Contributions to Geophysics and Geodesy Vol. 40/2, 2010 (121–147)

Inclusion of historical information in flood frequency analysis using a Bayesian MCMC technique: a case study for the power dam Orlík, Czech Republic

Ladislav GAÁL, Ján SZOLGAY, Silvia KOHNOVÁ, Kamila HLAVČOVÁ[1]

[1] Department of Land and Water Resources Management, Faculty of Civil Engineering, Slovak University of Technology, Radlinského 11, 813 68 Bratislava, Slovak Republic; e-mail: ladislav.gaal@stuba.sk, jan.szolgay@stuba.sk, silvia.kohnova@stuba.sk, kamila.hlavcova@stuba.sk

Alberto VIGLIONE[2]

[2]Vienna University of Technology, Karlsplatz 13/222, A-1040 Vienna, Austria
e-mail: viglione@hydro.tuwien.ac.at

Abstract: The paper deals with at-site flood frequency estimation in the case when also information on hydrological events from the past with extraordinary magnitude are available. For the joint frequency analysis of systematic observations and historical data, respectively, the Bayesian framework is chosen, which, through adequately defined likelihood functions, allows for incorporation of different sources of hydrological information, e.g., maximum annual flood peaks, historical events as well as measurement errors. The distribution of the parameters of the fitted distribution function and the confidence intervals of the flood quantiles are derived by means of the Markov chain Monte Carlo simulation (MCMC) technique.

The paper presents a sensitivity analysis related to the choice of the most influential parameters of the statistical model, which are the length of the historical period h and the perception threshold X_0. These are involved in the statistical model under the assumption that except for the events termed as 'historical' ones, none of the (unknown) peak discharges from the historical period h should have exceeded the threshold X_0. Both higher values of h and lower values of X_0 lead to narrower confidence intervals of the estimated flood quantiles; however, it is emphasized that one should be prudent of selecting those parameters, in order to avoid making inferences with wrong assumptions on the unknown hydrological events having occurred in the past.

The Bayesian MCMC methodology is presented on the example of the maximum discharges observed during the warm half year at the station Vltava-Kamýk (Czech Republic) in the period 1877–2002. Although the 2002 flood peak, which is related to the vast flooding that affected a large part of Central Europe at that time, occurred in the near past, in the analysis it is treated virtually as a 'historical' event in order to illustrate some crucial aspects of including information on extreme historical floods into at-site flood frequency analyses.

Key words: flood frequency analysis, historical floods, Bayesian inference, Markov chain Monte Carlo simulations, likelihood function, confidence intervals

1. Introduction

The hydrologic and hydrometeorological extremes and critical thresholds derived from complex hydrological and meteorological events for engineering design are usually obtained on the basis of single site characteristics (e.g., the annual maximum daily rainfall or discharge). Therefore, hydrological and meteorological frequency analyses have also mainly focused on one characteristic value (e.g., *Cunnane, 1987; Bobée and Rasmussen, 1994; Brunovský et al., 2009*). Various methods have been proposed to reduce the uncertainties of at-site flood frequency analyses and produce more robust flood quantile estimates based on larger sample sizes. Two main families of approaches can be distinguished *(Merz and Blöschl, 2008a,b)*:

(i) 'spatial extension' of information on floods through regional flood frequency methods based on aggregating statistically homogeneous data to build large data samples (e.g., *Hosking and Wallis, 1997*), and
(ii) 'temporal extension' of information on floods through at-site flood frequency studies on gauged streams extended by historical floods or paleofloods (e.g., *Reis and Stedinger, 2005*).

To overcome the problem of relatively short data series for frequency analysis the need to investigate extremes also spatially was traditionally widely acknowledged in the hydrological community. The very basic paradigm of this approach originates from the index flood method introduced by *Dalrymple (1960)* and it is commonly used to implement a regional frequency analysis for a particular variable of interest. The aim of regional frequency analysis is to increase the information content of the analysis and to reduce the uncertainty of the design values estimates by 'trading space for time'. To address this issue, spatial (regional) properties of extremes are studied and regional frequency analysis is typically applied. The regional approaches usually involve two major steps: the delineation of homogeneous regions (sometimes referred to as 'pooling groups') and the estimation of extreme value quantiles at the sites of interest using information from all sites in the region. Traditionally, homogeneous pooling groups were formed based on geographical position or administrative boundaries (*NERC, 1975; Beable and McKerchar, 1982; Wiltshire, 1986; Podolinská et al., 2005*). Therefore, *Acreman and Wiltshire (1989)* suggested a pooling approach with no need of having adjacent members in groups, i.e., groups defined in a flexible way. This concept was further developed in different ways: *Burn (1990a,b)* introduced the region of influence (ROI) focused pooling method, while various clustering techniques (e.g., *Burn, 1997; Burn and Goel, 2000*) and canonical correlation analysis *(Ouarda et al., 2001)* have been proposed to form homogeneous pooling groups. An important role among the various regional frequency analysis methods has the one based on the L-moments *(Hosking and Wallis, 1997)*. Numerous authors contributed to refinements of these directions of research (e.g., *Meigh et al., 1997; Institute of Hydrology, 1999; Solín, 2002; Merz and Blöschl, 2003; Castellarin, 2007; Guse et al., 2010*); thus, further details will not be given here.

A large part of our knowledge on extreme flood discharge values is based on inventories of data regarding extraordinary events (e.g., *UNESCO, 1976; Rodier and Roche, 1984; Mimikou, 1984; Costa 1987; Svoboda and Pekárová, 1998; Alcoverro et al., 1999; Herschy, 2005; Costa and Jarrett, 2008; Solín, 2008; Pekárová, 2009; Gaume et al., 2009*). Since such extraordinary events are important source of information on the flood extremes, attempts were undertaken to include these into at-site frequency analyses (e.g., *Hosking and Wallis, 1986; Stedinger and Cohn, 1986; Parent and Bernier, 2002*).

This is obtained in this paper by including past historical extreme values in flood frequency analyses by the Bayesian Markov chain Monte Carlo (MCMC) framework (*Kuczera, 1999; Reis and Stedinger, 2005*). Let us illustrate the principles of the inclusion of historic data in at-site flood frequency analyses. Suppose that the information on historical hydrological events consists of m extraordinary floods, and the joint data sample of systematic and historical observations is stationary. In order to properly account for the historical information, the evaluation of the m historical peak discharges is not sufficient. It is also important to consider the number of years h in which these events were the major floods and to evaluate the threshold X_0 which has certainly not been exceeded during this period by the other floods. In other words, the historical information consists not only in the m extreme discharge values but also in $h - m$ years of non-exceedance of the threshold X_0. The choice of h and X_0 should meet the criterion of "exhaustiveness" (i.e., no other major floods should have exceeded X_0 in the period of time h), which is a necessary condition for a proper statistical inference with censored data *(Leese, 1973; Gaume et al., 2010)*. The Bayesian MCMC procedure *(Reis and Stedinger, 2005)* is a flexible tool, which can handle the information both on the historical and systematic observations through adequately defined likelihood functions in a straightforward way, and more importantly, can account for uncertainties in the measurements of the hydrologic extremes, and provides estimates of confidence bounds for the estimated quantiles. It has been demonstrated that the inclusion of the historic period leads to a clear reduction of the confidence intervals *(Reis and Stedinger, 2005)*.

This paper aims at demonstrating that this method, although it may bring a significant added value in flood frequency analyses, is still not free from subjectiveness and both skill and care have to be exercised when applied. Several choices that have to be made are not trivial. The paper ends with some suggestion for the users of the method.

2. Methodology

Bayesian Markov chain Monte Carlo (MCMC) methods provide a computationally convenient way to fit frequency distributions for flood frequency analysis by using different sources of information as large flood records, historical floods, uncertainties (particularly measurement errors), regional information and other hydrologic information. They also provide an attractive and straightforward way to estimate the uncertainty in parameters and quantile metrics (e.g., *Robert and Casella, 2004*).

Since a flood frequency analysis with information on historical events included is a relatively complex methodology, in the next couple of subsections we will shed light on the most relevant details and settings of the procedure.

