





<https://doi.org/10.1590/2318-0331.272220220012>

## Assessing the effects of rating curve uncertainty in flood frequency analysis

### *Avaliação de efeitos das incertezas de curva-chave em análises de frequência de inundações*

Luan Marcos da Silva Vieira\* , Júlio César Lôbo Sampaio<sup>1</sup> , Veber Afonso Figueiredo Costa<sup>1</sup>  &  
Julian Cardoso Eleutério<sup>1</sup> 

<sup>1</sup>Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

E-mails: luan\_msv@hotmail.com (LMSV), juliolobosampaio@gmail.com (JCLS), veber@ehr.ufmg.br (VAFC), julian.eleuterio@gmail.com (JCE)

Received: February 06, 2022 - Revised: April 30, 2022 - Accepted: May 06, 2022

#### ABSTRACT

Maximum flows are often estimated from flood frequency analysis, by means of the statistical fitting of a theoretical probability distribution to maximum annual flow data. However, because of the limitations imposed by the practice of at-site flow measurement, empirical models are applied as the rating curve for estimating streamflow. These curves are approximations of the actual flows and incorporate different sources of uncertainty, especially in the extrapolation portions. These uncertainties are propagated in the frequency analysis and influence the estimated quantiles. For better understanding and describing the influence of the stage-discharge uncertainty in this process, the results of Bayesian rating curve modeling, which considers the physical knowledge of the gauging station as prior information, were combined with Bayesian flood frequency analysis under asymptotic extreme value theory. The method was applied to the Acorizal stream gauging station, located in the interior of the state of Mato Grosso - BR. The main results suggested that, although the uncertainties of the rating curve can be relevant in the estimation of maximum flow quantiles, the uncertainties arising from finite-sample inference might exert greater impacts on the flow credibility intervals even for moderate sample sizes.

**Keywords:** Rating curve; Bayesian inference; Flood frequency analysis; BaRatin.

#### RESUMO

As vazões máximas são frequentemente estimadas a partir de análises de frequência de inundação, por meio de ajustes estatísticos de uma distribuição teórica de probabilidades aos dados de vazões máximas anuais. Entretanto, em razão das limitações impostas pela prática da medição da vazão in loco, são aplicados modelos empíricos como a curva chave para estimar vazão. Essas curvas são aproximações das vazões reais e incorporam diferentes fontes de incertezas, em especial nas porções de extrapolação. Tais incertezas são propagadas nas análises de frequência e influenciam nos quantis estimados. Para melhor compreender e descrever a influência das incertezas da cota-descarga neste processo, foi combinado os resultados da modelagem Bayesiana da curva chave, que considera o conhecimento físico da estação de medição como conhecimento *a priori*, com a análise de frequência Bayesiana de inundação sob a teoria assintótica de valores extremos. O método foi aplicado na estação de fluviométrica de Acorizal, localizada no interior do estado de Mato Grosso – BR. Os principais resultados sugeriram que, embora as incertezas da curva chave possam ser relevantes na estimativa de quantis de vazões máximas, as incertezas decorrentes de séries históricas amostrais podem exercer maiores impactos nos intervalos de credibilidade da vazão mesmo para tamanhos de amostras moderadas.

**Palavras-chave:** Curva chave; Inferência Bayesiana; Análise de frequência de vazões máximas; BaRatin.

## INTRODUCTION

Floods are one of the main natural hazards to society and the environment (United Nations Office for Disaster Risk Reduction, 2019). In effect, in 2019, this phenomenon alone was responsible for 43% of the deaths related to natural disasters around the world (Centre for Research on the Epidemiology of Disasters, 2020). Floods have also entailed huge economic losses on a yearly basis, and these are expected to increase in view of the disorganized occupation of flooding-prone areas (McClymont et al., 2020). In this sense, flood risk assessment constitutes an essential task for mitigating negative impacts to the population, to economic development and to the cultural heritage.

Probabilistic flood modeling, or flood frequency analysis (FFA), is usually the first step for risk assessment and management. This expedient involves sampling the variable of interest, i.e., streamflow, either from a fixed time-span or through a point-process, and fitting a distributional model that summarizes its random behavior. FFA analysis implicitly acknowledges that the finite sample-based inference introduces errors to quantile estimates, but, for practical purposes, the streamflow observations are considered error-free (Naghetini, 2017).

However, streamflow data are inherently uncertain since they are, more often than not, indirectly obtained from models that relate water stages and discharges, namely, hydrometric rating curves. Rating curves summarize a reference flow condition that univocally relates discharges and stages and, hence, allows monitoring water levels instead of streamflow. This is considerably easier from an operational perspective but might not provide a suitable physical description of the flow rates.

Rating curves may be affected by both random and epistemic uncertainty. The random component is usually associated with equipment precision and measurement conditions, which might not be precisely known (Le Coz, 2012; Garcia et al., 2020). In fact, backwater effects or hysteresis may entail strong deviations from the reference stage-discharge relationship during a measurement expedient, but the occurrence of these conditions might not be determined *a priori* (Le Coz et al., 2014; McMillan & Westerberg, 2015).

The epistemic component may stem from sampling errors, deviations from the established reference regime due to non-steady flow, changes in the river cross-section or in the friction conditions due to seasonal variation in vegetation along the river reach, and extrapolation (Moges et al., 2021). Also, as with any model, rating curves are affected by structural uncertainty, which reflects their inability to fully describe the actual stage-discharge relationship (Sikorska & Renard, 2017) and entail systematic errors that might strongly affect streamflow prediction (Lang et al., 2010; Baldassarre et al., 2012; Steinbakk et al., 2016).

In recent years, a great deal of research effort (e.g., Kiang et al., 2018 and references therein) has been dedicated to understanding and quantifying the effects of rating curve uncertainty in hydrological applications. The propagation of rating curve uncertainty in flood frequency analysis (Steinbakk et al., 2016; Osorio & Reis Junior, 2016), rainfall-runoff modeling (Sikorska & Renard, 2017) and flood mapping (Kastali et al., 2021) has also comprised an active research field, and the explicit account of the rating curve errors has improved the definition of flood

management strategies (McMillan et al., 2017) and facilitated the communication of the cost-risk tradeoff to decision makers (Garcia et al., 2020).

In the context of FFA, a main source of uncertainty is related to rating curve extrapolation (Lang et al., 2010; Baldassarre et al., 2012). In effect, measurement procedures in high flow conditions are rare (Steinbakk et al., 2016) and, as a result, only a few (if any) sample points might be available for defining the stage-discharge relationship in flooding conditions. This might lead to large errors, which may amount more than 40% in the extrapolated portion of the rating curve (McMillan et al., 2012; Westerberg et al., 2020).

The Bayesian paradigm has offered suitable alternatives for properly accounting for the rating curve uncertainty and preserving the physical realism of the predicted steamflows in extrapolation conditions (Moyeed & Clarke, 2005; Renard et al., 2010; Reitan & Petersen-Øverleir, 2009; McMillan et al., 2010; Le Coz et al., 2014; Steinbakk et al., 2016; Garcia et al., 2020; Kastali et al., 2021). Bayesian methods allow introducing expert knowledge in inference procedures via informative prior distributions for the rating curve parameters under hydraulic considerations (friction, geometry, hydraulic controls) and, hence, model the rating curve on physical grounds.

