Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

Natural Hazards and Earth System Sciences
Open Access
Discussions

Decreasing uncertainty in flood frequency analyses by including historic flood events in an efficient bootstrap approach

Anouk Bomers[1], Ralph Schielen[1,2], and Suzanne Hulscher[1]

[1]University of Twente, Dienstweg 1, Enschede, The Netherlands
[2]Ministry of Infrastructure and Water Management-Rijskwaterstaat, Arnhem, The Netherlands

Correspondence: A. Bomers (a.bomers@utwente.nl)

Abstract. Flood frequency curves are usually highly uncertain since they are based on short data sets of measured discharges or weather conditions. To decrease the confidence intervals, an efficient bootstrap method is developed in this study. The Rhine river delta is considered as a case study. A hydraulic model is used to normalize historic flood events for anthropogenic and natural changes in the river system. As a result, the data set of measured discharges could be extended with approximately 600 years. The study shows that flood events decrease the confidence interval of the flood frequency curve significantly, specifically in the range of large floods. This even applies if the maximum discharges of these historic flood events are highly uncertain themselves.

1 Introduction

Floods are one of the main natural hazards to cause large economic damage and human casualties worldwide as a result of serious inundations with disastrous effects. Design discharges associated with a specific return period are used to construct flood defences to protect the hinterland from severe floods. These design discharges are commonly determined with the use of flood frequency analyses (FFA). The basic principle of an FFA starts with selecting the annual maximum discharges of the measured data set, or peak values that exceed a certain threshold (Schendel and Thongwichian, 2017). These maximum or peak values are then used to identify the parameters of a probability distribution. From this fitted distribution, discharges corresponding to any return period can be derived.

Return periods of design discharges are commonly in the order of 500 years or even more, while discharge measurements have been performed only for the last decades. For the Dutch Rhine river delta (used as a case study in this paper), water levels and related discharges have been registered since 1901 while design discharges have a return period up to 100,000 years (Van der Most et al., 2014). Extrapolation of these measured discharges to such return periods results in large confidence intervals of the predicted design discharges. As a result, uncertainties in FFA are generally much larger than acceptable for decision making purposes (Chbab et al., 2006). Uncertainty in the design discharges used for flood risk assessment can have major implications for national flood protection programs since it determines whether and where dike reinforcements are required. A too wide uncertainty range may lead to unnecessary investments.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

To obtain an estimation of a flood with a return period of e.g. 10,000 years with little uncertainty, a discharge data set of at least 100,000 years is required (Klemeš, 1986). Of course, such data sets do not exist. For this reason, many studies try to extend the data set of measured discharges with historic and/or paleo flood events. The most common methods in literature to include historical data into an FFA are based on the traditional methods of frequentist statistics (Frances et al., 1994; MacDonald et al., 2014; Sartor et al., 2010) and Bayesian statistics (O'Connell et al., 2002; Parkes and Demeritt, 2016; Reis and Stedinger, 2005). Most studies found that the confidence intervals of design discharges were reduced significantly by extending the systematic data set with historic events. This finding is beneficial for future flood reducing measures since these can be designed with less uncertainty.

Although previous studies extended the continuous data set of measured discharges with a few historic rare events in isolation, none of these studies created a continuous data set of measured discharges. Chbab et al. (2006) did create a continuous long data set of discharges with the use of a stochastic rainfall generator and flood routing models. The rainfall generator creates a long series (e.g. 1,000 years) of synthetic daily rainfall and temperature data with the same statistical properties as the measured data set. Next, a hydrologic HBV model was used to compute the annual maximum discharges of this extended data set. These discharges were used in an FFA to compute design discharges of the Rhine river for flood risk purposes. They found an increase in the discharge corresponding to a return period of 1,250 years of 2,000 m^3/s compared to the traditional FFA based on measured discharges. It should be noted that the method proposed by Chbab et al. (2006) is biased since it is based only on current climate conditions. This can be considered as a drawback of the method.

Toonen (2015) studied the effects of non-stationarity in flooding regimes over time on the outcome of an FFA. He extended the data set of measured discharges of the Rhine river at Lobith (German-Dutch border) with the use of water level measurements of surrounding sites based on a linear regression analysis. This resulted in a continuous data set of discharges starting from 1772. The fitted flood frequency relations for the period 1772-2012 were used to correspond the GEV parameters with an associated flood activity index. He extended this data set of flood activity indices back to 1350 based on historical information such as damage reports. A classification approach was used in which the geographical spread of reports and extent of damage and life-loss determined to which class a flood event belongs. These classes ranged from minor damage floods with a single dike breach to major damage floods with multiple dike breaches. To describe trends in flooding, the reconstructed flood classes were averaged over a 101-year period. The resultant flood activity indices, which were reconstructed back to 1350, were used to generate GEV parameters for the years before 1772. Monte Carlo simulations were used to generate a peak discharge for each year, creating a continuous data set starting from 1350. Toonen (2015) found that extension of the data set significantly reduced the confidence interval of the flood frequency (FF) curve. Moreover, the 1,250 years flood reduced with 17% if historic events were included in the analysis compared to the data set consisting of only measured discharges. This decrease contradicts the findings of Chbab et al. (2006) who found an increase of the design discharge corresponding to a return period of 1,250 years. A drawback of the method proposed by Toonen (2015) is that the maximum discharges of the continuous data set were not specifically considered since these are based on estimated flood activity indices and corresponding GEV-parameters.

This paper studies whether a continuous discharge data set starting from approximately 1300 AD can reduce the uncertainty in the prediction of discharges corresponding with large return periods. The methods of Chbab et al. (2006) and Toonen

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

Natural Hazards and Earth System Sciences
Discussions

(2015) are combined and extended to set up an efficient bootstrap method to include historic flood events in an FFA. The objective is to develop a straightforward method to consider historic flood events in an FFA, while the basic principles of an FFA remain unchanged. If historic flood events are added to the data set of measured discharges in isolation, confidence intervals are typically not symmetrical since a continuous data set is not present (Schendel and Thongwichian, 2017). Hence, a traditional FFA cannot be used to compute the confidence intervals of the flood frequency relation. This has recently led to the use of bootstrap approaches (e.g. Burn (2003); Kyselý (2008); Schendel and Thongwichian (2017)). This study is novel since a continuous data set is created and hence the confidence intervals are symmetrical. Consequently, a traditional FFA can be used to compute the 95% confidence intervals of the flood frequency relation. The use of a bootstrap approach to compute the confidence intervals are now redundant.