2.1. Parameter estimation by means of the maximum likelihood methodology

The maximum likelihood estimation (MLE) is a popular statistical method when a mathematical model is to be adjusted to the observed data. For a given data sample D (where I is the length of the data sample) and a probabilistic model $\ell\,(D|\theta)$ with parameters θ, the MLE aims at determining the parameters θ that maximize the probability (likelihood) of the data sample D. The function $\ell\,(D|\theta)$ is called likelihood function.

Under the assumption that the data are independent and identically distributed (iid), the likelihood function of the whole data sample $\ell\,(D|\theta)$ may be written as a product of the likelihood functions of the particular events d_i:

$$\ell\,(D|\theta) = \prod_{i=1}^{I} \ell\,(d_i|\theta). \tag{1}$$

Using a logarithmic transformation, the product in Eq. (1) changes to a sum

$$\log \ell\,(D|\theta) = \sum_{i=1}^{I} \log \ell\,(d_i|\theta). \tag{2}$$

The maximum value of the expression in Eq. (2) may then be found numerically, using any of a wide range of optimization algorithms.

2.2. Inclusion of historical information

For the estimation of parameters of a distribution function, *Stedinger and Cohn (1986)* presented a MLE methodology that – through properly defined likelihood functions – can also take into consideration information on historical hydrological events.

Suppose that the following hydrological information is available:

- s – length of the systematic observations X with realizations $\{x_1,\ ...,\ x_i,\ ...,\ x_s\}$,
- h – length of the historical period,
- n – total length of the analyzed time period: $n = s + h$,

- X_0 – a perception threshold, below which the $h-m$ non-recorded flood maxima are assumed to lie;
- m – number of events exceeding the threshold X_0 during the historical period h,
- c – number of events exceeding the threshold X_0 during the period with systematic records s,
- k – the total number of extraordinary events Y that exceed the threshold X_0 during the whole analyzed period, with realizations $\{y_1, \ldots, y_j, \ldots, y_k\}$; $k = c + m$.

A sketch of a hypothetical combination of systematic and historical information is presented in schematic figures below, with the following settings: $s = 50$ years, $h = 150$ years ($\Rightarrow$ $n = 200$ years), $X_0 = 1400$ m^3/s, $m = 3$, $c = 1$ ($\Rightarrow$ $k = 4$).

Furthermore, let $f_X(\cdot)$ and $F_X(\cdot)$ denote the probability distribution function (PDF) and the cumulative distribution function (CDF) of the variable X, respectively; θ is a parameter of $f_X(\cdot)$ and $F_X(\cdot)$, and D denotes the data. Moreover, let us suppose that both variables X and Y are iid, thus $f_X(\cdot)$ and $F_X(\cdot)$ may be used for taking into account the information on historical events Y.

Through two simple examples, we demonstrate how to combine systematic and historical information in the framework of likelihood estimation.

Example #1. Let us suppose that the magnitudes of the historical events y_j, $j = 1, \ldots, e$ are exactly known (Fig. 1). The joint likelihood function of the systematic and historical data $\ell(D|\theta)$ can then be expressed as a product of the following three terms:

a) ℓ_a – the likelihood of observing m historical events exceeding the threshold X_0,

b) ℓ_b – the likelihood of observing the other $h-m$ historical events below the perception threshold X_0, and

c) ℓ_c – the likelihood of observing the s systematic records:

$$\ell_a = \prod_{j=1}^{m} f_X(y_j), \tag{3}$$

$$\ell_b = [F_X(X_0)]^{h-m}, \tag{4}$$

$$\ell_c = \prod_{i=1}^{s} f_X(x_i), \tag{5}$$

$$\ell(D|\theta) = \ell_a \ell_b \ell_c = \prod_{j=1}^{m} f_X(y_j)\,[F_X(X_0)]^{h-m} \prod_{i=1}^{s} f_X(x_i). \tag{6}$$

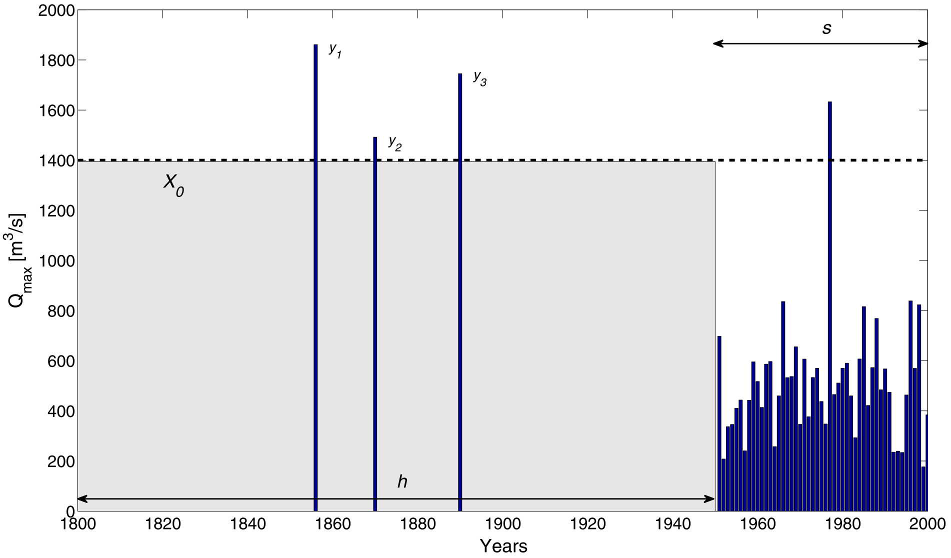

Fig. 1. A sketch of systematic and historical data where it is supposed that the latter ones are known exactly. Notation: s – length of the systematic observations; h – length of the historical period; y_1, y_2, y_3 – historical events; X_0 – perception threshold; Q_{max} – annual maxima of flood peaks.

Example #2. When compared to the previous example, the only change is that the magnitudes of the historical events are known with some uncertainty, i.e., they are bounded by the lower and upper limits y_{Lj} and y_{Uj}, $j = 1, \ldots, m$ (Fig. 2). Such a change affects the expression for ℓ_a, which takes a form

$$\ell_a = \prod_{j=1}^{m} [F(y_{Lj}) - F(y_{Uj})], \tag{7}$$

and accordingly

$$\ell\left(D|\theta\right)=\ell_a\ell_b\ell_c=\prod_{j=1}^{m}\left[F\left(y_{Lj}\right)-F\left(y_{Uj}\right)\right]\left[F_X\left(X_0\right)\right]^{h-m}\prod_{i=1}^{s}f_X\left(x_i\right). \tag{8}$$

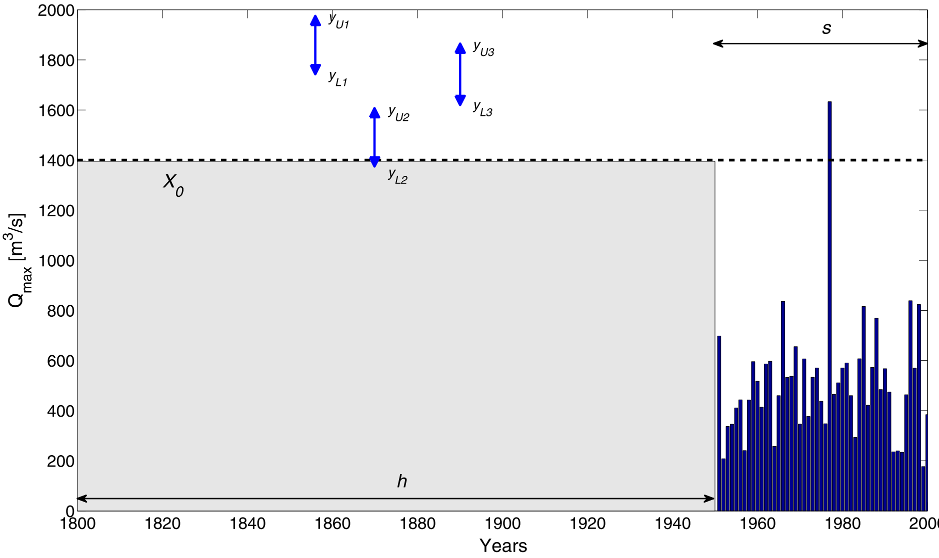

Fig. 2. A sketch of systematic and historical data where it is supposed that the latter ones are known with some uncertainty. Notation: s – length of the systematic observations; h – length of the historical period; y_{L1}, y_{L2}, y_{L3} (y_{L1}, y_{L2}, y_{L3}) – lower (upper) uncertainty bounds of historical events; X_0 – perception threshold; $\mathrm{Q}_{\max}$ – annual maxima of flood peaks.

