https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Using structured expert judgment to Estimate extreme river discharges: a case study of the Meuse River

Guus Rongen[1,2], Oswaldo Morales-Nápoles[1], and Matthijs Kok[1,3]

[1]Civil Engineering and Geosciences, Delft University of Technology, The Netherlands
[2]Pattle Delamore Partners Ltd., New Zealand
[3]HKV consultants, The Netherlands

Correspondence: Guus Rongen (g.w.f.rongen@tudelft.nl)

Abstract.

Accurate estimation of extreme discharges in rivers, such as the Meuse, is crucial for effective flood risk assessment. However, existing statistical and hydrological models that estimate these discharges often lack transparency regarding the uncertainty of their predictions, as evidenced by the devastating flood event that occurred in July 2021 which was not captured by the existing model for estimating design discharges. This article proposes an alternative approach with a central role for expert judgment, using Cooke's method. A simple statistical model was developed for the river basin, consisting of correlated GEV-distributions for discharges in upstream sub-catchments. The model was fitted to expert judgments, measurements, and the combination of both, using Markov chain Monte Carlo. Results from the model fitted only to measurements were accurate for more frequent events, but less certain for extreme events. Using expert judgment reduced uncertainty for these extremes but was less accurate for more frequent events. The combined approach provided the most plausible results, with Cooke's method reducing the uncertainty by appointing most weight to two of the seven experts. The study demonstrates that utilizing hydrological experts in this manner can provide plausible results with a relatively limited effort, even in situations where measurements are scarce or unavailable.

1 Introduction

Quantifying the uncertainty that comes with estimating the magnitude of extreme flood events is a difficult matter. This became clear once more on the 18th of July 2021, when the flood wave on the Meuse River, following a few days of rain in the Eiffel and Ardennes, reached its highest peak ever measured at Borgharen. Unprecedented rainfall volumes fell within a short period of time (Dewals et al., 2021). These caused flash floods with large loss of life and extensive damage in Germany, Belgium, and to a lesser extent also in the Netherlands(Task Force Fact-finding hoogwater 2021, 2021; Mohr et al., 2022). The discharge at the Dutch border exceeded the flood events of 1926, 1993, and 1995. Contrary to those events, this flood occurred during summer, a season that is (or was) often considered irrelevant for extreme discharges on the Meuse. The event was thus surprising in multiple ways. This might happen when we experience a new extreme, but given that Dutch flood risk has safety standards up to once per 100,000 years (Ministry of Infrastructure and Environment, 2016) one would have hoped this to be less of a surprise. This event underlines the importance of understanding the uncertainty that comes with extreme flood events.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Quantifying events that are more extreme than ever measured (i.e., with return levels that are longer than the time period of representative measurements), requires extrapolating from available data or knowledge. For the Meuse, this is traditionally done by fitting a probability distribution to extreme events and extrapolating from it (van de Langemheen and Berger, 2001). However, a statistical fit of extremes is often very sensitive to the most extreme events in the measured time series. Additionally, extreme events might be of a different type (statistical population) compared to regular events, and therefore badly described by a model that is based on historical measurements only. Advances in computational power allow to narrow this (hydrological) uncertainty by, in the first place, generating long synthetic time series of rainfall or discharges (Leander et al., 2005; Diederen et al., 2019). These can then be used to simulate a chain of models to approximate values of river discharges and potential flooding (e.g. Falter et al., 2015; Borgomeo et al., 2015). For the Dutch rivers Meuse and Rhine, the GRADE instrument is used for this. It generates 50,000 years of rainfall and discharges (Hegnauer et al., 2014). Advantages of such methods are that it can create spatially coherent results, which can correct for changes in the catchment or climate. It is however still based on "re-sampling" available measurements or knowledge. Moreover, the computational resources required to simulate longer time series make it tedious or costly to quantify uncertainty in the method, let alone the uncertainty that comes with modelling choices. This is illustrated by the fact that the July 2021 discharge was not exceeded once in the 50,000 years of summer discharges generated by GRADE. The event could have been more extreme than a once in 50,000 year event, but a more likely cause is that the re-sampling approach uses (only) historic rainfall measurements to create new rain events. Underestimating uncertainty is however not only an issue of GRADE. Meresa and Romanowicz (2017) show that hydrological model uncertainty can be larger than climate projection uncertainty and Winter et al. (2018) indicate that flood model uncertainties can lead up to a factor 3 difference in outcome (1.4 for flood frequency distribution). de Boer-Euser et al. (2017); Bouaziz et al. (2020) compared different hydrological modelling concepts for the Ourthe catchment (considered in this study as well), and showed the large differences that different models can give when comparing more characteristics than only stream-flow.

While most studies aimed at obtaining better estimates of discharge extremes use hydrological or statistical modeling, some follow the approach of using expert judgment (EJ). (Sebok et al., 2021) recently applied expert elicitation to reduce uncertainty in climate model predictions. (Kindermann et al., 2020) reproduced historical water levels using structured expert judgment (SEJ), and (Rongen et al., 2022b) recently applied SEJ to estimate dike failure probabilities for the Dutch part of the river Rhine. While examples are not abundantly available, using prior information to decrease the uncertainty and sensitivity for extrapolation, is not new. Mostly this information comes from EJ. Three examples in which a similar, Bayesian, approach was applied to limit the uncertainty in extreme discharge estimates are given by (Coles and Tawn, 1996; Parent and Bernier, 2003; Viglione et al., 2013). The mathematical approaches vary between the different studies, but the rationale for using EJ is the same: adding "soft" or uncertain information to available measurements can help in achieving more plausible extreme estimates. In this study, we applied expert judgment as well, to estimate the magnitude of discharge events for the Meuse River up to an annual exceedance probability of on average once per 1,000 years. We aimed to get credible estimates of extreme discharge events for the Meuse that would otherwise require statistical extrapolation or complex modelling. A relatively simple model was quantified both with observed data and with expert estimates, in which the latter serves to decrease uncertainty in the extrapolated range. By using Cooke's method for structured expert judgment (SEJ) (Cooke and Goossens, 2008), participants

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

can use their own approach to make their estimates, which are then combined based on the experts' performance in questions for which the analysts know the answer or will know the answer *post hoc*. Next, we investigate if this expert judgment-based method gives plausible results. For this, we compared the model results based on expert estimates only, on measured data only and on both. The differences show the added value of each component. This indicates how good the method works, both when measurements are available and when they are not, for example in data scarce areas.

This study shows that the proposed method for estimating extreme river discharges with expert judgment leads to credible estimates for extremes. While the method that combines discharge measurements and expert estimates works best, the approach with equally weighted expert judgments only leads to plausible estimates as well, albeit with higher uncertainty in the extrapolated range.

2 Study area and available data

Figure 1 shows an overview of the catchment of the Meuse River. The catchment areas that discharge trough the major tributaries are outlined with red. the three locations for which we are interested in extreme discharge estimates, Borgharen, Roermond, and Gennep, are shown in blue. We call these the 'downstream locations' throughout this study. The river continues further downstream until it flows into the North Sea at Rotterdam. This part of the river becomes increasingly intertwined with the Rhine River and more affected by the downstream sea water level. The water levels consequently can be ascribed decreasingly to the discharge from the upstream catchment. For this reason, we do not go further downstream than Gennep in this study.

The numbered dots indicate the locations along the tributaries where the discharge of the upstream sub-catchments are measured. The gauge locations' and catchments' names are shown on the lower left. The Semois catchment is part of the French Meuse catchment. The discharge estimates for this catchment are therefore only used for expert calibration, as the flow is part of the French Meuse flow.

Elevation is shown with the grey-scale. Data for this were obtained from EU-DEM ((Copernicus Land Monitoring Service, 2017)) and used to derive catchment delineation and tributary steepness. More hydrological characteristics, such as land use (from: (Copernicus Land Monitoring Service, 2018)), subsoil (Food and Agriculture Organization of the United Nations, 2003), rainfall statistics (Copernicus Land Monitoring Service, 2020) and hydrograph shapes (from discharge data, see below), that were provided to the experts, are shown in the supplementary information. This information was provided to the experts to help them make estimates. The discharge data needed for fitting the model to observations were obtained from Service public de Wallonie (2022) for the Belgian gauges, Waterschap Limburg (2021); Rijkswaterstaat (2022) for the Dutch gauges, and Land NRW (2022) for the German gauge. These discharge data are mostly derived from measured water levels and rating curves. During floods, water level measurements can be incomplete and rating curves inaccurate. Consequently, discharge data during extremes can be unreliable. Measured data were not provided to the experts, except in qualitative form as hydrograph shapes.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

