WATER RESOURCES RESEARCH, VOL. 49, 675–692, doi:10.1029/2011WR010782, 2013

Flood frequency hydrology: 3. A Bayesian analysis

Alberto Viglione,[1] Ralf Merz,[2] José Luis Salinas,[1] and Günter Blöschl[1]

Received 13 April 2011; revised 24 November 2012; accepted 27 November 2012; published 4 February 2013.

[1] Merz and Blöschl (2008a, 2008b) proposed the concept of flood frequency hydrology, which highlights the importance of combining local flood data with additional types of information: temporal information on historic floods, spatial information on floods in neighboring catchments, and causal information on the flood processes. Although most of the previous studies combined flood data with only one extra type of information, all three types are used here in a Bayesian analysis. To illustrate ways to combine the additional information and to assess its value, flood frequency analyses before and after the extraordinary 2002 flood event are compared for the 622 km^2 Kamp river in northern Austria. Although this outlier significantly affects the flood frequency estimates if only local flood data are used (60% difference for the 100 year flood), the effect is much reduced if all additional information is used (only 3% difference). The Bayesian analysis also shows that the estimated uncertainty is significantly reduced when more information is used (for the 100 year return period, the 90% credible intervals range reduces from 140% to 31% of the corresponding flood peak estimate). Further analyses show that the sensitivity of the flood estimates to the assumptions made on one piece of information is small when all pieces of information are considered together. While expanding information beyond the systematic flood record is sometimes considered of little value in engineering hydrology because subjective assumptions are involved, the results of this study suggest that the extra information (temporal, spatial, and causal) may outweigh the uncertainty caused by these assumptions.

Citation: Viglione, A., R. Merz, J. L. Salinas, and G. Blöschl (2013), Flood frequency hydrology: 3. A Bayesian analysis, *Water Resour. Res.*, *49*, doi:10.1029/2011WR010782.

1. Introduction

[2] The concept of flood frequency hydrology [*Merz and Blöschl*, 2008a, 2008b] highlights the importance of using a maximum of hydrologic information from different sources and a combination based on hydrological reasoning. In their framework, *Merz and Blöschl* [2008a, 2008b] propose to compile flood peaks at the site of interest plus three additional types of information: temporal, spatial, and causal information.

[3] Temporal information expansion is directed toward collecting information on the flood behavior before (or after) the period of discharge observations (systematic data period). Spatial information expansion is based on using flood information from neighboring catchments to improve flood frequency estimates at the site of interest. Causal information expansion analyzes the generating mechanisms of floods in the catchment of interest. For each of these types of information expansion, methods have been proposed in the literature. Formal methods exist on combining historical flood data (from flood marks and archives) and possibly paleofloods with available flood records [e.g., *Leese*, 1973; *Stedinger and Cohn*, 1986; *Cohn et al.*, 1997; *O'Connell et al.*, 2002; *England et al.*, 2003; *Reis and Stedinger*, 2005; *Benito and Thorndycraft*, 2005], which would be considered temporal information expansion. Methods of regional flood frequency analysis [e.g., *Dalrymple*, 1960; *Cunnane*, 1988; *Tasker and Stedinger*, 1989; *Bobée and Rasmussen*, 1995; *Hosking and Wallis*, 1997; *Merz and Blöschl*, 2005] would be considered spatial information expansion. Finally, the derived flood frequency approach [e.g., *Eagleson*, 1972; *Kurothe et al.*, 1997; *Fiorentino and Iacobellis*, 2001; *Sivapalan et al.*, 2005] or, more generally, rainfall-runoff modeling [e.g., *Pilgrim and Cordery*, 1993; *Wagener et al.*, 2004] would be considered causal information expansion.

[4] As discussed in *Merz and Blöschl* [2008b], it is vital to account for the respective uncertainties of the various pieces of information when combining them. In local flood statistics, a range of estimates may result from a reasonable fit of several distributions to the observed data or by accounting for the uncertainty associated with the estimated parameters of those distributions. Historical flood data may only allow us to give a range of estimates owing to large uncertainties. Spatial information may lead to a range of estimates when using several regionalization

[1]Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria.
[2]Department for Catchment Hydrology, Helmholtz Centre for Environmental Research, Halle, Germany.

Corresponding author: A. Viglione, Institut für Wasserbau und Ingenieurhydrologie, Technische Universität Wien, Karlsplatz 13/E222, 1040 Vienna, Austria. (viglione@hydro.tuwien.ac.at)

©2012. American Geophysical Union. All Rights Reserved.
0043-1397/13/2011WR010782

schemes or parameters of the regionalization schemes all of which may be consistent with the regional information. Causal information may result in a range of estimates due to using different methods, different data, and uncertainty in the expert judgment. In *Merz and Blöschl* [2008b], the final estimate was obtained by expert judgment, considering the relative uncertainties of the component sources of information [see also *Gutknecht et al*., 2006].

[5] The present paper is a follow up of *Merz and Blöschl* [2008a, 2008b] where, instead of reasoning in terms of ranges of estimates, we account for the uncertainty inherent to the different sources of information using the Bayesian framework. Bayesian methods provide a computationally convenient way to fit frequency distributions for flood frequency analysis by using different sources of information as systematic flood records, historical floods, regional information, and other hydrologic information along with related uncertainties (e.g., measurement errors). They also provide an attractive and straightforward way to estimate the uncertainty in parameters and quantiles metrics.

[6] In flood hydrology, Bayesian methods have been used often for pieces of information such as historic floods [e.g., *Stedinger and Cohn*, 1986; *O'Connell et al*., 2002; *Parent and Bernier*, 2003a; *Reis and Stedinger*, 2005; *Neppel et al*., 2010; *Payrastre et al*., 2011], regional information [e.g., *Wood and Rodriguez-Iturbe*, 1975; *Kuczera*, 1982, 1983; *Madsen and Rosbjerg*, 1997; *Fill and Stedinger*, 1998; *Seidou et al*., 2006; *Ribatet et al*., 2007; *Micevski and Kuczera*, 2009; *Gaume et al*., 2010], and less frequently for other information such as, for example, expert opinion [e.g., *Kirnbauer et al*., 1987; *Parent and Bernier*, 2003b]. In the literature, there are examples in which more than one piece of additional information was used in a Bayesian analysis. For example, *Vicens et al*. [1975] investigate information expansion from regional information (through regression models) or expert judgment (causal information from precipitation characteristics) but do not combine the two together. *Martins and Stedinger* [2001] use historical information jointly with the generalized maximum likelihood method, which can be thought as regional (or expert) information expansion. The aim of this paper is to illustrate by example how all three pieces of information (Figure 1, top row) can be combined in a Bayesian analysis and to assess the sensitivity of the final flood estimate to the assumptions involved. Obviously, there may be applications, where not all three types of information expansion (temporal, spatial, and causal) can be provided, e.g., no historic flood data and/or regional studies are available; so, we test the effect of each piece of information on the flood estimates separately.

[7] The sensitivity of the flood estimate to the flood peak sample at hand is also assessed by comparing two cases in a study catchment where a very large flood has occurred. In the first case, we assume that only the information before the big flood is available to mimic the situations where no large floods have been observed but may occur. In the second case, we include the large flood.

[8] The information expansion used in this study is not only diverse in terms of the temporal, spatial, and causal character of the additional information but also in the qualitative character of the information (Figure 1, bottom row): (1) additional data, (2) full information or, and (3) partial information on the prior distribution of parameters of the selected statistical model. In the example presented hereafter, historical floods are used as additional data; regional information provides an estimate of the full distribution of the model parameters, while the estimate with uncertainty of one flood peak quantile, obtained through expert judgment from a rainfall-runoff modeler, constitutes partial information on the model parameters. This may differ in other applications, where, e.g., spatial information expansion could provide only partial information while causal information expansion could provide full information on the prior distribution of the parameters of the selected statistical model.

2. Kamp at Zwettl

[9] The Kamp river at Zwettl is located in northern Austria and has a catchment area of 622 km^2. For the Kamp at Zwettl, annual flood peak data from 1951 to 2005 are available (Figure 2). The statistical analyses of the flood peaks are dominated by the extreme flood event in August 2002, which interested a large portion of Central Europe [see, e.g., *Choryński et al*., 2012]. Due to that flood, the Kamp catchment has been extensively studied in the last years [e.g., *Gutknecht et al*., 2002; *Komma et al*., 2007; *Blöschl et al*., 2008; *Reszler et al*., 2008; *Viglione et al*., 2010]. In August 2002, a Vb-cyclone [*Mudelsee et al*., 2004] carried

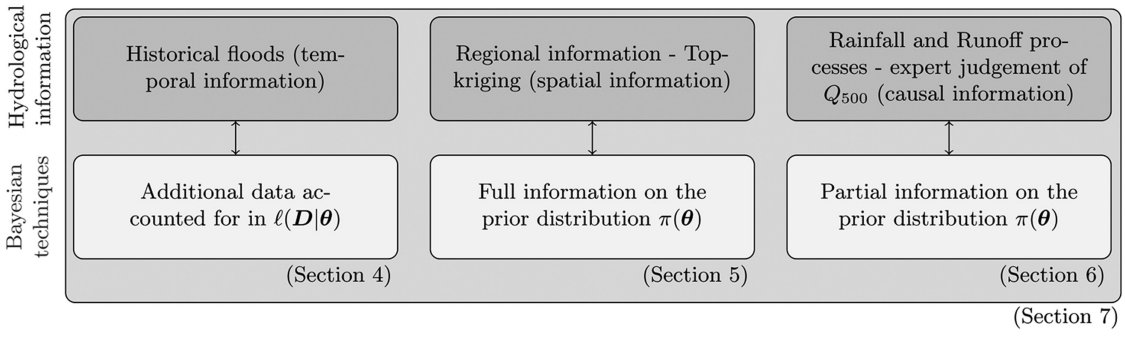

Figure 1. Use of information additional to the maximum annual flood peak systematic record. (top row) The hydrological pieces of information and (bottom row) the way they are used in the Bayesian framework. The terms $\ell(\mathbf{D}|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$ are the likelihood function and the prior distribution of the parameters of the selected statistical model, respectively, as in the Bayes' theorem (equation (1) in section 3). In brackets are indicated the sections of the paper where the types of information are discussed.

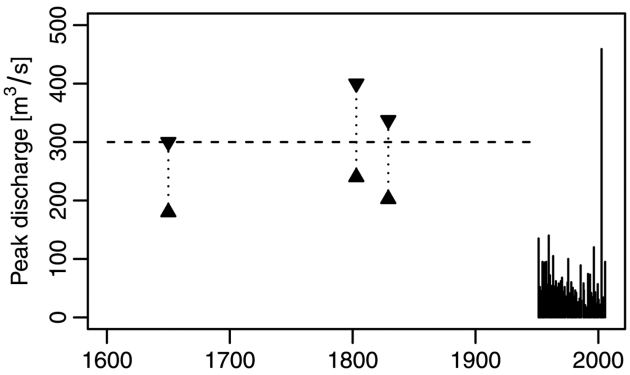

Figure 2. Time series of maximum annual peak discharges of the Kamp at Zwettl (622 km^2). Ranges for three historical floods and the threshold not exceeded (300 m^3/s) in the historic period (1600–1950), at least no more than in these three cases, are also shown.

warm moist air from the Adriatic region and caused persistent rainfall over the Kamp region. This resulted in an estimated peak flow of 460 m^3/s, which is three times the second largest flood in the 55 year record. In a humid climate such as Austria, this is an extraordinary event. Owing to the extreme event in 2002, the systematic sample mean annual flood (MAF), its coefficient of variation (CV), and skewness (CS) are 63 m^3/s, 0.98, and 5.21, respectively. A generalized extreme value (GEV) distribution, fitted by the method of L-moments, gives a 100 year flood runoff (Q_{100}) of 285 m^3/s. When extrapolating this flood frequency curve to large return periods, one would assign a return period of 340 years to the 2002 event.

[10] The 2002 flood departs significantly from the other events in the record and indeed statistical tests, such as threshold analyses [e.g., *Stedinger et al.*, 1993], identify the 2002 flood event as an outlier. Without the extreme event in 2002 (considering the record 1951–2001), the sample MAF, CV, and CS are 57 m^3/s, 0.51, and 1.14, respectively. In particular, the skewness is much smaller than that from the sample including the 2002 flood. The Q_{100} from the sample without the 2002 flood is 159 m^3/s (GEV distribution fitted with the method of L-moments). When extrapolating this flood frequency curve to large return periods, one would assign a return period greater than 100,000 years to the 2002 event. If one takes the samples (either with or without the 2002 flood) at face value, one assumes that they are representative of the population of extreme events. When including the 2002 flood, the statistically estimated return period of such an extreme event decreases dramatically, which implies that such extreme events occur regularly, while when excluding the event, the statistically estimated return period of such an extreme event is very high, which implies that such events occur very rarely. It is a similar situation as the one described more than 80 years ago by *Hazen* [1930], who investigated the effect of one large flood (June 1921 on the Arkansas River) in a short record.

[11] Of course, the extraordinary event of 2002 has to be included in the analysis because it reveals how extreme the floods can be in the Kamp catchment [*Laio et al.*, 2010]. From a practical perspective, one is particularly interested in cases where such an extraordinary flood has not yet occurred, although may occur in the future. We, therefore, compare two cases hereafter. In the first case, we assume that only the information until the end of 2001 is available, i.e., the situation of particular relevance for engineering design. In the second case, we include information until the end of 2005. The comparison is to assess how well a flood of the magnitude of the 2002 event could have been anticipated statistically prior to the occurrence of that event.

3. Bayesian Inference Using Systematic Data Only

[12] In flood frequency analysis, Bayesian inference is a method in which the Bayes' theorem is used to combine the information provided by the locally observed flood data with additional information independent from those data. For a flood frequency distribution with parameters $\boldsymbol{\theta}$, the Bayes' theorem states that

$$p(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{\ell(\boldsymbol{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Omega}} \ell(\boldsymbol{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto \ell(\boldsymbol{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{1}$$

where $p(\boldsymbol{\theta}|\boldsymbol{D})$ is the posterior distribution of the parameters $\boldsymbol{\theta}$, after having observed the data $\boldsymbol{D}$; $\ell(\boldsymbol{D}|\boldsymbol{\theta})$ is the likelihood function, i.e., the probability density function (pdf) of the data conditional on the parameters; $\pi(\boldsymbol{\theta})$ is the prior distribution of the parameters, which can be formulated from additional information that does not take into account any information contained in the observed data $\boldsymbol{D}$; and the integral in the denominator of equation (1), computed on the whole parameter space $\boldsymbol{\Omega}$, serves as a normalization constant to obtain a unit area under the posterior pdf $p(\boldsymbol{\theta}|\boldsymbol{D})$. In Bayesian inference, the parameters $\boldsymbol{\theta}$ are considered as random variables and the uncertainty associated with them can be explicitly modeled, thus allowing to assign credible intervals (which are the Bayesian analogs to the confidence intervals in frequentist statistics) to the estimated flood quantiles. These credible intervals reflect user perception instead of a frequentist assessment of the probability of the true value to fall between them [see, e.g., *Montanari et al.*, 2009].