Measured discharges at Lobith (1901-2018) and the continuous data set of Toonen (2015) (1772-1900) are used. This data is extended with historic flood events in Cologne reconstructed by Meurs (2006), which are routed towards Lobith. For this routing, a one dimensional-two dimensional (1D-2D) coupled hydraulic model is used to determine the maximum discharges during these historic events based on the current geometry. In such a way, the historic floods are corrected for anthropogenic interventions and natural changes of the river system, referred to as normalization in this study. Normalizing the historic events is of high importance since flood patterns most likely change over the years as a result of e.g. dike reinforcements, land use change or decrease of floodplain area (dike shifts). The normalized events almost always lead to a higher discharge than the historic event. In any case, it gives insight in the consequences of an event with the same characteristics of a historic flood event translated to present times. To create a continuous data set starting around 1300 AD, a bootstrap resampling technique is used. The results of the bootstrap method will be evaluated against an FFA based on solely measured annual maximum discharges (1901-2018 and 1772-2018). Specifically, the change in the design discharge and its 95% confidence interval of events with a return period of 100,000 years is considered. We are most concerned with estimating this frequency correctly because this design discharge corresponds with the highest safety level used in Dutch flood protection programs (Van Alphen, 2016).

In Section 2 the different data sets used to construct the continuous discharge data set are explained, as well as the 1D-2D coupled hydraulic model. Next, the bootstrap method and FFA are explained (Section 3 and Section 4 respectively). After that, the results of the FFA are given (5). The paper ends with a discussion (Section 6) and the main conclusions (Section 7).

2 Annual maximum discharges

2.1 Discharge measurements period 1901 - present

Daily discharge observations at Lobith have been performed since 1901 and are available at https://waterinfo.rws.nl. From this data set, the maximum annual discharges are selected in which the hydrologic time period, starting at the 1st of October and ending at the 30th of September, is used. Since changes to the system have been made the last century, Tijssen (2009) has normalized the measured data set from 1901-2008 for the year 2004. This data is used since no large alterations upstream of Lobith have occurred since 2004. For the period 2009-2018, the measured discharges are used for which normalization was not required.

During the discharge recording period, different methods have been used to perform the measurements. These different methods result in different uncertainties and must be included in the FFA to correctly predict the 95% confidence interval of the FF curve. From 1901 until 1950, discharges at Lobith were based on velocity measurements performed with floating sticks on the water surface. Since the velocity was only measured at the surface, extrapolation techniques were used to compute the total discharge. This resulted in an uncertainty of approximately 10% (Toonen, 2015). From 1950 until 2000, current meters were used to construct velocity-depth profiles. These profiles were used to compute the total discharge, having an uncertainty of approximately 5% (Toonen, 2015). Since 2000, Acoustic Doppler Current Profiles have been used of which the uncertainty is negligible with the exception of discharges which slightly exceed the bankfull discharge. If the discharge is slightly exceeding the bankfull discharge, the inundated floodplains have a too limited water depth to be accessible by boat. Therefore, the total discharge is estimated by interpolation resulting in an uncertainty of 5% (Toonen, 2015). This uncertainty applies to discharges in the range of between 4.000 to 8.000 m^3/s for the Lower Rhine (Middelkoop, 1997).

2.2 Water level measurements period 1772 - 1900

For the period 1772-1900, the data presented by Toonen (2015) is used. At Lobith, daily water level measurements are available since 1866. For the period 1772-1865 water levels were measured at the nearby gauging locations Emmerich, Pannerden and Nijmegen. Toonen (2015) used the water levels of these locations to compute the water level at Lobith with the use of linear regression between the different measurement locations. Subsequently, he translated these water levels, together with the measured water levels for the period 1866-1900, into discharges using stage-discharge relations at Lobith. This data set represents the computed maximum discharges at the time of occurrence and has not been normalized for changes in the river system. Toonen (2015) argues that, based on the work of Bronstert et al. (2007) and Vorogushyn and Merz (2013), the effect of recent changes in the river system on discharges of extreme floods of the Lower Rhine is small. Hence, it is justified to use the presented data set of Toonen (2015) in this study as normalized data. For more information about the translation of measured water levels to discharges, see Toonen (2015). Fig. 1 shows the annual maximum discharges for the period 1772-2018 and its 95% confidence interval. This data represents the systematic data set used in this study.

2.3 Reconstructed flood events period 1300 AD - 1772

Meurs (2006) has reconstructed maximum discharges during historic flood events near the city of Cologne (Germany). The oldest event dates back to 1342. The used method is described in detail by Herget and Meurs (2010), in which the 1374 flood event was used as a case study. Historic documents providing information about the maximum water levels during the flood event were combined with the reconstruction of the river cross section at that same time. Herget and Meurs (2010) calculated mean flow velocities near the city of Cologne at the time of the historic flood events with the use of the empirical Manning's equation. With this information, they were able to calculate maximum discharges of the specific historic flood events for a reconstructed channel and floodplain bathymetry near the city of Cologne.

In this way, in total 13 historic flood events were reconstructed that occurred before 1772. Two of the flood events occurred in 1651. Only the largest flood of these two is considered as data point. This results in 12 historic floods that are used to

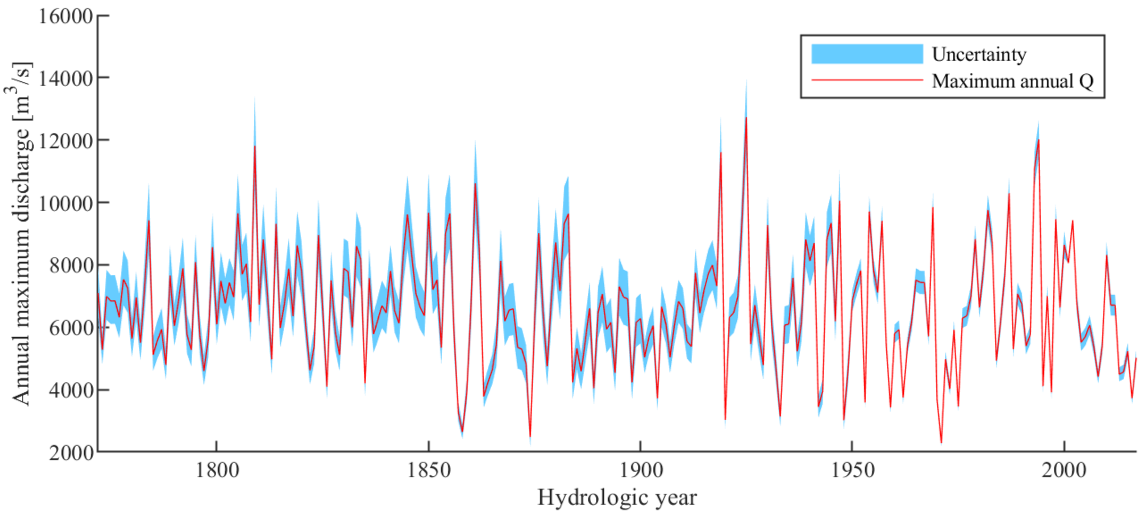

Figure 1. Maximum discharges (Q) and their 95% confidence interval during the systematic time period (1772-2018)

extend the systematic data set. The reconstructed maximum discharges at Cologne (Meurs, 2006), which are not normalized for anthropogenic interventions upstream of Cologne, are used to predict maximum discharges at Lobith with the use of a hydraulic model to normalize the data set. The model used to perform the hydraulic calculations is described in Section 2.3.1. The maximum discharges at Lobith of the 12 historic flood events are given in Section 2.3.2.