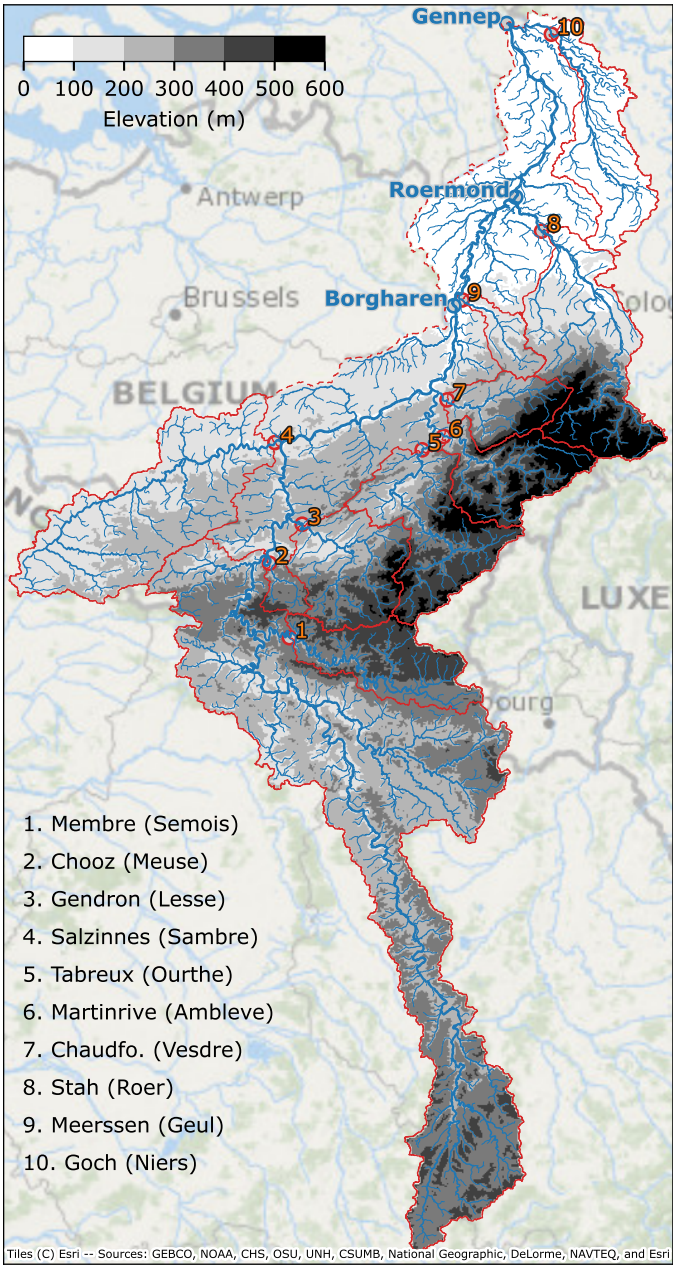

Figure 1. Map of the Meuse catchment considered in this study, with main river, tributaries, streams, and catchment bounds.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

3 Method for estimating extreme discharges with experts

3.1 Probabilistic model for estimating extreme discharges

To obtain estimates for downstream discharge extremes, experts needed to quantify a simple model that states that the downstream discharge is the sum of the tributary discharges, times a factor correcting for covered area and hydrodynamics:

$$Q_d = f_{\Delta t,u} \cdot \sum_j Q_u, \tag{1}$$

where Q_d is the peak discharge of a downstream location during an event, and Q_u the peak discharge of the u'th (upstream) tributary during that event. With the factor $f_{\Delta t,u}$ experts can compensate for inaccuracies in estimation. These are, for example, the time difference between discharge peaks and peak attenuation as the flood wave travels through the river (resulting in a lower factor), or rainfall on the Meuse catchment area that is not covered by one of the tributaries (leading to a higher factor). The factor can therefore be lower or higher than 1.0. The model is deliberately kept simple, so that the consequences of the experts' estimates remain traceable for them. Location d can be any location along the river where the discharge is assumed to be dependent mainly on rainfall in the upstream catchment. During an event, the tributary peak discharges Q_u are related; a rainfall event will likely span an area larger than the tributary, and therefore cause a lower or higher discharge in multiple tributaries. Additionally, hydrological characteristics of the catchments are similar for neighboring tributaries. These dependencies are modelled with a multivariate Gaussian distribution that is realized through Bayesian Networks estimated by the experts. The details of this concern the practical and theoretical aspects of eliciting dependence with experts and are beyond the scope of this article. They will therefore be presented in a separate article that is yet to be published. The resulting correlation matrices, presented in appendix C, are nonetheless used for calculating the discharge statistics in this study.

Each random variable of this multivariate model is a univariate (marginal) distribution. We use the generalized extreme value distribution (GEV) as a statistical model for extreme discharges (Jenkinson, 1955). This family of distributions is used because it is suitable to estimate the probabilities of extreme events and gives flexibility by varying between Frechet, Gumbel and Weibull distributions (i.e., heavy tailed, exponentially tailed, and light tailed, respectively) to fit specific catchment properties. Section 3.4 explains how downstream discharges are generated from these model components, including uncertainty bounds.

The model was described in (Rongen et al., 2022a) as well, where it was used in a solely data-driven context.

3.2 Assessing uncertainties with expert judgment

We use Cooke's method for structured expert judgment. Cooke's model is a method to elicit and combine expert judgments based on empirical control questions, with the aim to find a single, combined, estimate for the variable of interest (rational consensus). The approach is extensively described in (Cooke and Goossens, 2008), here we discuss some of the basic elements of the method.

The expert makes an uncertainty estimate for each question by estimating a number of percentiles. In this study (and in most others) these are the 5th, 50th and 95th percentile, that are combined into a probability density function (PDF) f_{exp} for each expert for each variable of interest. With expert's answers to seed or calibration variables (unknown to the experts at the moment

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

of the elicitation), Cooke's method assigns a weight to each participating expert. The expert weight, $w_\alpha(e)$, is calculated by multiplying the *calibration* and *information* scores, if the experts calibration score is above the chosen significance level:

$$w_\alpha(e) = 1_\alpha \times \text{calibration score}(e) \times \text{information score}(e). \quad (2)$$

The calibration score is calculated from the questions for which the answer is known by the researcher but not by the participants at the moment of the elicitation. These are referred to as seed or calibration variables. Calibration is a measure of the statistical accuracy of the expert. The information score expresses the precision of an expert's answers. Percentile estimates that are close together are more precise, more informative, and get a higher information score. In this study, the seed variables are the discharges that are exceeded on average once per 10 years, according to the measurements. An example of a seed question is: "What is the discharge that is exceeded on average once per 10 years, for the Vesdre at Chaudfontaine?". Discharges with a 10 year recurrence interval are exceptional, but can in general still be well approximated from the available data. The information score is calculated from all questions, and is a measure of the degree of uncertainty of the experts answer. The decision maker (DM) is a combination of the experts' estimates. The expert contributes to the DM by the assigned weight, the product of the calibration score and information score:

$$\text{DM}_\alpha(i) = \sum_e w_\alpha(e) f_{e,i} \Big/ \sum_e w_\alpha(e). \quad (3)$$

This is called the global (GL) DM. Alternatively, the experts can be given the same weight, which results in the equal weight (EQ) DM. This does not require eliciting seed variables but neither does it distinguish experts based on their performance, a key aspect of Cooke's method. Other methods for expert elicitation could have been used as well. A well-known alternative to Cooke's method is the Delphi method (Brown, 1968), in which experts estimate and discuss in rounds, until consensus is reached on the estimates. Another option is a Bayesian approach, as described in Hartley and French (2021). We chose Cooke's method because of its strong mathematical base and track record (Colson and Cooke, 2017), and the authors' familiarity with this method.

We created PDFs from the experts' quantile estimates by fitting a Metalog distribution trough the percentiles (Keelin, 2016). This distribution allows to fit any three percentile estimate without changing the estimates. For symmetric estimates, the distribution is bell-shaped. For asymmetric ones, it becomes left or right skewed. Normally, in Cooke's method, the PDF is created by assuming a uniform distribution in between the percentiles (minimum information). This leads to a piece-wise linear cumulative distribution, where the Metalog gives a smooth fit. An example of using the Metalog distribution in an expert elicitation study was described by Dion et al. (2020).

7 experts participated in the in-person elicitation that took place on the 4th of July 2022. The study and model where discussed before making the assessments to make sure that the study concepts and questions were clear. After this, an training exercise for the Weser catchment was done in which the experts needed to answer four questions that were discussed afterwards. This way, the experts could see how their answers compared to the realizations, and subsequently what their scores in Cooke's method were. Apart from the training exercise, the experts answered 26 questions, 10 questions for the discharge that is exceeded on average once per 10 years (1 for each tributary), 10 questions for the discharge exceeded on average once per 1000 years, and

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Table 1. List of experts with their affiliation and professional interests.

Name	Affiliation	Specialism
Alexander Bakker	Rijkswaterstaat & Delft University of Technology	Risk analysis for storm surge barriers, extreme value analyses, climate change and climate scenario's.
Eric Sprokkereef	Rijkswaterstaat	Coordinator crisis advisory group Rivers. Operational forecaster for Rhine and Meuse
Ferdinand Diermanse	Deltares	Expert advisor and researcher flood risk.
Helena Pavelková	Waterschap Limburg	Hydrologist
Jerom Aerts	Delft University of Technology	Hydrologist, focussed on hydrologic modelling on a global scale. PhD candidate.
Nicole Jungermann	HKV consultants	Advisor water and climate
Siebolt Folkertsma	Rijkswaterstaat	Advisor in the Team Expertise for the River Meuse

6 for the factors between upstream sum and downstream discharge (2 return periods $\times$ 3 locations). Note that we will use the shorthand 10-year and 1,000-year notation in the remainder of this article, when indicating the 'return period' of average recurrence interval of discharges. A list of the 7 participants and their affiliations is shown in table Table 1. The alphabetic order in which the experts are listed holds no relation to the number in which experts are labelled in the analysis carried out in this article.