[13] Since, in most of the cases, the integral in the denominator of equation (1) cannot be processed in closed form, simulation-based Monte Carlo techniques such as the Markov chain Monte Carlo (MCMC) approaches are used. MCMC methods (which include random walk Monte Carlo methods) are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution (in our case, the posterior probability model) as its equilibrium distribution [see, e.g., *Robert and Casella*, 2004; *Gelman et al.*, 2004]. The states of the chain after a large number of steps are then used as a sample from the desired distribution (the quality of the sample improves as a function of the number of steps). Several MCMC algorithms have been used in flood hydrology [e.g., *Kuczera*, 1999; *Reis and Stedinger*, 2005; *Ribatet et al.*, 2007]. We use the Metropolis-Hastings algorithm [*Chib and Greenberg*, 1995].

[14] When only systematic data are used, we write the likelihood function in equation (1) as

$$\ell_S(\boldsymbol{D}|\boldsymbol{\theta}) = \prod_{i=1}^{s} f_X(x_i|\boldsymbol{\theta}), \tag{2}$$

where $\boldsymbol{D}$ is the sample of annual discharge maxima systematically recorded $x_1, x_2, \ldots, x_s$ (in our case, $s = 55$ years) and $f_X(.)$ is the pdf of the variable X (representing the maximum annual peak discharges). We assume that the peaks x_i are known exactly but equation (2) could be generalized to account for uncertainties (random and independent) and/or systematic errors due, for example, to the construction of the rating curves [see, e.g., *Kuczera*, 1992, 1996; *O'Connell et al.*, 2002; *Reis and Stedinger*, 2005; *Neppel et al.*, 2010]. For the distribution function f_X, we assume the GEV distribution with parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$:

$$f_X(x|\boldsymbol{\theta}) = \frac{1}{\theta_2}\left[1 - \frac{\theta_3(x-\theta_1)}{\theta_2}\right]^{1/\theta_3 - 1} \times \exp\left\{-\left[1 - \frac{\theta_3(x-\theta_1)}{\theta_2}\right]^{1/\theta_3}\right\}, \tag{3}$$

where θ_1 denotes the location parameter, θ_2 the scale parameter, and θ_3 the shape parameter [e.g., *Grimaldi et al.*, 2011, p. 489]. An improper flat prior on the parameters is used in equation (1) that corresponds to no information other than the data, i.e., $\forall\boldsymbol{\theta}\ \pi(\boldsymbol{\theta}) = 1$. Note that other noninformative priors could be used [e.g., *Reis and Stedinger*, 2005] and that, in practice, general priors like the geophysical prior proposed by *Martins and Stedinger* [2000] (GML method) are often employed. In this section, we do not use the GML method, which restrict the shape parameter of the GEV distribution to a statistically/physically reasonable range, because it can be considered as a regional or expert information expansion.

[15] By applying the MCMC algorithm, one obtains the fit represented in Figure 3a, considering the data until 2001 (before the big 2002 event), and in Figure 3b, considering the data until 2005 (after the big 2002 event). The two graphs show the estimates for the flood frequency curves corresponding to the posterior mode (PM) (i.e., GEV with parameters $\boldsymbol{\theta}$ corresponding to the maximum of $p(\boldsymbol{\theta}|\boldsymbol{D})$) and the 90% credible bounds associated with them. Note that the MCMC algorithm is not needed to identify the posterior mode but to quantify the uncertainty. Because the posterior density is known exactly up to a scale normalization constant, one can find the maximum without knowing the normalization constant. The choice of the posterior mode has been made for consistency with the maximum

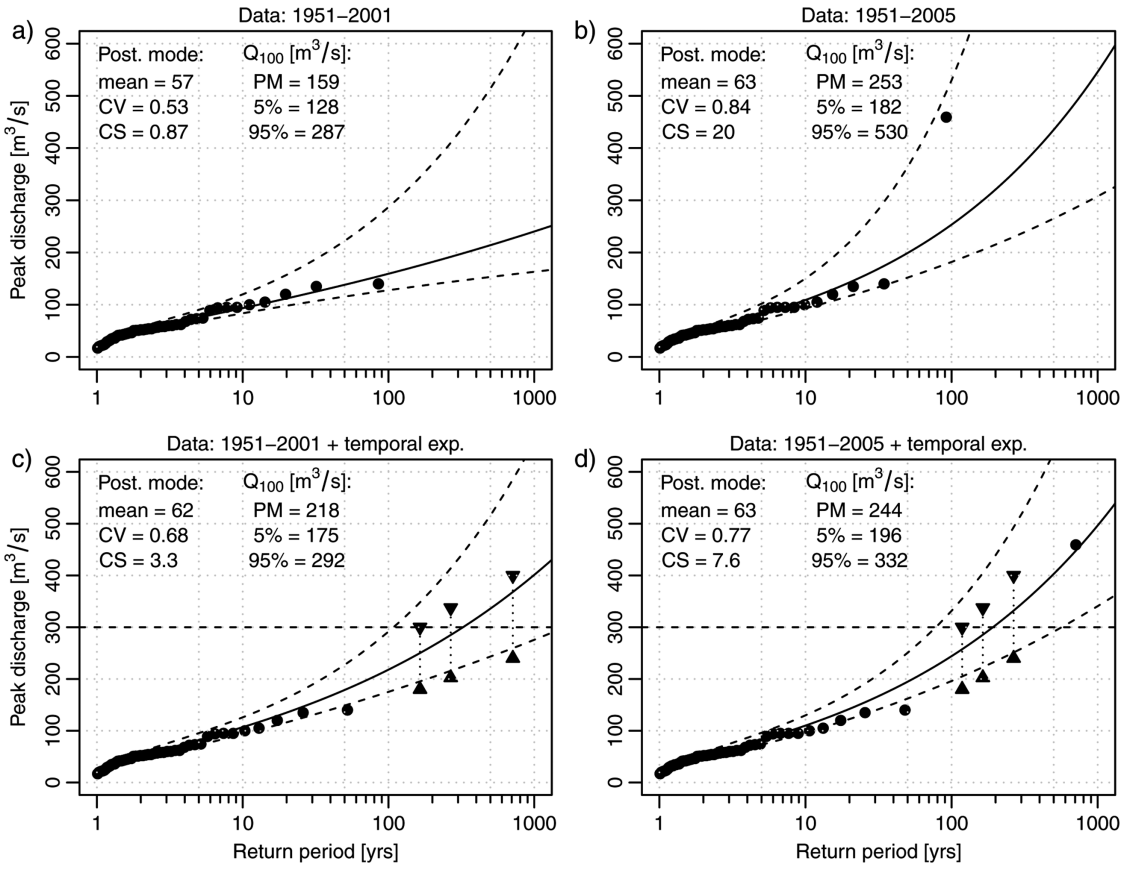

Figure 3. Bayesian fit of the GEV distribution to the data of the Kamp at Zwettl (a and c) before and (b and d) after the 2002 event. The following cases are shown: in Figures 3a and 3b, only the systematic data from 1951 are used; in Figures 3c and 3d, historic flood information is also included. The distribution corresponding to the posterior mode (PM) of the parameters is shown as continuous line, while the 5% and 95% credible bounds are shown as dashed lines. Mean, CV, and CS corresponding to the PM are indicated, as well as the 100 years return period quantile and its 5% and 95% credible values.

likelihood method, but other choices could have been made. For example, one could have taken the parameters corresponding to the mean of $p(\boldsymbol{\theta}|\boldsymbol{D})$, or, alternatively, one could use the Bayesian posterior predictive distribution, which is defined as the expected exceedence probability distribution of flood peak values with the expectation taken over all feasible values of the parameters (i.e., by integrating over their posterior distribution [see *Kuczera*, 1999, equation (9)]). The credible bounds have been constructed by reading off the 5% and 95% nonexceedance values from all quantiles corresponding to the MCMC generated parameters.

[16] Figure 3 also lists the values of MAF, CV, and CS for the GEV curve (corresponding to the PM) and the PM value of Q_{100} along with the 90% credible interval. The PM estimate of the 100 year flood runoff (Q_{100}) after 2002 is almost 100 m^3/s higher than the PM estimate before 2002 (the difference of the two is 60% of the value before 2002). The difference of the Q_{1000} is more than 100%. As can be seen in Figures 3a and 3b, the shape of the fitted curve is very different in the two cases, much more skewed if the analysis is done after the big 2002 event ($\theta_3 = -0.096$ before 2002 against $\theta_3 = -0.31$ after 2002). More remarkably, the 90% credible bounds are very different, i.e., the inclusion of the 2002 event increases considerably the uncertainty accounted for by the method. The 90% credible bounds range for Q_{100} is about 100% of Q_{100} before 2002 and 140% of Q_{100} after 2002. For Q_{1000}, the ranges are more than 200% and almost 300% of the PM estimates before and after 2002, respectively (and this does not change if a 20% random error is assumed for the 2002 event).

[17] Although this increase of uncertainty after 2002 would be expected from a statistical perspective, it is disturbing from a hydrological perspective as the large event has indeed revealed a lot of information on the hydrology of extreme events in this catchment (and one would hope that additional information allows to better constrain the estimates). Observing the 2002 event leads to an increase of knowledge on floods that could happen, therefore to a reduction of knowledge uncertainty (also called epistemic uncertainty in *Montanari et al*. [2009]). The widened credible bands after 2002 reflect the fact that a higher natural variability (also called structural uncertainty in *Montanari et al*. [2009]) has been correctly recognized by the statistical method. In the following sections, we compare the two estimates when additional information is used, as summarized in Figure 1.

4. Temporal Information Expansion

[18] To expand information into the past, historical flood information is used [*Brázdil et al*., 2006]. A survey of the local archives [*Wiesbauer*, 2004, 2007] reports that the three largest historical floods in the past 400 years occurred in 1655, 1803, and 1829 (Figure 2). The flood discharge of these events is highly uncertain but, for a historic analysis, the relative magnitudes as compared to the 2002 flood suffice. Information on inundation areas indicates that the water levels of the 1655 and 1829 events ranged around the 2002 event but these two events were caused by ice jams, so the discharges were likely smaller than those of the 2002 flood. The inundated area of the 1803 event in the downstream reach of the Kamp was much larger than in August 2002, but there were apparently backwater effects from the Danube, which were less pronounced in 2002, so that the associated flood discharges can be assumed to be smaller than for the 2002 event. These analyses, therefore, suggest that the 2002 event was probably the largest event since 1600. Based on the work of *Wiesbauer* [2004, 2007], the estimated peak discharges for the three historic events at Zwettl are those shown in Figure 2, where uncertainty bounds have been set at ±25% of the peak discharges from expert judgment on the basis of geometry, roughness, and potential changes in the river morphology. Moreover, for all the other years of the historic period (1600–1950), we assume that the threshold of 300 m^3/s was never exceeded. The threshold has been set equal to the highest possible value of the smallest of the three historic events (see Figure 2).

[19] Suppose that $x_1, x_2, \ldots, x_s$ is the sample of annual discharge maxima systematically recorded in $s = 55$ years (from 1951 to 2005), and $y_1, y_2, \ldots, y_k$ are $k = 3$ extraordinary flood discharges of the events of 1655, 1803, and 1829, which occurred during the historical period ($h = 350$ years). Further, let us suppose that $X_0 = 300$ m^3/s is the perception threshold, i.e., the threshold ensuring exhaustivity of the information above it. Finally, the magnitudes of the historical floods are known with uncertainty, for instance, with lower and upper bounds $[y_{Lj}, y_{Uj}]$ (Figure 2). All this information constitutes $\boldsymbol{D}$, the observed data. We write the joint probability of occurrence of recent and historical flood observations as in *Stedinger and Cohn* [1986]:

$$\ell(\boldsymbol{D}|\boldsymbol{\theta}) = \ell_S(\boldsymbol{D}|\boldsymbol{\theta}) \cdot \ell_H(\boldsymbol{D}|\boldsymbol{\theta}), \tag{4}$$

where $\ell_S(\boldsymbol{D}|\boldsymbol{\theta})$ is given by equation (2) and, indicating by $F_X(.)$, the cumulative of $f_X(.)$ given by equation (3):

$$\ell_H(\boldsymbol{D}|\boldsymbol{\theta}) = \binom{h}{k} F_X(X_0|\boldsymbol{\theta})^{(h-k)} \left\{ \prod_{j=1}^{k} \left[F_X\left(y_{Uj}|\boldsymbol{\theta}\right) - F_X\left(y_{Lj}|\boldsymbol{\theta}\right) \right] \right\}. \tag{5}$$

The likelihood function $\ell(\boldsymbol{D}|\boldsymbol{\theta})$ combines three terms: (1) the pdf of the s systematic data; (2) the probability of observing no events above X_0 for $h - k$ years; and (3) the probability of observing k historical events lying between the specified lower and upper bounds. The distribution describing the uncertainty on the historic flood peaks is assumed to be uniform (between y_{Lj} and y_{Uj}) but could be generalized to account for other error structures such as systematic errors due to hydraulic models used to reconstruct the discharges [*Neppel et al*., 2010]. In addition, the threshold is assumed without uncertainty for clarity although it is affected by the same uncertainty as the flood peaks. In section 8, the sensitivity of the estimates to the position of the perception threshold is examined.