Among the available Bayesian methods, this paper has utilized the BaRatin (Bayesian Rating Curve; Le Coz et al., 2014) for modeling the rating curve under a formal probabilistic framework. BaRatin builds a conditional likelihood based on uncertain gaugings (i.e., measurements with errors) and uses hydraulic controls for associating a theoretical flow model for the stage-discharge relationship, which allows well-posed inferences on the upper portion of the rating curve (Le Coz et al., 2014; Osorio & Reis Junior, 2016; Ocio et al., 2017; Sikorska & Renard, 2017; Garcia et al., 2020; Kastali et al., 2021). BaRatin also explicitly accounts for the distinct sources of uncertainty, which allows segregating measurement, sampling and structural errors, and is paramount for decomposing the predictive uncertainty.

In parallel, the propagation of the rating curve uncertainty into FFA has been investigated by several authors (Reitan & Petersen-Øverleir, 2009; Lang et al., 2010; Osorio & Reis Junior, 2016; Steinbakk et al., 2016). These previous works have highlighted the influence of the rating curve errors, in marginal terms (i.e., sampling and rating curve uncertainties were not jointly estimated), in flood risk assessment. However, a pragmatic approach for assessing the combined uncertainty of the rating curve modeling and the sampling errors in FFA has not been explored in depth and still calls for research effort.

In view of the foregoing, this paper proposes coupling BaRatin and Bayesian FFA under Extreme Value Theory for quantifying the joint influence of rating curve and parameter estimation in flood quantiles via a mixture model. Such an approach allows combining the distinct sources of uncertainty into a simple inference setup, albeit not explicitly segregating their effects. The proposed framework is utilized for estimating flood quantiles at a cross-section in the Cuiabá River, in the Brazilian state of Mato Grosso, and evaluating the contribution of the distinct sources of uncertainty under different scenarios of information regarding the rating curve (i.e., the number of gaugings) and the random sample of annual flood maxima. In this sense, the main

contribution of the paper is discussing a rigorous framework for formally quantifying and for properly interpreting how different sources of uncertainty affect the estimation of flood quantiles and the assessment of flood risk. These aspects are relevant with respect to the common practice in hydraulic design (i.e., ignoring streamflow errors) as they might underpin more robust decision-making expedients and potentially reduce costs in the implantation of large hydraulic structures for flood conveyance and mitigation (McMillan et al., 2017).

The remainder of the paper is organized as follows: Section 2 presents the statistical models utilized in the study, as well as their main simplifying assumptions. In Section 3, we present the case study, with a brief description of the study area, the elicitation of prior distributions and the main results for both rating curve modeling and FFA. Finally, in Section 4, we address the conclusions of the study and the envisaged research developments.

## STATISTICAL MODELS

In this paper, the Bayesian paradigm is utilized for modeling the rating curve under a formal probabilistic approach and for propagating the correspondent uncertainty into the estimation of the parameters of the Generalized Extreme Value distribution (GEV) for FFA. For estimating the rating curve, we have relied on the BaRatin framework, which was first introduced by Le Coz et al. (2014). In short, BaRatin is built upon theoretical flow equations, such as weirs or steady-state uniform flow, for establishing reference conditions for the stage-discharge relationship. Informative prior uncertainty distributions are then elicited for the parameters of the rating curve model for allowing some degree of deviation with respect to the reference conditions while preserving the physical realism in extrapolation.

As for FFA, we have resorted to the asymptotic Extreme Value Theory (EVT), under the block-maxima rationale, for estimating flood quantiles by incorporating the rating curve errors. Following the block-maxima approach, EVT deals with the limiting distribution of the (rescaled) maximum value of a large set of  $N$  random variables (e.g., daily streamflows), sampled at fixed time steps (e.g., a water year). According to the Fisher-Tippett-Gnedenko theorem (Fisher & Tippett, 1928; Gnedenko, 1943), if the distribution of the maximum value converges to a parametric form as the size of the as the block length increases (i.e.,  $N \rightarrow \infty$ ), this distribution should converge to the Generalized Extreme Value (GEV) model (Koutsoyiannis, 2021), which encompasses the three asymptotic forms for the upper tail decay (exponential or light tail, polynomial or heavy tail, and upper-bounded). This fact provides theoretical arguments for prescribing the GEV distribution in the modeling of annual maximum flows without the complete knowledge of the parent distribution of the daily flows (Naghetini, 2017).

The statistical models and their main assumptions of our approach are discussed in the next sections.

- Rating curve estimation with BaRatin

BaRatin allows the estimation of the parameters of the rating curve by taking into account the previous knowledge on the hydraulic behavior of the river reach and the river cross-section in which the gauging station is located. Such a knowledge is translated

into a conceptualized flow model that univocally relates stages and discharges (i.e., the reference condition) and whose parameters are estimated under uncertain measurements (Le Coz et al., 2014).

Mathematically, the rating curve may be expressed as a power function  $f_{RC}(h_t, \boldsymbol{\theta})$ , in which  $h_t$  is the stage at a given time  $t$ , and  $\boldsymbol{\theta}$  is a vector that encompasses model parameters. The predicted streamflow  $\hat{q}_t$ , at a given time  $t$ , is then given by

$$\hat{q}_t = f_{RC}(h_t, \boldsymbol{\theta}) \quad (1)$$

Alternatively, the predicted streamflow at time  $t$  may be expressed as

$$\hat{q}_t = q_t + e_t^{RC}(\boldsymbol{\beta}) \quad (2)$$

In which  $q_t$  denotes the real (unknown) streamflow and  $e_t^{RC}(\boldsymbol{\beta})$  corresponds to the structural error term, which is associated with the inability of the model in completely representing the true stage-discharge relationship.

Structural uncertainty is reported to stem from both random and epistemic errors (Juston et al., 2014; McMillan & Westerberg, 2015). Due to this fact, incorporating systematic errors in rating curve modeling is still a challenging task (Garcia et al., 2020). Hence, in this paper we assume that epistemic structural errors may be treated as random for inference purposes, although this assumption may aggregate bias to streamflow estimates. Following Sikorska & Renard (2017), we also assume that the structural errors are serially independent Gaussian variates, with zero mean and standard deviation linearly related to the predicted streamflow for accommodating heteroscedasticity. In formal terms

$$e_t^{RC}(\boldsymbol{\beta}) \sim N\left(0, (\beta_0 + \hat{q}_t \cdot \beta_1)^2\right) \quad (3)$$

In which  $\beta_0$  and  $\beta_1$  are parameters to be estimated from the data.

Under a similar rationale, the measurement errors may be incorporated into the inference by considering that the gaugings (i.e., the pairs of instantaneous measured values of stage  $h_t$  and discharge  $\tilde{q}_t$ ) comprise the real streamflows and purely random error terms  $e_t^{GP}$ . Formally

$$\tilde{q}_t = q_t + e_t^{GP} \quad (4)$$

By assuming that  $e_t^{GP}$  are also Gaussian variates, with zero mean and known standard deviation  $\delta_t$  (also related to the predicted streamflow), one may combine Eq. 2 and 4 and express the gaugings as

$$\tilde{q}_t = \hat{q}_t - e_t^{RC}(\boldsymbol{\beta}) + e_t^{GP} \quad (5)$$

Finally, by assuming that, conditioned on the stages, the gaugings are also Gaussian random variables, one may derive the likelihood function as follows

$$p(\tilde{\mathbf{q}}|\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{h}) = \prod_{i=1}^M N\left(\tilde{q}_{ii}; \hat{q}_{ii}; \left(\frac{\beta_0 + \hat{q}_{ii}}{\hat{q}_{ii} \cdot \beta_1}\right)^2 + \delta_{ii}^2\right) \quad (6)$$

In which  $N(a; \mu, \nu)$  denotes a Gaussian probability density function, with mean  $\mu$  and variance  $\nu$ , evaluated at a sample point  $a$  (which corresponds to the measured flows).