2.3.1 Model environment

In this study, the 1D-2D coupled modelling approach as described by Bomers et al. (2019) is used to normalize the data set of Meurs (2006). The study area stretches from Andernach to the Dutch cities of Zutphen, Rhenen and Druten (Fig. 2). In the hydraulic model, the main channels and floodplains are discretized by 1D profiles. The hinterland is discretized by 2D grid cells. The 1D profiles and 2D grid cells are connected by a structure corresponding with the dimensions of the dike that protects the hinterland from flooding. If the computed water level of a 1D profile exceeds the dike crest, water starts to flow into the 2D grid cells corresponding with inundations of the hinterland. A discharge wave is used as upstream boundary condition. Normal depths, computed with the use of the Manning's equation, were used as downstream boundary conditions. HEC-RAS (v. 5.0.3), developed by the Hydrologic Engineering Centre (HEC) of the US Army Corps of Engineers, is used to perform the computations (Brunner, 2016). For more information about the model set-up, see Bomers et al. (2019).

2.3.2 Normalization historic flood events

We use the hydraulic model to route the historical discharges at Cologne, as reconstructed by Meurs (2006), to Lobith. In these simulations, the potential dike breaches and upstream discharges from Meurs (2006) are included as random parameters in a Monte Carlo analysis (MCA). A detailed description of the method is given in Bomers et al. (2019). In particular, the dike

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

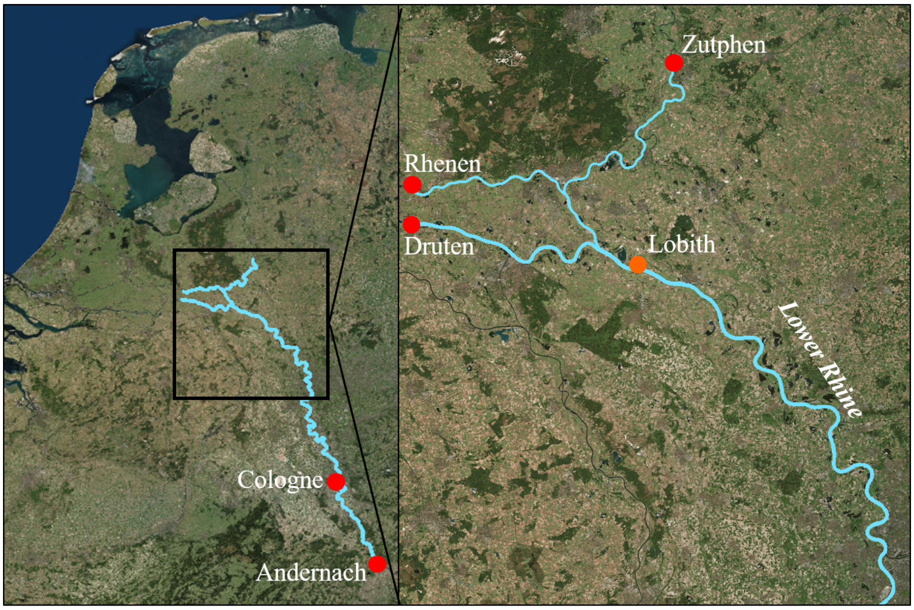

Figure 2. Model domain (blue river branches) of the 1D-2D coupled model

breach thresholds (i.e. the critical water level at which a dike starts to breach) are based on 1D fragility curves provided by the Dutch Ministry of Infrastructure and Water Management. A 1D fragility curve expresses the reliability of a flood defence as a function of the critical water level (Hall et al., 2003). The critical water levels influencing the timing of dike breaching are assumed as random parameters in the MCA. Moreover, the distributions of the final breach width and the breach formation time are based on literature and on historic data (Apel et al., 2008; Verheij and Van der Knaap, 2003). In this study, only the dike breach locations that result in significant overland flow are considered. On the other hand, overflow is possible to occur at every location in the model domain (Bomers et al., 2019). In case of overflow, no dike breach occurs and the water flows into the protected area.

The reconstructed historic discharges at Cologne (Meurs, 2006) have large confidence intervals. The severe 1374 flood, representing the largest flood of the last 1,000 years with a discharge of 23,000 m^3/s, even has an confidence interval of more than 10,000 m^3/s. To include the uncertainty as computed by Meurs (2006) in the analysis, the maximum upstream discharge is varied in the MCA based on its probability distribution. In addition, the upstream discharge shape is varied. The results of the MCA is then the maximum discharge at Lobith and its 95% confidence interval for each of the 12 historic flood events. Fig. 3 gives the reconstructed discharges and their 95% confidence intervals at Cologne as presented in Herget and Meurs (2010) and corresponding maximum discharges and confidence intervals at Lobith. This figure shows that the 1374 extreme flood with a maximum discharge between 18.800 m^3/s and 29.000 m^3/s at Cologne, reduces significantly in downstream direction as a result of overflow and dike breaches. Consequently, the maximum discharge at Lobith turns out to be between 13,825 and 17,753 m^3/s. This reduction in maximum discharge shows the neccesity to apply hydraulic modelling since the use of a