3.3 Determining model coefficients with Bayesian inference

We distinguished three approaches for fitting the model from Eq. 1,

- the 'data-only' approach, in which only measured discharges (the annual maxima per tributary that lead to a peak discharge at Borgharen) were used,
- the 'EJ-only' approach, in which only the expert's estimate for the 10-year and 1,000-year discharge is used, and
- the combined 'data and EJ' approach, in which the measured discharges are combined with the expert estimate for the 1,000-year discharge (not the 10-year discharge).

The fitting of the three options is performed using the Bayesian inference technique Markov-Chain Monte Carlo (MCMC). This results in an inference trace of the parameters of the GEV distribution (i.e., the PDF for the tributary discharges). This trace is the empirical joint probability distribution of the parameters (the GEV has three parameters) that is used for sampling (see Sect. 3.4). MCMC is the name commonly used for a group of algorithms used to sample from a PDF. One example of such algorithms is the Metropolis-Hastings algorithm Hastings (1970). This Bayesian technique combines *a-priori* distributions with observations to estimate *a-posteriori* distributions. We used a weakly informed prior for the GEV distribution, which is described in Appendix A in more detail. These were then updated with observations, with expert estimates, or with both.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

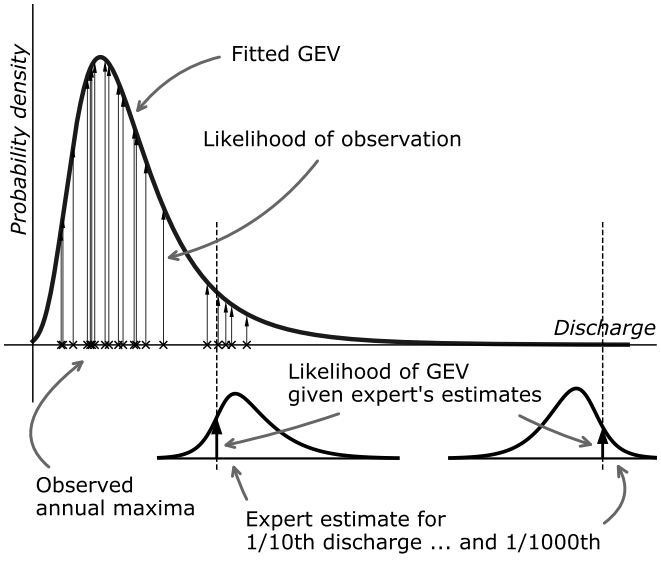

Figure 2. Conceptual visualization of elements in the likelihood-function of a GEV-distribution.

Just like most curve fitting algorithms, MCMC uses a (log-) likelihood-based criterion to find a good fit. Figure 2 shows how the likelihood of a specific GEV distribution is calculated, based on the observations and expert estimates. This calculation is implemented as a custom likelihood function in the used Python-package for Bayesian statistical modelling PyMC3 Salvatier et al. (2015). The top curve $f(Q|\theta)$ represents a fitted GEV-distribution with parameter vector θ. The log-likelihood of the parameters that give this specific GEV can be calculated with the product of the probability density function of the observations (i.e., the product of the length of the arrows in the figure):

$$\ell(\theta|q) = \sum_i \log\left(f_\theta(q_i)\right). \tag{4}$$

The best fit of the curve is the set of parameters θ for which the log-likelihood ℓ, given the observations, q is maximal. The MCMC sampling algorithm gives a (joint) probability density function of θ, instead of a single value as a maximum likelihood estimate would.

The log-likelihood of the expert's estimate is calculated as follows:

1. Given a GEV-distribution $f_\theta(Q)$, the discharges q for one or two annual exceedance probabilities are calculated (1/10 and 1/1,000 for EJ-only, 1/1,000 for data and EJ combined). These discharges correspond to an average recurrence interval of 10 and 1,000 year.

2. The expert is asked for an estimate of the discharge with this exceedance frequency resulting in the distribution f_{exp}. These estimates are displayed by the curves on the bottom of Fig. 2.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

3. The likelihood of the GEV-quantile can then be calculated in the expert estimated distribution:

$$\ell(\theta|exp) = \sum_j \log\left(f_{exp,j}(q_{p_j})\right) = \sum_j \log\left(f_{exp,j}\left(F_\theta^{-1}(1-p_j)\right)\right) \quad (5)$$

where p_j is the exceedance probability for quantile j, and q_{p_j} the discharge corresponding to that exceedance probability based on the GEV-distribution f_θ. F_θ is the cumulative distribution function (CDF), so its inverse, F_θ^{-1}, is the quantile function.

By summing the likelihood in equations 4 and 5, the likelihood of the distribution given both the observations and expert opinions is calculated. This is an unbalanced sum because there are many more observations contributing to that part of the likelihood, than there are expert estimates (only one when combining data and EJ). We therefore add a factor $10/N_{obs}$ to the likelihood function that weights the observations as if only 10 were measured. Note that 10 is still much more than the single expert estimate. The estimate is however made for the 1,000-year discharge, which gives it more weight due to the cantilever effect (i.e., a small parameter variation results in a large difference for the q_{1000}, and therefore the $\ell(\theta|exp)$ as well). We found that 10 gave a good balance between observations and expert estimates but we realize that it is somewhat subjective. Appendix B shows a sensitivity analysis of the MCMC-fit to substantiate the choice for this factor. The complete likelihood function that is used to fit a distribution to both data and expert estimates, is:

$$\ell(\theta|q, exp) = \frac{10}{N_i} \cdot \sum_i \log\left(f_\theta(q_i)\right) + \log\left(f_{exp}\left(F_\theta^{-1}(0.999)\right)\right) \quad (6)$$

For the factor between the tributaries' sum and the downstream discharge, we distinguished between observations and expert estimates as well. A log-normal distribution was fitted to the observations. This responds to a practical choice for a distribution of positive values with sufficient shape flexibility. The experts estimated a distribution for the factor as well, which is used directly for the experts-only fit. For the combined model fit, the log-normal was used up to the 10-year range, and the expert estimate for the 1,000-year factor. In between, the factor was interpolated, and for recurrence intervals larger than 1,000 years, the factor was extrapolated. During the experts session, a participant asked to make a different estimate for the factor at the 10-year event and 1,000-year event, a distinction that initially was not planned. Following the request, we changed the questionnaire such that a factor could be specified at both return periods. One expert used the option to make the distinction.

Finally, for the correlation matrix describing the dependence between tributary extremes, we have the observed correlations, used for the data-only option, and the expert-estimated correlations, used for the expert-only option. For the combined option, we simply take the average of the observed correlation matrix and the expert-estimated correlation matrix. Other possibilities for combining correlation matrices are available (e.g., Al-Awadhi and Garthwaite, 1998), however these go beyond the goal of this study. Notice also that in this study we investigate a "50-50" contribution from data and experts.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

3.4 Calculating the downstream discharges

With the fitted model components described in Section 3.1, we calculated the discharge at a downstream location. To calculate a single exceedance frequency curve, we used a Monte Carlo approach in which the following samples for the $N_T = 9$ tributaries, $N_Q = 10,000$ discharge events, and $N_M = 8,000$ MCMC parameter combinations were combined:

1. N_T draws from the dependence model transformed to the $[1, N_M]$ interval. These are used as index to pick a GEV parameter combination from each tributary's inference trace after these sorting these based on the once per 1,000 year discharge. Doing this leads to relatively similar (1,000-year) discharges for tributaries with a strong dependence.

2. $N_T \times N_Q$ draws from the dependence model. These events (on a standard normal scale) are transformed to the discharge realizations for each tributaries GEV parameter combination.

3. N_Q draws from the factor between upstream sum and downstream discharge, used to multiply the sum of the upstream discharges to get the downstream discharge.

Assigning exceedance frequencies to the N_Q discharges using plot positions gives an exceedance frequency curve (plot positions from (Bernard and Bos-Levenbach, 1955) were used). Repeating this procedure 2000 times, and calculating the percentiles per exceedance frequency, results in the uncertainty intervals of the exceedance frequency curves.

Note that in the sampling approach, the correlation model is not only used to model tributary dependence within individual events but also to model the dependence between different tributaries' GEV parameter combinations. With this we express our assumption that the correlation between tributaries is present as well in the uncertainty intervals. After all, the uncertainty in the fitted distributions is mainly the result from the largest observed events. Having a time series with (or without) high discharge(s) in multiple tributaries affects the fitted distributions in a similar way for these different tributaries.

4 Experts' performance and resulting discharge statistics

In this section, we first present the expert scores for Cooke's method, and the experts' rationale for answering the questions. After this, the extreme value results for the tributaries and downstream locations are presented.

4.1 Results for Cooke's method

The experts estimate three-percentiles (5^{th}, 50^{eth} and 95^{th}) for the 10 and 1,000-year discharges, for all larger tributaries in the Meuse catchment. The 10-year estimate is used for calibration. An overview of the answers is given in the supplementary material. Based on these estimates the scores for Cooke's method are calculated. The results are shown in table 2.

The calibration score, a measure for the expert's statistical accuracy, varies between $2.3 \cdot 10^{-8}$ for expert C to 0.683 for expert D. Two experts have a score above a significance level of 0.05. Figure 3 shows the position of each realization (answer) within the experts' estimates. There were 10 calibration variables assessed each with three percentiles. A well calibrated score would capture the realizations to these seed variables accordingly to the mass in each inter-quantile bin. Thus, one below the 5^{th}

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Table 2. Scores for Cooke's method, for the experts (top 7 rows) and decision makers (bottom 3).