The posterior distribution of the parameters of the rating curve and the structural error model is then given by

$$p(\boldsymbol{\theta}, \boldsymbol{\beta} | \tilde{\mathbf{q}}, \mathbf{h}) \propto p(\tilde{\mathbf{q}} | \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{h}) \cdot p(\boldsymbol{\theta}, \boldsymbol{\beta}) \quad (7)$$

The proposed model may be extended to multi-segmented rating curves by imposing continuity constraints to streamflows at the different activation stages (Steinbakk et al., 2016). Bayesian inference is performed with the software Stan, accessed by the R package RStan (Stan Development Team, 2022). For checking whether the prior assumptions on the structural error model hold after inference, i.e., validating the posterior rating curve model, we resort to diagnostic residual plots, as suggested by Le Coz et al. (2014) and Garcia et al. (2020). These plots encompass: i) scatterplots of the standardized residuals with respect to the measured flows, for assessing homoscedasticity; ii) frequency histograms and quantile-quantile plots, for evaluating normality; and iii) the partial autocorrelation function, for assessing serial independence.

- Propagation of rating curve uncertainty into FFA

For performing Bayesian FFA under uncertain streamflow information, we have utilized the Generalized Extreme Value (GEV) distribution. Under somewhat general conditions, the GEV model arises as the asymptotic form for the distribution of the maximum value of a set of  $N$  random variables. The distribution function of the GEV model is given by (Coles, 2001)

$$F(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{\frac{-1}{\xi}} \right\} \quad (8)$$

In which  $\mu$ ,  $\sigma$  e  $\xi$  are the location, scale, and shape parameters, respectively. Eq. (8) aggregates the three asymptotic behaviors for the distribution upper tail: for  $\xi > 0$ , the extreme value type II or Fréchet distribution, with domain  $\mu - \frac{\sigma}{\xi} < x < \infty$  and polynomial decay when  $x \rightarrow \infty$ , arises; when  $\xi < 0$ , the extreme value type III or Weibull distribution, with upper bound  $\mu + \frac{\sigma}{\xi}$ , is obtained; for  $\xi = 0$ , the extreme value type I or Gumbel distribution, with domain  $-\infty < x < \infty$  exponential decay when  $x \rightarrow \infty$ , stems (Costa & Sampaio, 2021).

For evaluating the propagation of the rating curve uncertainty into FFA, three models were considered: the first model, denoted M1, aggregates the combined uncertainty of the rating curve and that of the sampling errors. For this, we simulate 100 rating curve samples from the joint posterior distribution obtained under the BaRatin framework. Next, for each rating curve, we compute the maximum annual streamflow series  $\chi^{(i)}$ ,  $i = 1, \dots, 100$ . Then, we estimate the posterior distributions of the GEV parameters for each of these series and, finally, following Steinbakk et al. (2016), we summarize the overall uncertainty through a mixture distribution, which is given by

$$f(\mu, \sigma, \xi | \chi) = \frac{1}{N} \sum_{i=1}^N f(\mu, \sigma, \xi | \chi^{(i)}) \quad (9)$$

Which corresponds to an approximation to the posterior predictive distribution of the annual maxima, given the gaugings

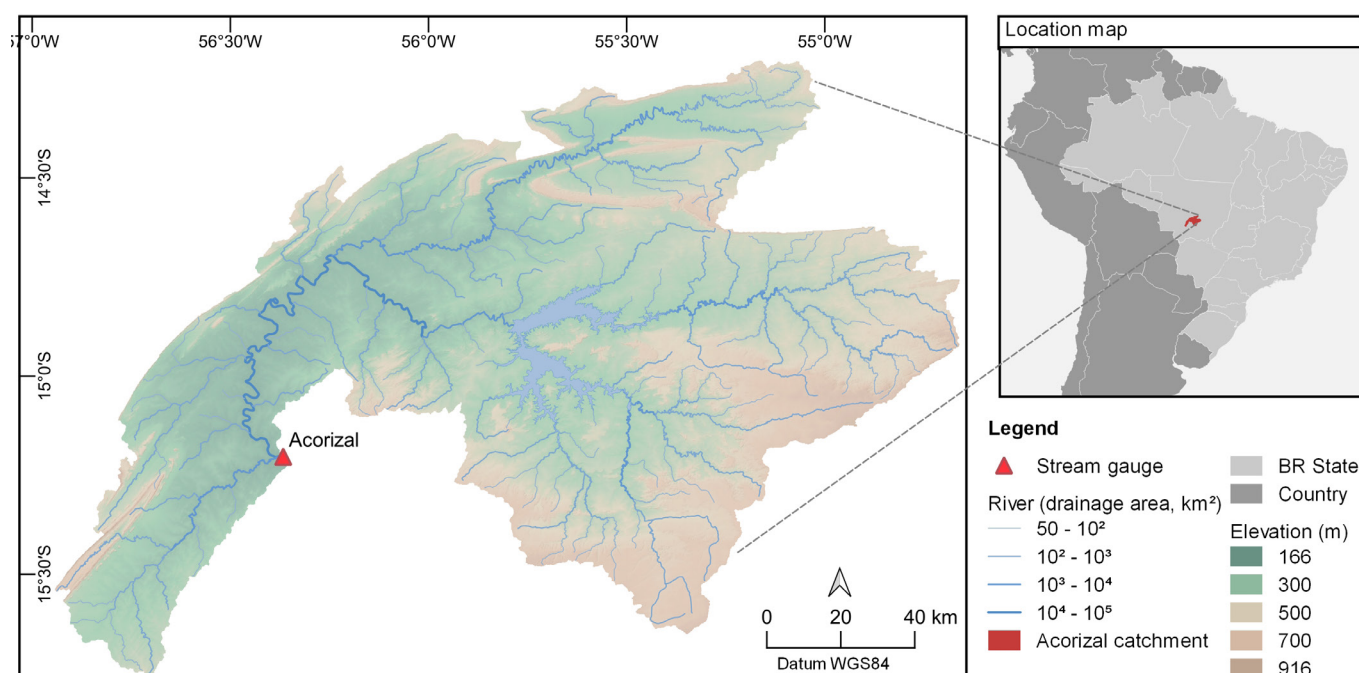
and the series of stages which are converted in annual maximum streamflow via rating curves. The mixture distribution presented in our study, which averages flood quantiles over different rating curve models while accounting for sampling uncertainty, is rarely discussed in the hydrologic literature as it does not present a closed quantile function (Thorarinsdottir et al., 2018). However, the paper of Steinbakk et al. (2016) has suggested that the rating curve inter-model variance might inflate the posterior distribution of the GEV tail index, enhancing the credible intervals of the quantile curve and affecting the definition of design return levels in FFA. Hence, albeit Equation (9) comprises an approximation for the predictive posterior distribution, it might be useful for quantifying the influence of the rating curve in the decay of the GEV upper tail.

The second model, M2 is based on the median posterior rating curve and ignores the rating curve uncertainty, only accounting for the sampling counterpart. Finally, the third model, M3, only accounts for the rating curve uncertainty and summarizes it by computing the maximum likelihood estimates of the GEV parameters for each of the block-maxima series. For assessing the influence of the information content regarding the rating curve and the number of sample points, we evaluate three scenarios for each model: A) the entire dataset of gaugings and annual maximum floods; B) all annual peaks and a reduced number of gaugings, aggregating a single flow measurement per year and excluding those measurements at the upper portion of the rating curve (i.e., above the bankfull stage); and C) the second half of the flood peaks sample and the same gaugings as in (B). These scenarios are intended to extend the current analysis to a broader range of data availability (stage and flow measurements) and assess the generalization abilities of the proposed model for FFA. Bayesian inference of the GEV parameters, for each outlined scenario, is also performed with R package RStan (Stan Development Team, 2022).