[20] Considering the case represented in Figure 2, applying the MCMC algorithm and assuming a GEV distribution and the same improper flat prior used in section 3 (in equation (1), $\forall\boldsymbol{\theta}\ \pi(\boldsymbol{\theta}) = 1$), one obtains the fit represented in Figure 3c, considering the data until 2001, and in Figure 3d, considering the data until 2005. The PM estimate of Q_{100} after 2002 is just 12% higher than the PM estimate before 2002 (the difference between the two estimates is only 26 m^3/s). For Q_{1000}, the difference is about 25%. Compared to the case shown in Figures 3a and 3b (use of

the systematic data only), the difference of estimation of flood quantiles in the two cases is much reduced, i.e., the weight of the 2002 event in the estimation exercise has been reduced by accounting for other extreme events. Still, for high return periods, the PM curves differ, as the estimated CS after 2002 is still much higher than before 2002 ($\theta_3 = -0.22$ before 2002 against $\theta_3 = -0.28$ after 2002). Accounting for the historic information also narrows the 90% credible bounds considerably (e.g., the range for Q_{100} is about 54% of Q_{100} before 2002 and 56% after 2002).

5. Spatial Information Expansion

[21] Quantitative estimates of flood frequencies based on neighboring catchments can be obtained by various formal regionalization schemes. In the work by *Merz and Blöschl* [2005], the predictive performance of various types of automatic regionalization methods was examined on the basis of a jack-knifing comparison for 575 Austrian catchments, indicating that a geostatistical method outperforms other methods such as regressions and the region of influence approach. The geostatistical regionalization method known as topkriging [*Skøien et al.*, 2006] takes both catchment area and the river network structure into account and provides regional estimates of streamflow statistics at each point of the network as well as estimates of the uncertainty related to them (variances of estimation). In principle, one could use topkriging to regionalize the GEV parameters, estimated at the gauged stations, but we use the results of *Merz et al.* [2008] who regionalize the maximum annual flood moments in Austria. In the following, we indicate these moments as $\boldsymbol{\mathcal{M}} = [\mathrm{MAF}_\alpha, \mathrm{CV}, \mathrm{CS}]$. For ease of comparison of catchments of different size, the MAF peak has been standardized as $\mathrm{MAF}_\alpha = \mathrm{MAF} \times A^{\beta-1}\alpha^{-\beta}$, where A is the catchment area, $\alpha = 100\ \mathrm{km}^2$ is a standard catchment area and β is obtained from a regional analysis [*Merz and Blöschl*, 2005, 2008a, 2008b]. Topkriging provides estimates for the means $E(\boldsymbol{\mathcal{M}})$ and variances $\mathrm{Var}(\boldsymbol{\mathcal{M}})$ for the Kamp at Zwettl in cross-validation mode, i.e., without using the local data.

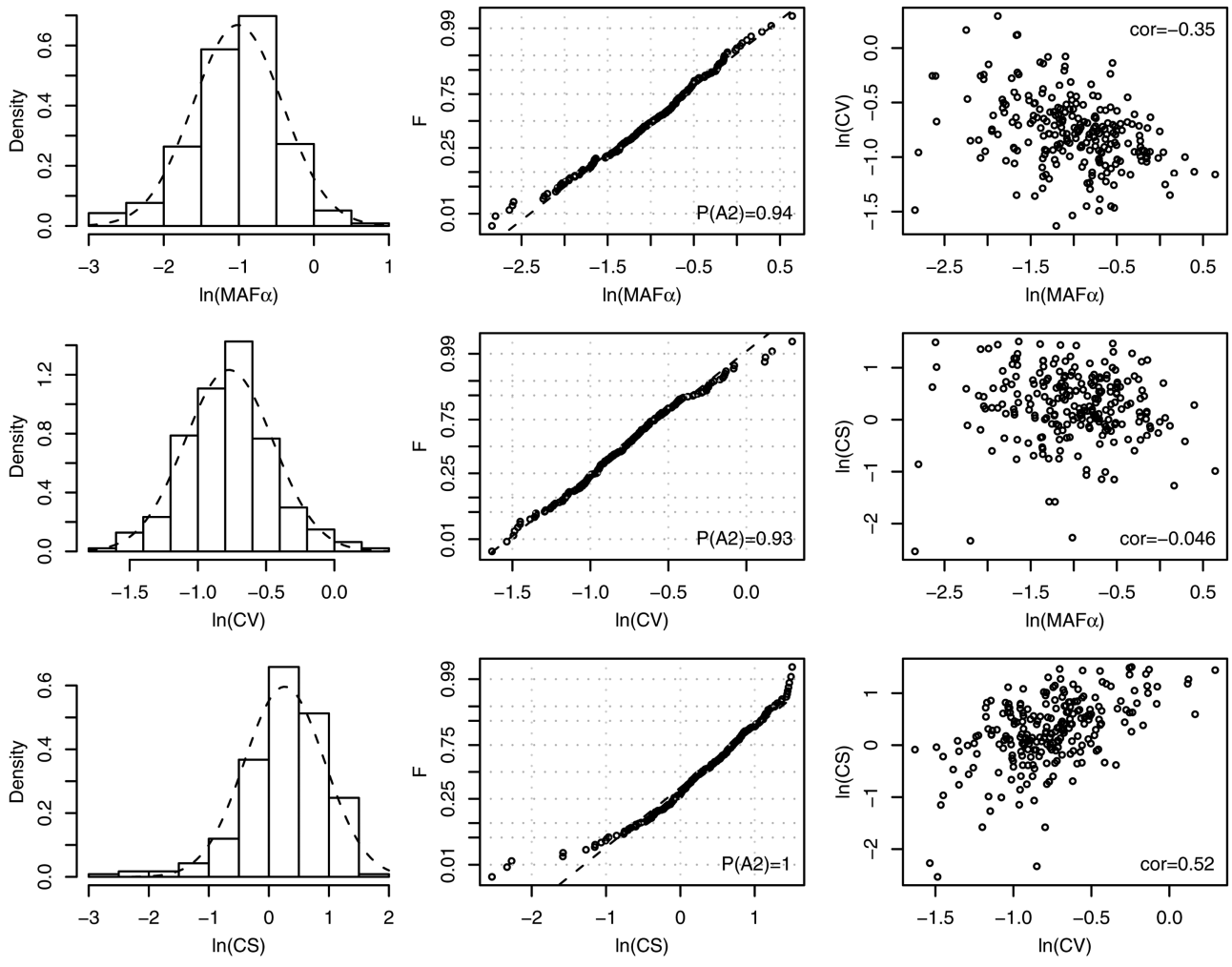

Figure 4. Distribution of the estimated $\mathrm{MAF}_\alpha = \mathrm{MAF} \times A^{\beta-1}100^{-\beta}$ [m^3/s/km^2], CV, and CS of the maximum annual floods in Austria (stations with more than 40 years of data). The natural logarithm transformation is shown. (left column) The histogram is compared with a fitted normal density function. (middle column) The data are represented in normal probability plots and the result of the Anderson-Darling normality test is shown as $P(A2)$. (right column) The correlation between the natural logarithms of the three moments is calculated.

[22] If the three-variate distribution of the maximum annual flood moments $p(\mathcal{M})$ is given, then the prior distribution of the GEV parameters $\pi(\boldsymbol{\theta})$ to be used in the Bayesian procedure can be uniquely defined. Unfortunately, topkriging does not provide an estimate of the entire trivariate distribution $p(\mathcal{M})$. What is missing is the correlation between the moments (cotopkriging could in principle be used to provide this but would require to fit three-dimensional variograms, which is very difficult) and the type of joint distribution.

[23] To obtain an understanding of the shape of the marginals and the correlation between moments, a regional analysis was conducted here of the sample MAF_{α}, CV, and CS for all stations in Austria. In Figure 4, the natural logarithms of the estimated MAF_{α}, CV, and CS are plotted for all Austrian sites with more than 40 years of data. The first column shows histograms of the flood moments and, for comparison, a normal distribution fitted to them. The second column in Figure 4 represents the same data in normal probability plots and the result of the Anderson-Darling normality test, $P(A2)$, is also shown (see *D'Agostino and Stephens* [1986] and *Laio* [2004] for details on the test). For $\ln(\mathrm{MAF}_{\alpha})$ and $\ln(\mathrm{CV})$, the test does not reject normality at the 5% level ($P(A2) = 0.94$ and 0.93, respectively), while it does for $\ln(\mathrm{CS})$. Also, in principle, CS should be allowed to be negative. However, visual inspection of Figure 4 suggests that $\ln(\mathrm{CS})$ is close to normal (closer than the nontransformed CS) and only nine catchments of 575 in Austria have a local negative CS (corresponding to catchments where anthropogenic effects are important, which is not the case for the Kamp catchment). For simplicity, we therefore assume that the marginal distribution of all three moments, in this particular case study, is lognormal and that their joint distribution $p(\mathcal{M})$ is three-variate lognormal. Note that alternative distributions could be used without changing the thrust of the analysis.

[24] The topkriging moment means and standard deviations at Zwettl in cross-validation mode and the sample correlations between the natural logarithms of the three moments estimated from the regional data (see the third column of Figure 4) are used to estimate $p(\mathcal{M})$ at Zwettl as described in Appendix A (equations A1–A4). The result is plotted in Figure 5 for the period before the 2002 event (first column) and for the whole period (second column). Mean and CV do not differ very much for the two cases, while the estimated CS is slightly higher when considering the data until 2005 because the neighboring stations also experienced the 2002 event although it was less extreme than at the Kamp. Figure 5 represents the spatial information that we want to include into the Bayesian analysis. The prior distribution of GEV parameters is calculated as $\pi(\boldsymbol{\theta}) = p(\mathcal{M}) \cdot |\boldsymbol{J}(\boldsymbol{\theta})|$, where $|\boldsymbol{J}(\boldsymbol{\theta})|$ is the Jacobian of the transformation $\boldsymbol{\theta} \rightarrow \mathcal{M}$ (see Appendix A, equations A5–A7). Because information is provided on all three parameters, we consider this as full information on the prior distribution $\pi(\boldsymbol{\theta})$ (Figure 1).

[25] The results of the Bayesian fit using the MCMC algorithm are shown in Figures 6a and 6b. The estimates of the CV with the regional information are slightly smaller than the estimates from the temporal information expansion (Figures 3c and 3d). The regionally estimated CS (i.e., $\theta_3 = -0.15$ before 2002 against $\theta_3 = -0.22$ after 2002) is also smaller than CS from temporal information expansion. The difference between the two PM estimates in Figures 6a and 6b, before and after 2002, is 38 m^3/s, which is about 20%. For Q_{1000}, the estimate after is about 35% higher than the estimate before the 2002 event (see also Figure 10 in section 9). For the 90% credible bounds, the range for Q_{100} is about 53% of Q_{100} before 2002 and 45% after 2002. For Q_{1000}, the ranges are about 80% and 60% of the PM estimates before and after 2002, respectively. The width of the credible bounds for highest return periods obtained from the regional information is narrower than the one obtained with the temporal expansion.

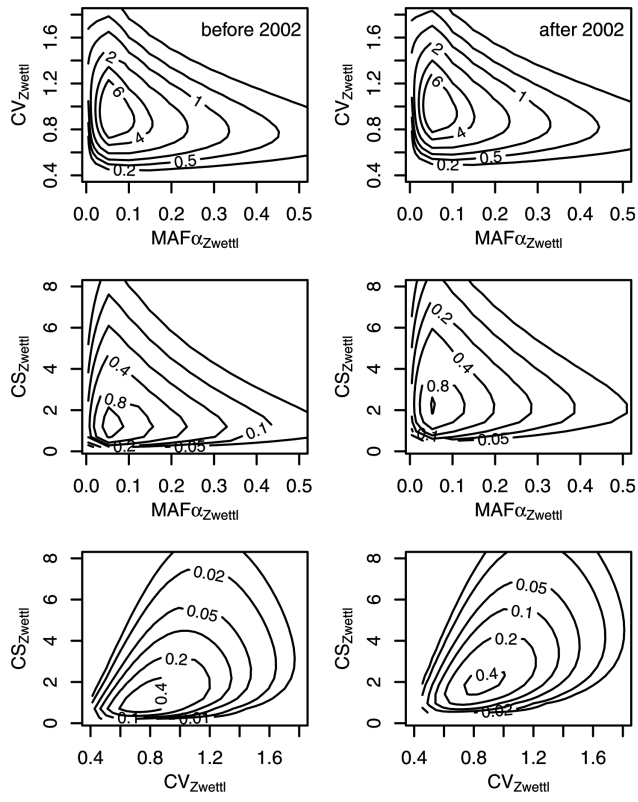

Figure 5. Bivariate marginals of the annual flood peak moment distribution $p(\mathcal{M})$ at Zwettl, based on topkriging in cross-validation mode (left) before and (right) after the 2002 event and the correlations between moments from the regional analysis.