	Calibration score	Information score		Comb. score
		All	Calibr.	
Exp A	0.000799	1.605	1.533	0.00123
Exp B	0.000456	1.576	1.633	0.000745
Exp C	$2.3 \cdot 10^{-8}$	1.900	1.868	$4.4 \cdot 10^{-8}$
Exp D	0.683	0.711	0.626	0.427
Exp E	0.192	1.395	1.263	0.242
Exp F	0.000456	1.419	1.300	0.000593
Exp G	0.00629	1.302	1.232	0.00775
GL (opt)	0.683	0.659	0.670	0.458
GL	0.683	0.648	0.661	0.452
EQ	0.493	0.537	0.551	0.271

percentile, 4 in between the 5th and the median, four between the median and the 95th and one above the 95th. A concentration of dots on both ends indicates overconfidence (too narrow estimates, resulting in realizations outside of the bounds.) We can see that most experts tend to underestimate, since most realizations are higher than their estimated 95th percentile.

The information scores show less variation, as is usually the case. The expert with the highest calibration score (expert D) also has the lowest information score. Expert E, who has a high calibration as well, provided more concentrated answers, resulting in a higher information score.

The variation in the three decision makers (DMs) shown is limited. Optimizing the DM (i.e., excluding experts based on calibration score to improve the DM-score) has a limited effect: Only expert D and E remain, resulting in more or less the same results as when including all experts even when some of them contribute with "marginal" weights. The equal weights DM results in a good outcome; a high calibration score with a slightly lower information score compared to the other two DMs. The item weights DM, which allocates more weight to precise answers, results in the same DM calibration score as the global weights DM. We chose not to use it as it assumes more precise (i.e., confident) answers are better, something we did not want to assume for this case study.

We present the model results by fitting it to i) only data, ii) only expert estimates, and iii) the two combined as described in Section 3.3. We used the global weights DM for the data and experts option (iii). This means the experts' estimates for the 10-year discharges were used to assess the value of the 1,000-year answer. For the experts-only option, we used the equal weights DM, because using the global weights emphasizes estimates matching the measured data in the 10-year range. This would indirectly lead to including the measured data in the fit. By using equal weights, we ignore the relevant calibration questions, a situation that could ultimately be used when differential weighting of expert judgments is not considered.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

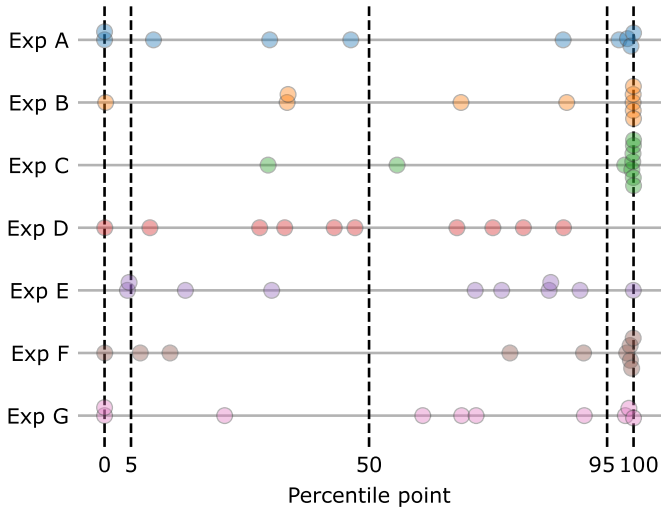

Figure 3. Seed questions realizations' compared to each expert's estimates. The position of each realization is displayed as percentile point in the expert's distribution estimate.

4.2 Rationale for estimating tributary discharges

We asked the experts to briefly describe the procedure they followed for making their estimates. From their responses we distinguished three approaches. The first was making, or thinking, of a simple conceptual hydrological model, in which the discharge follows from catchment characteristics like (a subset of) area, rainfall, evaporation and transpiration, precipitation-hydrograph response, land-use, subsoil, slope, or the presence of reservoirs. Most of this information was provided to the experts, and if not so, they made estimates for it themselves. A second approach was to compare the catchments to others that are known by the expert, and possibly adjusting the outcomes based on specific differences. A third approach was using rules of thumb, such as the expected discharge per square kilometer of catchment or a 'known' factor between an upstream tributary discharge and a downstream discharge (of which the statistics are better known). For estimating the 1,000-year discharge, the experts had to do some kind of extrapolation. Some experts scaled with a fixed factor, while others tried to extrapolate the rainfall, for which empirical statistics where provided.

Figure 4 gives an impression of how the different approaches lead to different answers per tributary. It compares the 50th percentile of the discharge estimates per tributary, by dividing them through the catchment area.

For estimating the factor between the tributaries' sum and the downstream discharge, experts mainly took into consideration that not 100% of the area is covered by (i.e., it flows through) the locations for which the discharge-estimates were made and that there is a time difference between the downstream 'arrival' of the tributary peaks. Additional aspects noted by the experts were the effects of flood peak attenuation and spatial dependence between tributaries and rainfall.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

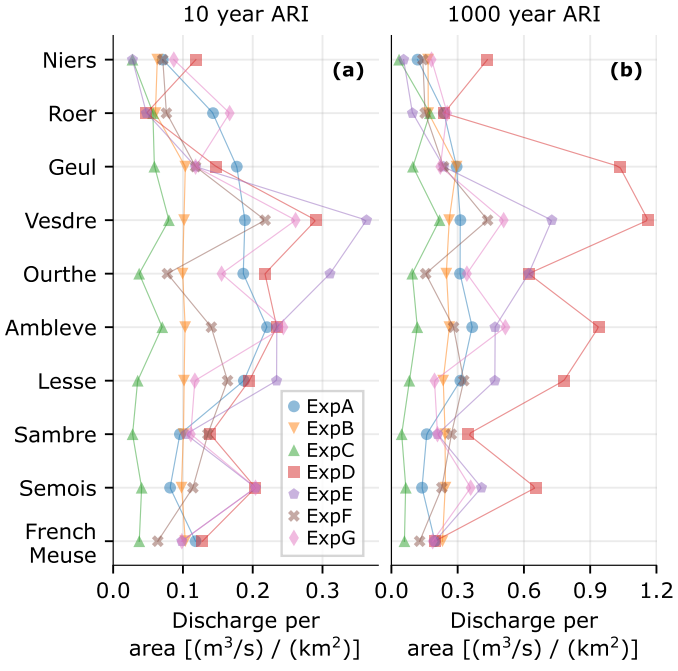

Figure 4. Discharge per area for each tributary and experts, based on the estimate for the 50th percentile. **(a)** for the 10-year, and **(b)** for the 1,000-year discharge. The lines are displayed to help distinguish overlapping markers.

285 The last part of the estimations required in order to calculate discharges at Borgharen is the dependence structure (correlations) between the tributaries. Experts estimated a correlation matrix for this by using a Non-parametric Bayesian Network. The resulting correlation matrices are shown in appendix C.

4.3 Extreme discharges for tributaries

Based on the procedures described in this paper including experts' estimates we calculated the extreme discharge statistics for 290 each of the tributaries. Figure 5 shows the results for Chooz and Chaudfontaine (left and middle column), a larger not to steep tributary, and a smaller steep tributary. The results for the other tributaries are shown in the supplementary information for all experts and DMs.

The top row **(a, d, g)** in Fig. 5 shows the uncertainty interval of these distributions when fitted only to the discharge measurements. The outer colored area is the 95% interval, the inner, darker, area the 50% interval, and the thick line the median value. 295 The second row **(b, e, h)** shows the fitted distributions when only expert estimates are used. The bottom row **(c, f, i)** shows the combination of expert estimates and data. The data-only option closely matches the data in the return period range where data are available, but the uncertainty interval grows for return periods further outside sample. Opposite to this, the experts-only option shows much more variation in the "in sample" range, while the out of sample return periods are more constrained. The

