering for different countries, Water (Switzerland), 9, https://doi.org/10.3390/w9030152, 2017.

Ries, K. G.: The national streamflow statistics program: A computer program for estimating streamflow statistics for ungaged sites, in: Hydrologic Analysis and Interpretation Section A, Statistical Analysis, p. 37, 2007.

Rogger, M., Kohl, B., Pirkl, H., Viglione, A., Komma, J., Kirnbauer, R., Merz, R., and Blöschl, G.: Runoff models and flood frequency statistics for design flood estimation in Austria - Do they tell a consistent story?, Journal of Hydrology, 456-457, 30–43, https://doi.org/10.1016/j.jhydrol.2012.05.068, 2012.

Sælthun, N. R.: The "Nordic" HBV Model. Description and documentation of the model version developed for the project Climate Change and Energy Production, NVE Publication 7, Norwegian Water Ressources and Energy Administration, Oslo, p. 26, 1996.

Salazar, S., Salinas, J. L., García-bartual, R., and Francés, F.: A flood frequency analysis framework to account flood-generating factors in Western Mediterranean catchments, in: STAHY2017, September, p. 2017, 2017.

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, Hydrol. Earth Syst. Sci, 17, 2637–2652, https://doi.org/10.5194/hess-17-2637-2013, www.hydrol-earth-syst-sci.net/17/2637/2013/, 2013.

Schaefer, M. and Barker, B.: Stochastic Event Flood Model (SEFM), in: Mathematical models of small watershed hydrology and applications, edited by Singh, V. P. and Frevert, D., chap. 20, p. 950, Water Resources Publications, Colorado, USA, 2002.

Skaugen, T. and Onof, C.: A rainfall-runoff model parameterized from GIS and runoff data, Hydrological Processes, 28, 4529–4542, https://doi.org/10.1002/hyp.9968, 2014.

Svensson, C., Kjeldsen, T. R., and Jones, D. a.: Flood frequency estimation using a joint probability approach within a Monte Carlo framework, Hydrological Sciences Journal, 58, 8–27, https://doi.org/10.1080/02626667.2012.746780, http://www.tandfonline.com/doi/abs/10.1080/02626667.2012.746780, 2013.

Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, Water Resources Research, 43, https://doi.org/10.1029/2005WR004723, 2007.

Vormoor, K. and Skaugen, T.: Temporal Disaggregation of Daily Temperature and Precipitation Grid Data for Norway, Journal of Hydrometeorology, 14, 989–999, https://doi.org/10.1175/JHM-D-12-0139.1, http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-12-0139.1, 2013.

Wilson, D., Fleig, A., Lawrence, D., Hisdal, H., Petterson, L., and Holmqvist, E.: A review of NVE's flood frequency estimation procedures, 9, 2011.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

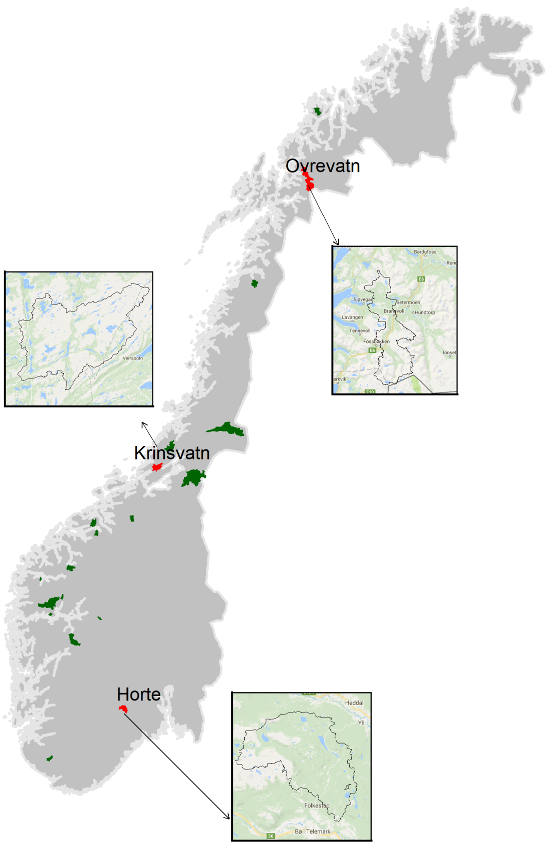

Figure 1. Location of the selected catchments, the catchments Hørte, Øvrevatn and Krinsvatn for which we show the method in more detail are plotted in red