6. Causal Information Expansion

[26] The most obvious way of causal information expansion is to derive flood frequencies from rainfall information and modeling of rainfall-runoff processes. The main benefit of using rainfall information is that available rainfall records are often much longer than the flood records (e.g., 100 years as opposed to flood records of 50 years in Austria). Modeling the rainfall-runoff processes assists in assessing the processes that have occurred and could occur, including nonlinearities and threshold effects [*Gutknecht et al.*, 2002; *Komma et al.*, 2007; *Blöschl et al.*, 2008; *Rogger et al.*, 2012b]. In this paper, we use the expert opinion of one modeler, Jürgen Komma, who has much experience with the Kamp area,

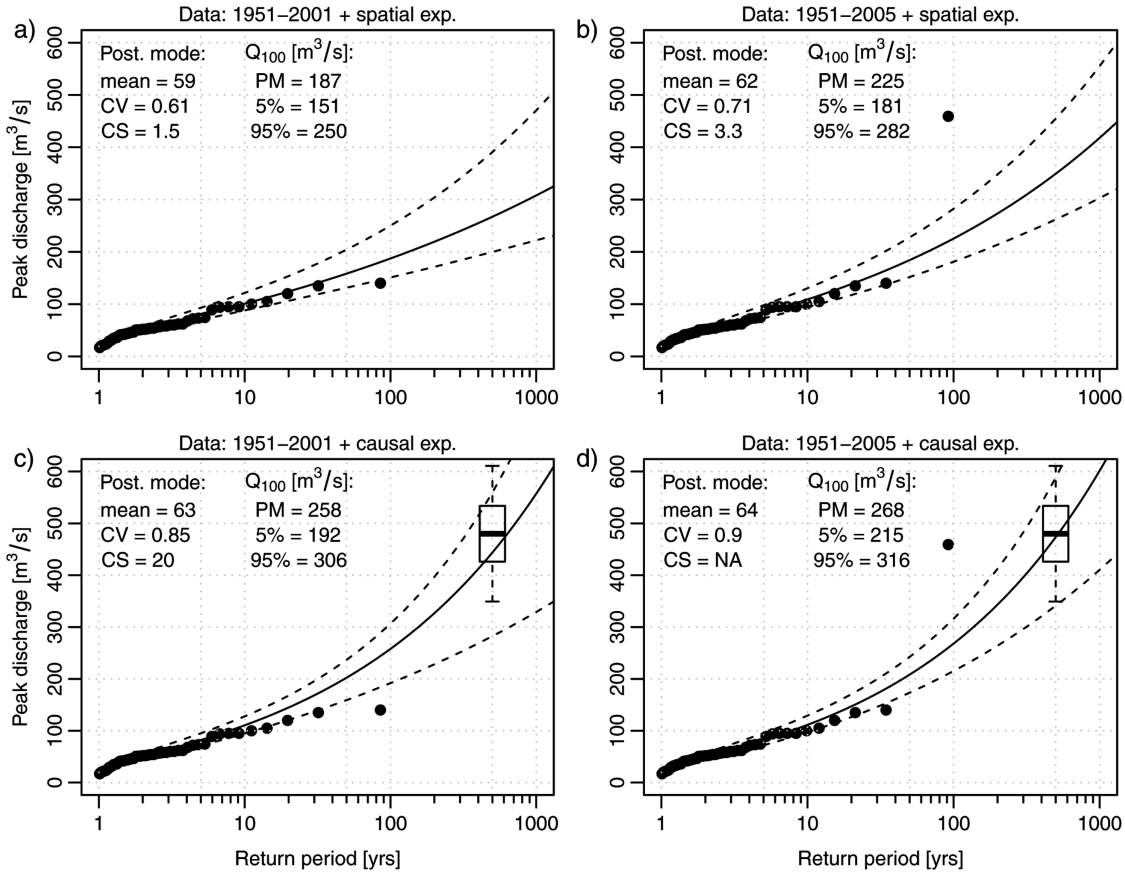

Figure 6. Bayesian fit of the GEV distribution to the data of the Kamp at Zwettl (a and c) before and (b and d) after the 2002 event. The following cases are shown: in Figures 6a and 6b, the regional information of topkriging is used along with the systematic data; in Figures 6c and 6d, expert guess for Q_{500} is used along with the systematic data. (See caption of Figure 3.)

being one of the developers of the operation flood forecasting model in the region, a continuous-simulation model, which he parameterized on the basis of hydrological data and field information [see *Komma et al*., 2007, for details]. We believe that, though subjective, his understanding of the physical mechanisms of rainfall generation and rainfall-runoff transformation can provide a valuable prior information. After discussing with the modeler, the floods that happened in the past, the mechanisms leading to them, the way he modeled the local hydrology, and his opinion on the uncertainties involved, we asked him to formulate quantitatively his beliefs on extremal behavior of floods in the area. As would be expected, it was not possible for the expert to formulate these beliefs in terms of GEV parameters, but in terms of extreme quantiles associated with large return periods, "a scale on which the expert has familiarity" [*Coles and Tawn*, 1996, p. 467]. His guess for the 500 year flood peak was of 480 m^3/s $\pm 20\%$. The guessed value (480) is the result of model simulations with artificial rainfall series while the uncertainty is based on his experience with the model in the Kamp and other catchments in Austria. Based on previous studies [*Blöschl et al*., 2008; *Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall* (*DWA*), 2012], and together with the expert, we refined the guess on the 500 year flood peak to the assumption that

$$h(Q_{500}) = N(\mu_{500}, \sigma_{500}), \tag{6}$$

where $\mu_{500} = 480$ m^3/s, $\sigma_{500} = 80$ m^3/s, and N is the normal distribution. This is shown in Figure 7, which represents the causal information that we want to include into the Bayesian analysis.

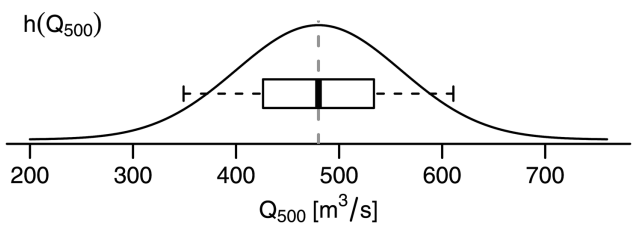

Figure 7. Expert guess for Q_{500} at Zwettl, modeled by a Gaussian distribution with mean $\mu_{500} = 480$ m^3/s and standard deviation $\sigma_{500} = 80$ m^3/s. The box-plot shows the 5%, 25%, 50%, 75%, and 95% quantiles.

[27] In contrast to *Kirnbauer et al*. [1987] and *Coles and Tawn* [1996], we did not ask the expert for information on more moments/quantiles to specify entirely the prior $\pi(\boldsymbol{\theta})$. Information on Q_{500} is, therefore, a partial information on the GEV parameters $\boldsymbol{\theta}$ (Figure 1). We use the information in equation (6) as part of the MCMC process. During the

MCMC random walk, triplets of values $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$ are extracted at each step, corresponding to the following GEV quantile with 500 year return period:

$$Q_{500} = g(\boldsymbol{\theta}) = \theta_1 + \frac{\theta_2}{\theta_3}\left[1 - \left(\ln \frac{500}{500-1}\right)^{\theta_3}\right]. \tag{7}$$

During the MCMC random walk, at each step, the likelihood term is multiplied to $h(g(\boldsymbol{\theta}))$ (equation (6)) to calculate a posterior distribution of the parameters, which is consistent with the reasonable range for Q_{500}. In other words, the function used by the MCMC analysis to compute the posterior distribution of the parameters is $\ell(\boldsymbol{D}|\boldsymbol{\theta}) \cdot h(g(\boldsymbol{\theta}))$. The fact that $\pi(\boldsymbol{\theta})$ is substituted by $h(g(\boldsymbol{\theta}))$ could suggest that this is an arbitrary and nonrigorous choice, but it is not. In Appendix B, we demonstrate that $h(g(\boldsymbol{\theta})) = \pi(\theta_1|\theta_2, \theta_3)$, i.e., the pdf of the location parameter conditional on the other two GEV parameters. The information is partial; we cannot specify the entire $\pi(\boldsymbol{\theta})$ from $h(Q_{500})$ but only one of the three GEV parameters conditional on the other two. In Appendix B, we show that $\pi(\theta_2|\theta_1, \theta_3)$ and $\pi(\theta_3|\theta_1, \theta_2)$ also can be related to $h(g(\boldsymbol{\theta}))$ but in different ways (see equations (B1) and (B2)). There is therefore a subjective choice involved (i.e., for what parameter is the information on Q_{500} used) and one would expect this choice to affect the results. However, we found out that, with the given conditions (e.g., Kamp data, GEV distribution, etc.), this subjective choice has no significant effects on the fitted flood frequency curves (not shown here). Because the information on one parameter conditional to the other two is used as part of the MCMC process, the joint likelihood function, the information provided as to reasonable values of Q_{500}, and the prior distribution for the parameters all work together to describe the likelihood of the triplets as a whole. In particular, the information provided as to reasonable values of Q_{500} boosts the posterior probability of triplets that are consistent with the specified reasonable range for Q_{500} and decreases the posterior probability of other sets of parameters.

[28] The results, obtained using $h(g(\boldsymbol{\theta}))$ as prior, are shown in Figures 6c and 6d. The estimated CS is much higher than CS from temporal and spatial information expansion, which can be explained by the high guess of Q_{500} obtained by the rainfall-runoff model ($\theta_3 = -0.31$ before 2002 against $\theta_3 = -0.334$ after 2002). In Figure 6d, the value of CS is marked as NA (not available) because the coefficient of skewness of a GEV distribution with shape parameter $\theta_3 < -1/3$ does not exist. The PM estimate of Q_{100} after 2002 is only 4% higher than the PM estimate before 2002 (about 10 m^3/s), and Q_{1000} is about 7% higher. The 90% credible bounds range for Q_{100} is about 44% of Q_{100} before 2002 and 38% after 2002. For Q_{1000}, the ranges are about 65% of both the PM estimates before and after 2002. The difference between the estimates before and after 2002 is very small for high return periods since the additional information on Q_{500} is related to the most extreme floods. Citing *Coles and Tawn* [1996, p. 476]: "if the prior is specified for the extreme tail and we focus attention on the posterior distribution of a quantile which remains above the maximum observed datum as the sample size increases to infinity, then the data necessarily fail to dominate the prior."

7. Combination of All Information

[29] One of the most appealing aspects of the Bayesian method is the possibility to account for all the different pieces of information together. Among the many different ways that could be used to combine prior information (spatial and causal), the most natural way is by multiplication, which is symmetric to the multiplication of the prior and the likelihood to derive the posterior. Equation (1) then becomes

$$p(\boldsymbol{\theta}|\boldsymbol{D}) \propto [\ell_S(\boldsymbol{D}|\boldsymbol{\theta})\ell_H(\boldsymbol{D}|\boldsymbol{\theta})][\pi_S(\boldsymbol{\theta})\pi_C(\boldsymbol{\theta})]. \tag{8}$$

This corresponds to give equal weight to the four pieces of information: (1) $\ell_S(\boldsymbol{D}|\boldsymbol{\theta})$, which is the likelihood function accounting for systematic data only (see section 3); (2) $\ell_H(\boldsymbol{D}|\boldsymbol{\theta})$, which is the likelihood function accounting for historic floods (see section 4); (3) $\pi_S(\boldsymbol{\theta})$, which is the prior obtained from the spatial information (topkriging, see section 5); and (4) $\pi_C(\boldsymbol{\theta})$, which is the prior obtained from causal information (in this example, just partially defined by using $h(g(\boldsymbol{\theta}))$ from the expert judgment on Q_{500} in the MCMC procedure, see section 6).

[30] The results obtained by the MCMC algorithm by using equation (8) are shown in Figures 8a and 8b. Using in this way, all information together provides the highest agreement between the estimates before and after the 2002 event. The difference for the 100 and 1000 year quantiles is of the order of 3% of Q_{100} (about 8 m^3/s) and 5% for Q_{1000}. The CS are very close, and the shape parameters of the GEV distributions are $\theta_3 = -0.23$ before 2002 and $\theta_3 = -0.24$ after 2002. The combination of all information provides the narrowest credible bounds as well: the 90% credible bounds range for Q_{100} is about 34% of Q_{100} before and 31% of Q_{100} after 2002; for Q_{1000}, the ranges are about 45% and 40% of the PM estimates before and after 2002, respectively. Of course, inclusion of the extreme 2002 event still affects the flood frequency curve but, if one compares Figure 8 to Figures 3a and 3b, it is clear that the influence of the single extreme event is now very small. The width of the credible bounds is smaller than what was obtained when using any of the single pieces of information (temporal, spatial, or causal), especially for high return periods.

[31] Many other ways of combining priors exist. *Genest and Zidek* [1986] discuss several pooling models, which are based on different assumptions. The type of combination one chooses is a choice (a model of combination) rather than the result of some probabilistic algebra (as is the multiplication of the prior and the likelihood to derive the posterior). The combination $(\pi_S \cdot \pi_C)$ is a special case of what *Genest and Zidek* [1986] call "geometric (or logarithmic) opinion pool," i.e., $\pi \propto \pi_1^{w_1} \cdot \ldots \cdot \pi_n^{w_n}$ with weights w_i assumed equal to one (other weights are considered in the sensitivity analysis of section 8). The rationale behind multiplying the priors is that each prior is individually considered representative of the entire population of the GEV parameters. One fundamentally different choice is to sum up the prior distributions $(\pi_S + \pi_C)$, which corresponds to what *Genest and Zidek* [1986] call "linear opinion pool" (i.e., $\pi \propto w_1\pi_1 + \cdots + w_n\pi_n$). The rationale behind summing up the priors is that each prior individually represents a different part but together they are representative of the entire population of the GEV parameters. As an illustration,

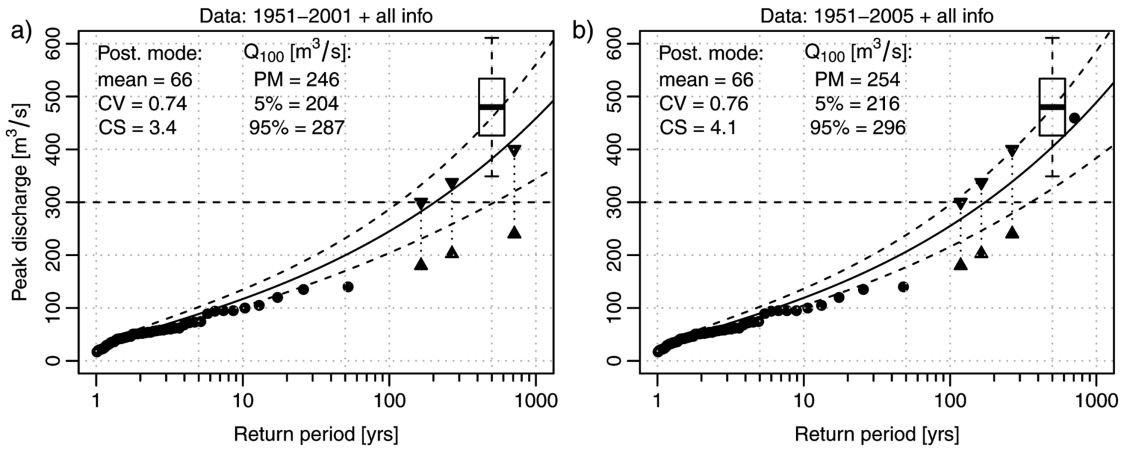

Figure 8. Bayesian fit of the GEV distribution to the data of the Kamp at Zwettl (a) before and (b) after the 2002 event, when all available information is used. (See caption of Figure 3.)

if one is interested in estimating the distribution of weights of Italian people and has prior information based on two studies conducted one in Rome and the other in Turin, one would combine the pieces of information by multiplying these priors. If, instead, one has prior information based on two studies conducted in Rome, one on males and the other on females, one would combine the pieces of information by summing up these priors. Even though in our study we assume that the spatial and causal pieces of information are both individually representative of the flood frequency in the Kamp region, and therefore only the multiplication of priors should be considered, in section 8 we also show the results we would obtain by summing up (and eventually weighting) the priors. The summation of priors would be definitively an option in studies where different pieces of information can be associated to different aspects of the population of floods, for instance, to different flood types.