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

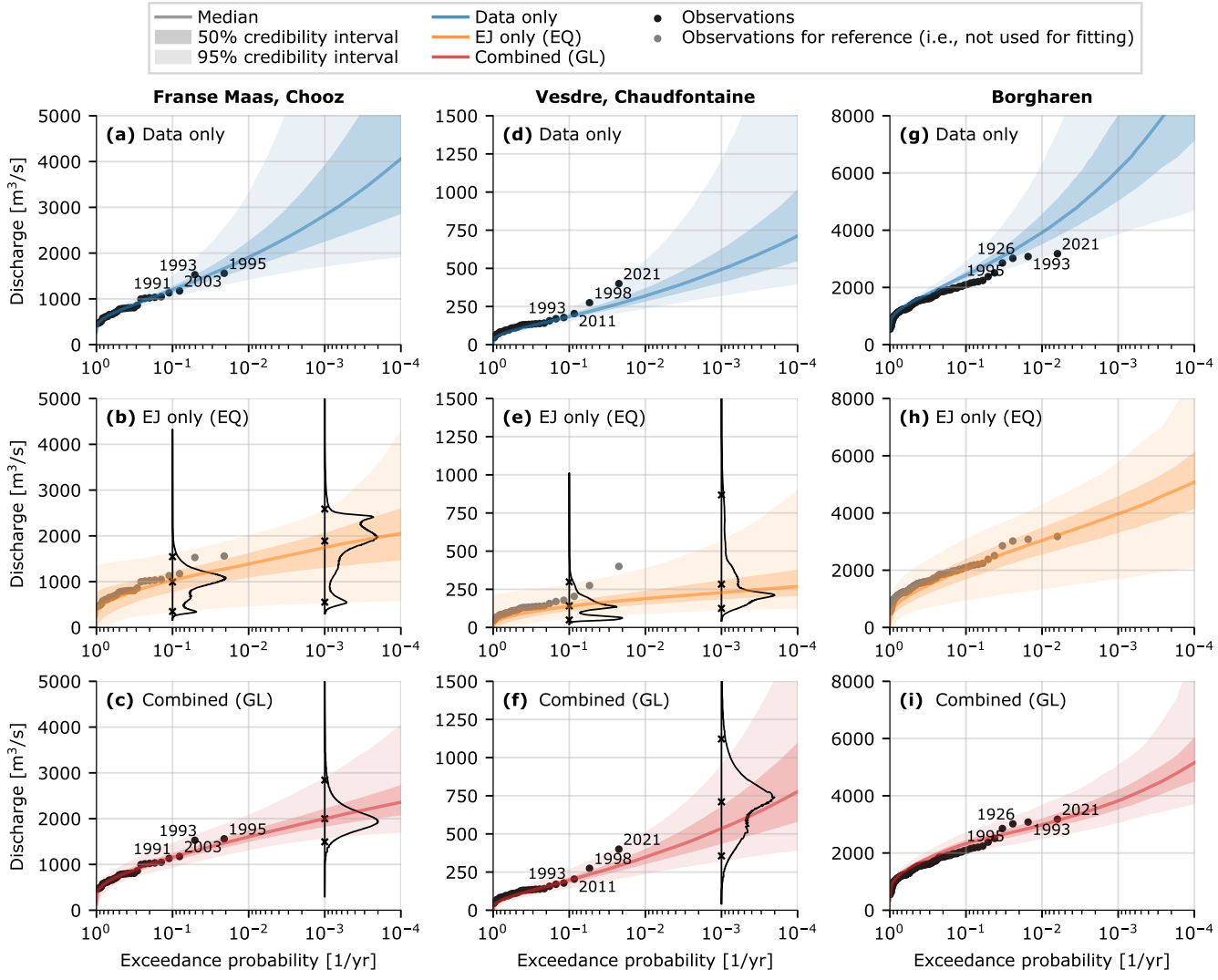

Figure 5. Extreme discharge statistics for Chooz **(a, b, c)**, Chaudfontaine **(d, e, f)** and Borgharen **(g, h, i)**. **(a, d, g)** represent data only, **(b, e, h)** expert judgment only, and **(c, f, i)** the data and expert judgment combined.

combined option is accurate in the “in sample" range, while the influence of the DM estimates is visible in the 1,000 year return period range.

4.4 Extreme discharges for Borgharen

Combining all the marginal (tributary) statistics with the factor for downstream discharges and the correlation models estimated by the experts, we get the discharge statistics for Borgharen. The results for this are shown in Fig. 5 (g, h, i).

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

As with the statistics of the tributaries, we observe high accuracy for the data-only estimates in the in sample range, constrained uncertainty bounds for EJ-only in the range with higher return periods, and both when combined. The data-only results deviate from the observations for frequent events as well. This is the result of the wide uncertainty in the marginal distributions; a single large event at one of the larger tributaries (e.g. the French part of the Meuse) can by itself cause a large discharge event. Sampling from these wide uncertainty bounds will therefore (too) often result in a high discharge event. The combined results match the historical observations very well. Note that this is not self-evident as the distributions were not fitted directly to the observed discharges at Borgharen but rather obtained through the dependence model for individual catchments and equation 1. The EJ-only estimate gives a much wider uncertainty estimate, but the median ('best estimate') matches the observations surprisingly well given that the model was not directly fitted to the data.

Zooming in on the discharge statistics for the downstream location Borgharen, we consider the 10, 100, and 1,000-year discharge. Figure 6 shows the probability distributions (smoothed with a kernel density estimate) for the discharges at this location.

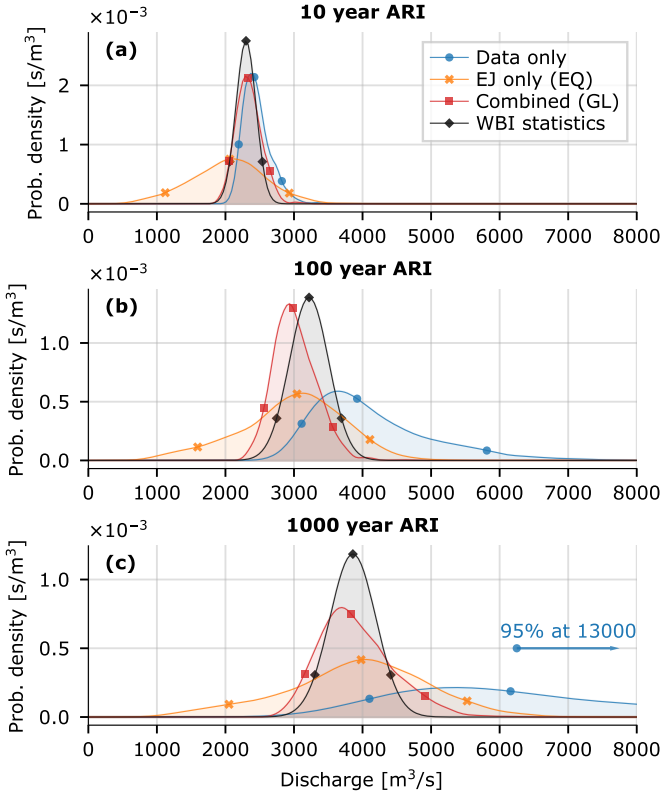

Figure 6. Kernel density estimates for the 10-year **(a)**, 100-year **(b)**, and 1,000-year **(c)** discharge for Borgharen. The dots indicate the 5, 50th and 95th percentile.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Comparing the same three variants, we see the data-only option is much too uncertain, with a 95% credibility interval of 6,000 to around 11,000 m^3/s for the 1,000-year discharge. A Meuse-discharge of 4,000 m^3/s will likely flood large stretches along the Meuse in the Dutch province Limburg, while a discharge of 5,000 m^3/s also floods large areas further downstream (Rongen, 2016). For discharges higher than 6,000 m^3/s the simple sum-model (Eq. 1) should be reconsidered, as the hydrodynamic properties of the system change due to upstream flooding.The combined results are surprisingly close to the currently used GRADE-statistics for dike assessment; the uncertainty is slightly larger but the median is close. The EJ-only results are less precise, but the median values ('best estimate') are similar to the combined results and GRADE-statistics. The large uncertainty is largely the results of equally weighting all experts, instead of assigning most weight to expert D and E (the global DM). The experts' results for Roermond and Gennep are similar to those for Borgharen, and therefore not presented here. They are however displayed in the supplementary material.

5 Discussion

The discharge estimates that result from this study show the value of fitting a relatively simple statistical model to both data and expert estimates. The predictive power of such models usually diminishes in the extrapolated range, but this is greatly improved by combining it with expert estimates. The data themselves help to increase the precision in the frequent range and it can point out the statistically accurate experts to improve the extrapolated range.

To test the model without data, we used an equal weight decision maker and left out the data in the fitting procedure. Note that the equal weight DM is a conservative choice, as the experts' statistical accuracy could still be determined based on another catchment where data are available. While the marginal distributions present wide bandwidths, the final results for Borgharen gave an accurate result, albeit with a large bandwidth (low precision).

The discharge statistics at Borgharen currently used for dike assessment give a lower and less uncertain range for the 1,000-year discharge (Hegnauer and Van den Boogaard, 2016). The estimates given in this study present larger uncertainty bands and indicate higher extreme discharges. This might be a consequence of the fact that we did not show the measured tributary discharges to the experts. This was a choice made in order to still be able to draw conclusions regarding the method without data. These measurements could have helped the experts in making "less uncertain" estimates. On the other hand, if the currently used statistics would be derived again including the July 2021 event, it is not unlikely that they would appoint more probability to slightly higher discharges.

Using expert judgment to provide answers for a model (like we did) can still give the analyst a large influence in the results. We try to keep the model transparent and provide the experts with unbiased information, but by defining the model on beforehand and choosing which information we provide, the participants are steered towards a certain way of reasoning. Every step in the method; such as the choice for a GEV-distribution, the dependence model, or the choice for Cooke's method', affects the end result. By presenting the method and providing background information explicitly, we hope to make this transparent and show the use of the method for similar applications.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

6 Conclusions

This study sets out to estimate extreme river discharges with expert judgment in a case study of the River Meuse. Experts' estimates of tributary discharges for large return periods were combined with measured high river discharges in ranges that are commonly "in sample". We combined the different tributary discharges with a multivariate correlation model describing their dependence, and compared the results for three approaches, i) data only, ii) expert judgment only, and iii) the combination. We used Cooke's method for structured expert judgment, in which the experts estimated the discharge that is exceeded on average once per 10 and 1,000 years. The once per 10 year estimate is in the observed range and was therefore used as calibration question for Cooke's method.