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

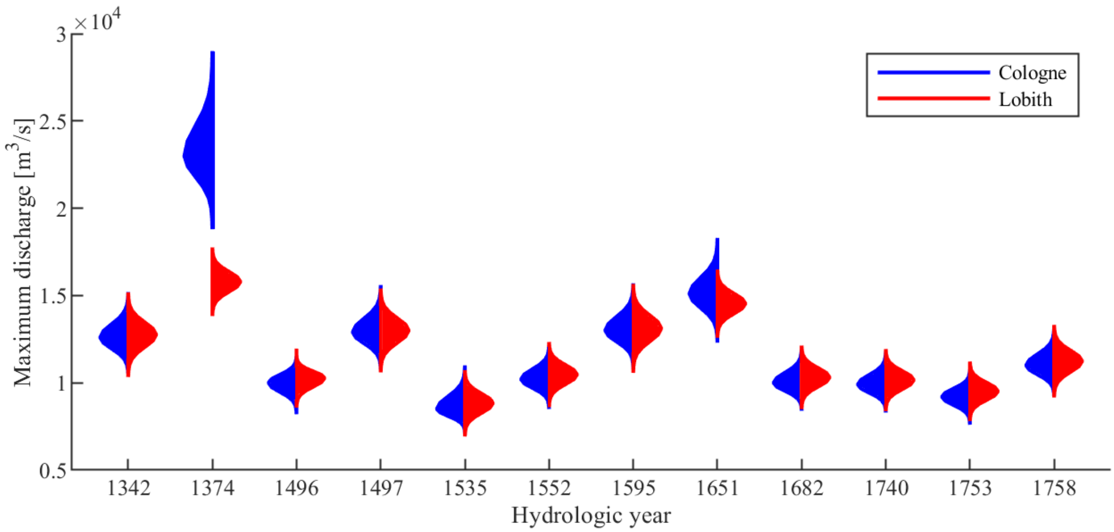

Figure 3. Maximum discharges and their 95% confidence intervals of the reconstructed historic floods at Cologne (Herget and Meurs, 2010) and simulated maximum discharges and their 95% confidence intervals at Lobith for the 12 historic flood events

linear regression analysis based on measured discharges between Cologne and Lobith will result in a much larger maximum discharge at Lobith.

The reconstructed discharges at Lobith are used to extend the systematic data set presented in Sections 2.1 and 2.2. In the next section, these discharges are used in an FFA with the use of a bootstrap method.

3 Bootstrap method

The data set of annual maximum discharges (1901-2018) is extended with the 1772-1900 data set of Toonen (2015), which is based on daily water level measurements at Lobith and surrounding sites, and with 12 reconstructed historic flood events (1300-1772). To create a continuous data set, a bootstrap method based on sampling with replacement is used. The continuous systematic data set (1772-2018) is resampled over the missing years from the start of the historical period to the start of the systematic record. Two assumptions must be made such that the bootstrap method can be applied:

1. The start of the continuous discharge series since the length of the true historical period is not known.

2. The perception threshold over which floods were recorded in the historical times before water level and discharge measurements were conducted.

Assuming that the historical period starts with the first known flood (in this study: 1342) will significantly underestimate the true length of this period. This underestimation influences the shape of the FF curve (Hirsch and Stedinger, 1987; Schendel and Thongwichian, 2017). Therefore, Schendel and Thongwichian (2017) proposed the following equation to determine the length of the historical period:

$$M = L + \frac{L+N-1}{k} \tag{1}$$

where M represents the length of the historical period (years), L the number of years from the first historic flood to the start of the systematic record (431 years), N the length of the systematic record (247 years) and k the number of floods exceeding the perception threshold in both the historical period as well as in the systematic record (28 in total). Using equation 1 results in a historical period from 1317-1771.

The perception threshold is considered to be equal to the discharge of the smallest flood present in the historic period, representing the 1535 flood with an expected discharge of 8,826 m^3/s (Fig. 3). We follow the method of Parkes and Demeritt (2016) assuming that the perception threshold was fairly constant over the historical period. However, the maximum discharge of the 1535 flood is uncertain and hence also the perception threshold is uncertain. Therefore, the perception threshold is treated as a random uniformly distributed parameter in the bootstrap method which boundaries are based on the 95% confidence interval of the 1535 flood event.

The bootstrap method consist of creating a continuous discharge series from 1317-2018. The method includes the following steps (Fig. 4):

1. Combine the 1772-1900 data set with the 1901-2018 data set to create a systematic data set.
2. Select the flood event with the lowest maximum discharge present in the historic time period. Randomly sample a value of between 6,928 and 10,724 m^3/s, representing the 95% confidence interval of the lowest flood event. This value is used as perception threshold.
3. Compute the start of the historical time period (equation 1).
4. Of the systematic data set, select all discharges that have an expected value lower than the sampled perception threshold.
5. Use the data set created in Step 4 to create a continuous discharge series in the historical time period based on bootstrap sampling with replacement.
6. Since both the reconstructed as the measured discharges are uncertain due to e.g. measurement errors, these uncertainties must be included in the analyses. Therefore, for each discharge present in the systematic data set as well as in the historical data set, its value is randomly sampled based on its 95% confidence interval.
7. Combine the data sets of Step 5 and 6 to create a continuous data set starting from 1317-2018.

The presented steps in the bootstrap method are repeated 5,000 times in order to create 5,000 continuous discharge data sets resulting in convergence in the FFA. The FFA procedure itself is explained in the next section.

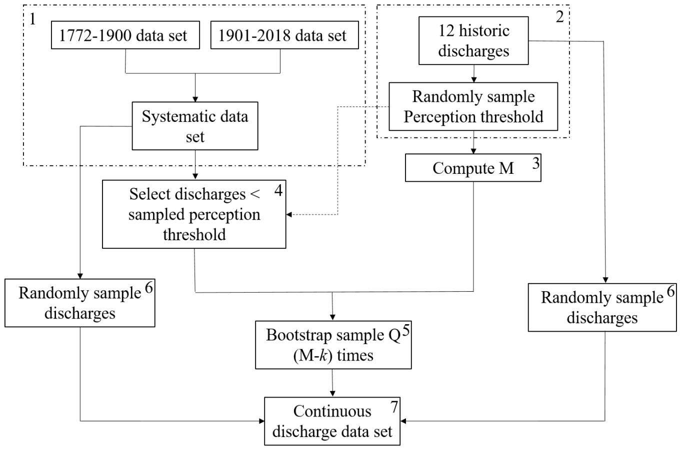

Figure 4. Bootstrap method to create a continuous discharge series

4 Flood frequency analysis

An FFA is performed to determine the FF relation of the different data sets (e.g. systematic record, historical records). A probability distribution function is used to fit the annual maximum discharges to its probability of occurrence. However, the available goodness-of-fit tests for selecting an appropriate distribution function are often inconclusive. Those tests are more appropriate for the central part of the distribution than for the tail (Chbab et al., 2006), where we are interested in since the tail determines the investments required for future flood protection measures. We restrict our analysis to the use of a Generalized Extreme Value (GEV) distribution since this is commonly used in literature to perform an FFA (Parkes and Demeritt, 2016; Haberlandt and Radtke, 2014; Gaume et al., 2010). Additionally, several studies have shown the applicability of this distribution on the flooding regime of the Rhine river (Toonen, 2015; Chbab et al., 2006; Te Linde et al., 2010). The GEV distribution is capable of flattening off at extreme values by having a flexible tail. Therefore, the FF relation converges towards a maximum value for extremely large return periods. This value represents the maximum discharge that is capable of occurring at the location of interest (the location for which the FFA is performed, in this study at Lobith). The GEV distribution is described with the following equation:

$$F(x) = exp\{-[\xi\frac{x-\mu}{\sigma}]^{\frac{1}{\xi}}\}$$

where μ represents the location parameter indicating where the origin of the distribution is positioned, σ the scaling parameter describing the spread of the data, and ξ represents the shape parameter controlling the skewness and kurtosis of the distribution, both influencing the upper tail and hence the upper bound of the system. The maximum likelihood method is used to determine the values of the three parameters of the GEV distribution (Stendinger and Cohn, 1987; Reis and Stedinger, 2005). Although many distributions and fitting methods exist, only the GEV distribution in combination with the maximum likelihood method is considered, since we focus on the influence of extending the data set of measured discharges on the reduction in uncertainty of the FF relations rather than on the suitability of the different distributions and fitting methods.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