8. Sensitivity Analysis

[32] The flood estimates of the Bayesian method, obviously, depend on the parameters used and, to some extent, on the assumptions made. To assess what is the relative role of the different pieces of information accounted for in equation (8) and the importance of their reliability, Figure 9 shows the result of a sensitivity analysis of some of the parameters used and the assumptions made in the previous sections. For clarity, Q_{100} and Q_{1000} are considered as estimates

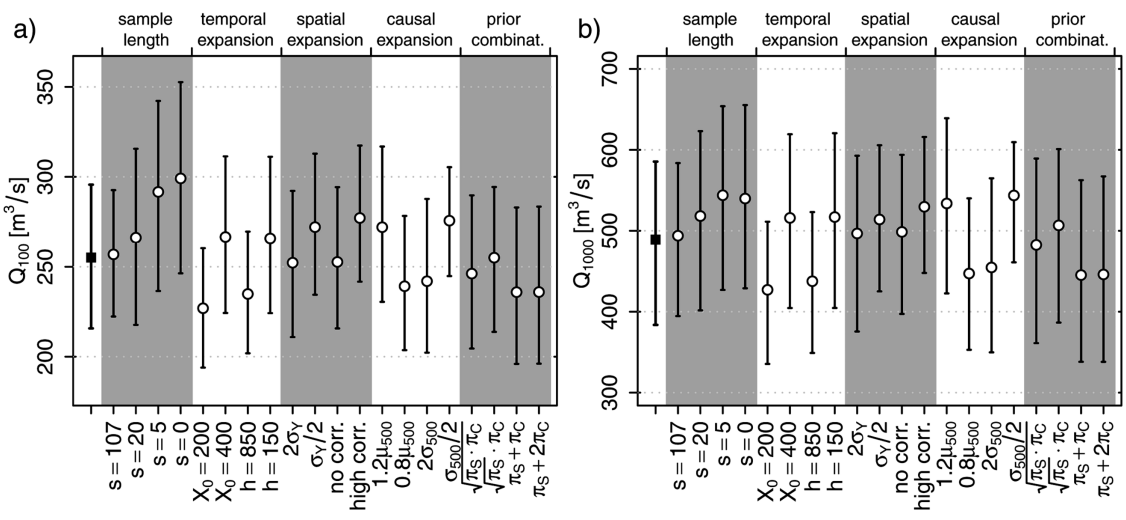

Figure 9. Sensitivity of the flood estimates on the parameters used and the assumptions made. Posterior mode estimation (points) and 90% credible bounds (segments) of (a) Q_{100} and (b) Q_{1000} at the Kamp catchment using all pieces of information, after the big 2002 event. The far left bars (with the black square point) show estimates as in Figure 8; s is the length of the systematic record, X_0 is the perception threshold, h is the length of the historical period (equation (8)), σ_Y is the standard deviation of the topkriging estimates of the moments of the maximum annual peaks, μ_{500} and σ_{500} are the mean and standard deviation of Q_{500} given by the expert rainfall-runoff modeler, respectively, and π_S and π_C are the prior distributions of the GEV parameters considering the spatial and causal information expansion, respectively.

after the 2002 event only. The first whisker with the black square point in the left part of the two graphs represents the PM estimate and 90% credible bounds obtained using all information as in Figure 8.

8.1. Sample Length

[33] The first four whiskers, with a circular point and with a gray background shade, show the effect on the estimates of Q_{100} and Q_{1000} of the length of the systematic data period, when all other pieces of information are also considered (as in section 7). The first of these whiskers (denoted by $s = 107$) is obtained by a longer systematic data sample reconstructed from water stage data [*Gutknecht et al.*, 2002] for the period 1896 to 1947. These additional data were not used in our previous analyses (i.e., sections 4 and 7) to account for the information added by historical floods only. The reconstructed data indicate that a number of large floods occurred in the first half of the twentieth century, while floods tended to be lower in the second half. Considering these additional data as systematic recordings, the likelihood function is again expressed by equation (2), where the number of observations is now $s = 107$ for the period 1896–1947 and 1951–2005. The PM estimate does not change much (it is slightly higher because of the large floods in the first half of the century), and the credible bounds are only slightly narrower. The second whisker in the gray region ($s = 20$) is obtained by considering only the last 20 years of data, the third ($s = 5$) considering only the last 5 years, and the fourth ($s = 0$) by excluding the systematic data from the analysis (in this last case, for the 2002 event, we have just assumed that it was over the threshold of 300 m^3/s used for the historical period). The PM values increase for a decreasing number of systematic years, but the increase is moderate. In the worst case of no systematic data (ungauged site, $s = 0$), the PM estimate of Q_{100} and Q_{1000} increase by circa 20% and 10%, respectively. The width of the credible bounds also increases for a decreasing number of systematic years, especially for Q_{100}. This last effect is more evident for low return period quantiles (not shown in the figure) since the systematic data give much more information for the small floods than for the big ones.

8.2. Temporal Expansion

[34] The subsequent four whiskers in Figures 9a and 9b, with white background, result from changes in the assumptions made for the temporal expansion of information. The first of them is obtained if one assumes the threshold X_0 of equation (4) to be 200 m^3/s instead of 300 m^3/s. Of course, the PM estimate is then lower because one assumes that, in the 350 years before the systematic period, all maximum annual floods were relatively small (except for the three historic events). Also, the credible bounds are narrower. This is because, reducing the threshold, one constrains the range of values for the flood peaks in the past. This would be a valuable result, provided that the threshold was really exceeded only three times. Normally, the less the information one has on the historic period, the highest the threshold X_0 is, which is usually chosen to be slightly below the most extreme events reported for the presystematic period. The neighboring whisker is obtained by considering $X_0 = 400$ m^3/s and presents, as expected, a higher PM value (but not as much as it was lowered by the $X_0 = 200$ m^3/s case) and slightly wider credible bounds. Increasing the threshold corresponds to providing less information on floods in the historic period.

[35] The following two whiskers show the case where the historic period is considered to be 500 years longer ($h = 850$ years) or 200 years shorter ($h = 150$ years). In the first case, this is like increasing the information content and the result is similar to the one obtained by decreasing X_0. The PM value is lower and the credible bounds narrower because now we assume that the threshold $X_0 = 300$ m^3/s was exceeded only three times in 850 years instead of 350 years. Analogously, for a shorter historic period, the PM estimate is higher and the credible bounds larger similarly as for a higher threshold. Similar sensitivity analyses were presented in *Gaume et al.* [2010] showing analogous results. We also analyzed the effect of increasing or reducing the uncertainty of estimation of the three historic peaks (not shown in Figure 9). By setting the historical peak discharge ranges to be $\pm 45\%$ of the peak discharges estimates (more uncertainty, if compared to $\pm 25\%$ used before), Q_{100} is circa 1% and 2% lower before and after 2002, respectively. By setting the historical peak discharge ranges to be $\pm 5\%$ of the peak discharges estimates (less uncertainty), Q_{100} is circa 1% higher both before and after 2002. The same considerations apply to the 90% credible bounds. Because the historic period is long compared with the systematic sample, the level of uncertainty of the historic flood peak estimates has only a limited impact on the statistical inference results [see also *Stedinger and Cohn*, 1986; *Martins and Stedinger*, 2001; *Payrastre et al.*, 2011]. This might also be due to the assumption that these errors are independent, i.e., due to the imprecise knowledge of the water levels. Systematic errors, such as rating curve errors, might exert more leverage on the inference [e.g., *Kuczera*, 1996; *Neppel et al.*, 2010].

8.3. Spatial Expansion

[36] The four whiskers in the central part with gray background of Figures 9a and 9b result from a sensitivity analysis on the assumptions made for the spatial expansion of information. The first two whiskers are obtained by doubling or halving the standard deviation of the topkriging estimates of the moments of the maximum annual peaks ($2\sigma_Y$ and $\sigma_Y/2$, respectively). Surprisingly, the width of the credible bounds is not much affected by that, but the position of the estimates is. By doubling the standard deviations, the estimates of Q_{100} and Q_{1000} are smaller. The reason of that is related to the asymmetric shape of the prior distribution of the moments (Figure 5). For instance, increasing the standard deviation σ_Y decreases $\mu_{\ln(Y)}$ because of equation (A3) (see Appendix A). For the same reason, halving the standard deviations increases the estimates of Q_{100} and Q_{1000}.

[37] The third whisker (no corr.) is obtained by considering no correlation in the prior distribution of the moments of Q_T given by topkriging. In the example provided here, the effect of neglecting the correlation among moments is very small. Visible effects on the estimates of the flood frequency curve are obtained only if the degree of correlation among moments is very high (high corr.), i.e., $\mathrm{cor}[\ln(\mathrm{MAF}_\alpha), \ln(\mathrm{CV})] = -0.9$, $\mathrm{cor}[\ln(\mathrm{MAF}_\alpha), \ln(\mathrm{CS})] = -0.9$, and $\mathrm{cor}[\ln(\mathrm{CV}), \ln(\mathrm{CS})] = 0.9$, when the PM estimates increase and the credible bounds become slightly narrower,

especially for Q_{1000}. Assuming high correlation of the moments (or parameters) corresponds to reducing the degrees of freedom for the fitted model. It is like providing prior information on the type of distribution with less than three free parameters.

8.4. Causal Expansion

[38] The four whiskers, in the central-right part with white background of Figures 9a and 9b, show the effect of increasing or decreasing the mean estimate of Q_{500} by $\pm 20\%$ in the causal expansion, as well as doubling or halving the standard deviation. Departures of $\pm 20\%$ in the expert guess (a priori) cause departures in the PM estimates (a posteriori) of less than $\pm 10\%$ (for both Q_{100} and Q_{1000}), because of the other data taken into account. Doubling the standard deviation increases the width of the credible intervals and produces a lower PM estimate of Q_{100}. Halving the standard deviation has the opposite effect. The results are consistent with those of the sensitivity analysis on the spatial expansion, but the reasons are different. In this case, the prior distribution of Q_{500} is symmetric and produces an increase of the PM estimation of Q_{100} and Q_{1000} with respect to the other pieces of information. By reducing the confidence in Q_{500} (i.e., doubling the standard deviation), the increase of the PM estimates of Q_{100} and Q_{1000} is reduced, while the opposite happens if one has more confidence in the estimates of Q_{500} provided by the expert. Also, the effects are much more pronounced than when doubling and halving the standard deviations of the topkriging estimates of the moments of Q_T, especially for Q_{1000}. This is because the causal information expansion controls the highest return period quantiles in this example.

8.5. Prior Combination

[39] The last four whiskers, in the right part with gray background of Figures 9a and 9b, show the effect of combining the spatial and causal information in different ways. As discussed in section 7, we multiply the priors $(\pi_S \cdot \pi_C)$, which is symmetric to the multiplication of the prior and the likelihood to derive the posterior and which is based on the assumption that each prior is individually representative of the entire population of the GEV parameters. To investigate the sensitivity to this assumption, we also try different combinations. For instance, the first whisker uses the combination $\sqrt{\pi_S \cdot \pi_C}$, which corresponds to smoothing the prior distribution used in section 7 (i.e., augmenting its variance). In fact, compared with the simple multiplication (first whiskers with the square black point), the credible bounds are wider. The posterior mode is a little bit lower because less weight is given to the prior, which, because of the causal information, tends to increase the estimates for high quantiles. That is the reason of the higher position of the second whisker, which illustrates the case of using a higher weight for the causal than for the spatial information (i.e., $\sqrt{\pi_S} \cdot \pi_C$).

[40] As discussed in section 7, one fundamentally different choice is to sum up the prior distributions $(\pi_S + \pi_C)$. The rationale behind summing up the priors is that each prior individually represents a different part but together they are representative of the entire population of the GEV parameters. Even though we argue that this is not a reasonable assumption in our case, the third whisker shows the effect of this choice on the results, for illustration purposes. The credible bounds are wide because they embrace those of the spatial expansion (which gives lower flood estimates) and of the causal expansion (which gives higher flood estimates) together. The fourth and last whisker illustrates the effect of giving more weight to one of the pieces of information, in this case the causal information, on the flood estimates. Using the combination $(\pi_S + 2\pi_C)$ does not change the results. The result would be the same if $(2\pi_S + \pi_C)$ were used (not shown here).

9. Discussion

[41] In this paper, we illustrate by example how three pieces of information (temporal, spatial, and causal) can be combined with the systematic flood data in a Bayesian analysis and compare the flood frequency estimations before and after the extraordinary 2002 flood event in the Kamp catchment in Austria. Figure 10 shows the comparison of the PM estimates (i.e., Bayesian analogues to the maximum likelihood estimates) and 90% credible intervals of Q_{100} and Q_{1000} if the analyses are made before or after 2002 and considering the different pieces of information. The figure allows to directly compare the results of Figures 3, 6, and 8.

[42] By looking at Figure 10, one notices that, when compared to the temporal and causal expansions, the spatial information expansion gives (1) the smallest PM estimates of Q_{100} and Q_{1000}; (2) a higher difference between the estimates before and after 2002; and (3) narrower confidence bounds, specially for Q_{1000}. The reason for the low estimates obtained by the spatial information expansion has to do with the regional data in northern Austria. For most of the gauging stations of the Kamp region, the local discharges (including the 2002 flood) are higher than the regional estimates. Because of the smaller number of outliers outside the Kamp region, the quantiles obtained with the regional information are smaller than those obtained with the temporal or causal information. Another reason is that, by deriving the prior distribution of the GEV parameters from the moments, the GEV shape parameter cannot be lower than $-1/3$, for which CS does not exist. This could be avoided by regionalizing the GEV parameters directly using topkriging. The highest difference between the estimates before and after 2002 is due to the fact that we use the topkriging estimates before and after the 2002 event (we respectively ignore or account for the 2002 event in the other catchments of the Kamp area), while for the temporal and causal expansion, we use the same additional information in the two cases (before and after 2002).