In the case when flood events exceeding the perception threshold X_0 occur among the systematic data (i.e., $c \neq 0$, analogously to the sketch in Fig. 2), the events are virtually removed from the period s and are treated as historical data (cf. *Bayliss and Reed, 2001*). Having this situation, Eqs. (6) and (8) are slightly modified in the following way:

$$\ell\left(D|\theta\right)=\prod_{j=1}^{k}f_X\left(y_j\right)\left[F_X\left(X_0\right)\right]^{h-m}\prod_{i=1}^{s-c}f_X\left(x_i\right), \tag{9}$$

and

$$\ell\left(D|\theta\right)=\prod_{j=1}^{k}\left[F\left(y_{Lj}\right)-F\left(y_{Uj}\right)\right]\left[F_X\left(X_0\right)\right]^{h-m}\prod_{i=1}^{s-c}f_X\left(x_i\right). \tag{10}$$

2.3. Bayesian inference

The Bayesian approach is a branch of statistical analysis that is based on a unique philosophy: the statistical inference is drawn in the way that the initial beliefs on the subject of the interest are modified according to the observed data. Thus, the Bayesian inference combines two kinds of information: (i) the prior knowledge (belief, hypothesis) on the unknown parameters that may come from other data sets, logical intuition or the past experiences of the analyst and (ii) the information encapsulated in the observed data, which are represented by the likelihood function (*Reis and Stedinger, 2005*).

A Bayesian inference is based on the application of the Bayes' theorem:

$$P\left(\theta|D\right)=\frac{P\left(D|\theta\right)\ P\left(\theta\right)}{P\left(D\right)}, \tag{11}$$

where $P\left(\theta\right)$ is the prior (marginal) distribution of the parameters θ (it does not take into account any information contained in the observed data D); $P\left(\theta|D\right)$ is the posterior distribution of the parameters θ, having the data D observed; $P\left(D|\theta\right)$ is the conditional probability of the data, given the parameters θ; and $P\left(D\right)$ is the prior (marginal) distribution of the data D. $P\left(D\right)$ only serves as a normalization constant in order to obtain a unit area under the posterior PDF $P\left(\theta|D\right)$.

The conditional probability where the second argument is considered as a parameter is also called likelihood function: $p\left(\theta|D=d\right)$ or $\ell\left(D|\theta=\Theta\right)$ (see also Sect. 2.1). Bayes' theorem, using the aforementioned notation and having a continuous variable θ can be rewritten as follows:

$$p\left(\theta|D\right)=\frac{\ell\left(D|\theta\right)\ p\left(\theta\right)}{\int_{\Omega}\ell\left(D|\theta\right)\ p\left(\theta\right)\ d\theta}, \tag{12}$$

where integral in the denominator is computed through the entire parameter space Ω.

Eqs. (11–12) are the mathematical formulation of the way the hypothesis of the statistical properties of the parameters θ (the existing/prior beliefs) is updated by having observed the data D (in the light of new pieces of knowledge). Naturally, the final posterior distribution may serve as a prior distribution in a further Bayesian inference.

One of the main advantages of the Bayesian inference is that it results in a full posterior probability of the parameters: one may easily derive the credible intervals (counterparts of the traditional confidence intervals) of the parameters or any of their functions. Compared to this, the traditional methods of statistical analysis that are usually based on asymptotical assumptions, look at the parameters of a distribution function as fixed (unknown) constants, and the result of a statistical analysis is usually a point estimate of the parameters (i.e., *Coles, 2001*).

2.4. Markov Chain Monte Carlo simulations

In general, an analytical computation of the integral in the denominator of Eq. (12) is very hard if not impossible; thus, that is the main reason why statistical methods based on the Bayesian inference have not been used frequently in the past (e.g., *Coles, 2001; Gelman et al., 2004; Reis and Stedinger, 2005*). The rapid development of computers in the past 2–3 decades, however, has opened wide perspectives for numerical evaluation of complex mathematical problems. Markov chain Monte Carlo (MCMC) methods represent a class of such algorithms. By means of Monte Carlo simulations, the Metropolis-Hastings algorithm (e.g., *Tierney, 1994; Gelman et al., 2004*) generates a Markov chain, which results in a sample, distribution of which converges to the posterior distribution $P(\theta|D)$. In other words, MCMC simulation draws samples from the posterior distribution of the parameters without having computed the normalization constant of Eq. (12) analytically *(Reis and Stedinger, 2005).* The quality of the sample improves as a function of the number of steps of the Markov chain. The resulting data sample then serves for the estimation of marginal distributions of the joint probability distribution function, mean values, standard deviations, and confidence intervals not only for the parameters themselves but also for their arbitrary functions such as the required quantiles (design values) of the analyzed hydrological extremes *(Reis and Stedinger, 2005).*

2.5. Plotting position formulae

There are several plotting position formulae (PPF) that are used in frequency analysis to get a quick glance on the empirical distribution of the data sample analyzed: to check whether they follow a particular distribution, if there are some errors or outliers etc. (e.g., *Rao and Hamed, 1999*). One of the most frequently used PPFs is the one of Cunnane:

$$p_i = \frac{i - 0.4}{n + 0.2}, \tag{13}$$

where n denotes the sample size, i is the rank of the observations in an ascending order, and p_i is the cumulative probability of non-exceedance of the i-th data.

In the case of a joint frequency analysis of systematic and historical data, the PPF (Eq. 13) should be slightly modified according to the number of the historical events k that occurred during the whole n-year period analyzed *(Bayliss and Reed, 2001)*. For the historical events, one should apply

$$p_j = \frac{k}{n}\frac{j - 0.4}{k + 0.2}, \quad j = 1,\ ...,\ k, \tag{14}$$

and for the systematic data

$$p_i = 1 - \frac{n-k}{n}\left(1 - \frac{i - 0.4}{s - c + 0.2}\right), \quad i = 1,\ ...,\ s - c, \tag{15}$$

where the notation is in accordance with the one introduced in Sect. 2.2 as well as with the formulae presented by *Bayliss and Reed (2001, p. 34)*. In the paper, we apply Eqs. (14–15) for the visualization of the observed flood peak data on the probability plots (Sect. 5).

2.6. Probability distribution function

To demonstrate different aspects of joint probability modelling of the systematic and the historical data, we employ the 3-parameter log-normal distribution (LN3). The selected distribution function is one of the most frequently used statistical models in flood frequency analyses (e.g., *Hosking and Wallis, 1997*).

The cumulative distribution function of the LN3 distribution is

$$F(x) = \Phi(y), \tag{16}$$

where Φ is the CDF of the standard normal distribution and

$$y = \begin{cases} -k^{-1} \log\left[1 - k(x-\xi)/\alpha\right], & k \neq 0 \\ (x-\xi)/\alpha, & k = 0, \end{cases} \tag{17}$$

where ξ, α and k are the location, scale and shape parameters, respectively *(Hosking and Wallis, 1997, p. 197)*. The special case with $k = 0$ yields the standard normal distribution.

3. Data

The frequency analysis is performed at the station Vltava-Kamýk, which is situated at the Vltava River on the south of the Czech Republic. One of the most important features of the station Vltava-Kamýk is the fact that it is located near the hydroelectric power dam Orlík, which is the largest one in the country, and which also plays a key role in the flood prevention of Prague, the capital city of the country.