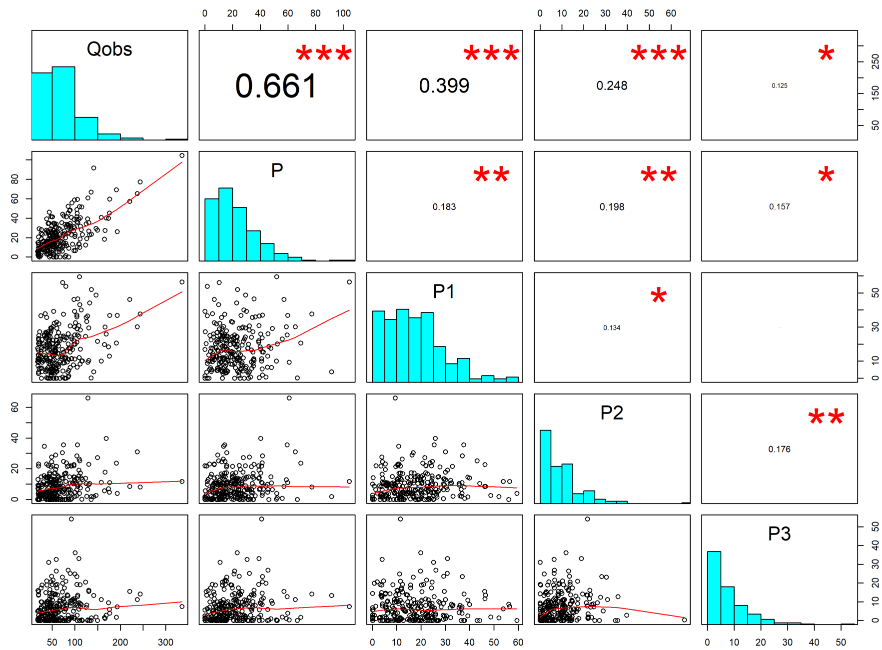

Figure 2. Critical duration for Krinsvatn, stars represent significant correlation between *Qobs* and *P* at p=0.01. The critical duration is set to two days because the correlation is over 0.25 between Qobs and P and P1

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

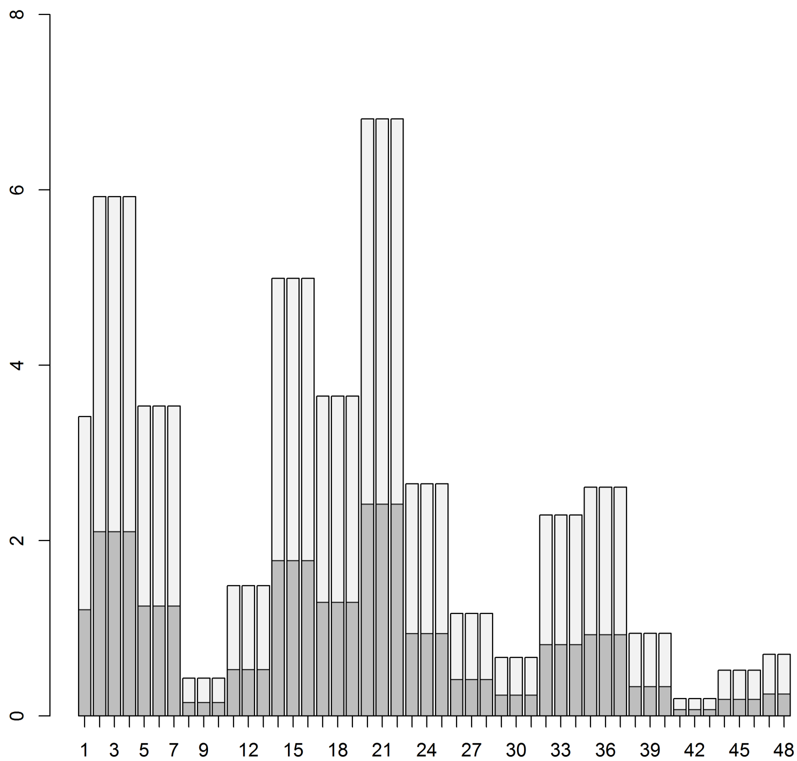

Figure 3. Storm pattern scaling for the simulated events, sampled event is shown in dark grey and the simulated storm event in light grey