Natural Hazards and Earth System Sciences
Open Access
Discussions

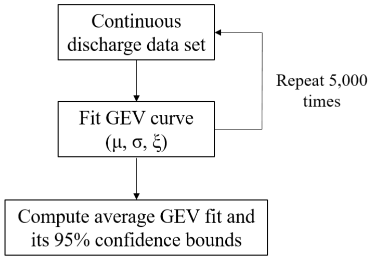

Figure 5. FFA method to construct an averaged flood frequency curve of the MCA

Table 1. Discharges [m^3/s] and their 95% confidence interval corresponding to several return periods for the 1901, 1772 and 1317 data sets and the data set of Toonen (2015)

Data	Q_{10}	Q_{100}	$Q_{1,000}$	-2σ	$Q_{1,250}$	$+2\sigma$	-2σ	$Q_{100,000}$	$+2\sigma$
1901-2018	9,274	12,075	14,121	10,631	14,289	20,836	11,354	16,781	29,631
1772-2018	9,148	11,535	13,149	11,144	13,276	16,263	11,980	15,035	19,996
1317-2018	8,914	11,628	13,727	12,567	13,905	15,493	14,508	16,687	19,487
Toonen (2015)	8,880	10,760	11,580	11,180	11,920	12,800	-	-	-

The FFA is performed for each of the 5,000 continuous discharge data sets created with the bootstrap method (Section 3), resulting in 5,000 fitted GEV curves. The average of these relations is taken to get the final FF curve and its 95% confidence interval. The results are given in the next section.

5 Results

5.1 Flood frequency relations

In this section the FFA results (Fig. 6) of the following data sets are presented:

- Period 1901-2018; measured discharges
- Period 1772-2018; as above and extended with discharges based on measured water levels at Lobith and surrounding sites
- Period 1317-2018; as above and extended with 12 reconstructed historic discharges and the bootstrap resampling method to create a continuous discharge series.

We find a large reduction in the confidence interval of the FF curve if the data set of maximum discharges is extended (Fig. 6 and Table 1). Only extending the data set with discharges based on measured water levels (1772 data set) reduced this confidence interval with 5,200 m^3/s for the floods with a return period of 1,250 years. Adding the reconstructed historic

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

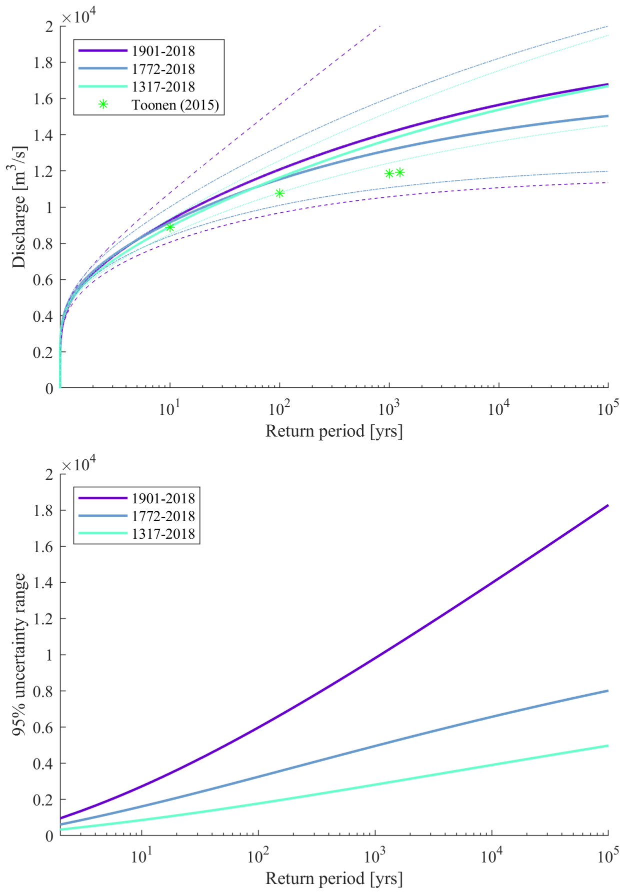

Figure 6. Fitted GEV curves and their 95% confidence intervals of the continuous 1901, continuous 1772 and bootstrap 1317 data sets. The stars represent the data points found by Toonen (2015) based on his 1350-2011 data set

flood events to the data set in combination with a bootstrap method to create a continuous data set, results in an even larger reduction in the confidence interval of 7,400 m^3/s. For the discharges with a return period of 100,000 years, we find an even larger reduction in the confidence intervals (Table 1).

Furthermore, we find that using only the 1901-2018 data set results in larger design discharges compared to the two extended data sets. This is in line with the work of Toonen (2015). Surprisingly however, we find that the 1772-2018 data set predicts the lowest discharges for return periods $>$ 100 years, while we would expect that the 1317-2018 data set predicts the lowest values (Table 1). Toonen (2015) showed that extending the data set with historic flood reconstructions (1350-1772) resulted in a reduction in design discharges corresponding to extreme floods compared to both the 1772 and 1901 data sets. This discrepancy might be explained by the fact that in this paper the 1772-2018 data set was resampled to create a continuous data set starting from 1317. Toonen (2015) assumed a period of low-flood activity between 1350-1772. As a result, his average annual maximum discharge in the period 1350-1772 is lower compared to our data set. Hence, the FF curve of Toonen (2015) is shifted in downward direction compared to our constructed FF curve (Table 1 and Fig. 6). Another explanation of the relatively low positioning of the FF curve of Toonen (2015) is that the large floods present in the 1350-1772 period might be underestimated by Toonen (2015) since these are based on estimated GEV-parameters. The annual maximum discharges of these events are thus not specifically considered in the analysis, since the GEV-parameters were determined with the use of estimated flood activity indices.