We analyze the peak discharge values measured in the summer season (May – October) that are available in the period from 1877 to 2002. The 2002 event, however, is an exceptional one in the data records since it is one of the largest floods of the last decade that affected not only the Czech Republic but also the Central Europe, and which resulted in a number of fatalities and huge infrastructural damages (e.g., *Ulbrich et al., 2003; Kundzewicz et al., 2005*). The exceptional magnitude of the 2002 flood was also confirmed by analyses of the specialists of the Czech Hydrometeorological Institute *(Drbal et al., 2003; Boháč and Kulasová, 2005)*. Note that since an antecedent flood wave filled up all the smaller reservoirs upstream the Vltava River, the extraordinary flood wave on 13 August 2002 was practically unaffected.

Due to the fact that the peak discharge value of this extraordinary 2002 event at station Vltava-Kamýk has nearly exceeded two times the maximum of the rest of the flood peak records from the period 1877–2001, we treat this event as a historical one.

4. Settings of the frequency model

In Sect. 2.2, we introduced a number of parameters to characterize various features of a hypothetical data set consisting both of systematic measurements and historical data. Herein, we apply this notation on the records of flood peaks from the station Vltava-Kamýk as follows:

- The length of the systematic observations is unequivocally given: $s =$ 126 years.
- The 2002 flood event is exceptional in the light of the other flood peaks observed during the whole period with the systematic observations; thus, we consider this flood as historical one. Moreover, this is the only extraordinary event that appears in the analysis; therefore, $c = 1$.
- The perception threshold X_0 is unknown. Nevertheless, we suppose that X_0 lies somewhere between the absolute maximum of the systematic records (4390 m^3/s, recorded in 2002) and the secondary maximum (2309 m^3/s, recorded in 1890). Since a) we do not have any information on flood peaks having occurred before the systematic observations, and b) we would not like to put much restriction on these unknown events, we set the value of the perception threshold level X_0 just below the magnitude of the historical event, i.e., let $X_0 = 4000$ m^3/s. We then also perform a sensitivity analysis related to the choice of this value (Sect. 5.2).
- The length of the historical period h is also unknown. We only suppose that the extraordinary 2002 event has not been exceeded during the whole analyzed period $n = h + s$. In other words, we have to make an initial estimate of the return period of this extraordinary event, and assume that the magnitudes of all the unknown events that have occurred in the 'historical' period do not exceed the perception threshold X_0 (i.e., $m = 0$). A sensitivity analysis regarding the choice of h is also carried out in the paper (Sect. 5.1).
- Since the magnitude of the historical flood event is very well known, it is incorporated into the statistical model using the methodology presented in 'Example # 1' in Sect. 2.1. Uncertainties in the magnitude of this event (similarly to 'Example # 2') are not considered in the analysis herein.

- The Bayesian approach allows for choosing any prior distribution $p(\theta)$ of the parameters θ of the selected distribution function (Sect. 2.3, Eq. 12). Since an arbitrary choice is allowed, we choose the simplest solution: we do not put any stress on the priors of θ, i.e., we use uniform initial distribution $p(\theta) = \text{const.}$, which does not have effects on the evaluation of Eq. (12).

5. Results

All the simulations were carried out within the R environment, which is an open-source software for statistical computing and visualization (`http://www.r-project.org/`), and using the library *nsRFA (Viglione, 2009)*, which has specifically been developed for flood frequency analyses.

The main results of the frequency analysis are presented in terms of figures (probability plots) and tables, where the main focus was set to the flood quantiles corresponding to return periods T = 100, 1000 and 10000 years. The latter value, although it may seem unreasonably high, is justified by the fact that the flood quantile corresponding to the return period of T = 10000 years has been defined as the critical design value for the safety of the power dam Orlík.

In the next sub-sections, we present sensitivity analyses, which examine the effects of the choice of the most important parameters of the statistical model (the length of the historical period h, the threshold X_0 and the length of the systematic records s) on the quantile estimates and the width of their confidence intervals.

5.1. Selection of the length of the historical period

The way the choice of the length of the historical period h affects the quantile estimates of peak discharges in a joint flood frequency analysis of systematic and historical flood events is presented in Fig. 3 and Table 1. The panel of Fig. 3 consists of 6 subplots. The plot in the top left corner displays the case with the historical event excluded ('No historical information' in its title), i.e., a frequency analysis only based on the records from the period 1877–2001. The other plots are related to a joint analysis of systematic and historical information with different estimates of the parameter h, namely $h = 500$, 1000, 1500, 2000 and 3000 years. The solid line in the middle of the plots is the fitted distribution function, while the dashed line above (below) it is the upper (lower) bound of the 5-95% confidence intervals of flood quantiles. The systematic data are represented by empty circles, while the historical information is depicted by a solid black circle. The perception threshold has been set to $X_0 = 4000$ m^3/s in all simulations related to Fig. 3.

Comparing the first subplot with no historical information included, and the rest of the plots corresponding to different h values it is discernible that the inclusion of the historical event with an extraordinary magnitude reduces the width of the confidence bounds. This fact is underpinned by the numerical evaluation of the confidence intervals (CIs) of the representative flood quantiles: both the absolute and relative widths of the CI (ΔCI and $\Delta CI/Q_T$, respectively) for the highest value of the parameter h are approximately half of the widths of the corresponding CI for the alternative with no historical information involved (Table 1). This holds for all three return periods considered. From Fig. 3 it is also clear that the value of h controls the slope of the fitted CDF: higher values of h are associated with less steeper CDFs.

The inclusion of historical information evidently reduces the uncertainty in the quantile estimates. Nevertheless, one of the most important questions still remains open: which of the selected h values should be preferred? In a traditional flood frequency analysis that is only based on systematic data, different goodness-of-fit tests (such as the test of χ^2, test of Kolmogorov and Smirnov or the Z-test; see, e.g., *Wilks, 1995; Hosking and Wallis, 1997*) can be applied in a simple way to assess the 'closeness' of the fitted distribution to the observed data. Nevertheless, as soon as any historical information is incorporated, testing the goodness of the fit becomes a difficult business. As far as we know, there are no methods in the literature to accomplish such a task. Therefore, we decided to use a visual inspection method. We check whether 90% of the observed data visualized through the modified plotting position formulae (Eq. 15) lie or do not lie within the 90% confidence interval estimated by the MCMC simulation procedure. Although the method is subjective, it can indicate with respect to the CDF where the fit is not sufficiently good, the presence of changes of slopes (e.g., due to threshold