The results showed that extreme river discharges resulting from the combined expert-and-data approach are most plausible. Using only data gives relatively small uncertainty bounds in the range of lower return periods, while using only experts does constrains the uncertainty in the range of higher return periods. Note that results with smaller uncertainty bands does not mean they are also correct (in order to assess correctness we would need to observe thousands of years of measurements in an unchanging environment), but they seem credible when compared to the most extreme discharges we have observed.

In conclusion, we found that with the method presented in this study we were successfully able to derive credible extreme discharges for the River Meuse. The combined data-EJ approach performed best for estimating extreme river discharges, while the experts-only approach performed satisfactory as well, albeit with a larger uncertainty. This indicates that the method can be applied as well to river systems where measurement data are scarce or absent. A case study for a different river could verify these findings. The credible results, together with the relatively limited effort needed, makes the presented method an attractive alternative for a more complex hydrological model-study.

Appendix A: Prior distribution for GEV inference

A1 A weakly informed prior

Section 3.3 described how Markov-Chain Monte Carlo (MCMC) was used to derive credibility bounds for the GEV fit. MCMC is an algorithm for Bayesian inference, meaning it updates a-priori distribution with observations to an a-posteriori distribution. The central theme of this paper is using structured expert judgment to quantify a discharge model, so we wanted the prior to be unbiased regarding the expert estimates. We chose that this meant using a prior for the GEV that gives a uniform distribution between 0 and 10,000 m^3/s for the 1,000-year discharge. This range is wide enough to cover the plausible range for any expert or data fit (remember that we are only fitting tributary discharges). As this is not truly uninformative, we call it a weakly informative prior.

A2 Deriving the prior joint distribution

The GEV-distribution needs three parameters, a location parameter μ, scale-parameter σ, and shape parameter ξ. Ambivalence about the three parameters (i.e, a uniform distribution $\mathcal{U}$ with range $(-\infty, \infty)$ for μ and ξ, and with range $(0, \infty)$ for σ) does

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

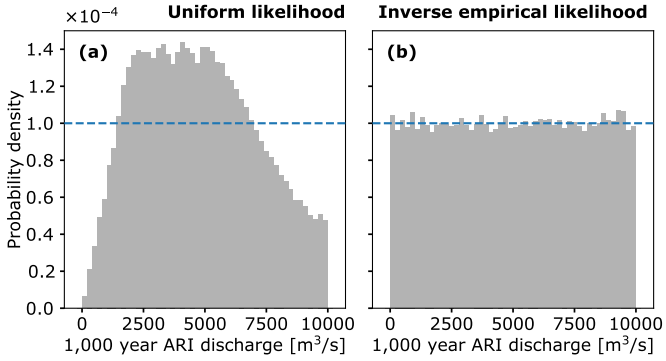

Figure A1. Histograms of 1,000-year discharges resulting from the MCMC prior traces. **(a)** With a uniform probability density as likelihood function, **(b)** with the inverse of the left densities as likelihood function.

not lead to a uniform 1,000-year discharge, so we needed to derive the joint distribution of the three parameters that does give the required discharge distribution. We did this by using MCMC just like we fitted the expert estimates (i.e., with the likelihood function of Eq. 5), but then with a uniform probability density between 0 and 10,000 m^3/s:

$$\begin{aligned}\ell(\theta|exp) &= \sum_j \log\left(f_{\mathcal{U}_{[0,10000]}}(q_{p_j})\right) \\ &= \sum_j \log\left(f_{\mathcal{U}_{[0,10000]}}(F_\theta^{-1}(1-p_j))\right)\end{aligned} \tag{A1}$$

By using the weakly informed priors $\mu = \mathcal{U}_{[0,2000]}$, $\sigma = \mathcal{U}_{[0,10000]}$, and $\xi = \mathcal{U}_{[-2,2]}$ and doing the inference, we get the distribution for the 1,000-year discharge shown in Fig. A1 (a). The probability density is limited in between 0 and 10,000 m^3/s, but the probability density is not uniform. Therefore, we repeat the above procedure, but now with an empirical probability density, of 1 divided by the densities shown in Fig. A1 **(a)**. This results in a sufficiently uniform pattern, Fig. A1 **(b)** shows.

The inference-trace can now be used as empirical prior. However, it still has a problem: The degrees a freedom within the (μ, σ, ξ)-combinations lead to too much prevalence of light-tailed distributions in the prior (i.e., a horizontal curve). Figure A2 **(a)** shows this. When using this prior to fit the expert estimates without data, the results become unfeasible, because the two expert distributions leave to much freedom for fitting the GEV shape. To solve this, we first did the inference for the data-only fits (Sect. 3.3, Eq. 4), and fitted a beta-distribution to the shape parameters. This distribution was then used as prior for ξ, with which we repeated the just described procedure. The wide tail shape that follows from the data (see for example Fig. 5 **a.** or **d.**) leaves enough freedom for fitting differently shaped GEV-distributions to data and expert estimates. A sample of GEV-curves that results from this final prior, is shown in Fig. A2 **b.**.

To summarize, we followed these steps for deriving the GEV-prior:

1. Fit the GEV-distribution with the likelihood function from Eq. 4 to the observed tributary discharges, using the weakly informed priors $\mu = \mathcal{U}_{[0,2000]}$, $\sigma = \mathcal{U}_{[0,10000]}$, and $\xi = \mathcal{U}_{[-2,2]}$.

2. Fit a beta-distribution through the resulting shape parameters ξ.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

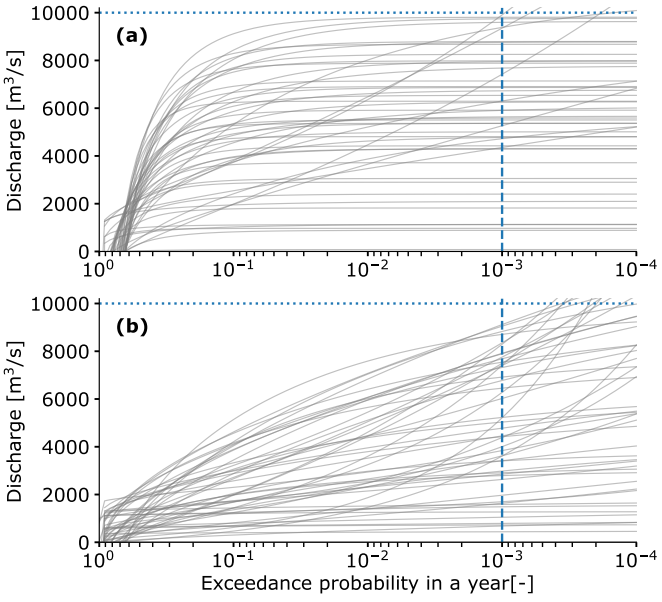

Figure A2. Example curves drawn from the priors. **(a)** Drawn from the prior with uninformed shape parameter ξ. **(b)** Using an informed shape parameter ξ.

3. Sample from $\mu = \mathcal{U}_{[0,2000]}$, $\sigma = \mathcal{U}_{[0,10000]}$, and the fitted $\xi = \text{Beta}(5.35, 3.72)$, ranging from -0.85 to 0.75, with a log-likelihood $\ell = \log\left(f_{\mathcal{U}_{[0,10000]}}(Q_{1000y})\right)$ for the 1,000-year discharge.

4. Sample again from the same priors, with an adjusted log-likelihood $\ell = \log\left(1/f_{obs}(q)\right)$

5. Sample again from these priors, with an adjusted log-likelihood function, to obtain a uniform probability density for the 1,000-year discharge.

6. Use this sample as prior for all GEV-fits.

A3 Sampling from the prior

The MCMC-inference gives a trace that is used as prior. A heatmap of the parameter-combinations is shown in Fig. A3. To use this empirical distribution with MCMC, we sample from a uniform distribution for each of the parameters μ, σ, and ξ and transform these to the parameter-space, taking into account the dependencies:

$$
\begin{aligned}
x_\mu &= F^{-1}_{X_\mu}(u_\mu) \\
x_\sigma &= F^{-1}_{X_\sigma}(u_\sigma | X_\mu = x_\mu) \\
x_\xi &= F^{-1}_{X_\xi}(u_\xi | X_\mu = x_\mu, X_\sigma = x_\sigma)
\end{aligned}
\tag{A2}
$$

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

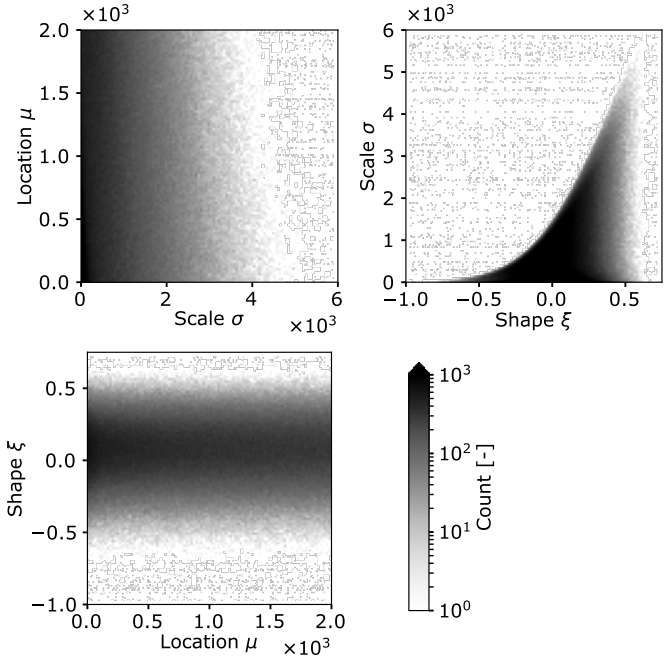

Figure A3. Joint prior distribution of the (μ, σ, ξ) combinations, plotted for each pair.