Moreover, the relatively low positioning of the FF curve constructed with the 1772 data, compared to our other two data sets, is explained by the fact that the 1772-1900 data set of Toonen (2015) has not been normalized. This period has a relative high flood intensity (Fig. 1). However, only two flood events exceeded 10,000 m^3/s. A lot of dike reinforcements along the Lower Rhine have been executed during the last century. Therefore, it is likely that before the 20^{th} century, flood events with a maximum discharge exceeding 10,000 m^3/s resulted in dike breaches and overflow upstream of Lobith. As a result, the maximum discharge of such an event decreased significantly. Although Toonen (2015) mentions that the effect of recent changes in the river system on discharges of extreme floods of the Lower Rhine is small, we argue that it does influence the flood events with maximum discharges slightly lower than the current main channel and floodplains capacity. Currently, larger floods are possible to flow in downstream direction without the occurrence of inundations compared to the 19^{th} century. This is because the crest levels were lower in the 19^{th} century compared to current crest levels. Therefore, it is most likely that the 1772-1900 data set of Toonen (2015) underestimates the flooding regime of that specific time period. Hence, the continuous data set of Toonen (2015) starting from 1350 is not reliable to predict modern design discharges. However, the aim of Toonen (2015) was to demonstrate the effects of non-stationarity in flooding regimes on the outcome of an FFA. Therefore, non-normalized annual maximum discharges were used in his study.

5.2 Resampling measured data

The results so far have shown that extending the 1901 data set with water level measurements and reconstructed historic floods can significantly reduce the confidence intervals of the FF curves, specifically for the extreme floods. This is in line with the work of O'Connell et al. (2002) who claim that the length of the discharge data set is the single most important factor influencing flood frequency relations. However, the question arises whether it is required to reconstruct historic floods or whether it would also be sufficient to resample the measured data set in order to extend its length. Reconstructing historic floods is quite time consuming, especially if these floods are normalized with a hydraulic model. Therefore, we use the bootstrap method presented in Section 3 to create a data set of approximately 700 years (equal to the 1317-2018 data set) based on solely measured discharges from the period 1901-2018. The perception threshold is assumed to be equal to the lowest measured discharge such that the entire data set of measured discharges is used during the bootstrap resampling. Again, 5,000 discharge data sets are created to reach convergence in the FFA.

We find that the use of the Bootstrap 1901 data set results in lower uncertainties of the FF curve compared to the continuous 1901 data set (Fig. 7). This is because the length of the measured data set is increased through the resampling method. However, although the confidence interval of the 1901 data set decreases significantly after resampling, its interval is still larger compared to the 1317 data set (Fig. 7). In addition, the Bootstrap 1901 FF curve is now based on a relatively short data set of measured

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

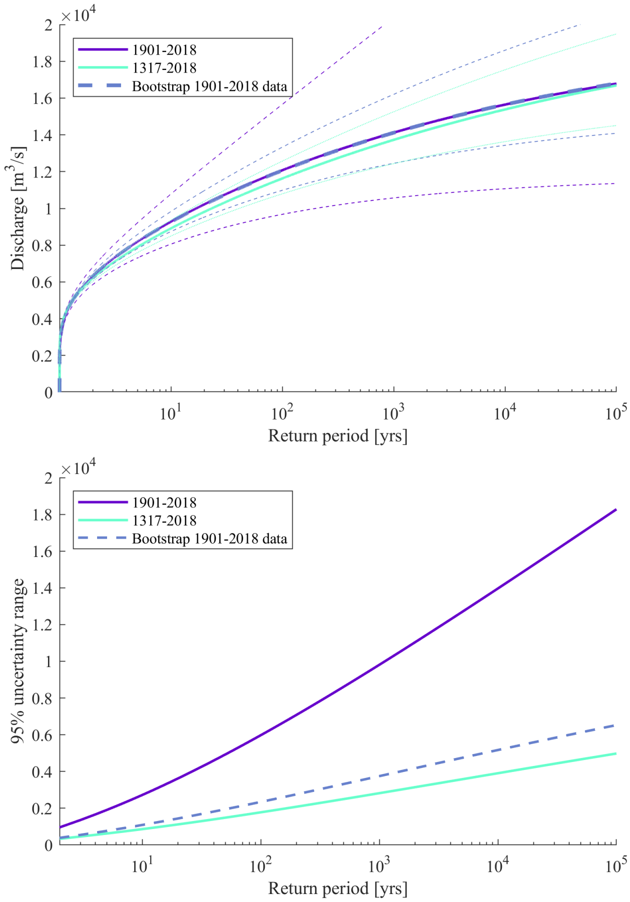

Figure 7. Fitted GEV curves of the continuous 1901, bootstrap 1901 and bootstrap 1317 data sets

discharges and hence only based on the climate conditions of this period. Extending the data set with historic flood events gives a better representation of the long term climatic variability in flood events and hence gives a more accurate prediction of the FF curve and its uncertainty. Therefore, we conclude that reconstructing historic events, even if their uncertainty is large, is worth the effort since it reduces the confidence intervals of rare flood events. Hence, accurate prediction of return periods of rare events is crucial for flood protection policy-making.

6 Discussion

In this study, we developed an efficient bootstrap method to include historic flood events in an FFA. We used a 1D-2D coupled hydraulic model to normalize the data set of Meurs (2006) for modern topography. An advantage of the proposed method is that any kind of historical information (e.g. flood marks, sediment deposits) can be used to extend the data set of annual maximum discharges as long as the information can be translated into discharges.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

The historic flood events reconstructed by Meurs (2006) had a large uncertainty range, especially for the most extreme flood events (Fig. 3). The use of the 1D-2D coupled model reduced this uncertainty range of the maximum discharge at Lobith for most flood events as a result of the overflow patterns and dike breaches along the Lower Rhine. However, the uncertainty was still quite large (at least 3.380 m^3/s). The inclusion of these historic flood events in combination with a bootstrap method to create a continuous data set, resulted in a decrease of the 95% confidence interval of 72% for the discharges at Lobith corresponding to a return period of 100,000 years. Therefore, we can state that adding historical information about rare events with a large uncertainty range in combination with a bootstrap method has the potential to significantly decrease the confidence interval of design discharges of extreme events.

A great advantage of the proposed method is the computation time to create the continuous data sets and to fit the GEV distributions. The entire process is completed within several minutes. Furthermore, it is easy to update the analysis if more historical information about flood events would become available. The shape of the constructed FF curve strongly depends on the climate conditions of the period considered. If the data set is extended with a period which only has a small number of large flood events, this will result in a significant shift of the FF curve in downward direction. This shift can be overestimated if the absence of large flood events only applies to the period used to extend the data set. This study showed that extending the data set from 1901 to 1772 resulted in a shift in downward direction of the FF curve. This is because in the period 1772-1900, a relatively small number of floods exceeded a discharge larger than 10,000 m^3/s. Extending the data set to 1317 resulted in a shift in upward direction again since some rare flood events occurred in this historical period with discharges larger than ever measured. The 1772-1900 data set was taken from Toonen (2015) and has not been normalized for current conditions. Since the crest levels of many flood defences along the Lower Rhine were heightened in the 20th century, larger discharges are possible to occur at Lobith nowadays. The flooding regime in the 1772 data set are thus underestimated. Therefore, it is important that the measured data set is extended with normalized reconstructed discharges that capture the long-term climatic variability. In this way, the FF curve is predicted more accurately and can be used for future flood protection programs.