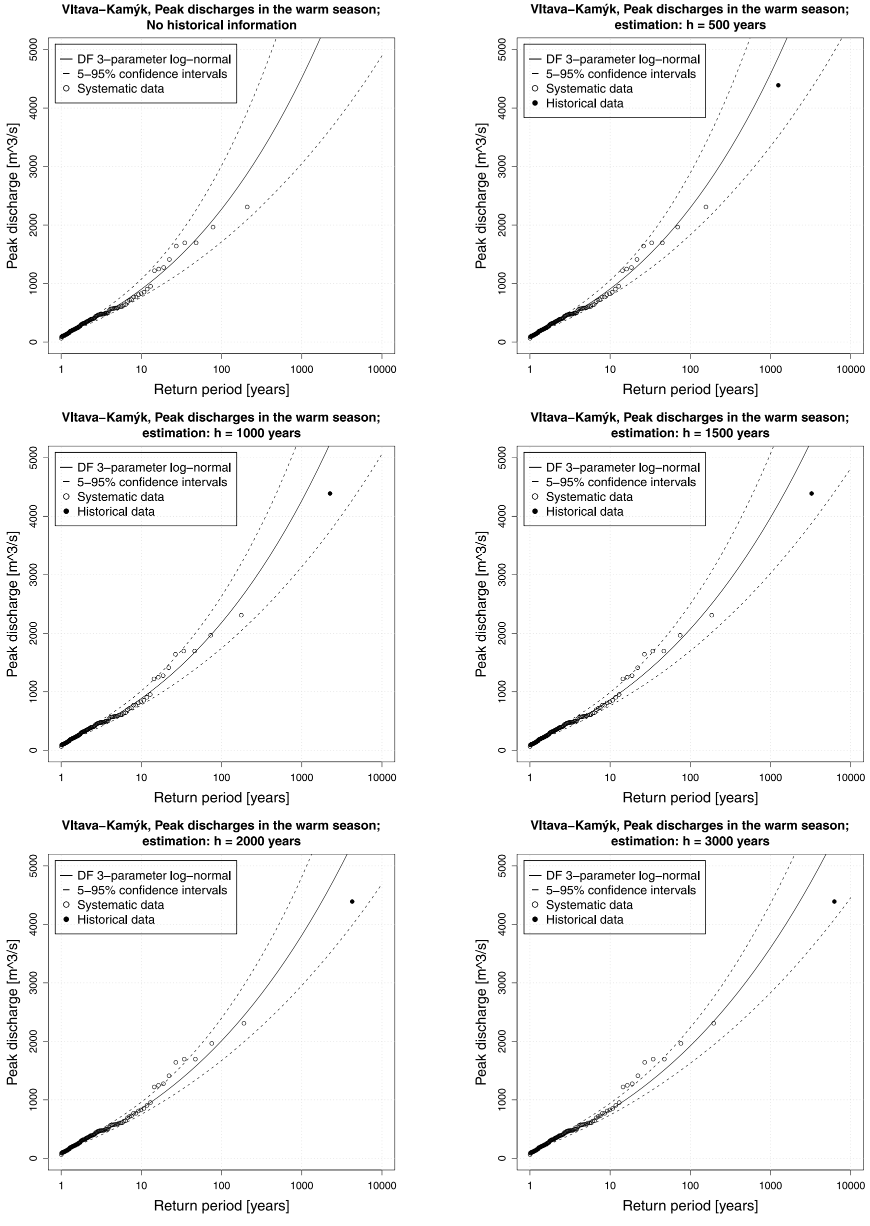

Fig. 3. Three-parameter log-normal distribution function (LN3) fitted to the warm-season maxima of peak discharges at the Vltava-Kamýk (Orlík) station: sensitivity analysis related to the selection of the length of the historical period h. DF stands for distribution function.

Table 1. Estimation of the discharge quantiles Q_T and their confidence intervals corresponding to the return periods T = 100, 1000 and 10000 years at the station Vltava-Kamýk, with no historical information involved, and various assumptions concerning the length of the historical period h, respectively. $CI_{0.05}$ ($CI_{0.95}$) is the 5% (95%) confidence limit of the estimates Q_T, $\Delta CI = CI_{0.95} - CI_{0.05}$

Return period *T* [years]	MCMC settings	Q_T [m³/s]	$CI_{0.05}$ [m³/s]	$CI_{0.95}$ [m³/s]	ΔCI [m³/s]	$\Delta CI/Q_T$ [%]
100	no hist. information	2234.6	1737.4	3069.4	1331.9	59.6
	h = 500 years	2300.0	1793.2	2901.3	1108.0	48.2
	h = 1000 years	2139.6	1760.1	2611.4	851.3	39.8
	h = 1500 years	2076.9	1711.1	2496.5	785.3	37.8
	h = 2000 years	1999.1	1677.9	2388.9	711.0	35.6
	h = 3000 years	1922.8	1639.1	2254.6	615.5	32.0
1000	no hist. information	4414.9	3115.7	6664.3	3548.5	80.4
	h = 500 years	4593.9	3241.2	6175.8	2934.6	63.9
	h = 1000 years	4165.2	3176.7	5386.6	2209.8	53.1
	h = 1500 years	3994.3	3053.5	5075.4	2021.9	50.6
	h = 2000 years	3786.6	2974.8	4758.9	1784.1	47.1
	h = 3000 years	3575.2	2875.8	4403.7	1527.9	42.7
10000	no hist. information	7761.7	5019.2	12686.7	7667.5	98.8
	h = 500 years	8150.0	5280.2	11610.9	6330.7	77.7
	h = 1000 years	7233.6	5161.4	9886.4	4725.0	65.3
	h = 1500 years	6867.0	4920.7	9202.5	4281.8	62.4
	h = 2000 years	6427.0	4745.9	8463.5	3717.6	57.8
	h = 3000 years	5973.8	4526.9	7708.2	3181.3	53.3

effects) or whether another type of distribution would be more suitable. For these reasons, as for other visual techniques, our visual inspection method could be used in the engineering practice.

Based on the visual inspection, we eliminated the two highest estimates of h (2000 and 3000 years; see two bottom plots in Fig. 3) since in these two cases, a larger number of data points (which belong to the highest flood peaks observed during the period with systematic records, and therefore, are of enhanced importance) get outside the confidence bounds. In order to find the only acceptable h value, we also fitted another distribution functions (e.g., generalized extreme value, 3-parameter log-Pearson, 2-parameter log-normal etc.) to the same data set with the same settings of the statistical model *(Szolgay et al., 2008)*. The results are not reported herein; however, taking into consideration the visual check of other distribution functions we conclude that the most acceptable estimate of the length of the historical period is about $h = 1000$–1500 years.

5.2. Selection of the perception threshold

Due to the facts that a) the magnitude of the historical plod peak is 4390 m^3/s, and b) the maximum of the rest of the flood peak records is 2309 m^3/s, we selected 6 different values for the sensitivity analysis related to the threshold X_0:

- 4300 m^3/s,
- 4000 m^3/s (i.e., the 'basic' estimate of X_0, used also in Sect. 5.1),
- 3600 m^3/s,
- 3200 m^3/s,
- 2800 m^3/s,
- 2400 m^3/s.

On the other hand, the length of the historical period is set to a constant value, $h = 1000$ years.

The graphical outputs of the analysis are presented through a panel of plots in Fig. 4 where the different threshold values are accentuated by a corresponding horizontal line. The characteristics of the selected quantiles and their CIs are summarized in Table 2.

The outcomes indicate that the gradual lowering of the perception level results in two effects: a) narrower uncertainty bounds, and b) less steep slope of the fitted distribution function. Both effects can easily be explained. Case a) means that by setting the threshold X_0 lower and lower, one assumes that the unknown past values were not particularly high, i.e., they varied in a narrower range. At the same time, case b) means that having a lower threshold X_0, the magnitudes of the unknown peak discharges that have occurred in the past are supposed to be lower, therefore, the distribution function is also fitted towards the lower values. While the first effect could, in principle, be beneficial for the quantile estimates by having narrower CIs, this theoretical advantage is outweighed by the second effect, which – similarly to the one discussed in Sect. 5.1– puts several data points outside the confidence bounds. Based on these considerations we conclude that the