## CASE STUDY

For applying the Bayesian approach described in the previous section, we have selected the Acorizal gauging station (Brazilian National Water and Sanitation Agency - ANA code 66255000), which is located in the Cuiabá River, downstream of the Manso Hydropower plant, in the Brazilian state of Mato Grosso (Figure 1). At this outlet, the catchment drains an area of about 19,700 km<sup>2</sup>. On average, the annual rainfall in the catchment is approximately 1,900 mm (Serviço Geológico do Brasil, 2011), which is concentrated from October to March.

For Bayesian rating curve modeling and FFA, gauging data (stage and discharge measurements) and mean daily water levels at the Acorizal gauging station were collected from the Brazilian Water and Sanitation Agency (ANA) digital platform (<http://www.snirh.gov.br/hidroweb>). Gauging data comprised a collection of 410 measurements, spanning from 1966 to 2012. Mean daily water level information was available from 1965 to 2016. However, as the Manso Hydropower plant started operating in the end of 1999, we have excluded data recorded after this event in FFA. As a result, the annual maximum series used in FFA comprised



**Figure 1.** Location of the stream gauge and its catchment utilized in the study.

32 records. The available records of stages and streamflow are shown in Figure S1 (supplemental material).

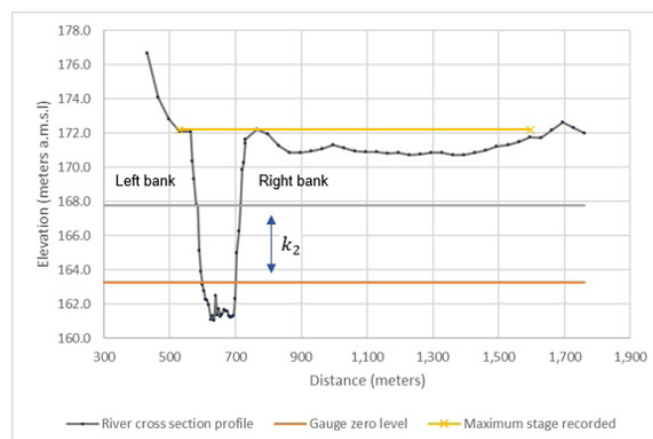
From a physiographic perspective, the slope of the Cuiabá River, in the reach that encompasses the Acorizal gauging station, is approximately 23.3 cm/km (Paes, 2011). The width of the main channel, in turn, is approximately 165 m.

Estimates for the Manning roughness coefficients are 0.035 for the main channel and 0.050 for the floodplains, obtained with the Cowan method (Arcement Junior & Schneider, 1989). The components of the Cowan method were computed by combining the theoretical limits proposed by Arcement Junior & Schneider (1989), field surveys provided by Furnas' Department of Hydrometeorology and Hydropower Generation Programming, and satellite images. Channel bed in the studied reach is mainly composed of sand, whereas the banks comprise clay soil and undergrowth vegetation.

The cross section of Acorizal, presented in Figure 2, suggests that the stage-discharge relationship cannot be represented by only one rating curve, due to the alteration in the shape of the cross section. The main channel and the floodplain cross sections can be approximated by two rectangles. For elevations higher than the bankfull stage (165.145 meters a.m.s.l.), the hydraulic radius and the vegetation abruptly change, thus requiring two distinct models to represent the stage-discharge relationship. Hence, for modeling purposes, a two-segment rating curve was proposed as follows (Sikorska & Renard, 2017)

$$f_{RC}(h_t, \theta) = \begin{cases} a_1(h_t - b_1)^{c_1} & \text{if } k_1 < h_t < k_2 \\ a_1(h_t - b_1)^{c_1} + a_2(h_t - b_2)^{c_2} & \text{if } k_2 < h_t \end{cases} \quad (10)$$

In which the activation stage  $k_1 = b_1$  refers to the cease-to-flow level, the activation stage  $k_2 = b_2$  corresponds to the transition level between the main channel and the floodplain (Figure 2), and  $a_1$ ,  $a_2$ ,  $c_1$  and  $c_2$  are model parameters.



**Figure 2.** Estimated activation stage ( $k_2$ ) in the Acorizal section.

In order to understand the behavior of the flows in the Acorizal section, a 1-D hydraulic simulation was performed, considering cross sections upstream and downstream of the gauging station. The boundary conditions were defined as a hydrograph of observed flows at the former and the normal depth at the latter.

The hydraulic simulation was carried out in HEC-RAS, from which only a small backwater effect was observed at the Acorizal cross-section, which suggests that the assumption of uniform flow regime is acceptable for deriving the rating curve. Hence, under the steady-state uniform flow equations, one may assume that parameters  $a_1$  and  $a_2$  are functions of the cross-section width  $B_{w,i}$ , the Manning roughness coefficient  $n$ , and the slope of the river reach  $S$ . In formal terms, the theoretical values of  $a_1$  and  $a_2$  may be obtained as

$$a_i(B_{w,i}, n_i, S) = \frac{1}{n_i} B_{w,i} \sqrt{S}, \quad i = 1, 2 \quad (11)$$

Parameters  $c_1$  and  $c_2$ , in turn, comprise the theoretical exponents of the Manning-Strickler equation for rectangular cross-sections – being both equal to  $\frac{5}{3}$  in this case study. Finally, the activation stages were estimated on the basis of the original design of the staff gauge ( $k_1 = -2.0$ ) and the floodplain level ( $k_2 = 4.48$ ), both with respect to the zero level of the staff gauge (163.265 meters a.m.s.l; see Figure 2).

For prescribing the prior uncertainty distributions of the rating curve parameters, we have assumed Gaussian distributions, centered at their theoretical values. The prior variances of parameters  $a_1$  and  $a_2$  were estimated according to the recommendations of the Joint Committee for Guides in Metrology (Joint Committee for Guides in Metrology, 2010) for propagating uncertainty. For this, we have assumed that  $B_{w,1}$ ,  $B_{w,2}$  and  $S$  are uniformly distributed with ranges  $165 \pm 16.5 \text{ m}$ ,  $1,050 \pm 105 \text{ m}$ , and  $0.000221 \pm 0.000245$ , respectively. The Manning roughness coefficients for the main channel and the floodplains, in turn, were assumed to follow lognormal distributions  $LN(0.035; 0.247^2)$  and  $LN(0.050, 0.282^2)$  (US Army Corps of Engineers, 1986).

For parameters  $c_1$  and  $c_2$ , the prior standard deviations were fixed at 0.025, following the original formulation of the BaRatin framework (Le Coz et al., 2014). Finally, the standard deviations for the activation stages were fixed on 1 m as a means for prescribing fairly imprecise prior distributions. Table 1 summarizes the prior distributions of the rating curve parameters.

For completing the inference setup, the standard deviation of the measurement errors was assumed as  $\delta_t \mu = 0.07 \hat{q}_t$  (Le Coz et al., 2014), as site-specific information was not available to address this parameter under a more formal approach (e.g., World Meteorological Organization, 2017), and the prior distributions of the parameters of the structural error model were elicited as  $\beta_0 \sim N_+(0, 100)$  and  $\beta_1 \sim N_+(0, 1)$ .

For obtaining samples of the joint posterior distribution of the parameters of the rating curve and the structural error model, 4 Markov chains were run in parallel, with 1,000 iterations each. Half the samples were discarded as warm-up and convergence was assessed by means of the Brook-Gelman-Rubin statistic (Gelman et al., 2013, p. 288), which must be lower than 1.1 for all parameters. Figure 3 presents frequency histograms depicting the posterior distributions of the rating curve parameters and the error model, along with the prior counterparts. One may notice that the posterior distributions are unimodal and approximately symmetric, with exception of that of  $\beta_1$ , which resembled an exponential distribution, indicating that some outliers were simulated during the MCMC procedure. A narrow posterior range was obtained for  $c$ , with small coefficient of variation, due to the prescribed informative prior distributions, which reflect the physical phenomena, i.e., the section characteristics and the hydraulic control conditions (Le Coz et al., 2014).