[43] The causal information places the PM estimates of Q_{100} and Q_{1000} above the values obtained by all other information. The rainfall-runoff model used by the expert gives relatively large peak discharges because the catchment consists of porous soils, which fill up above a rainfall threshold of about 70 mm and significantly increase the runoff coefficients [*Gutknecht et al.*, 2002; *Komma et al.*, 2007; *Blöschl et al.*, 2008]. In addition, with event-based models, widely used in hydrologic practice, it is quite common that the design flood method gives significantly larger peak flows than flood statistics for the same catchment, which is related to the usual assumptions made in the design flood method [*Viglione and Blöschl*, 2009; *Viglione et al.*, 2009; *Rogger et al.*, 2012a]. The nature of

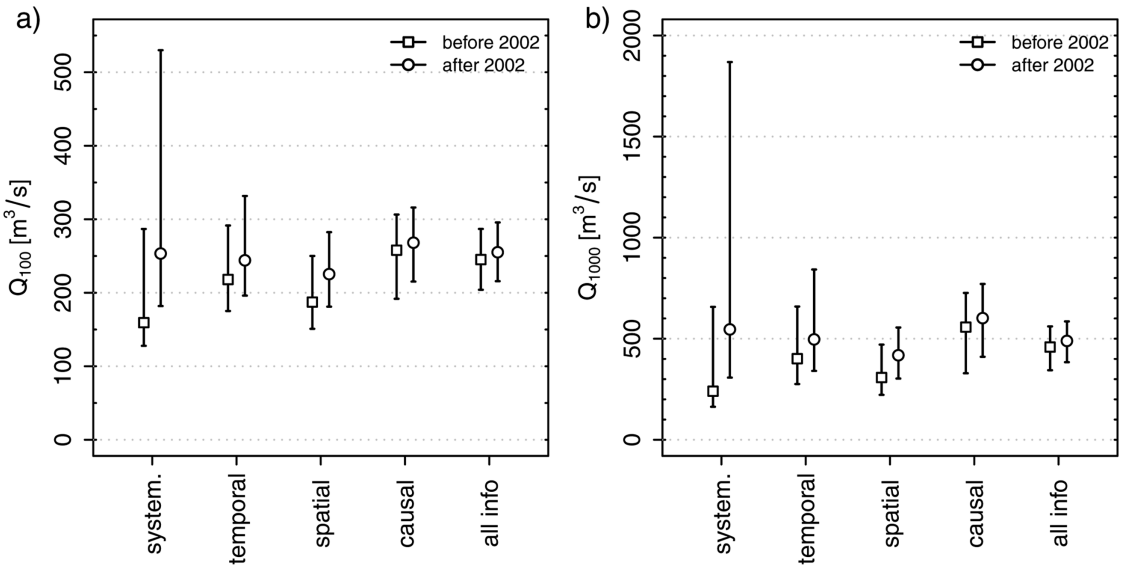

Figure 10. Posterior mode estimation (points) and 90% credible bounds (segments) of (a) Q_{100} and (b) Q_{1000} at the Kamp before and after the big 2002 event. In the first case (system) just the systematic flood data are used. The following cases show the estimates using different pieces of information in addition to the systematic flood data.

the judgment of the expert also plays a role. With a similar Bayesian analysis, but for extreme rainfall, *Coles and Tawn* [1996] obtain estimates of high return period quantiles higher than those obtained with maximum likelihood, since the expert opinion essentially is that events more extreme than those observed at the site can happen in the region with nonnegligible probability. In our example, the causal information expansion is the one that gives the smallest difference between Q_{100} and Q_{1000} estimated before and after 2002. This is because the expert guess for Q_{500} is the same in the two cases (we interviewed him only once) and affects the estimation of high return period quantiles.

[44] Despite the differences in Figure 10, the estimates of Q_{100} and Q_{1000} from the three types of additional information, all fall within a range that is narrower than that from flood frequency analysis with systematic data alone. Under the assumptions made, adding information to the systematic data reduces the difference between the two estimates substantially (increase the stability of estimation). Temporal, spatial, and causal expansions, in the example considered in this paper, all provide a significant reduction of the credible bounds (increase the precision). What has not been analyzed in this paper is the validity of estimation of the flood frequency curve and the uncertainty associated with it, i.e., how close is the estimate to the truth. In the real case example treated here, we cannot answer this question. Simulations of 'synthetic realities' could have been used instead, as for example those in *Kuczera* [1982] or *Seidou et al*. [2006], to compare the validity of the method and investigate the influence of prior specifications (or misspecifications) on the estimates, but we considered a real world case to be more insightful for the purposes of this paper: i.e., for conveying the message that combining information from many sources is valuable not only in theory but also in practice.

[45] The considerations made so far are valid under the assumptions made, as stated many times in the paper. It is, therefore, important to discuss these assumptions, at least the principal ones, and refer to the literature dealing with them explicitly. For example, regarding the historic floods, we have assumed that the uncertainties were given by independent, nonsystematic errors. The evaluation of the sources of uncertainties is a delicate issue, as shown by *Neppel et al*. [2010], who analyzed recent and historical maxima, accounting for additive and multiplicative errors affecting discharge values due to random errors in water-level readings and systematic rating curve errors, respectively. Their finding is that, as in our case, independent discharge errors stemming from imprecise knowledge of the water levels does not appear to impact much on the results, while the prior for systematic rating curve errors exerts a significant influence on the estimated quantiles.

[46] When applying the Bayes theorem with historical floods, we have implicitly assumed stationarity in time, i.e., we have assumed that the climate and catchment processes leading to floods in the historic period are the same as those leading to floods today. There are ways of incorporating information on changes in time in nonstationary models. For example, *Perreault et al*. [2000a, 2000b] consider change-points (in the mean and variance) in the Bayesian analysis of hydropower energy inflow time series; *Renard et al*. [2006] consider change-point and linear trend models for flood frequency analysis; *El Adlouni et al*. [2007] and *Ouarda and El Adlouni* [2011] generalize the generalized maximum likelihood method of *Martins and Stedinger* [2000] to the nonstationary case; and *Lima and Lall* [2010] account for nonstationarity of annual maximum floods and monthly streamflows in a regional context with a Bayesian method. Similar to stationarity in time, an assumption of homogeneity in space has been made here, i.e., when applying topkriging (and the same would hold for any other regional model), we assume that the processes leading to floods in the neighboring catchments are analogous to those leading to floods in the catchment of interest.

[47] In this paper, the GEV distribution is used as statistical model for the flood peaks, which has been shown to

reproduce the sample frequency distributions of hydrological extremes around the world [see, e.g., *Stedinger et al*., 1993; *Vogel and Wilson*, 1996; *Robson and Reed*, 1999; *Castellarin et al*., 2001; *Coles et al*., 2003]. However, if mixed processes and thresholds occur [*Gutknecht et al*., 2002; *Komma et al*., 2007; *Blöschl et al*., 2008; *Rogger et al*., 2012b], other models should be preferred. In the Kamp catchment, of the maximum annual floods in the systematic period 37% were long rain floods, 11% short rain floods, 4% flash floods, 41% rain on snow floods, and 7% of snow melt floods [*Merz and Blöschl*, 2003], while two of the three historic floods were caused by ice jams associated with heavy rain and snow melt. A mixture of models could be used to account for the diversity of processes, but the data samples stratified by flood processes would be too short. Recent literature exists in which model mixture and/or uncertainty in model selection is accounted for in a Bayesian framework [e.g., *Wood and Rodriguez-Iturbe*, 1975; *Perreault et al*., 2000a, 2000b; *Renard et al*., 2006; *Montanari*, 2011].

[48] Another issue is the independence of the pieces of information. Even if in general there are ways to combine non-independent pieces of information into priors [*Genest and Zidek*, 1986], in Bayesian analysis, priors have to be specified independently from the local data. The topkriging estimates after 2002 at Zwettl, even if in cross-validation mode, are not fully independent from the 2002 event happening in the neighboring catchments (see *Stedinger* [1983], *Stedinger and Tasker* [1985], *Hosking and Wallis* [1988], and *Castellarin et al*. [2005, 2008] for a discussion on the effects of intersite correlation in regional frequency analysis). Taking these correlations into account may slightly increase the width of the credible bounds. However, we expect this effect to be small, given the similarity between the estimations with regional information before and after 2002. In addition, for the causal information expansion, the independence assumption may not fully apply. When running the rainfall-runoff model, the expert does not ignore the observed floods at Zwettl, since the model was calibrated. However, the calibration of the continuous model considers the whole runoff time series, including minor events and dry periods. Therefore, the assumption of independence from the maximum annual peak data is reasonable.

[49] The consistency of information is another assumption made in the Bayesian approach [*Gelman et al*., 2004; *Laio and Tamea*, 2007]. Data (a likelihood function) that are in conflict with prior information will result in a posterior that is a compromise between the two disparate sources of information but is in disaccord with both. This is because likelihood and prior distribution are multiplied. Nonconsistencies have to be recognized when checking the results of the analysis. In the sensitivity analysis, for example, we checked the effect of departures in the causal information (by varying μ_{500} in Figure 9). In the Kamp example, all pieces of information are reasonably consistent, as can be inferred by comparing the ranges in Figure 10. In the case that different pieces of prior information are inconsistent because they are representative of different parts of the population of the variable of interest, pooling methods exist that allow to combine them properly (e.g., by linear pooling as discussed in sections 7 and 8).

[50] In the example reported in this paper, temporal and spatial information expansions can be considered as "objective", as their use (in the likelihood and as prior) are based on procedures whose hypotheses can be listed and openly discussed and scrutinized. The causal information expansion instead is "subjective" as its specification is based on personal judgment that is far more difficult to formalize. Of course, objective causal information could be obtained from rainfall-runoff model predictions, ideally not mediated by expert judgment, by rigorously quantifying the associated uncertainties (in the inputs, model parameters, and structure). This is a challenge, and there is ongoing research in the hydrologic literature [see *Montanari et al*., 2009, and references therein]. Subjective information, in form of personal experience, is nevertheless very valuable. However, to use personal experience in a Bayesian analysis, the experts have to be "trained" to make probabilistic statements. More qualified experts may be able to provide more accurate statements, in the sense that their prior mean is closer to the true value they are trying to predict compared with less qualified experts, but they may have a tendency to underestimate uncertainty. This is what *Taleb* [2007] calls "epistemic arrogance", i.e., the difference between what someone actually knows and how much he thinks he knows. Also, if many experts are interviewed [e.g., *Kirnbauer et al*., 1987], it is not trivial to decide how to combine their opinions [*Genest and Zidek*, 1986].

10. Conclusions

[51] The main message of this paper is in line with the messages of the companion papers of *Merz and Blöschl* [2008a, 2008b] in the flood frequency hydrology series: combining information from many sources is very useful. While, in the previous papers, the usefulness of combining many pieces of information is demonstrated by heuristic reasoning, it is demonstrated here by Bayesian analysis. Even if a number of assumptions have been made in this paper, we believe that the ability of the Bayesian approach to use all pieces of information in conjunction is a major advantage over other methods. The way to formalize the hydrological information into a Bayesian framework is not straightforward, but possible, as has been shown here through examples. Expanding information beyond the systematic flood record is sometimes considered of little value in engineering hydrology because subjective assumptions are involved. In the past, most of the national guidelines [e.g., *Natural Environment Research Council* (*NERC*), 1975; *Deutscher Verband für Wasserwirtschaft und Kulturbau* (*DVWK*), 1999] have therefore involved prescriptive procedures with little choice for the hydrologists who applies them, while few have encouraged the use of additional information [see e.g. *U.S. Interagency Advisory Committee on Water Data*, 1982, p. 20, Bulletin 17B]. More recent versions of these guidelines [e.g., *DWA*, 2012] explicitly focus on temporal, spatial, and causal information to complement the systematic flood data (see also *Stedinger and Griffis* [2008], for a discussion on how Bulletin 17B will be revised).

[52] The results of this study suggest that the additional information can significantly improve the confidence in flood frequency estimates. With all the information combined together, our estimated 100 year flood peak at Zwettl is 250 m^3/s with a 90% credible region of $\pm 15\%$. This estimate would have been practically the same in 2001, before

observing the extraordinary 2002 flood event. The return period of this event (460 m^3/s) would be estimated as 1000 years before the event and 800 years after it with 90% credible bounds of 500 to 4500 years and 450 to 2300, respectively. This contrasts the estimates of >100,000 years and 340 years with the systematic data alone. While a range of assumptions need to be made, the final flood estimate is not much affected by assumptions varying in a reasonable range because the information provided through combining independent sources (temporal, spatial, and causal) may outweigh the uncertainty introduced by these assumptions. It is suggested that complementing systematic flood data by temporal, spatial, and causal information should become the standard procedure for estimating large return period floods.

Appendix A: From the GEV Moments to the GEV Parameters

[53] In section 5, it is argued that, in Austria, the logarithms of the maximum annual peak moments $\ln(\boldsymbol{\mathcal{M}}) = [\ln(\mathrm{MAF}_\alpha), \ln(\mathrm{CV}), \ln(\mathrm{CS})]$ follow the three-variate normal distribution:

$$p(\ln(\boldsymbol{\mathcal{M}})) = \frac{1}{(2\pi)^{3/2}|\boldsymbol{\Sigma}[\ln(\boldsymbol{\mathcal{M}})]|^{1/2}} \times \exp\left\{-\frac{1}{2}(\ln(\boldsymbol{\mathcal{M}}) - E[\ln(\boldsymbol{\mathcal{M}})])'(\boldsymbol{\Sigma}[\ln(\boldsymbol{\mathcal{M}})])^{-1}(\ln(\boldsymbol{\mathcal{M}}) - E[\ln(\boldsymbol{\mathcal{M}})])\right\}, \tag{A1}$$

with mean $E[\ln(\boldsymbol{\mathcal{M}})] = [E[\ln(\mathrm{MAF}_\alpha)], E[\ln(\mathrm{CV})], E[\ln(\mathrm{CS})]]$ and covariance matrix:

$$\boldsymbol{\Sigma}[\ln(\boldsymbol{\mathcal{M}})] = \begin{pmatrix} \mathrm{Var}[\ln(\mathrm{MAF}_\alpha)] & \mathrm{Cov}[\ln(\mathrm{MAF}_\alpha), \ln(\mathrm{CV})] & \mathrm{Cov}[\ln(\mathrm{MAF}_\alpha), \ln(\mathrm{CS})] \\ \mathrm{Cov}[\ln(\mathrm{MAF}_\alpha), \ln(\mathrm{CV})] & \mathrm{Var}[\ln(\mathrm{CV})] & \mathrm{Cov}[\ln(\mathrm{CV}), \ln(\mathrm{CS})] \\ \mathrm{Cov}[\ln(\mathrm{MAF}_\alpha), \ln(\mathrm{CS})] & \mathrm{Cov}[\ln(\mathrm{CV}), \ln(\mathrm{CS})] & \mathrm{Var}[\ln(\mathrm{CS})] \end{pmatrix}. \tag{A2}$$

Because *Merz et al.* (2008) provided $E[\boldsymbol{\mathcal{M}}]$ and $\mathrm{Var}[\boldsymbol{\mathcal{M}}]$, and not their logarithms, we used the transformation [see *Kottegoda and Rosso*, 1997, pp. 225–227]:

$$E[\ln(\boldsymbol{\mathcal{M}})] = \ln\left(\frac{E[\boldsymbol{\mathcal{M}}]}{\sqrt{\mathrm{Var}[\boldsymbol{\mathcal{M}}]/E[\boldsymbol{\mathcal{M}}]^2 + 1}}\right) \tag{A3}$$

$$\mathrm{Var}[\ln(\boldsymbol{\mathcal{M}})] = \ln\left(\mathrm{Var}[\boldsymbol{\mathcal{M}}]/E[\boldsymbol{\mathcal{M}}]^2 + 1\right).$$

While for the covariance terms, we used the sample correlations between the natural logarithm of the three moments estimated from regional data and indicated in Figure 4. The trivariate lognormal distribution of the maximum annual peak moments $\boldsymbol{\mathcal{M}}$ is obtained transforming the normal distribution of equation (A1) to

$$p(\boldsymbol{\mathcal{M}}) = \frac{p(\ln(\boldsymbol{\mathcal{M}}))}{|\mathrm{MAF}_\alpha \cdot \mathrm{CV} \cdot \mathrm{CS}|}, \tag{A4}$$

whose bivariate marginals are plotted in Figure 5.