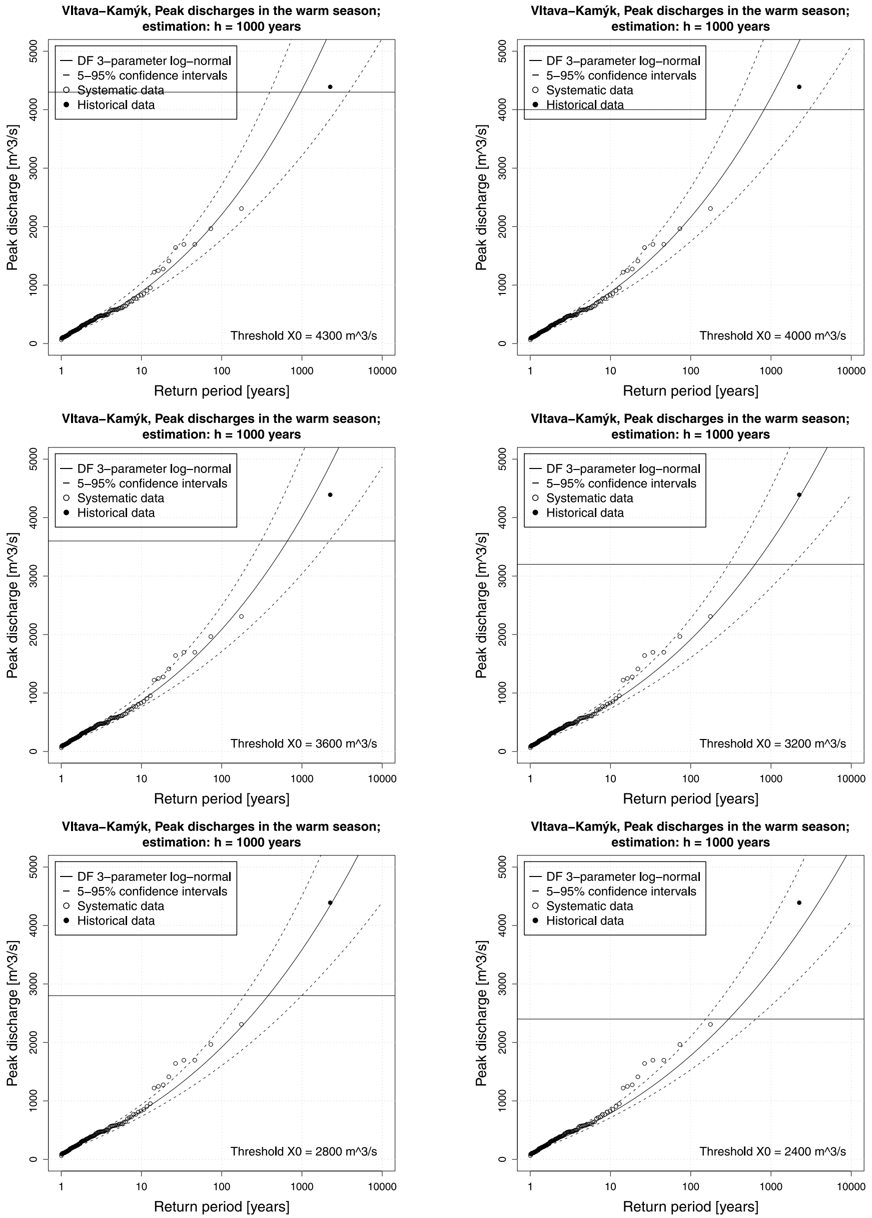

Fig. 4. Three-parameter log-normal distribution function (LN3) fitted to the warm-season maxima of peak discharges at the Vltava-Kamýk (Orlík) station: sensitivity analysis related to the selection of the perception threshold X_0. DF stands for distribution function.

Table 2. Estimation of the discharge quantiles Q_T and their confidence intervals corresponding to the return periods T = 100, 1000 and 10000 years at the station Vltava-Kamýk, with various assumptions concerning the perception threshold X_0. $CI_{0.05}$ ($CI_{0.95}$) is the 5% (95%) confidence limit of the estimates Q_T, $\Delta CI = CI_{0.95} - CI_{0.05}$

Return period T [years]	MCMC settings	Q_T [m³/s]	$CI_{0.05}$ [m³/s]	$CI_{0.95}$ [m³/s]	ΔCI [m³/s]	$\Delta CI/Q_T$ [%]
100	X_0 = 4300 m³/s	2186.7	1757.2	2696.1	938.9	42.9
	X_0 = 4000 m³/s	2150.5	1761.8	2651.2	889.4	41.4
	X_0 = 3600 m³/s	2092.2	1695.8	2516.7	820.9	39.2
	X_0 = 3200 m³/s	2008.4	1666.6	2391.6	725.0	36.1
	X_0 = 2800 m³/s	1912.1	1615.7	2259.2	643.5	33.7
	X_0 = 2400 m³/s	1790.5	1526.7	2096.3	569.5	31.8
1000	X_0 = 4300 m³/s	4284.8	3148.3	5602.3	2454.1	57.3
	X_0 = 4000 m³/s	4187.1	3184.4	5500.5	2316.1	55.3
	X_0 = 3600 m³/s	4031.5	3012.9	5136.1	2123.2	52.7
	X_0 = 3200 m³/s	3818.2	2934.3	4785.9	1851.6	48.5
	X_0 = 2800 m³/s	3582.1	2834.6	4482.8	1648.3	46.0
	X_0 = 2400 m³/s	3270.1	2607.2	4032.6	1425.4	43.6
10000	X_0 = 4300 m³/s	7481.5	5098.7	10300.4	5201.7	69.5
	X_0 = 4000 m³/s	7271.5	5173.3	10146.0	4972.7	68.4
	X_0 = 3600 m³/s	6941.8	4822.0	9308.9	4486.9	64.6
	X_0 = 3200 m³/s	6501.3	4663.6	8578.5	3914.9	60.2
	X_0 = 2800 m³/s	6024.6	4473.2	7939.5	3466.3	57.5
	X_0 = 2400 m³/s	5383.6	4033.2	6968.4	2935.2	54.5

initial selection of the perception threshold ($X_0 = 4000$ m^3/s) is acceptable for the recent analysis. It also can be higher; however, it definitely should not be set considerably lower.

5.3. The effect of the record length on the results

In the report by *Szolgay et al. (2008)*, the collective of authors raised concerns about the appropriateness of the flood peak measurements made in the last decades of the 19$^{\text{th}}$ century (possible inhomogeneities present in the data series; details not reported herein). Due to this fact we decided to exclude the problematic part of the data series and restrict the analysis to the supposedly homogeneous data set belonging to the period 1900–2002. The results of this analysis are presented in Fig. 5 and Table 3.

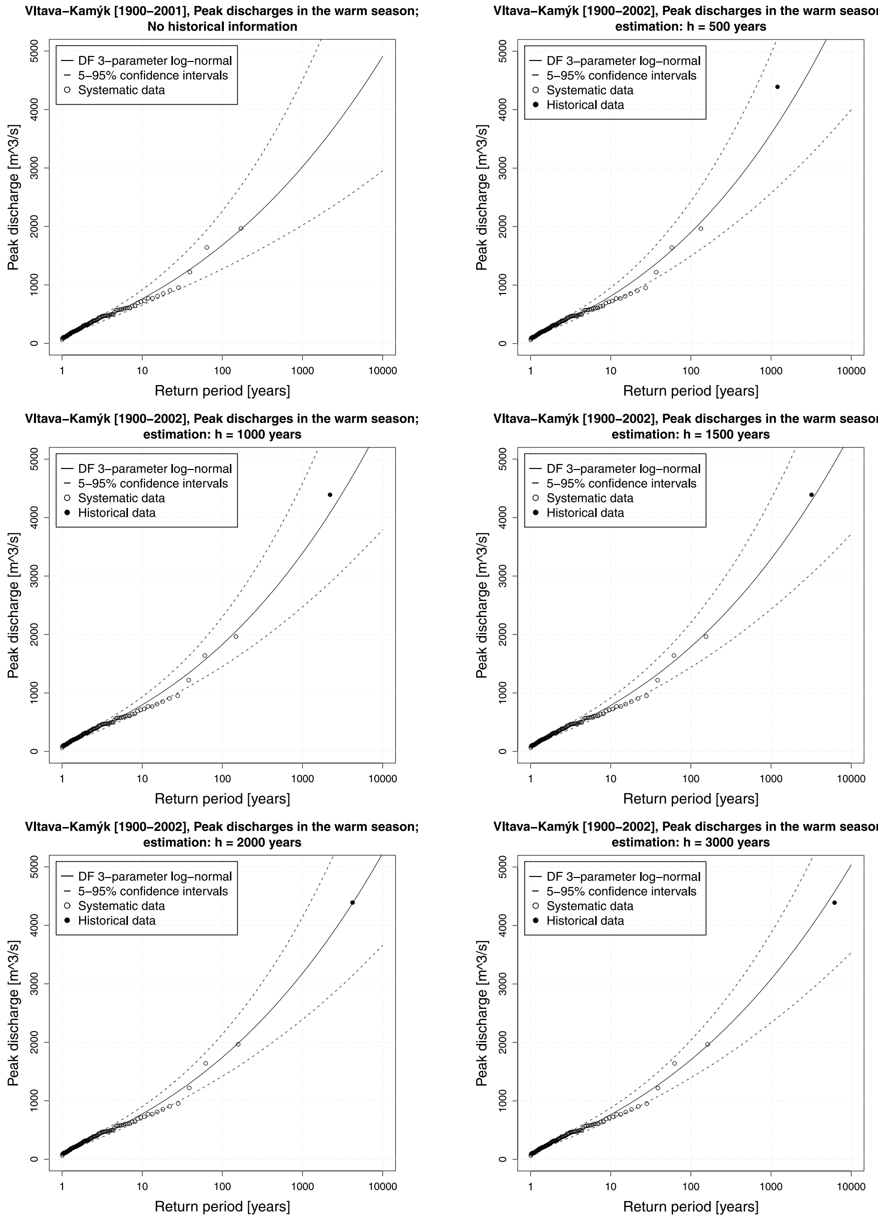

Fig. 5. Three-parameter log-normal distribution function (LN3) fitted to the warm-season maxima of peak discharges at the Vltava-Kamýk (Orlík) station: analysis based on a shorter period of systematic observations s (1900–2002). DF stands for distribution function.