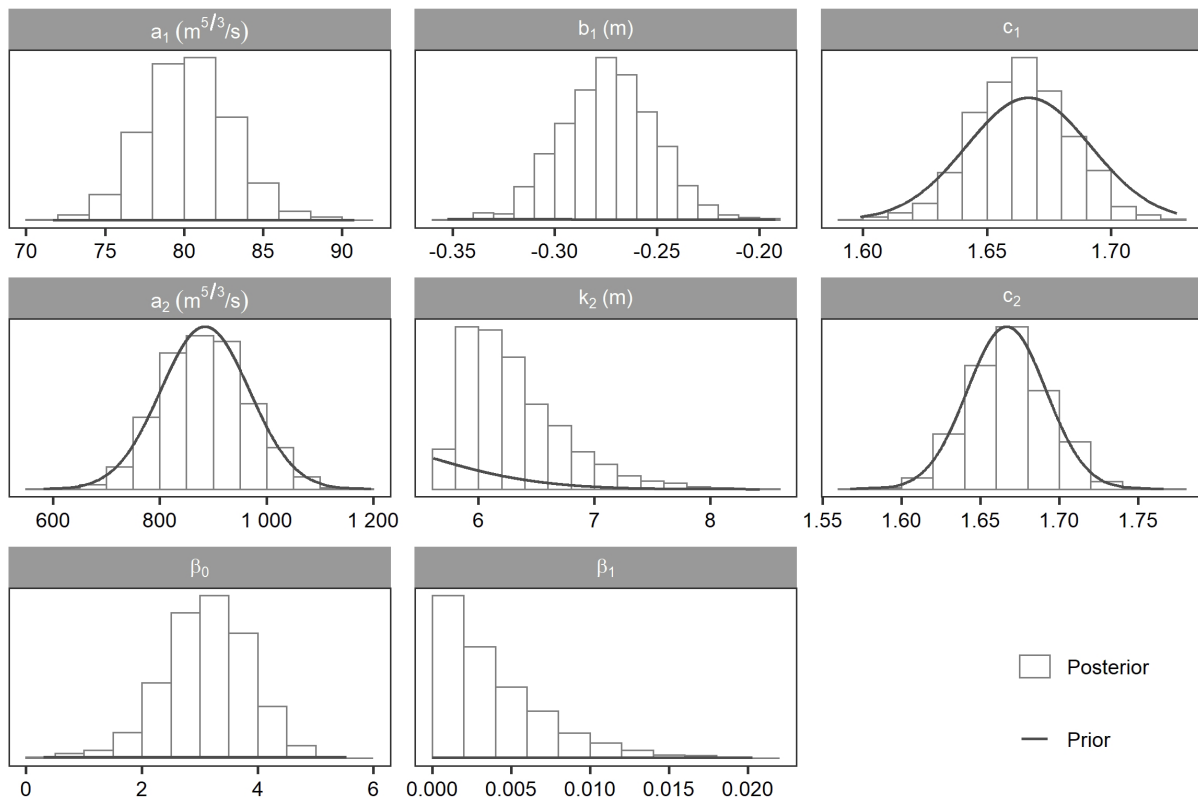
The inferred rating curve is presented in Figure 4. In general, the median estimates of the rating curve are close to the discharge measurements for small-to-mid elevations. In this range of stages, the credible intervals are narrow, and the measurement uncertainty was dominant - the large number of gaugings utilized in inference, along with the informative prior distributions, have led to reduced widths for the credible intervals in this portion of the rating curve. However, the estimated intervals were sufficiently wide for accommodating some portions of the measurement error bars (i.e., the uncertainty of the measured streamflows). It is also possible to note that, in the lower tail, the parametric uncertainty is a small part of the total uncertainty, and the structural uncertainty prevails. In effect, the adopted error model is not designed for reproducing the scatter of the lower flows, which might increase the structural errors in this range of stages. In the upper range of the rating curve, in which data are scarce, the credible intervals are wide and the total uncertainty (almost entirely covered by the parametric uncertainty at the high stages) increases, mainly due to the parametric uncertainty – because of the low information content for high stages, the prior and posterior distributions of the rating curve parameters within this range are similar. Exception is made to those of  $k_2$ , which, due to continuity constraints in the predicted streamflows, benefits from the information aggregated by the gaugings in lower stages. Similar results were obtained in the works of Le Coz et al. (2014), Sikorska & Renard (2017) and Garcia et al. (2020).

Diagnostics were performed to check whether the assumptions of the behavior of the residuals held after inference. For this, we have utilized the median rating curve for estimating the errors with respect to the gaugings. Figure 5 presents the diagnostic plots, which suggest that the residuals are slightly skewed-to-left, but do not deviate considerably from the theoretical quantiles from mid-to-high streamflows. Also, the dispersion of the residuals increases for the smaller streamflows, which indicates the heteroscedasticity was not entirely removed with the proposed structural error model. We note, however, that, although more complex models may be more effective for dealing with heteroscedasticity along the entire range of stages (e.g., Garcia et al., 2020), the increased computational effort for properly reproducing low flows might not be justified when the main objective of inference is estimating flood quantiles. At last, the partial autocorrelation function (PACF) shows a positive serial correlation, which indicates the residuals are not independent. Also, the PAFC suggests that a first order autoregressive model would be suitable to model the residuals dependence.

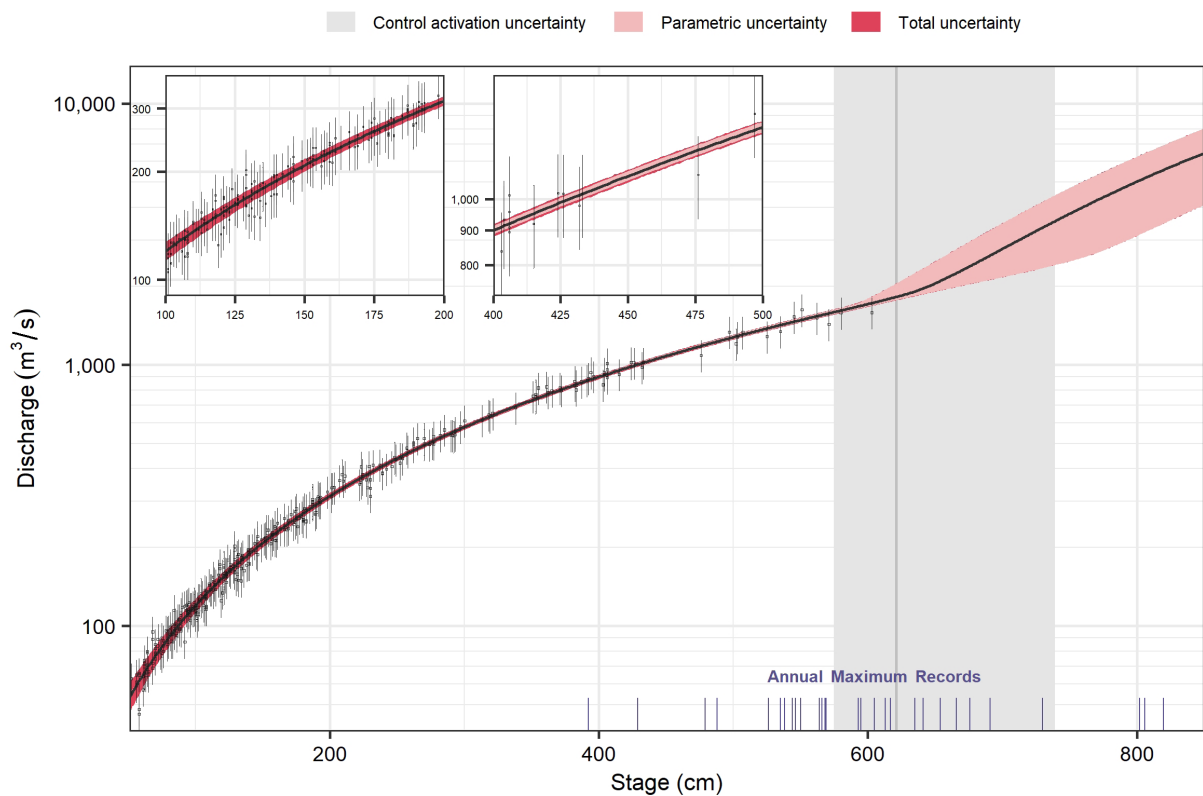
Once the samples of rating curves were obtained, we proceeded to Bayesian FFA. Following the indications from the literature, we have resorted to distinct strategies for eliciting prior distributions for the GEV parameters relying. For the location ( $\mu$ ) and scale ( $\sigma$ ) parameters, weakly informative prior distributions