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

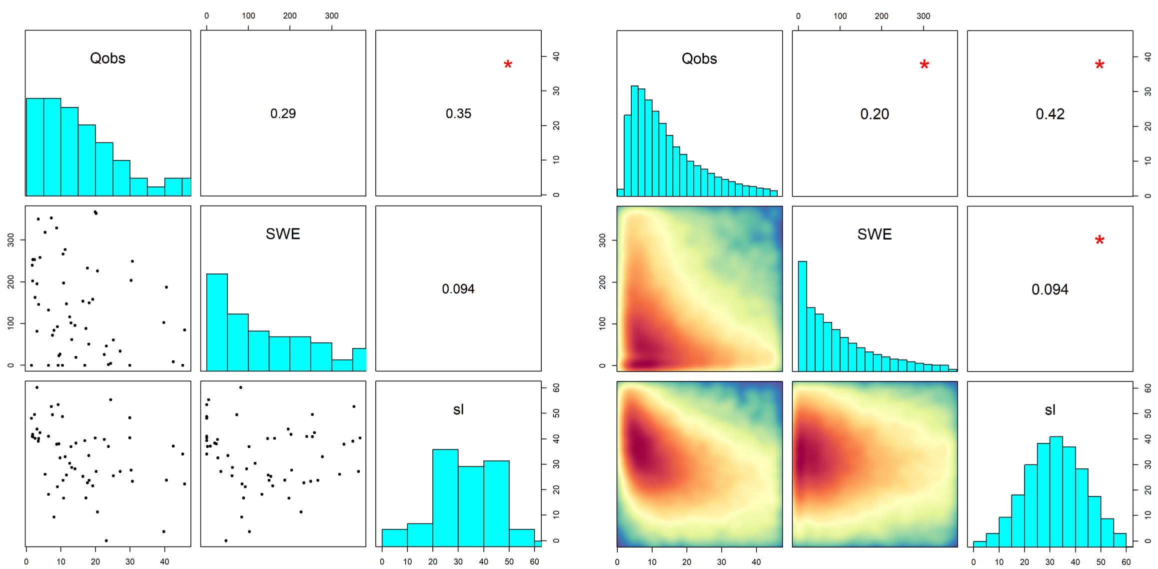

Figure 4. Correlation scatterplot for initial condition (Snow water equivalent, soil moisture deficit, initial discharge) for POT events (flood events over 0.9 quantile), observed on the left , simulated on the right, stars represent significant correlation