Table 3. Estimation of the discharge quantiles Q_T and their confidence intervals corresponding to the return periods T = 100, 1000 and 10000 years at the station Vltava-Kamýk, based on a shorter based on a shorter period of systematic observations s (1900–2002). $CI_{0.05}$ ($CI_{0.95}$) is the 5% (95%) confidence limit of the estimates Q_T, $\Delta CI = CI_{0.95} - CI_{0.05}$

Return period *T* [years]	MCMC settings	Q_T [m³/s]	$CI_{0.05}$ [m³/s]	$CI_{0.95}$ [m³/s]	ΔCI [m³/s]	$\Delta CI/Q_T$ [%]
100	no hist. information	1649.3	1290.6	2339.2	1048.6	63.6
	h = 500 years	1896.6	1484.3	2451.0	966.6	51.0
	h = 1000 years	1830.0	1462.3	2309.2	846.8	46.3
	h = 1500 years	1797.5	1440.1	2196.5	756.4	42.1
	h = 2000 years	1757.8	1423.5	2114.5	691.0	39.3
	h = 3000 years	1703.6	1381.8	2040.5	658.6	38.7
1000	no hist. information	2951.4	2063.1	4718.4	2655.3	90.0
	h = 500 years	3579.6	2547.3	5031.9	2484.6	69.4
	h = 1000 years	3394.9	2499.8	4613.1	2113.3	62.3
	h = 1500 years	3314.7	2447.3	4314.8	1867.5	56.3
	h = 2000 years	3210.2	2395.4	4101.4	1706.0	53.1
	h = 3000 years	3080.9	2303.4	3887.8	1584.4	51.4
10000	no hist. information	4777.7	3031.0	8510.1	5479.1	114.7
	h = 500 years	6059.6	3951.5	9120.8	5169.3	85.3
	h = 1000 years	5663.9	3858.5	8211.8	4353.3	76.9
	h = 1500 years	5502.2	3752.1	7577.9	3825.7	69.5
	h = 2000 years	5286.0	3666.0	7120.9	3454.9	65.4
	h = 3000 years	5031.1	3486.6	6675.8	3189.2	63.4

Keeping the settings of the statistical model applied so far, and having a shorter period of systematic observations, the MCMC simulations lead to lower quantile estimates compared to the results based on the whole data records (Sect. 5.1, Table 1). It is likely that this is only a sampling effect since the eliminated period 1877–1899 contains a number of flood events of a relatively high magnitude. On the other hand, the effect of the shorter period s on the uncertainty of the estimated quantiles is unclear (Table 3). Lower quantile estimates Q_T are accompanied with narrower absolute width of the confidence intervals ΔCI; however, these effects result in enhanced relative width of the CIs $\Delta CI/Q_T$ (cf. Table 3 and Table 1).

At the present stage of the analysis it is evident that the two parameters h and X_0 are the most dominant controls of the width of the CIs. In order to derive further information on the role of s in the statistical model, a deeper analysis is needed (possibly accompanied by further Monte Carlo simulations), and such a task is beyond the scope of the recent study.

6. Summary and conclusions

In this paper, a method how to incorporate historical floods into the at-site flood frequency analysis has been reviewed. It is based on Bayesian inference where a likelihood function is built to properly handle the information on historical floods.

Despite of the fact that subjective choices cannot be avoided when applying the method and that some of these (which are necessary to conduct the computations) may be questionable, the method is transparent to the users. Most of the subjective choices and hypotheses are explicitly formulated, have hydrological meaning (e.g., m historical peak discharges the number of years h in which these m events were the major floods and the threshold X_0 which has certainly not been exceeded during this period by the other floods) and their effect can be evaluated, discussed and modified based on sensitivity analyses, which can be conducted to test the influence of these hypotheses on the results, as it was shown in this paper.

As the particular results of the case study indicate, among the number of parameters of the statistical model to be chosen, the length of the historical period h and the perception threshold X_0 have the most remarkable influence on the width of the confidence intervals of the estimated flood quantiles. The higher the assumption of h (i.e., analysis reaches further back to the past) and/or the lower the assumption of X_0 (i.e., the unknown peak discharges from the past are assumed to be generally low), the narrower are the confidence bounds of the estimated quantiles. Nevertheless, selecting high values of h and/or low values of X_0 for the model yields a great risk that the analyst may be wrong of not considering all the (unknown) extreme events that might have occurred during the period of length h in the past. It is therefore advised to be rather prudent in selecting the model parameters to have confidence intervals only moderately narrower compared to the alternative with no historical information, instead of having considerably narrower confidence intervals but making wrong inference on “unknown” historical data that are not well estimated.

A particularly satisfying result of this case study was that the outcomes of the method are relatively robust when considering the uncertainties in the estimated historical flood discharges and the subjective choices in the parameterizations of the method.

In a similar study *(Gaume et al., 2010)*, the idea of inclusion of extraordinary flood events into frequency analysis is further discussed: based on the analogy with a temporal extension of at-site data with historical flood extremes, a methodology for a spatial extension of regionally pooled data with flood extremes observed in ungauged catchments is developed. *Gaume et al. (2010)* emphasizes that the choice of the parameters h and X_0 is crucial also in that case.

Acknowledgments. The study was supported by the Slovak Grant Agency VEGA under Project No. 1/0103/10 and 1/0496/08, and the Austrian-Slovak Co-operation in Science and Education "Aktion".

References

Acreman M. C., Wiltshire S. E., 1989: The regions are dead; long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis. IAHS Publ. 187, 175–188.

Alcoverro J., Corominas J., Gómez M., 1999: The Barranco de Arás flood of 7 August 1996 (Biescas, Central Pyrenees, Spain). Engineering Geology, 51, 237–255.

Bayliss A. C., Reed D. W., 2001: The use of historical data in flood frequency estimation. Report to MAFF. CEH Wallingford, 87 p. Available at: `http://nora.nerc.ac.uk/8060/1/BaylissRepN008060CR.pdf`, accessed: 15-04-2010.

Beable M. E., McKerchar A. I., 1982: Regional flood estimation in New Zealand. In: Water and Soil Division Technical Publication No. 20, Water and Soil Division, Ministry of Works and Development, Wellington, New Zealand.

Bobée B., Rasmussen P. F., 1994: Statistical analysis of annual flood series. In: Menon J. (Ed.): Trends in Hydrology (1). Council of Scientific Research Integration, India, 117–135.

Boháč M., Kulasová B., 2005: Hydrologická studie pro VD Orlík: Průběhy teoretických povodňových vln s kulminačním průtokem s pravděpodobnosti překročení pQ = 0.0001 a s podmíněnými pravděpodobnostmi překročení objemu. Czech Hydrometeorological Institute. 14 p. (in Czech).

Brunovský P., Lapin M., Melicherčík I., Somorčík J., Ševčovič D., 2009: Risks due to variability of K-day extreme precipitation totals and other K-day extreme events. Journal of Hydrology and Hydromechanics, **57**, 4, 250–263.

Burn D. H., 1990a: An appraisal of the 'region of influence' approach to flood frequency analysis. Hydrological Sciences Journal, **35**, 2, 149–165.

Burn D. H., 1990b: Evaluation of regional flood frequency analysis with a region of influence approach. Water Resources Research, **26**, 10, 2257–2265.