After the 1993 and 1995 flood events of the Rhine river, the FF relation was recalculated taking into account the discharges of these events. All return periods were adjusted. The design discharges with a return period of 1,250 years, which was the most important return period at that time, was increased with 1,000 m^3/s (Parmet et al., 2001). Such an increase in the design discharge requires more investments in the near future. Parkes and Demeritt (2016) found similar results for the river Eden, UK. They showed that the inclusion of the 2015 flood event had a significant effect on the upper end of the FF curve, even though their data set was extended from 1967 to 1800. This extension was done by adding 21 reconstructed historic events to the data set of measured data. Therefore, Schendel and Thongwichian (2017) argues that if the flood frequency relation alters after a recent flood, and if this change can be ambiguously attributed to this event, the data set of measured discharges must expanded since otherwise the FF results will be upward biased.

We are interested in how a future flood event will change the FF curve using the proposed bootstrap method compared to a traditional FF based on measured discharges. Both the 1317 and 1901 data set are extended from 2018 to 2019 with a hypothesized flood in 2019. We assume that in 2019 a flood event has occurred that equals the highest measured discharge so far. This corresponds with the 1926 flood event, having a maximum discharge of 12,600 m^3/s. Since measurement errors

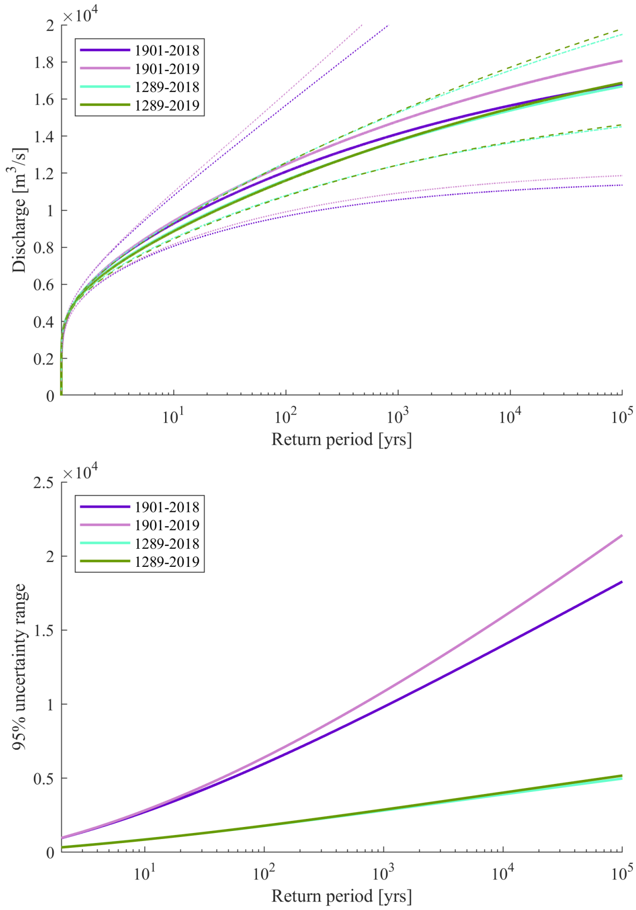

Figure 8. Fitted GEV curves and their 95% confidence intervals of the continuous 1901 and bootstrap 1317 data set if they are extended with a future flood event

can be neglected nowadays, no uncertainty of this event is included in the analysis. Fig. 8 shows that the FF curve based on the 1901 data set changes significantly as a result of this hypothesized 2019 flood. We calculate an increase in the discharge corresponding with a return period of 100,000 years of 1,280 m^3/s. Contrarily, the 2019 flood has almost no effect on the extended 1317 data set. The discharge corresponding to a return period of 100,000 years only increased slightly with 180 m^3/s. Therefore, we conclude that the extended data set is more robust to changes in FF relations as a result of future flood events. Hence, we expect that the changes in FF relations after the occurrence of the 1993 and 1995 flood events would be less severe if the analysis was performed with an extended data set as presented in this study. Consequently, less investments were required to cope with the new flood safety standards in the near future. Therefore, we state that flood managers can be less nervous about the occurrence of future flood events and their impact on flood frequency relations.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-83
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 18 March 2019

7 Conclusions

Design discharges are commonly determined with the use of flood frequency analyses (FFA) in which measured discharges are used to fit a probability distribution function. However, discharge measurements have been performed only for the last 10-15 decades. This relatively short data set of measured discharges results in large uncertainties in the prediction of design discharges corresponding to rare events. Therefore, this study presents an efficient bootstrap method to include historic flood events in an FFA. The proposed method is efficient in terms of computational time and set-up. Additionally, the basic principles of the traditional FFA remain unchanged.

The proposed bootstrap method was applied to the discharge series at Lobith. The data sets based on discharges (1901-2018) and water levels (1772-1900) were extended with the 12 most extreme historic flood events. It was found that extending the data set of measured discharges with historic flood events significantly decreases the 95% confidence interval of the FF curve. This specifically applies for rare events (in this study with 72% for the discharge with a return period of 100,000 years). Even though some of the historic events were highly uncertain themselves, including these in the data set still resulted in a large reduction in the 95% confidence intervals of the predicted flood frequency relations. Since correct prediction of flood frequency relations with little uncertainty is of high importance for future national flood protection programs, we recommend to use historical information in FFA. Additionally, extending the data set with historic events makes the flood frequency relation less sensitive to future flood events. The proposed method to include historical discharges into a traditional FFA can be easily implemented in flood safety assessments because of its simple nature in terms of mathematical computations as well as of its computational efforts.

Acknowledgements. This research is supported by the Netherlands Organisation for Scientific Research (NWO, project 14506) which is partly funded by the Ministry of Economic Affairs and Climate Policy. Furthermore, the research is supported by the Ministry of Infrastructure and Water Management and Deltares. This research has benefited from cooperation within the network of the Netherlands Centre for River studies NCR (www.ncr-web.org).

The authors would like to thank the Dutch Ministry of Infrastructure and Water Management, Prof. Dr. Herget (University of Bonn) and Dr. Toonen (KU Leuven) for providing the data. Furthermore, the authors would like to thank Dr. Toonen (KU Leuven) for his valuable suggestions that improved the manuscript.

References

Apel, H., Merz, B., and Thieken, A. H.: Quantification of uncertainties in flood risk assessments, International Journal of River Basin Management, 6, 149–162, https://doi.org/10.1080/15715124.2008.9635344, 2008.