[54] The transformation of the three-variate distribution $p(\boldsymbol{\mathcal{M}})$ into the prior distribution of the GEV parameters $\pi(\boldsymbol{\theta})$ is

$$\pi(\boldsymbol{\theta}) = p(\boldsymbol{\mathcal{M}}(\boldsymbol{\theta})) \cdot |\boldsymbol{J}(\boldsymbol{\theta})|, \tag{A5}$$

where $|\boldsymbol{J}(\boldsymbol{\theta})|$ is the Jacobian of the transformation $\boldsymbol{\theta} \to \boldsymbol{\mathcal{M}}$. The transformation $\boldsymbol{\theta} \to \boldsymbol{\mathcal{M}}$ is [*Stedinger et al.*, 1993, p. 18.17]

$$\mathrm{MAF}_\alpha(\boldsymbol{\theta}) \propto \theta_1 + \theta_2(1 - \Gamma(1+\theta_3))/\theta_3 \quad \text{for} \quad \theta_3 > -1$$

$$\mathrm{CV}(\boldsymbol{\theta}) = \frac{\sqrt{\theta_2^2\left(\Gamma(1+2\theta_3) - \left(\Gamma(1+\theta_3)^2\right)\right)/\theta_3^2}}{\mathrm{MAF}} \quad \text{for} \quad \theta_3 > -1/2$$

$$\mathrm{CS}(\boldsymbol{\theta}) = \mathrm{sign}(\theta_3)\left[\frac{-\Gamma(1+3\theta_3) + 3\Gamma(1+\theta_3)\Gamma(1+2\theta_3)}{\left(\Gamma(1+2\theta_3) - (\Gamma(1+\theta_3))^2\right)^{3/2}} + -\frac{2(\Gamma(1+\theta_3))^3}{\left(\Gamma(1+2\theta_3) - (\Gamma(1+\theta_3))^2\right)^{3/2}}\right] \quad \text{for} \quad \theta_3 > -1/3, \tag{A6}$$

where the proportionality in the first equation is due to the standardization of the MAF values. The corresponding Jacobian is

$$|\boldsymbol{J}(\boldsymbol{\theta})| \propto \left|\frac{(g_2 - g_1^2)\sqrt{\theta_2^2(-g_1^2 + g_2)/\theta_3^2}}{\theta_2(\theta_1 + (\theta_2(1-g_1))/\theta_3)(-g_1^2 + g_2)} \times \left\{\frac{-3(-2g_1^3 + 3g_1g_2 - g_3)(-2g_1^2d_1 + 2g_2d_2)}{2(-g_1^2 + g_2)^{5/2}} + \frac{-6g_1^3d_1 + 3g_1g_2d_1 + 6g_1g_2d_2 - 3g_3d_3}{(-g_1^2 + g_2)^{3/2}}\right\}\right|, \tag{A7}$$

with $g_i = \Gamma(1 + i\theta_3)$ and $d_i = \Psi(1 + i\theta_3)$, where $i = 1, 2, 3$, $\Gamma(x)$ is the gamma function, and $\Psi(x) = \frac{d}{dx}\ln\Gamma(x)$ is the digamma function.

Appendix B: Partial Information on the GEV Parameters

[55] In section 6, we argued that the information on Q_{500} correspond to partial information on the GEV parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$. The density $h(Q_{500})$ provides information

on one parameter given the other two, e.g., on $\pi(\theta_1|\theta_2,\theta_3)$, where

$$\pi(\theta_1|\theta_2,\theta_3) = \frac{\pi(\theta_1,\theta_2,\theta_3)}{\pi(\theta_2,\theta_3)}.$$

Because the functional form relating Q_{500} to the GEV parameters ($g(\theta_1,\theta_2,\theta_3)$ in equation (7)) is assumed to be known with certainty, then

$$\begin{aligned}\pi(\theta_1|\theta_2,\theta_3) &= \int_{\text{all } Q_{500}} \pi(\theta_1|\theta_2,\theta_3,Q_{500}) \cdot h(Q_{500}) dQ_{500} \\ &= \int_{\text{all } Q_{500}} \delta\left[\theta_1 - \theta_1^0(Q_{500})\right] \cdot h(Q_{500}) dQ_{500}\end{aligned},$$

where $\delta[x]$ is the Dirac delta function and $\theta_1^0(Q_{500})$ is the value of the scale parameter for which, given θ_2 and θ_3, $g\left(\theta_1^0,\theta_2,\theta_3\right) = Q_{500}$. Because $\delta[f(x)] = \delta[x - x_0]/|f'(x)|_{x_0}$, where f has a single root x_0, then

$$\begin{aligned}\pi(\theta_1|\theta_2,\theta_3) = \int_{\text{all } Q_{500}} &\delta[g(\theta_1,\theta_2,\theta_3) - Q_{500}] \\ &\cdot \left|\frac{\partial g(\theta_1,\theta_2,\theta_3)}{\partial\theta_1}\right|_{\theta_1^0(Q_{500})} \cdot h(Q_{500}) dQ_{500}.\end{aligned}$$

The two terms after the δ Dirac function depend on Q_{500} only; therefore, considering the properties of the δ Dirac, the integral is equal to evaluating these two terms on $g(\theta_1,\theta_2,\theta_3)$, i.e., noting that $\theta_1^0(g(\theta_1,\theta_2,\theta_3)) = \theta_1$,

$$\pi(\theta_1|\theta_2,\theta_3) = \left|\frac{\partial g(\theta_1,\theta_2,\theta_3)}{\partial\theta_1}\right| \cdot h(g(\theta_1,\theta_2,\theta_3))$$

Because $\partial g(\theta_1,\theta_2,\theta_3)/\partial\theta_1 = 1$, $\pi(\theta_1|\theta_2,\theta_3) = h(g(\theta_1,\theta_2,\theta_3))$.

[56] Analogously,

$$\begin{aligned}\pi(\theta_2|\theta_1,\theta_3) &= \left|\frac{\partial g(\theta_1,\theta_2,\theta_3)}{\partial\theta_2}\right| \cdot h(g(\theta_1,\theta_2,\theta_3)) \\ &= \left|\frac{1-c^{\theta_3}}{\theta_3}\right| \cdot h(g(\theta_1,\theta_2,\theta_3))\end{aligned} \tag{B1}$$

and

$$\begin{aligned}\pi(\theta_3|\theta_1,\theta_2) &= \left|\frac{\partial g(\theta_1,\theta_2,\theta_3)}{\partial\theta_3}\right| \cdot h(g(\theta_1,\theta_2,\theta_3)) \\ &= \left|\frac{\theta_2}{\theta_3}\left(\frac{1-c^{\theta_3}}{\theta_3} + c^{\theta_3}\ln c\right)\right| \cdot h(g(\theta_1,\theta_2,\theta_3))\end{aligned}, \tag{B2}$$

where $c = \ln(500/(500-1))$.

[57] During the MCMC random walk, θ_1, θ_2, and θ_3 are given at each step and either $\pi(\theta_1|\theta_2,\theta_3)$, $\pi(\theta_2|\theta_1,\theta_3)$, or $\pi(\theta_3|\theta_1,\theta_2)$ can be used by the MCMC analysis to compute the posterior distribution of the parameters. There is, therefore, a subjective choice involved, i.e., for what parameter is the information on Q_{500} used. However, by applying the procedure to the case study considered in this paper, we found that this choice has no significant effect on the results (not shown here).

[58] In the analyses in this paper, we have used $\pi(\theta_1|\theta_2,\theta_3) = h(g(\theta_1,\theta_2,\theta_3))$. Although the information is partial, one can see $h(g(\boldsymbol{\theta}))$ as a whole prior (although an improper one) by noting that $\pi(\theta_1,\theta_2,\theta_3) = \pi(\theta_1|\theta_2,\theta_3)\cdot\pi(\theta_2,\theta_3) = h(g(\theta_1,\theta_2,\theta_3)) \cdot 1 \cdot 1$, because flat priors are used for θ_2 and θ_3.

[59] **Acknowledgments.** Financial support for the project "Mountain floods—regional joint probability estimation of extreme events" funded by the Austrian Academy of Sciences, the Doctoral program DK-plus W1219-N22 funded by the Austrian Science Funds, and the "Flood-Change" project funded by the European Research Council are acknowledged. We would like to thank John England, Dmitri Kavetski, Francesco Laio, Alberto Montanari, Jery Stedinger, and an anonymous reviewer, whose comments enabled us to significantly improve the paper.

References

Benito, G., and V. Thorndycraft (2005), Palaeoflood hydrology and its role in applied hydrological sciences, *J. Hydrol.*, *313*(1–2), 3–15, doi:10.1016/j.jhydrol.2005.02.002.

Blöschl, G., C. Reszler, and J. Komma (2008), A spatially distributed flash flood forecasting model, *Environ. Model. Software*, *23*(4), 464–478.

Bobée, B., and P. F. Rasmussen (1995), Recent advances in flood frequency-analysis, *Rev. Geophys.*, *33*(S1), 1111–1116, doi:10.1029/95RG00287.

Brázdil, R., Z. W. Kundzewicz, and G. Benito (2006), Historical hydrology for studying flood risk in Europe, *Hydrol. Sci. J.*, *51*(5), 739–764, doi:10.1623/hysj.51.5.739.

Castellarin, A., D. H. Burn, and A. Brath (2001), Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, *J. Hydrol.*, *241*, 270–285.

Castellarin, A., R. M. Vogel, and N. Matalas (2005), Probabilistic behavior of a regional envelope curve, *Water Resour. Res.*, *41*, W06018, doi:10.1029/2004WR003042.

Castellarin, A., D. H. Burn, and A. Brath (2008), Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem?, *J. Hydrol.*, *360*, 67–76.

Chib, S., and E. Greenberg (1995), Understanding the Metropolis-Hastings algorithm, *Am. Stat.*, *49*(4), 327–335.

Choryński, A., I. Pińskwar, W. Kron, G. R. Brakenridge, and Z. W. Kundzewicz (2012), Catalogue of large floods in Europe in the 20th century, in *Changes in Flood Risk in Europe, IAHS Spec. Publ.*, vol. 10, edited by Z. W. Kundzewicz, chap. 3, pp. 27–54, CRC Press, Boca Raton, Fla.

Cohn, T. A., W. Lane, and W. Baier (1997), An algorithm for computing moments-based flood quantile estimates when historical flood information is available, *Water Resour. Res.*, *33*(9), 2089–2096.

Coles, S., L. Pericchi, and S. Sisson (2003), A fully probabilistic approach to extreme rainfall modeling, *J. Hydrol.*, *273*(1–4), 35–50, doi:10.1016/S0022-1694(02)00353-0.

Coles, S. G., and J. A. Tawn (1996), A Bayesian analysis of extreme rainfall data, *J. R. Stat. Soc., Ser. C Appl. Stat.*, *45*(4), 463–478.

Cunnane, C. (1988), Methods and merits of regional flood frequency analysis, *J. Hydrol.*, *100*(1–3), 269–290, doi:10.1016/0022-1694(88)90188-6.

D'Agostino, R., and M. Stephens (1986), *Goodness-of-Fit Techniques, Statistics: Textbooks and Monographs*, vol. 68, Dep. of Stat., South. Methodist Univ., Dallas, Tex.

Dalrymple, T. (1960), Flood frequency analyses, Water Supply Paper, vol. *1543-A*, *U. S. Geol. Surv.*, Reston, Va.

Deutscher Verband für Wasserwirtschaft und Kulturbau (DVWK) (1999), Statistische Analyse von Hochwasserabflüssen, *Tech. Rep. H. 251*, Bonn, Germany.

Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall (DWA) (2012), Ermittlung von Hochwasserwahrscheinlichkeiten, *Tech. Rep. DWA-M. 552*, Hennef, Germany.

Eagleson, P. S. (1972), Dynamics of flood frequency, *Water Resour. Res.*, *8*(4), 878–898.

El Adlouni, S., T. B. Ouarda, X. Zhang, R. Roy, and B. Bobée (2007), Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, *43*, W03410, doi:10.1029/2005WR004545.

England, J. F., J. D. Salas, and R. D. Jarrett (2003), Comparisons of two moments-based estimators that utilize historical and paleoflood data for

the log Pearson type III distribution, *Water Resour. Res.*, *39*(9), 1243, doi:10.1029/2002WR001791.

Fill, H. D., and J. R. Stedinger (1998), Using regional regression within index flood procedures and an empirical Bayesian estimator, *J. Hydrol.*, *210*(1–4), 128–145, doi:10.1016/S0022-1694(98)00177-2.

Fiorentino, M., and V. Iacobellis (2001), New insights about the climatic and geologic control on the probability distribution of floods, *Water Resour. Res.*, *37*(3), 721–730, doi:10.1029/2000WR900315.