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

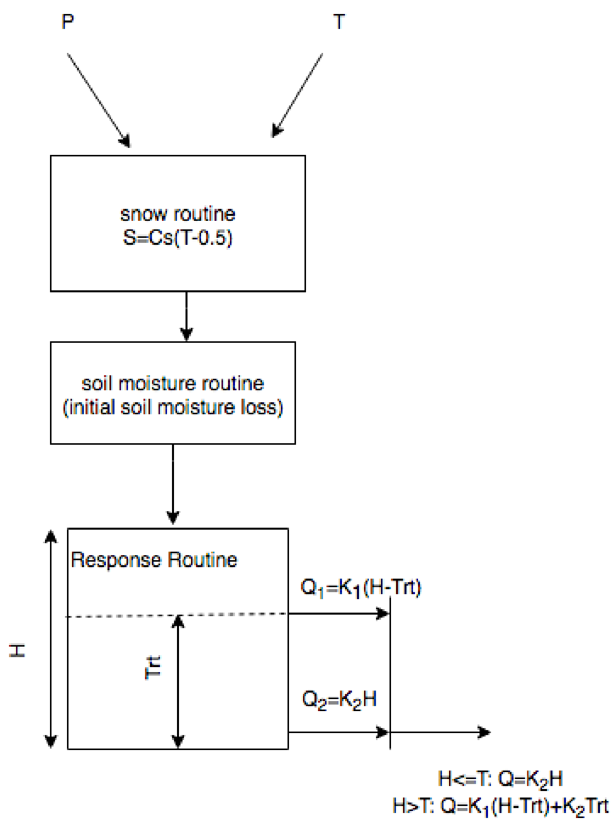

Figure 5. Structure of the PQRUT model

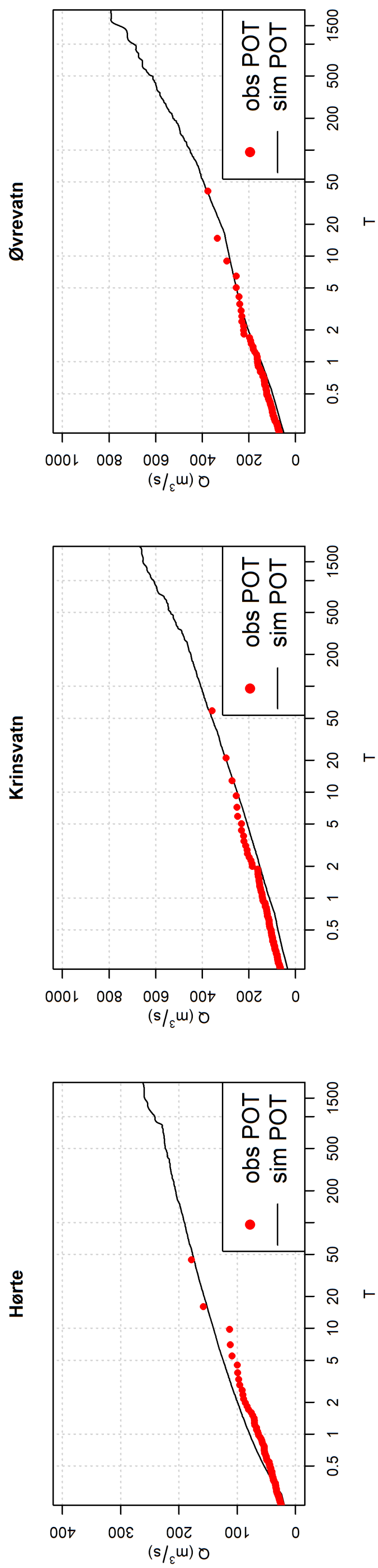

Figure 6. Comparison between the observed and simulated flood frequency curve for Hørte, Krinsvatn and Øvrevatn