Bomers, A., Schielen, R. M. J., and Hulscher, S. J. M. H.: The effect of dike breaches on downstream discharge partitioning, in: Abstract from NCR Days 2019: Land of Rivers, p. 28, Utrecht, The Netherlands, 2019.

Bronstert, A., Bardossy, A., Bismuth, C., Buiteveld, H., Disse, M., Engel, H., Fritsch, U., Hundecha, Y., Lammersen, R., Niehoff, D., and Ritter, N.: Multi-scale modelling of land-use change and river training effects on floods in the Rhine basin, River research and applications, 23, 1102–1125, https://doi.org/10.1002/rra.1036, 2007.

Brunner, G. W.: HEC-RAS, River Analysis System Hydraulic Reference Manual, Version 5.0, Tech. Rep. February, US Army Corp of Engineers, Hydrologic Engineering Center (HEC), Davis, USA, 2016.

Burn, D. H.: The use of resampling for estimating confidence intervals for single site and pooled frequency analysis, Hydrological Sciences Journal, 48, 25–38, https://doi.org/10.1623/hysj.48.1.25.43485, 2003.

Chbab, E. H., Buiteveld, H., and Diermanse, F.: Estimating Exceedance Frequencies of Extreme River Discharges Using Statistical Methods and Physically Based Approach, Osterreichse Wasser- und Abfallwirtschaft, 58, 35–43, 2006.

Frances, F., Salas, J. D., and Boes, D. C.: Flood frequency analysis with systematic and historical or paleoflood data based on the two-parameter general extreme value models, Water Resources Research, 30, 1653–1664, https://doi.org/10.1029/94WR00154, 1994.

Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., and Blöschl, G.: Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites, Journal of Hydrology, 394, 101–117, https://doi.org/10.1016/j.jhydrol.2010.01.008, 2010.

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.

Hall, J. W., Dawson, R. J., Sayers, P. B., Rosu, C., Chatterton, J. B., and Deakin, R.: A methodology for national-scale flood risk assessment, Proceedings of the Institution of Civil Engineers - Water and Maritime Engineering, 156, 235–247, https://doi.org/10.1680/wame.2003.156.3.235, http://www.icevirtuallibrary.com/doi/10.1680/wame.2003.156.3.235, 2003.

Herget, J. and Meurs, H.: Reconstructing peak discharges for historic flood levels in the city of Cologne, Germany, Global and Planetary Change, 70, 108–116, https://doi.org/10.1016/j.gloplacha.2009.11.011, 2010.

Hirsch, R. M. and Stedinger, J. R.: Plotting positions for historical floods and their precision, Water Resources Research, 23, 715–727, https://doi.org/10.1029/WR023i004p00715, 1987.

Klemeš, V.: Dilettantism in hydrology: Transition or destiny?, Water Resources Research, 22, 177–188, https://doi.org/10.1029/WR022i09Sp0177S, 1986.

Kyselý, J.: A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models, Journal of Applied Meteorology and Climatology, 47, 3236–3251, https://doi.org/10.1175/2008JAMC1763.1, 2008.

MacDonald, N., Kjeldsen, T. R., Prosdocimi, I., and Sangster, H.: Reassessing flood frequency for the Sussex Ouse, Lewes: The inclusion of historical flood information since AD 1650, Natural Hazards and Earth System Sciences, 14, 2817–2828, https://doi.org/10.5194/nhess-14-2817-2014, 2014.

Meurs, H.: Bestimmung der Spitzenabflusse historischer Hochwasser in Koln, PhD thesis, 2006.

Middelkoop, H.: Embanked floodplains in the Netherlands. Geomorphological evolution over various time scales, Netherlands Geographical Studies, 224, 341, 1997.

O'Connell, D. R. H., Ostenaa, D. A., Levish, D. R., and Klinger, R. E.: Bayesian flood frequency analysis with paleohydrologic bound data, Water Resources Research, 38, 1058–1071, https://doi.org/10.1029/200WR000028, 2002.

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.

Parmet, B., van de Langemheen, W., Chbab, E., Kwadijk, J., Diermanse, F., and Klopstra, D.: Analyse van de maatgevende afvoer van de Rijn te Lobith, Tech. rep., RIZA, Arnhem, The Netherlands, 2001.

Reis, D. S. and Stedinger, J. R.: Bayesian MCMC flood frequency analysis with historical information, Journal of Hydrology, 313, 97–116, https://doi.org/10.1016/j.jhydrol.2005.02.028, 2005.

Sartor, J., Zimmer, K.-h., and Busch, N.: Historische Hochwasserereignisse der deutschen Mosel, Wasser Abfall, 10, 46–51, 2010.

Schendel, T. and Thongwichian, R.: Considering historical flood events in flood frequency analysis: Is it worth the effort?, Advances in Water Resources, 105, 144–153, https://doi.org/10.1016/j.advwatres.2017.05.002, 2017.

Stendinger, J. R. and Cohn, R. A.: Flood frequency analysis with historical and paleoflood information, Water Resources Research, 22, 785–793, 1987.

Te Linde, A. H., Aerts, J. C., Bakker, A. M., and Kwadijk, J. C.: Simulating low-probability peak discharges for the Rhine basin using resampled climate modeling data, Water Resources Research, 46, 1–19, https://doi.org/10.1029/2009WR007707, 2010.

Tijssen, A.: Herberekening werklijn Rijn in het kader van WTI2011, Tech. rep., Deltares, Delft, the Netherlands, 2009.

Toonen, W. H.: Flood frequency analysis and discussion of non-stationarity of the Lower Rhine flooding regime (AD 1350-2011): Using discharge data, water level measurements, and historical records, Journal of Hydrology, 528, 490–502, https://doi.org/10.1016/j.jhydrol.2015.06.014, 2015.

Van Alphen, J.: The Delta Programme and updated flood risk management policies in the Netherlands, Journal of Flood Risk Management, 9, 310–319, https://doi.org/10.1111/jfr3.12183, 2016.

Van der Most, H., De Bruijn, K. M., and Wagenaar, D.: New Risk-Based Standards for Flood Protection in the Netherlands, in: 6th International Conference on Flood Management, pp. 1–9, https://doi.org/10.1017/CBO9781107415324.004, 2014.

Verheij, H. J. and Van der Knaap, F. C. M.: Modification breach growth model in HIS-OM. H.J. Verheij, Tech. rep., WL | Delft Hydraylics, Delft, The Netherlands, 2003.

Vorogushyn, S. and Merz, B.: Flood trends along the Rhine: The role of river training, Hydrology and Earth System Sciences, 17, 3871–3884, https://doi.org/10.5194/hess-17-3871-2013, 2013.