Gaume, E., L. Gaál, A. Viglione, J. Szolgay, S. Kohnova, and G. Blöschl (2010), Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites, *J. Hydrol.*, *394*, 101–117, doi:10.1016/j.jhydrol.2010.01.008.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (2004), *Bayesian Data Analysis*, 2nd ed., Chapman and Hall, New York.

Genest, C., and J. V. Zidek (1986), Combining probability distributions: A critique and an annotated bibliography, *Stat. Sci.*, *1*(1), 114–135, doi:10.1214/ss.

Grimaldi, S., S.-C. Kao, A. Castellarin, S. M. Papalexiou, A. Viglione, F. Laio, H. Aksoy, and A. Gedikli (2011), Statistical hydrology, in *Treatise on Water Science*, vol. 2, edited by P. Wilderer, chap. 2.18, pp. 479–517, Elsevier, New York, doi:10.1016/B978-0-444-53199-5.00046-4.

Gutknecht, D., C. Reszler, and G. Blöschl (2002), Das Katastrophenhochwasser vom 7. August 2002 am Kamp—Eine erste Einschätzung [The August 7, 2002—flood of the Kamp—A first assessment], *Elektrotech Informationstech*, *119*(12), 411–413.

Gutknecht, D., G. Blöschl, C. Reszler, and H. Heindl (2006), Ein "Mehr-Standbeine"-Ansatz zur Ermittlung von Bemessungshochwässern kleiner Auftretenswahrscheinlichkeit [A "multi-pillar"-approach to the estimation of low probability design floods], *Österreichische Wasser Abfallwirtsch.*, *58*(H 3/4), 44–50.

Hazen, A. (1930), *Flood Flows: A Study of Frequencies and Magnitudes*, 199 pp., John Wiley and Sons, New York.

Hosking, J., and J. Wallis (1988), The effect of intersite dependence on regional flood frequency analysis, *Water Resour. Res.*, *24*(4), 588–600, doi:10.1029/WR024i004p00588.

Hosking, J., and J. Wallis (1997), *Regional Frequency Analysis: An Approach Based on L-Moments*, 224 pp., Cambridge Univ. Press, Cambridge, U. K.

Kirnbauer, R., S. Schnatter, and D. Gutknecht (1987), Bayesian estimation of design floods under regional and subjective prior information, in *Probability and Bayesian Statistics*, edited by R. Viertl, pp. 285–294, Plenum Press, New York.

Komma, J., C. Reszler, G. Blöschl, and T. Haiden (2007), Ensemble prediction of floods—catchment non-linearity and forecast probabilities, *Nat. Hazards Earth Syst. Sci.*, *7*(4), 431–444, doi:10.5194/nhess-7-431-2007.

Kottegoda, N. T., and R. Rosso (1997), *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, 735 pp., McGraw-Hill, New York.

Kuczera, G. (1982), Combining site-specific and regional information: An empirical Bayes Approach, *Water Resour. Res.*, *18*(2), 306–314, doi:10.1029/WR018i002p00306.

Kuczera, G. (1983), Effect of sampling uncertainty and spatial correlation on an empirical Bayes procedure for combining site and regional information, *J. Hydrol.*, *65*(4), 373–398, doi:10.1016/0022-1694(83)90088-4.

Kuczera, G. (1992), Uncorrelated measurement error in flood frequency inference, *Water Resour. Res.*, *28*(1), 183–188, doi:10.1029/91WR02269.

Kuczera, G. (1996), Correlated rating curve error in flood frequency inference, *Water Resour. Res.*, *32*(7), 2119–2127.

Kuczera, G. (1999), Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference, *Water Resour. Res.*, *35*(5), 1551–1557.

Kurothe, R., N. Goel, and B. Mathur (1997), Derived flood frequency distribution for negatively correlated rainfall intensity and duration, *Water Resour. Res.*, *33*(9), 2103–2107, doi:10.1029/97WR00812.

Laio, F. (2004), Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters, *Water Resour. Res.*, *40*, W09308, doi:10.1029/2004WR003204.

Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecast of continuous hydrological variables, *Hydrol. Earth Syst. Sci.*, *11*(4), 1267–1277.

Laio, F., P. Allamano, and P. Claps (2010), Exploiting the information content of hydrological "outliers" for goodness-of-fit testing, *Hydrol. Earth Syst. Sci.*, *14*, 1909–1917, doi:10.5194/hess-14-1909-2010.

Leese, M. N. (1973), Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series, *Water Resour. Res.*, *9*(6), 1534–1542, doi:10.1029/WR009i006p01534.

Lima, C. H., and U. Lall (2010), Spatial scaling in a changing climate: A hierarchical Bayesian model for non-stationary multi-site annual maximum and monthly streamflow, *J. Hydrol.*, *383*(3–4), 307–318, doi:10.1016/j.jhydrol.2009.12.045.

Madsen, H., and D. Rosbjerg (1997), Generalized least squares and empirical Bayes estimation in regional partial duration series index-flood modeling, *Water Resour. Res.*, *33*(4), 771–781.

Martins, E., and J. R. Stedinger (2000), Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resour. Res.*, *36*(3), 737–744, doi:10.1029/1999WR900330.

Martins, E., and J. R. Stedinger (2001), Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series, *Water Resour. Res.*, *37*(10), 2559–2567.

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

Merz, R., and G. Blöschl (2005), Flood frequency regionalisation—Spatial proximity vs. catchment attributes, *J. Hydrol.*, *302*(1–4), 283–306.

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

Merz, R., and G. Blöschl (2008b), Flood frequency hydrology: 2. Combining data evidence, *Water Resour. Res.*, *44*, W08433, doi:10.1029/2007WR006745.

Merz, R., G. Blöschl, and G. Humer (2008), National flood discharge mapping in Austria, *Nat. Hazards*, *46*, 53–72, doi:10.1007/s11069-007-9181-7.

Micevski, T., and G. Kuczera (2009), Combining site and regional flood information using a Bayesian Monte Carlo approach, *Water Resour. Res.*, *45*, W04405, doi:10.1029/2008WR007173.

Montanari, A. (2011), Uncertainty of hydrological predictions, in *Treatise on Water Science*, vol. 2, edited by P. Wilderer, chap. 2.17, pp. 459–478, Elsevier, New York.

Montanari, A., C. A. Shoemaker, and N. van de Giesen (2009), Introduction to special section on uncertainty assessment in surface and subsurface hydrology: An overview of issues and challenges, *Water Resour. Res.*, *45*, W00B00, doi:10.1029/2009WR008471.

Mudelsee, M., M. Börngen, G. Tetzlaff, and U. Grünewald (2004), Extreme floods in central Europe over the past 500 years: Role of cyclone pathway "Zugstrasse Vb," *J. Geophys. Res.*, *109*, D23101, doi:10.1029/2004JD005034.

Neppel, L., et al. (2010), Flood frequency analysis using historical data: Accounting for random and systematic errors, *Hydrol. Sci. J.*, *55*(2), 192–208, doi:10.1080/02626660903546092.

Natural Environment Research Council (NERC) (1975), Flood studies report, technical report, London.

O'Connell, D. R., D. A. Ostenaa, D. R. Levish, and R. E. Klinger (2002), Bayesian flood frequency analysis with paleohydrologic bound data, *Water Resour. Res.*, *38*(5), 1058, doi:10.1029/2000WR000028.

Ouarda, T. B., and S. El Adlouni (2011), Bayesian nonstationary frequency analysis of hydrological variables, *J. Am. Water Resour. Assoc.*, *47*(3), 496–505, doi:10.1111/j.1752-1688.2011.00544.x.

Parent, E., and J. Bernier (2003a), Bayesian POT modeling for historical data, *J. Hydrol.*, *274*(1–4), 95–108, doi:10.1016/S0022-1394(02)00396-7.

Parent, E., and J. Bernier (2003b), Encoding prior experts judgments to improve risk analysis of extreme hydrological events via POT modeling, *J. Hydrol.*, *283*(1–4), 1–18, doi:10.1016/S0022-1694(03)00080-5.

Payrastre, O., E. Gaume, and H. Andrieu (2011), Usefulness of historical information for flood frequency analyses: Developments based on a case study, *Water Resour. Res.*, *47*, W08511, doi:10.1029/2010WR009812.

Perreault, L., J. Bernier, B. Bobée, and E. Parent (2000a), Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited, *J. Hydrol.*, *235*(3–4), 221–241, doi:10.1016/S0022-1694(00)00270-5.

Perreault, L., J. Bernier, B. Bobée, and E. Parent (2000b), Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting, *J. Hydrol.*, *235*(3–4), 242–263, doi:10.1016/S0022-1694(00)00271-7.

Pilgrim, D. H., and I. Cordery (1993), Flood runoff, in *Handbook of Hydrology*, edited by D. R. Maidment, chap. 9, p. 42, McGraw-Hill, New York.

Reis, D. S. J., and J. R. Stedinger (2005), Bayesian MCMC flood frequency analysis with historical information, *J. Hydrol.*, *313*(1–2), 97–116.

Renard, B., M. Lang, and P. Bois (2006), Statistical analysis of extreme events in a non-stationary context via a Bayesian framework: Case study with peak-over-threshold data, *Stochastic Environ. Res. Risk Ass.*, *21*(2), 97–112, doi:10.1007/s00477-006-0047-4.

Reszler, C., G. Blöschl, and J. Komma (2008), Identifying runoff routing parameters for operational flood forecasting in small to medium sized catchments, *Hydrol. Sci. J.*, *53*(1), 112–129.

Ribatet, M., E. Sauquet, J. Grésillon, and T. B. Ouarda (2007), Usefulness of the reversible jump Markov chain Monte Carlo model in regional flood frequency analysis, *Water Resour. Res.*, *43*, W08403, doi:10.1029/2006WR005525.

Robert, C. P., and G. Casella (2004), *Monte Carlo Statistical Methods*, Springer, New York.

Robson, A. J., and D. W. Reed (1999), Statistical procedures for flood frequency estimation, in *Flood Estimation Handbook*, vol. 3, p. 338, Inst. of Hydrol. Crowmarsh Gifford, Wallingford, U. K.

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

Rogger, M., H. Pirkl, A. Viglione, J. Komma, B. Kohl, R. Kirnbauer, R. Merz, and G. Blöschl (2012b), Step changes in the flood frequency curve: Process controls, *Water Resour. Res.*, *48*, W05544, doi:10.1029/2011WR011187.

Seidou, O., T. B. Ouarda, M. Barbet, P. Bruneau, and B. Bobée (2006), A parametric Bayesian combination of local and regional information in flood frequency analysis, *Water Resour. Res.*, *42*, W11408, doi:10.1029/2005WR004397.

Sivapalan, M., G. Blöschl, R. Merz, and D. Gutknecht (2005), Linking flood frequency to long-term water balance: Incorporating effects of seasonality, *Water Resour. Res.*, *41*, W06012, doi:10.1029/2004WR003439.

Skøien, J. O., R. Merz, and G. Blöschl (2006), Top-kriging—Geostatistics on stream networks, *Hydrol. Earth Syst. Sci.*, *10*(2), 277–287.

Stedinger, J. R. (1983), Estimating a regional flood frequency distribution, *Water Resour. Res.*, *19*, 503–510, doi:10.1029/WR019i002p00503.

Stedinger, J. R., and T. A. Cohn (1986), Flood frequency analysis with historical and paleoflood information, *Water Resour. Res.*, *22*(5), 785–793.

Stedinger, J. R., and V. Griffis (2008), Flood frequency analysis in the United States: Time to update, *J. Hydrol. Eng.*, *13*(4), 199–204.

Stedinger, J. R., and G. D. Tasker (1985), Regional hydrologic analysis, 1, Ordinary, weighted, and generalized least squares compared, *Water Resour. Res.*, *21*(9), 1421–1432.

Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou (1993), Frequency analysis of extreme events, in *Handbook of Hydrology*, edited by D. R. Maidment, chap. 18, p. 66, McGraw-Hill, New York.

Taleb, N. N. (2007), *The Black Swan—The Impact of the Highly Improbable*, 366 pp., Random House, New York.

Tasker, G. D., and J. R. Stedinger (1989), An operational GLS model for hydrologic regression, *J. Hydrol.*, *111*(1–4), 361–375, doi:10.1016/0022-1694(89)90268-0.

U. S. Interagency Advisory Committee on Water Data (1982), *Guidelines for Determining Flood Flow Frequency: Bulletin 17B of the Hydrology Subcommittee*, Off. of Water Data Coordination, Reston, Va.

Vicens, G. J., I. Rodriguez-Iturbe, and J. C. J. Schaake (1975), A Bayesian framework for the use of regional information in hydrology, *Water Resour. Res.*, *11*(3), 405–414, doi:10.1029/WR011i003p00405.

Viglione, A., and G. Blöschl (2009), On the role of storm duration in the mapping of rainfall to flood return periods, *Hydrol. Earth Syst. Sci.*, *13*(2), 205–216, doi:10.5194/hess-13-205-2009.

Viglione, A., R. Merz, and G. Blöschl (2009), On the role of the runoff coefficient in the mapping of rainfall to flood return periods, *Hydrol. Earth Syst. Sci.*, *13*(5), 577–593, doi:10.5194/hess-13-577-2009.

Viglione, A., G. B. Chirico, J. Komma, R. A. Woods, M. Borga, and G. Blöschl (2010), Quantifying space-time dynamics of flood event types, *J. Hydrol.*, *394*, 213–229, doi:10.1016/j.jhydrol.2010.05.041.

Vogel, R. M., and I. Wilson (1996), Probability distribution of annual maximum, mean, and minimum streamflows in the United States, *J. Hydrol. Eng.*, *1*(2), 69–76.

Wagener, T., H. V. Gupta, and H. S. Wheater (2004), *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*, Imperial College Press, London.

Wiesbauer, H. (2004), Historische Überflutungen am Kamp, technical report, Wasser Niederösterreich, St. Pölten, Österreich, Germany.

Wiesbauer, H. (2007), Historische Überflutungen am Kamp, in *Extreme Abflussereignisse, Wiener Mitt.*, vol. 206, edited by D. Gutknecht, pp. 43–58, Inst. für Wasserbau und Ingenieurhydrol., Vienna, Austria.

Wood, E. F., and I. Rodriguez-Iturbe (1975), Bayesian inference and decision making for extreme hydrologic events, *Water Resour. Res.*, *11*(4), 533–542, doi:10.1029/WR011i004p00533.