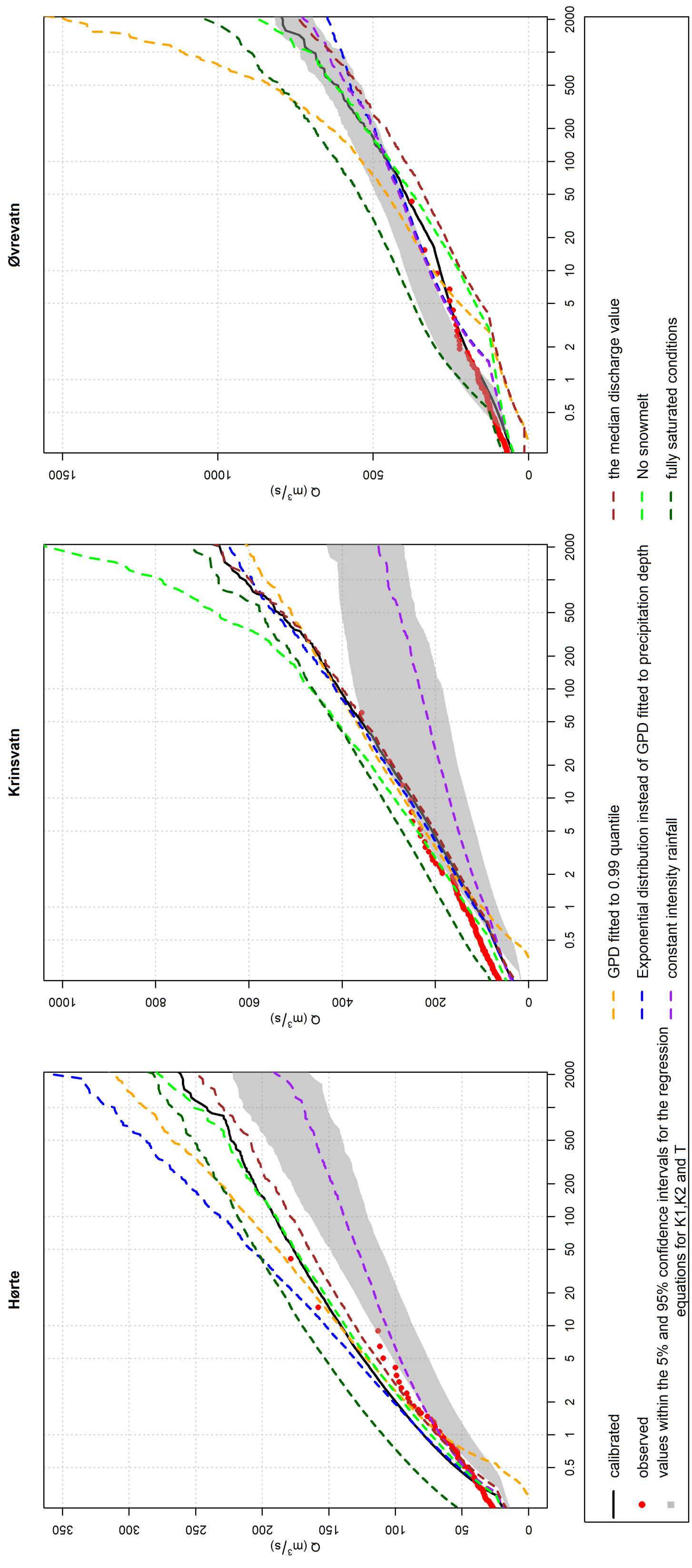

Figure 7. Sensitivity analysis of the Stochastic PQRUT to the initial conditions, rainfall and the parameters of the hydrological model

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

Figure 8. Boxplots of Quantile score, calculated for the Flood frequency analysis and Stochastic PQRUT model,catchments are represented by different colours