**Table 1.** Prior uncertainty distributions for the rating curve parameters.

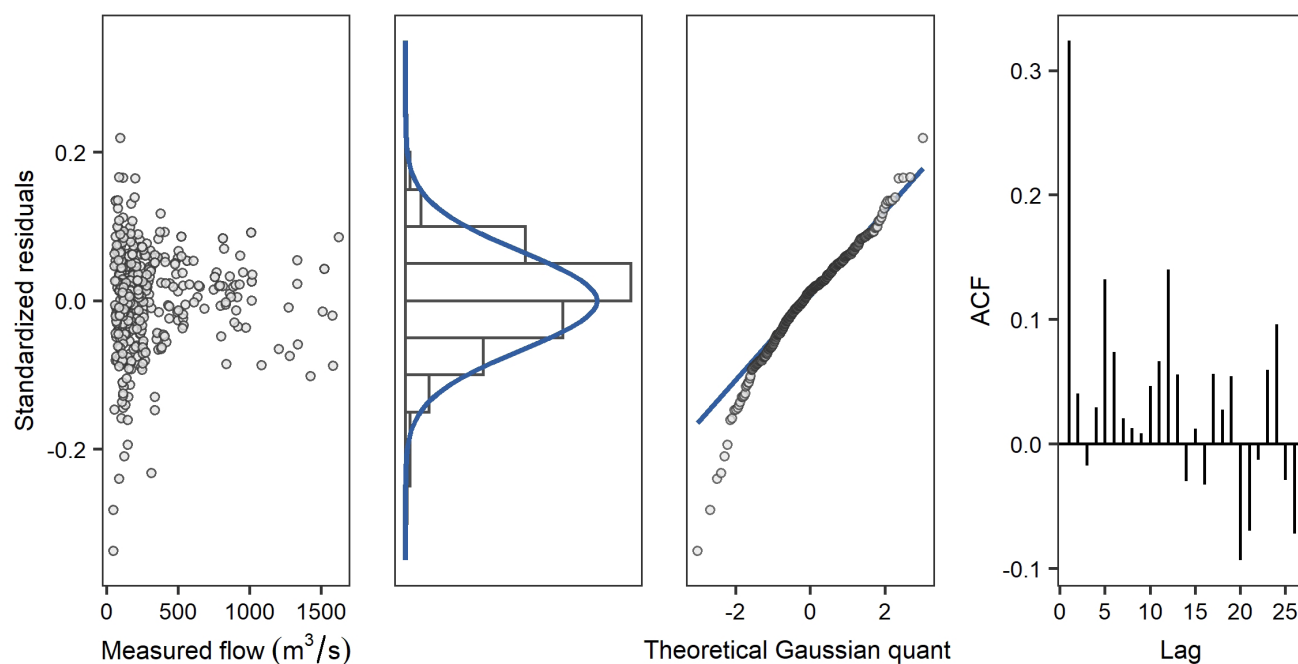
Cross-section element	$\pi(a_i)$	$\pi(k_i)$	$\pi(c_i)$
Main channel ( $i=1$ )	$N(164.7, 19^2)$	$N(-2.0, 1.0^2)$	$N(1.67, 0.025^2)$
Floodplains ( $i=2$ )	$N(884.7, 83.2^2)$	$N(4.48, 1.0^2)$	$N(1.67, 0.025^2)$



**Figure 3.** Marginal posterior distributions of the rating curve and error model parameters. The panels show the range of the posterior distribution parameters (the prior distributions are cropped within this range).



**Figure 4.** Estimated rating curve at Acorizal gauging station. The points and error bars show the measured stage-discharge at the field and its uncertainty (two standard deviation).



**Figure 5.** Diagnostic plots of the residuals for the error model.

were elicited as  $\mu \sim N(1600, 500^2)$  and  $\sigma \sim N_+(700, 300^2)$ , centered at values obtained by linear regression with catchments' drainage areas (Lima et al., 2016, Sampaio & Costa, 2021), and fairly large variances to account for potential physical and climatic differences among the catchment in study and those from the literature. The shape parameter ( $\xi$ ) is often difficult to estimate with reliability, especially for small samples (Martins & Stedinger, 2000; Bracken et al., 2016). Hence, a Gaussian prior distribution, centered at 0 and with a standard deviation of 0.3, was used to restrict the MCMC sampler of exploring values that would entail unreasonable heavy tails to the distribution (Renard, 2011). We note that this distribution entails a similar prior range as compared to the geophysical model proposed by Martins & Stedinger (2000) but requires less computational effort for allowing convergence.

The quantile estimates from model M1 to M3 were assessed in terms of the quantile curve with the annual maximum flows. The Bayesian model for FFA was also implemented using Rstan, with 4 Markov chains running in parallel with 1,500 iterations each (from which 750 were utilized for warm-up). A larger number of iterations was necessary, as compared to the rating curve inference, because the Stan sampler indicated low effective sample size at the tails of the parameters' distributions.