$F_{X_\mu}^{-1}$ is the inverse empirical CDF (i.e., the percentile point function) for parameter the location parameter X_μ. x_μ is the realization of X_μ in parameter space, and u_μ its realization in $\mathcal{U}_{[0,1]}$ space.

Numerically, x_μ, x_σ, and x_ξ are determined by discretizing the sampled values into 100 equally spaced bins per variable. The cumulative sum of the count per bin, divided by the total number of variables (i.e., normalized), gives the empirical CDF. x_μ is determined by interpolating u_μ within the normalized, cumulative bin count, and returning the bin x-values. Subsequently, x_σ is determined in a similar way, but now the empirical CDF is created from the values where $X_\mu \in bin(x_\mu)$ ($bin(x_\mu)$ is the bin that contains x_μ). Finally, x_ξ is determined by interpolating u_ξ in the empirical CDF created from the values where $X_\mu \in bin(x_\mu)$ and $X_\sigma \in bin(x_\sigma)$.

Appendix B: Sensitivity of observation to expert judgment factor

When fitting a GEV-distribution to both observations and expert judgments, like we do in this study, the contribution of each to the likelihood function affects the resulting fit. Normally, the more evidence (i.e., observations) are available, the closer the solution should follow these observations. We however strive for a balance between the two, irrespective of the number of observations, because the two sources are used for fitting the distribution to different ranges of return periods.

Equation 6 shows the combined likelihood function, in which the factor $10/N_i$ gives the weight of the observations relative to the expert judgment. The 10 in this fraction means that the observations have a weight like (only) ten events were considered.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Without the factor, the fraction would be 1.0, or N_i/N_i, while an equal weight between observation and expert judgment would give a factor $1/N_i$. Fig. B1 shows the resulting solution for the five options $1/N_i$, $5/N_i$, $10/N_i$, $20/N_i$, and N_i/N_i, for expert F's estimates for the tributaries Semois and Niers. Most experts underestimated and overestimated these two tributaries' discharges respectively. The comparison shows that $10/N_i$ gives a middle ground between observations and expert estimates in these two illustrative cases.

The sensitivity of the outcome to the factor becomes less when expert judgment and observations are more in line with a single GEV distribution. This is expected to be the case for the high scoring experts but not necessarily, as the expert weights were determined from the on average once per 10-year discharge estimate rather than the 1000-year estimate shown in Fig. B1.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

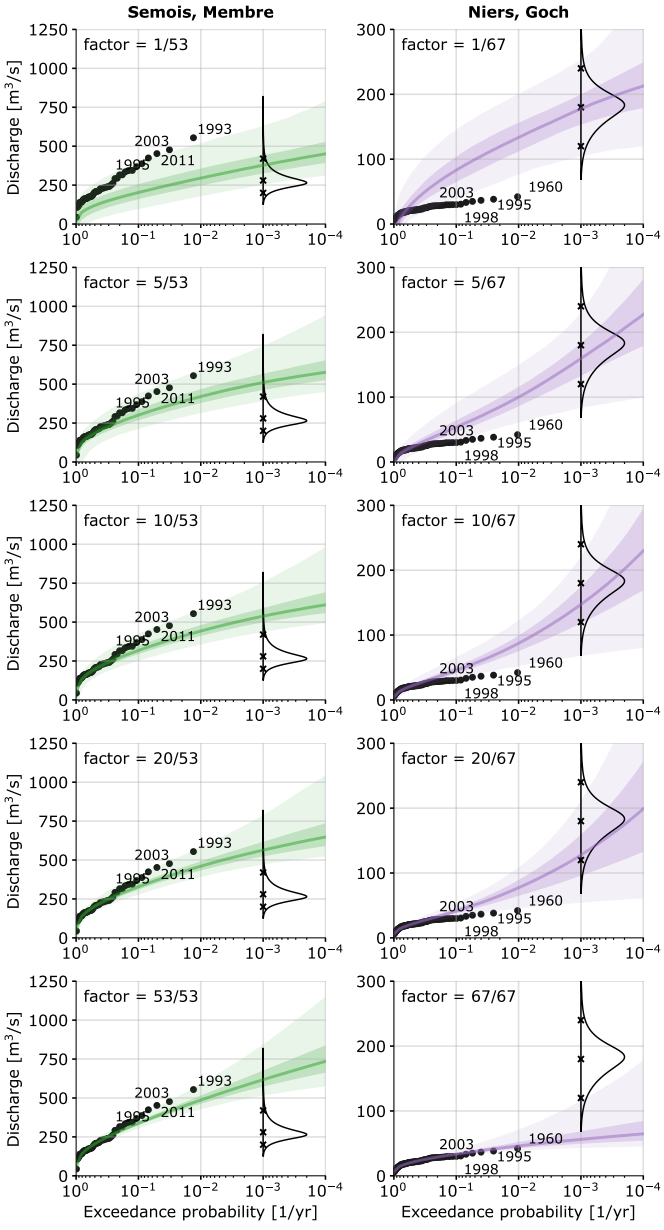

Figure B1. Sensitivity of the fitted GEV for different observations-to-EJ weights. A underestimated tributary, Semois (left), as well as an overestimated tributary, Niers (right), are shown.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Appendix C: Expert and DM correlation matrices

Figure C1 shows the correlation matrices estimated by the experts. The DM correlation matrices are weighted combinations of the expert matrices, based on the weights from Table 2.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

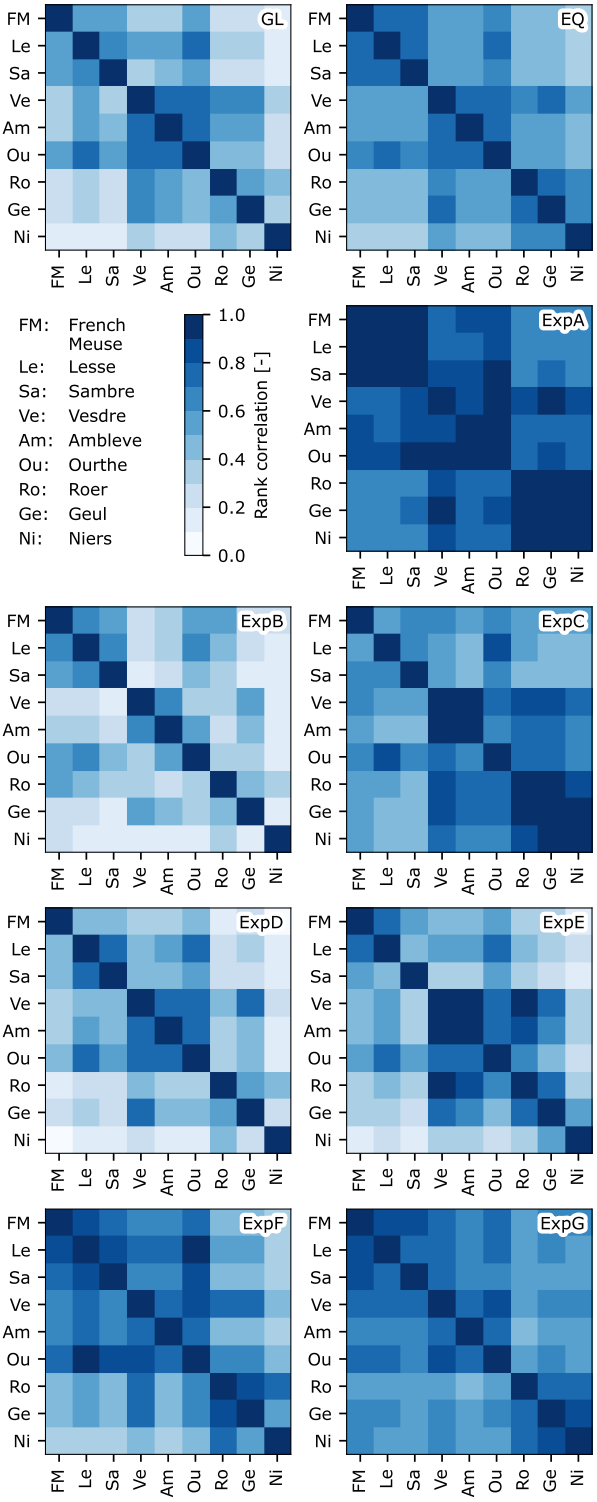

Figure C1. Correlation matrices estimated by the expert

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Code and data availability. All raw data and the code used to process them can be provided by the corresponding author upon request.

Author contributions. GR, OMN, and MK planned the study. GR prepared and carried out the expert elicitation, processed and analyzed the results, and wrote the manuscript draft. OMN and MK reviewed and edited the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to thank all experts that participated in the study, Alexander, Eric, Ferdinand, Helena, Jerom, Nicole, and Siebolt, for their time and effort in making this research possible. Secondly, we thank Dorien Lugt en Ties van der Heijden, who's hydrological and statistical expertise greatly helped in preparing the study through test rounds.

This research was funded by the TKI project EMU-FD. This research project is funded by Rijkswaterstaat, Deltares and HKV consultants.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

References

Al-Awadhi, S. A. and Garthwaite, P. H.: An elicitation method for multivariate normal distributions, Communications in Statistics-Theory and Methods, 27, 1123–1142, 1998.