Burn D. H., 1997: Catchment similarity for regional flood frequency analysis using seasonality measures. Journal of Hydrology, 202, 212–230.

Burn D. H., Goel N. K., 2000: Formation of groups for regional flood frequency analysis. Hydrological Sciences Journal, 45, 97–112.

Castellarin A., 2007: Probabilistic envelope curves for design flood estimation at ungauged sites. Water Resources Research, 43, W04406, doi: 10.1029/2005WR004384.

Coles S., 2001: An introduction to statistical modeling of extreme values. Springer Verlag, London, 208 p.

Costa J. E., 1987: A comparison of the largest rainfall-runoff floods in the United States with those of the People's Republic of China and the World. Journal of Hydrology, 69, 101–115.

Costa J. E., Jarrett R. D., 2008: An evaluation of selected extraordinary floods in the United States reported by the U.S. Geological Survey and implications for future advancement of flood science. Scientific Investigations Report 2008–5164, U.S. Geological Survey, Reston, Virginia, USA.

Cunnane C., 1987: Review of statistical models for flood frequency estimation. In: Singh V. P. (Ed.): Hydrologic frequency modelling. Reidel, Dordrecht, The Netherlands, 49–95.

Dalrymple T., 1960: Flood frequency analyses. Water Supply Paper 1543-A. U.S. Geological Survey, Reston,Virginia, USA.

Drbal K., Hlavínek J., Ošlejšková J., 2003: Evaluation of catastrophic flood in August 2002 - Analysis of effect of waterworks on the flood. (Vyhodnocení katastrofální povodně v srpnu 2002 – Posouzení vlyvu vodních děl na průběh povodně). Research Report. VÚV T.G. Masaryka, 2003, 79 p. (in Czech).

Gaume E., Bain V., Bernardara P., Newinger O., Barbuc M., Bateman A., Blaškovičová L., Blöschl G., Borga M., Dumitrescu A., Daliakopoulos I., Garcia J., Irimescu A., Kohnová S., Koutroulis A., Marchi L., Matreata S., Medina V., Preciso E., Sempere-Torres D., Stancalie G., Szolgay J., Tsanis J., Velasco D., Viglione A., 2009: A collation of data on European flash floods. Journal of Hydrology, 367, 70–78 (doi: 10.1016/j.jhydrol.2008.12.028).

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

Gelman A., Carlin J. B., Stern H. S., Rubin D. B., 2004: Bayesian data analysis. Chapman & Hall / CRC, 696 p.

Guse B., Thieken A. H., Castellarin A., Merz B., 2010: Deriving probabilistic regional envelope curves with two pooling methods. Journal of Hydrology, 380, 14–26.

Herschy R. W., 2005: The world's maximum observed floods. Flow Measurement and Instrumentation, 13, 231–235.

Hosking J. R. M., Wallis J. R., 1986: Paleoflood hydrology and flood frequency analysis. Water Resources Research, **22**, 4, 543–550.

Hosking J. R. M., Wallis J. R., 1997: Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge, 224 p.

Institute of Hydrology, 1999: Flood Estimation Handbook, Part 3: Statistical procedures for flood frequency estimation. Institute of Hydrology, Wallingford, UK.

Kuczera G., 1999: Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. Water Resources Research, **35**, 5, 1551–1557.

Kundzewicz Z. W., Ulbrich U., Brucher T., Graczyk D., Kruger A., Leckebusch G. C., Menzel L., Pinskwar I., Radziejewski M., Szwed M., 2005: Summer floods in central Europe – Climate change track? Natural Hazards, 36, 165–189.

Leese M. N., 1973: Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series. Water Resources Research, **9**, 6, 1534–1542.

Meigh J. R., Farquharson F. A. K., Sutcliffe J. V., 1997: A world-wide comparison of regional flood estimation methods and climate. Hydrological Sciences Journal, **42**, 2, 225–244.

Merz R., Blöschl G., 2003: A process typology of regional floods. Water Resources Research, **39**, 12, 1340 (`doi:10.1029/2002WR001952`).

Merz R., Blöschl G., 2008a: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. Water Resources Research, 44, W08432 (`doi:10.1029/2007WR006744`).

Merz R., Blöschl G., 2008b: Flood frequency hydrology: 2. Combining data evidence. Water Resources Research, 44, W08433 (`doi:10.1029/2007WR006745`).

Mimikou M., 1984: Envelope curves for extreme flood events in north-western and western Greece. Journal of Hydrology, 67, 55–66.

NERC (Natural Environment Research Council), 1975: Flood Studies Report, Vol. I.: Hydrological Studies. NERC, London, UK.

Ouarda T. B. M. J., Girard C., Cavadias G. S., Bobée B., 2001: Regional flood frequency estimation with canonical correlation analysis. Journal of Hydrology, **254**, 1-4, 157–173.

Parent E., Bernier J., 2002: Bayesian POT modeling for historical data. Journal of Hydrology, 274, 95–108.

Pekárová P., 2009: Flood regime of rivers in the Danube River basin (Režim povodní v povodí rieky Dunaj). Journal of Hydrology and Hydromechanics, **57**, 2, 142–150 (in Slovak).

Podolinská J., Šipikalová H., Škoda P., Blaškovičová L., Demeterová B., 2005: N-year maximum discharges on the Slovakia rivers. Conference 'Hydrological days', Bratislava.

Rao A. R., Hamed K. H., 1999: Flood frequency analysis. CRC Press, Boca Raton, Florida, USA, 350 p.

Reis D. S., Stedinger J. R., 2005: Bayesian MCMC flood frequency analysis with historical information. Journal of Hydrology, **313**, 1-2, 97–116.

Robert C. P., Casella G., 2004: Monte Carlo statistical methods. Springer, New York, 2nd edition, 645 p.

Rodier J. A., Roche M., 1984: World catalogue of maximum observed floods. IASH Publication 143, IASH Press.

Solín L., 2002: Identification of physical regional types for regional flood frequency analysis. In: Weingartner R., Sperafico M. (Eds.): Proceedings of international conference on flood estimation, CHR Report II-17, Bern, Switzerland, 687–697.

Solín L., 2008: Analysis of floods occurrence in Slovakia in the period 1996–2006 (Analýza výskytu povodňových situácií na Slovensku v období rokov 1996–2006). Journal of Hydrology and Hydromechanics, 56, 2, 95–107 (in Slovak).

Stedinger J. R., Cohn T. A., 1986: Flood frequency analysis with historical and paleoflood information. Water Resources Research, **22**, 5, 785–793.

Svoboda A., Pekárová P., 1998: The catastrophic flood of July 1998 in the Malá Svinka catchment – its simulation (Katastrofálna povodeň z júla 1998 v povodí Malej Svinky - simulácia jej priebehu). Journal of Hydrology and Hydromechanics, **46**, 6, 356–365 (in Slovak).

Szolgay J., Kohnová S., Hlavčová K., Gaál L., Bacigál T., 2008: Assessment of hydrological design values for the hydro-electric power plant Orlík [Final Report] (Posúdenie hydrologických návrhových veličín vodného diela Orlík [Záverečná správa]). Department of Land and Water Resources Management, Faculty of Civil Engineering Slovak University of Technology, Bratislava, 126 p. (in Slovak).

Tierney L., 1994: Markov Chains for exploring posterior distributions. The Annals of Statistics, **22**, 4, 1701–1728.

Ulbrich U., Brücher T., Fink A. H., Leckebusch G. C., Krüger A., Pinto J. G., 2003: The central European floods of August 2002. Part I: Rainfall periods and flood development. Weather, 58, 371–377.

UNESCO, 1976: World catalogue of very large floods. The UNESCO Press, 424 p.

Viglione A., 2009. nsRFA: Non-supervised regional frequency analysis. R package. Version 0.6-9. Available at (`http:cran.r-project.org/web/packages/nsRFA`), published 19-08-2009, accessed 15-04-2010.

Wilks D., 1995: Statistical methods in the atmospheric sciences. Academic Press, San Diego, 467 p.

Wiltshire S. E., 1986: Identification of homogeneous regions for flood frequency analysis. Journal of Hydrology, 84, 287–302.