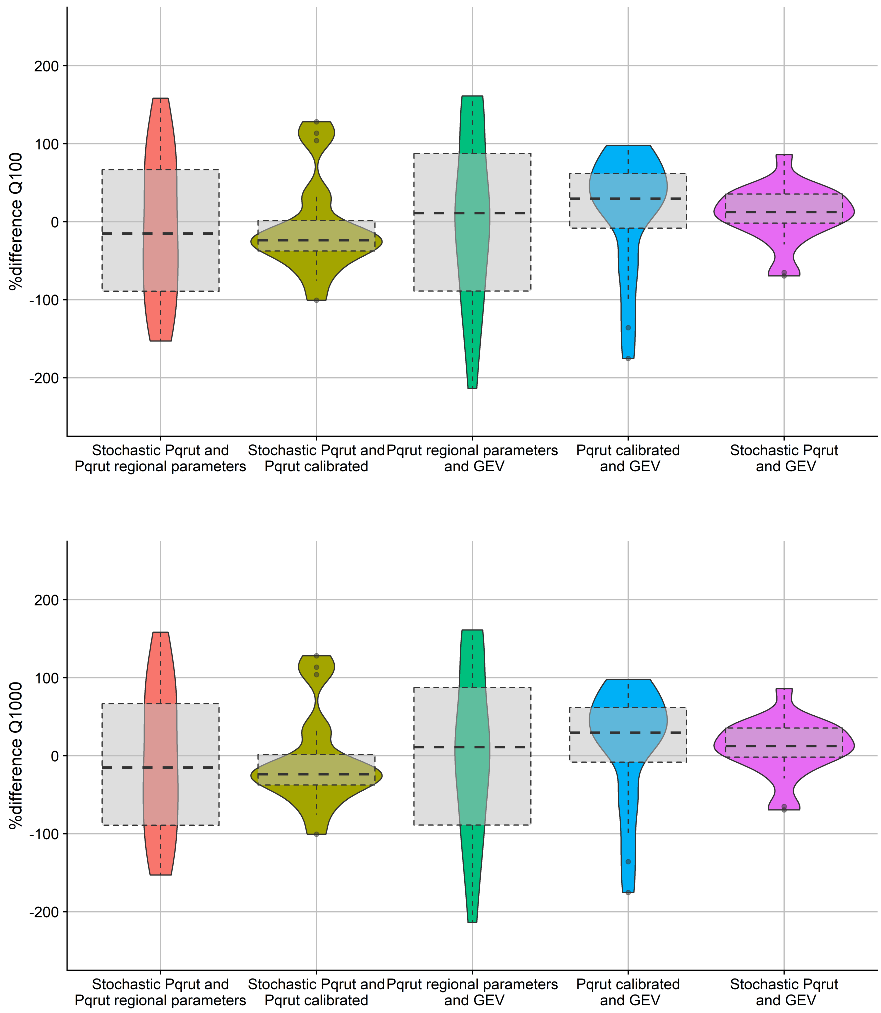

Figure 9. Violin and boxplots showing the distribution of the differences between the Stochastic PQRUT, PQRUT and GEV for the 100- and 1000 – year return level