Bernard, A. and Bos-Levenbach, E.: The plotting of observations on probability-paper, Stichting Mathematisch Centrum. Statistische Afdeling, 1955.

Borgomeo, E., Farmer, C. L., and Hall, J. W.: Numerical rivers: A synthetic streamflow generator for water resources vulnerability assessments, Water Resources Research, 51, 5382–5405, 2015.

Bouaziz, L. J., Thirel, G., de Boer-Euser, T., Melsen, L. A., Buitink, J., Brauer, C. C., De Niel, J., Moustakas, S., Willems, P., Grelier, B., et al.: Behind the scenes of streamflow model performance, Hydrology and Earth System Sciences Discussions, 2020, 1–38, 2020.

Brown, B. B.: Delphi process: a methodology used for the elicitation of opinions of experts, Tech. rep., Rand Corp Santa Monica CA, 1968.

Coles, S. G. and Tawn, J. A.: A Bayesian analysis of extreme rainfall data, Journal of the Royal Statistical Society: Series C (Applied Statistics), 45, 463–478, 1996.

Colson, A. R. and Cooke, R. M.: Cross validation for the classical model of structured expert judgment, Reliability Engineering & System Safety, 163, 109–120, 2017.

Cooke, R. M. and Goossens, L. L.: TU Delft expert judgment data base, Reliability Engineering and System Safety, 93, 657–674, https://doi.org/10.1016/j.ress.2007.03.005, 2008.

Copernicus Land Monitoring Service: EU-DEM, https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1/view, © European Union, Copernicus Land Monitoring Service 2018, European Environment Agency (EEA). Date accessed: 12 October 2021, 2017.

Copernicus Land Monitoring Service: CORINE Land Cover, https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=download, © European Union, Copernicus Land Monitoring Service 2018, European Environment Agency (EEA). Date accessed: 16 March 2022, 2018.

Copernicus Land Monitoring Service: E-OBS, https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-gridded-observations-europe?tab=overview, © European Union, Copernicus Land Monitoring Service 2018, European Environment Agency (EEA). Date accessed: 19 May 2022, 2020.

de Boer-Euser, T., Bouaziz, L., De Niel, J., Brauer, C., Dewals, B., Drogue, G., Fenicia, F., Grelier, B., Nossent, J., Pereira, F., et al.: Looking beyond general metrics for model comparison–lessons from an international model intercomparison study, Hydrology and Earth System Sciences, 21, 423–440, 2017.

Dewals, B., Erpicum, S., Pirotton, M., and Archambeau, P.: Extreme floods in Belgium. The July 2021 extreme floods in the Belgian part of the Meuse basin, 2021.

Diederen, D., Liu, Y., Gouldby, B., Diermanse, F., and Vorogushyn, S.: Stochastic generation of spatially coherent river discharge peaks for continental event-based flood risk assessment, Natural Hazards and Earth System Sciences, 19, 1041–1053, 2019.

Dion, P., Galbraith, N., and Sirag, E.: Using expert elicitation to build long-term projection assumptions, in: Developments in demographic forecasting, pp. 43–62, Springer, Cham, 2020.

Falter, D., Schröter, K., Dung, N. V., Vorogushyn, S., Kreibich, H., Hundecha, Y., Apel, H., and Merz, B.: Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain, Journal of Hydrology, 524, 182–193, 2015.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Food and Agriculture Organization of the United Nations: Digital Soil Map of the World, https://data.apps.fao.org/map/catalog/srv/eng/catalog.search?id=14116#/metadata/446ed430-8383-11db-b9b2-000d939bc5d8, source: Land and Water Development Division, FAO, Rome. Date accessed: 20 June 2022, 2003.

Hartley, D. and French, S.: A Bayesian method for calibration and aggregation of expert judgement, International Journal of Approximate Reasoning, 130, 192–225, 2021.

Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, Biometrika, 57, 97–109, https://doi.org/10.1093/biomet/57.1.97, 1970.

Hegnauer, M. and Van den Boogaard, H.: GPD verdeling in de GRADE onzekerheidsanalyse voor de Maas, Tech. rep., Deltares, Delft, 2016.

Hegnauer, M., Beersma, J., Van den Boogaard, H., Buishand, T., and Passchier, R.: Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse basins. Final report of GRADE 2.0, Tech. rep., Deltares, Delft, 2014.

Jenkinson, A. F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Quarterly Journal of the Royal Meteorological Society, 81, 158–171, https://doi.org/https://doi.org/10.1002/qj.49708134804, 1955.

Keelin, T. W.: The metalog distributions, Decision Analysis, 13, 243–277, 2016.

Kindermann, P. E., Brouwer, W. S., van Hamel, A., van Haren, M., Verboeket, R. P., Nane, G. F., Lakhe, H., Prajapati, R., and Davids, J. C.: Return level analysis of the hanumante river using structured expert judgment: a reconstruction of historical water levels, Water, 12, 3229, 2020.

Land NRW: ELWAS-WEB, https://www.elwasweb.nrw.de/elwas-web/index.xhtml#, land NRW, dl-de/by-2-0 (www.govdata.de/dl-de/by-2-0) https://www.elwasweb.nrw.de. Date accessed: 5 August 2022, 2022.

Leander, R., Buishand, A., Aalders, P., and Wit, M. D.: Estimation of extreme floods of the River Meuse using a stochastic weather generator and a rainfall, Hydrological Sciences Journal, 50, 2005.

Meresa, H. K. and Romanowicz, R. J.: The critical role of uncertainty in projections of hydrological extremes, Hydrology and Earth System Sciences, 21, 4245–4258, 2017.

Ministry of Infrastructure and Environment: Regeling veiligheid primaire waterkeringen 2017 no IENM/BSK-2016/283517, https://wetten.overheid.nl/BWBR0039040/2017-01-01, 2016.

Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., Ehmele, F., Feldmann, H., Franca, M. J., Gattke, C., et al.: A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe. Part 1: Event description and analysis, Natural Hazards and Earth System Sciences Discussions, pp. 1–44, 2022.

Parent, E. and Bernier, J.: Encoding prior experts judgments to improve risk analysis of extreme hydrological events via POT modeling, Journal of Hydrology, 283, 1–18, 2003.

Rijkswaterstaat: Waterinfo, https://waterinfo.rws.nl/#!/kaart/Afvoer/Debiet___20Oppervlaktewater___20m3___2Fs/, source: Rijkswaterstaat Waterinfo (https://waterinfo.rws.nl) Date accessed: 11 March 2022, 2022.

Rongen, G.: The effect of flooding along the Belgian Meuse on the discharge and hydrograph shape at Eijsden, Master's thesis, Delft University of Technology, 2016.

Rongen, G., Morales-Nápoles, O., and Kok, M.: Extreme Discharge Uncertainty Estimates for the River Meuse Using a Hierarchical Non-Parametric Bayesian Network, in: Proceedings of the 32th European Safety and Reliability Conference (ESREL 2022), edited by Leva, M. C., Patelli, E., Podofillini, L., and Wilson, S., pp. 2670–2677, Research Publishing, https://doi.org/10.3850/978-981-18-5183-4_S17-04-622-cd, 2022a.

https://doi.org/10.5194/egusphere-2023-39
Preprint. Discussion started: 7 March 2023

Rongen, G., Morales-Nápoles, O., and Kok, M.: Expert judgment-based reliability analysis of the Dutch flood defense system, Reliability Engineering & System Safety, 224, 108 535, 2022b.

Salvatier, J., Wiecki, T., and Fonnesbeck, C.: Probabilistic Programming in Python using PyMC, PeerJ Computer Science, 2, https://doi.org/10.7717/peerj-cs.55, 2015.

Sebok, E., Henriksen, H. J., Pastén-Zapata, E., Berg, P., Thirel, G., Lemoine, A., Lira-Loarca, A., Photiadou, C., Pimentel, R., Royer-Gaspard, P., et al.: Use of expert elicitation to assign weights to climate and hydrological models in climate impact studies, Hydrology and Earth System Sciences Discussions, pp. 1–35, 2021.

Service public de Wallonie: Annuaires et statistiques, http://voies-hydrauliques.wallonie.be/opencms/opencms/fr/hydro/Archive/annuaires/index.html, source: Voies Hydraulique Wallonie (http://voies-hydrauliques.wallonie.be/opencms/opencms/fr) Date accessed: 26 June 2022, 2022.

Task Force Fact-finding hoogwater 2021: Hoogwater 2021 - Feiten en duiding, Tech. rep., Expertisenetwerk Waterveiligheid (ENW), Delft, 2021.

van de Langemheen, W. and Berger, H.: Hydraulische randvoorwaarden 2001: maatgevende afvoeren Rijn en Maas, Tech. rep., RIZA, 2001.

Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G.: Flood frequency hydrology: 3. A Bayesian analysis, Water Resources Research, 49, 675–692, 2013.

Waterschap Limburg: Discharge Measurements, source: Waterschap Limburg (https://www.waterstandlimburg.nl/Home) Historical time series from personal communication. Date retrieved: 16 August 2021, 2021.

Winter, B., Schneeberger, K., Huttenlau, M., and Stötter, J.: Sources of uncertainty in a probabilistic flood risk model, Natural Hazards, 91, 431–446, 2018.