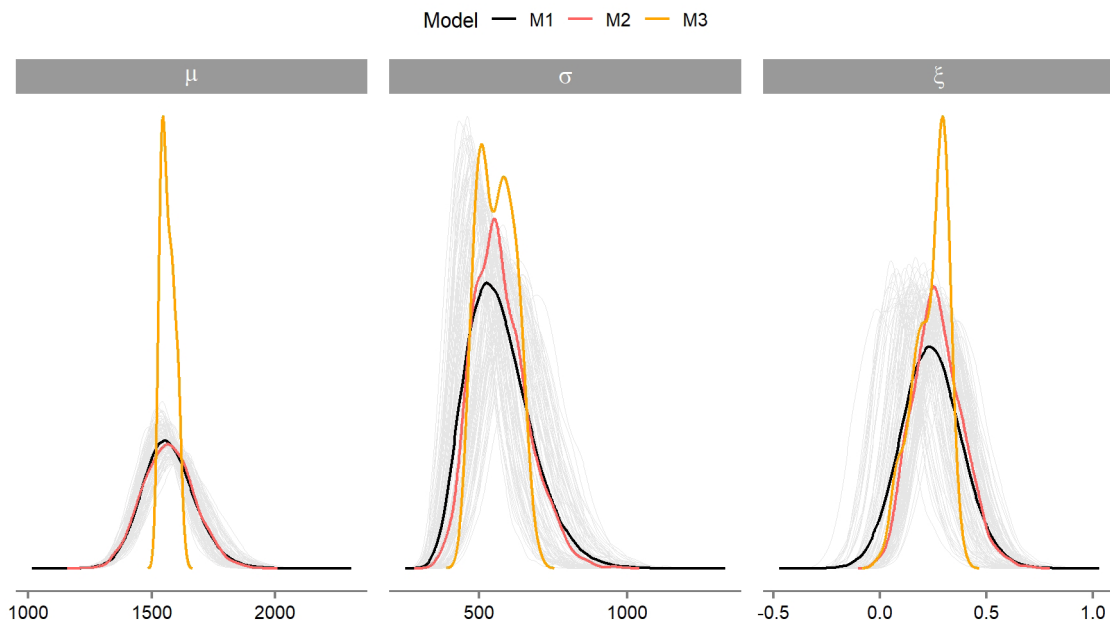
The posterior distribution of the GEV parameters for models M1, M2 and M3 are presented in Figure 6. First, it should be noted that the distributions are unimodal and approximately symmetric for most models - exception is made for the distribution of  $\sigma$  in model M3. As for the parameter  $\xi$ , similar point posterior estimates were obtained for the three models, with values of 0.24, 0.27 and 0.24 for M1, M2 and M3, respectively. On the other hand, the upper tail of model M1 is heavier than those of models M2 and M3. Also, the posterior variances of the three models

were considerably different, with a larger spread being verified for model M1 (the posterior coefficients of variation were equal to 0.59, 0.43 and 0.35 for models M1, M2 and M3, respectively). This probably occurred because of the rating curve uncertainty - while propagating the streamflow errors to FFA, larger values for the shape parameter (in absolute value) were simulated in the inference of M1, which would, in turn, enhance the uncertainty of the quantiles, as can be verified in Figure 7.

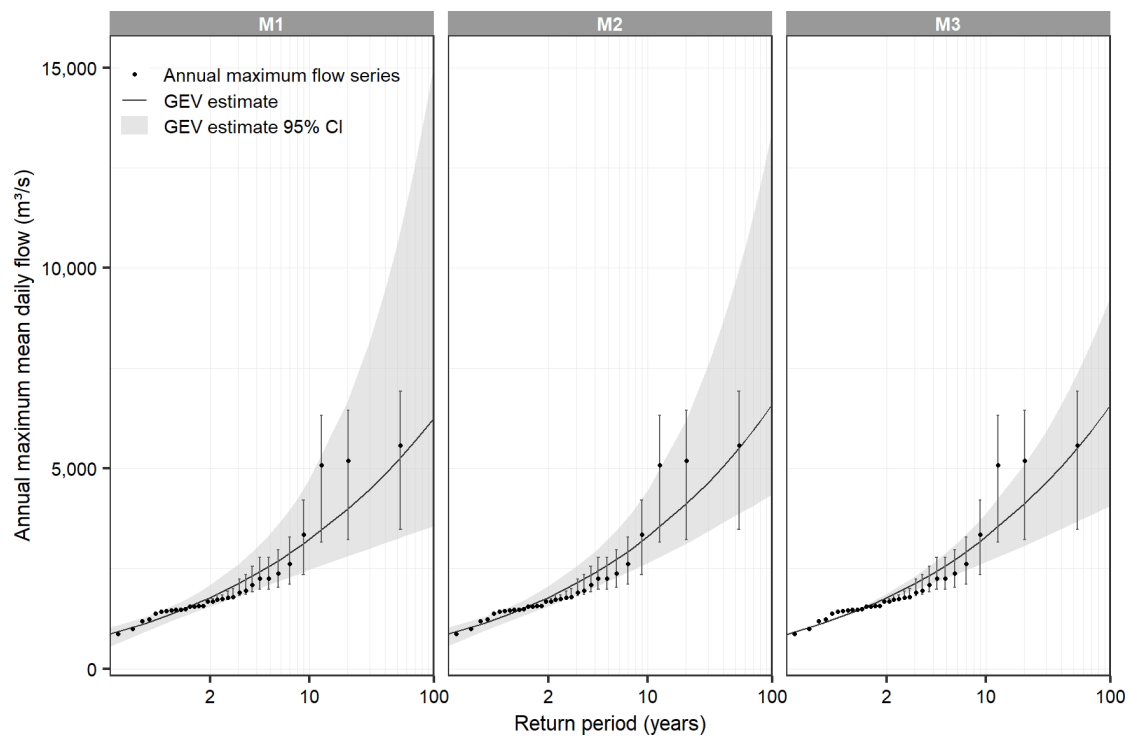
To further investigate the effects of rating curve uncertainty in the flow quantiles for different return periods, the quantile curves and their credibility intervals of 95% were compared (Figure 7). As the return periods increased, the uncertainties in all three models increased, as expected. It is also worth noting, that as the return times increase, the uncertainty due to the rating curve also increases, directly influencing the credible intervals. In effect, for the 100-year return level, the inclusion of the rating curve uncertainty in model M1 (Figure 7, left panel) led to an increase of 26% in the width of the credible interval and to higher positive skewness, as compared to model M2 (Figure 7, central panel). For model M3 (Figure 7, right panel), the credible intervals were considerably narrower, which suggests that the sampling errors play a major role on the overall uncertainty. On the other hand, the credibility intervals of the M3 model are approximately symmetric, while in the others are strong skewed-to-the-right, possibly due to the influence of the shape parameter  $\xi$ . Nonetheless, the effects of the rating curve uncertainty in quantile estimation are not negligible, as the credible interval for the 100-year flood quantile ranges from 6,256 to 6,602 m<sup>3</sup>/s.

Under conditions of limited stage-discharge information, the flood quantile curve was remarkably different. In effect, when pooling fewer flow measurements for inferring the rating curve





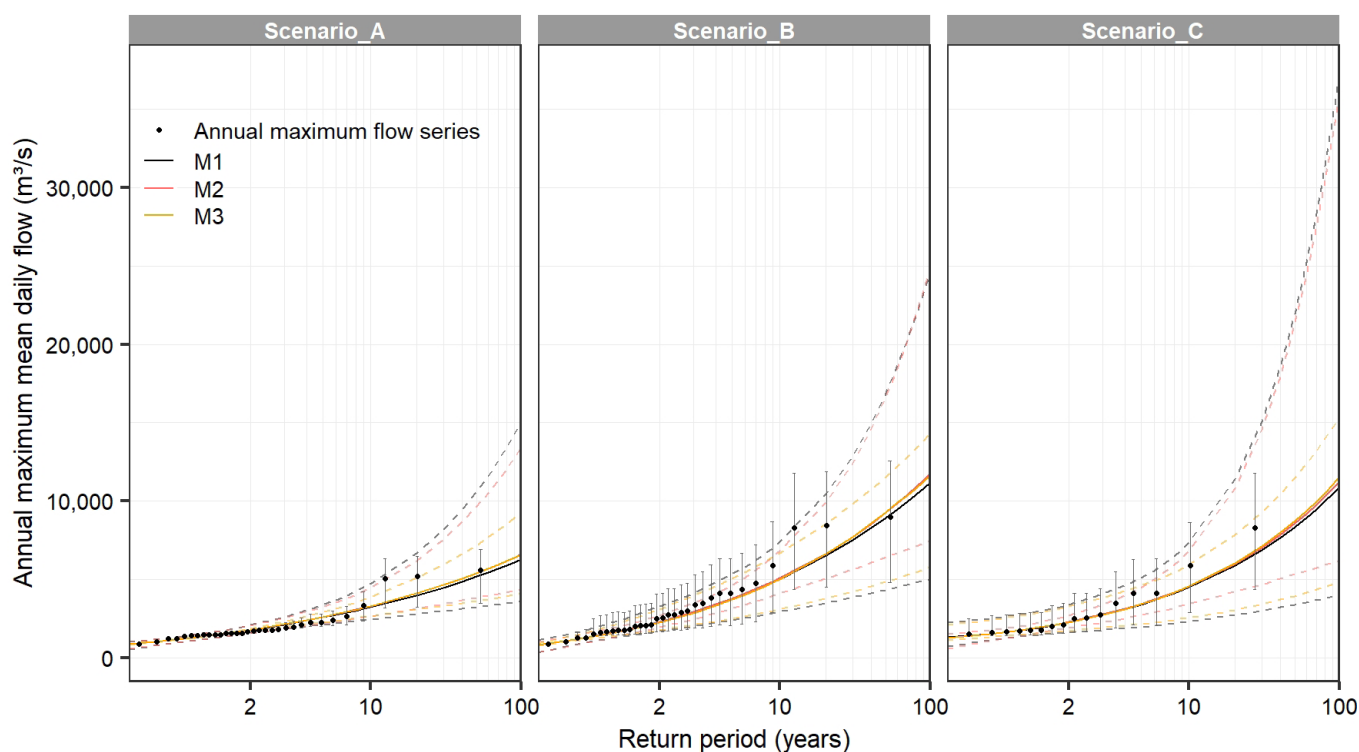
**Figure 6.** Posterior estimates of the GEV parameters  $\mu, \sigma$  and  $\xi$ . Light gray lines show the individual posterior distribution of each 100 rating curves that comprise model M1 mixture distribution.