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

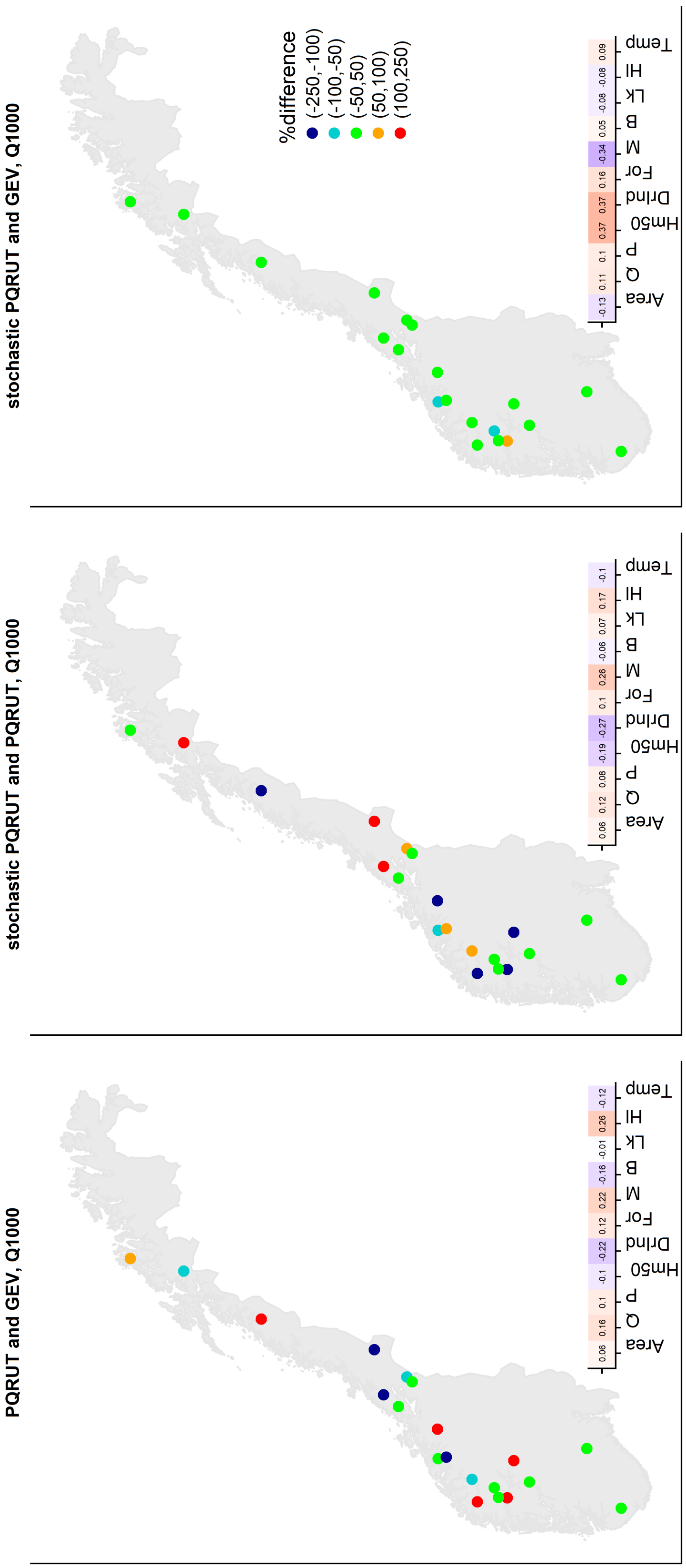

Figure 10. Results of the comparison between stochastic PQRUT, PQRUT and GEV for the values of the 1000-year return level. The absolute differences are correlated with catchment properties,red colour represents positive correlation and blue colour-negative correlation

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

Table 1. Properties for the selected catchments

Station	Area	Q	P	Hm50	For	M	B	Lk	Temp	Season of AMAX
Hørte	157	961	1261	501	73	3	18	0.3	2.89	SON
Krinsvatn	207	1890	2354	348	20	9	57	1.1	4	DJF
Øvrevatn	526	1448	1558	564	35.2	2.5	52	0.6	-0.14	JJA

Table 2. Precipitation and initial conditions for Q100, range corresponds to 5 and 95 percentile

Station name	P mean	P me-dian	P range	Q mean	Q me-dian	Q range	Snowmelt mean	Snowmelt median	Snowmelt range	Soil Moisture mean	Soil Moisture median	Soil Moisture range	Season
Krinsvatn	157.4	143.2	110.1-248.9	18.5	13.6	4 - 42.8	-5.3	0	-29.3-9.6	23	25.5	3.7-39.9	SON/DJF
Hørte	70	66.1	51-94	11.3	9.4	1.9- 23.2	6.8	0	0 - 25.2	8.9	5.2	1.0- 26.9	SON
Øvrevatn	55.73	52.5	35-95	66.5	62.6	33-106.7	29	35.8	-6.11 -42.9	13.6	13	2.34-21.6	JJA

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-174
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 25 June 2018

Table 3. Percent difference between the model runs of the sensitivity analysis to the calibrated model

Catchments		Hørte	Krinsvatn	Øvrevatn	Hørte	Krinsvatn	Øvrevatn	Hørte	Krinsvatn	Øvrevatn
setup	return period	Q10	Q10	Q10	Q100	Q100	Q100	Q1000	Q1000	Q1000
1	100 values, sampled using latin hypercube within the 5% and 95% confidence intervals for the regression equations for K1, K2 and T (min value)	-38.4	-46.9	11.3	-36.8	-54.5	-1.6	-38.2	-58.5	-14.8
1	100 values, sampled using latin hypercube within the 5% and 95% confidence intervals for the regression equations for K1,K2 and T (max value)	-17.0	4.7	39.0	-14.1	-8.4	18.7	-12.6	-34.1	4.2
2	GPD was fitted to 0.99 quantile	0.7	7.4	7.8	10.0	1.1	18.6	19.3	-6.9	60.5
3	Exponential distribution instead of GP distribution	8.2	3.0	10.7	21.3	1.8	-0.3	28.9	-2.8	-13.5
4	Disaggregate precipitation depth using uniform distribution (constant intensity) instead of using temporal patterns	-24.2	-32.0	12.3	-27.5	-42.9	1.2	-29.8	-50.5	-7.5
5	median discharge instead of randomly generated	-7.5	-2.6	-24.0	-6.3	-2.0	-12.3	-5.1	-2.1	-8.8
6	no snowmelt modelled	-2.4	15.1	-21.2	-0.7	14.4	-2.1	4.1	29.1	1.2
7	fully saturated conditions	20.6	25.3	48.1	14.4	13.8	36.2	10.5	8.9	29.1