**Figure 7.** GEV quantile curve for models M1 to M3 - scenario A. The error bars comprise the annual maximum flow series uncertainty (95% Credible Interval - CI) as stemming from the rating curve.

parameters (scenarios B and C), the data and the prior distributions alone could not precisely estimate the stage-discharge relationship at high stages (i.e., the extrapolated portion of the rating curve; see Figure S2). For the highest recorded gauging (39% higher than the second maximum record), the rating curve bias, with respect to the measured flow, was 71% in scenarios B and C, whereas,

in scenario A, it amounted about 9%. Also, an average increase of 123% in the width of the credible intervals was observed as compared to the complete gauging dataset. These facts highlight the importance of at least a few flow measurements in flooding conditions for properly describing the high flow regime, despite the physical basis for extrapolation provided in BaRatin.



**Figure 8.** GEV quantile curve for models M1 to M3 - all scenarios. The error bars comprise the annual maximum flow series uncertainty (95% Credible Interval - CI) as stemming from the rating curve. The dashed lines denote the 95% CI from the fitted quantile curve.

The bias introduced by the use of censored gauging information propagated into the FFA, shifting the modes of the GEV parameters towards higher values and increasing their posterior variances in scenarios B and C (Figure S3). As a result, the point estimates of the 100-year quantile in all models were almost twice higher than their counterparts in scenario A (Figure 8). The increase in the rating curve uncertainty due to the partial pool of the stage-discharge measurements also inflated the credible intervals, which were 64% to 91% higher among the models from scenario B, as compared to their counterparts in scenario A for the 100-year quantile (see Table S1). Finally, the interactions among the different uncertainty sources appear to be much more complex when the upper portion rating curve is poorly defined. In fact, the comparison between models M1 and M2 in scenario B suggests that it is difficult to distinguish the marginal effects of sampling errors from those stemming from the rating curve - because of the larger scatter of the higher flows in this case, the widths of the credible intervals are much more similar than those in scenario A, probably due to the higher similarity of the tail index posterior distributions between M1 and M2. Nonetheless, it is still possible to perceive the influence of the rating curve uncertainty in extreme quantile estimation as the posterior distribution of  $\xi$  in model M1 presents a slightly heavier tail than that of M2, which entails wider credible intervals (Figure 8).

At last, scenario C, which combines the incremental uncertainty of the censored rating curve and a shorter sample, illustrates the difficulties for performing FFA in situations of scarce flow information. In effect, for this scenario the posterior distributions of the tail index for models M1 and M2 virtually

overlap, which likely results from the larger influence of the informative prior distribution over the information aggregated by the data. In addition, the variance of the posterior distributions of  $\xi$  are much larger than in the previous scenarios, which caused a considerable increase in the widths of the credible interval for all models. Finally, we note that M1 presents wider credible intervals than M2, possibly due to the shift of the GEV location parameter towards lower values for the former model (Figure S3). However, in both models, the rating curve uncertainty appears to play a secondary role in extreme quantile estimation, as compared to that of the sampling errors.

## CONCLUSIONS

In this paper, the effects of rating curve uncertainty on the estimation of maximum flow quantiles were investigated by combining a Bayesian rating curve model, which takes into account physical knowledge for defining prior distributions, and Bayesian frequency analysis model under the GEV distribution. With respect to the analysis, a few considerations can be mentioned. (i) similarly to Steinbakk et al. (2016), we observed that, for our case study, the sampling uncertainty presented a larger contribution in increasing the uncertainty in the estimation of quantiles, as compared to the uncertainty arising from the rating curve, which may possibly be associated with the size of the sample of maximum annual flows; (ii) an increase in the influence of rating curve uncertainty was observed as the return periods increased; and, (iii) in this case study, the bias introduced in the quantile estimates by fewer

discharge measurements over shorter annual maximum series was more pronounced.

The use of Bayesian approaches that make the representation of physical processes more realistic, can reduce potential uncertainties. Moreover, the inclusion of different sources of uncertainty in the definition of flood quantiles and the assessment of their contributions, either individually or in combination, can support and guide decision makers in risk management, allocation of resources, and prioritization of actions for damage minimization, control, and prevention.

We believe that the proposed approach may certainly be further deepened by addressing some of its limitations, such as residual evaluations that include uncertainties, consideration of autocorrelation in the residuals, and a higher level of knowledge of the study area and hydrometric measurement, which may stem from more accurate Digital Terrain Models (DEM), identification of hydraulic controls, definition of roughness coefficients and survey reports of the flow measurements. This might allow the elucidation of more informative prior distributions for the parameters of the rating curve, with the intention of incorporating larger levels of physical realism into the models.

## ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. The authors also acknowledge the anonymous reviewers and the editors for their valuable comments and suggestions, which greatly helped improve the paper.

## REFERENCES

- Arcement Junior, G. J., & Schneider, V. R. (1989). *Guide for selecting Manning's roughness coefficients for natural channels and floodplains*. Denver, CO: United States Government Printing Office.
- Baldassarre, G., Laio, F., & Montanari, A. (2012). Effects of observation errors on the uncertainty of design floods. *Physics and Chemistry of the Earth*, 42-44, 85-90.
- Bracken, C., Rajagopalan, B., Cheng, L., Kleiber, W., & Gangopadhyay, S. (2016). Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain. *Water Resources Research*, 52(8), 6643-6655.
- Centre for Research on the Epidemiology of Disasters. United States Agency for International Development (2020). *Disaster\* year in review 2019*. Retrieved in 2021, September 23, from <https://reliefweb.int/report/world/cred-crunch-newsletter-issue-no-58-april-2020-disaster-2019-year-review>
- Coles, S. G. (2001). *An introduction to statistical modeling of extreme events*. London: Springer.
- Costa, V., & Sampaio, J. (2021). Bayesian approach for estimating the distribution of annual maximum floods with a mixture model. *Journal of Hydrologic Engineering*, 26(6), 04021017.
- Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180-190.
- García, R., Costa, V., & Silva, F. (2020). Bayesian rating curve modeling: alternative error model to improve low-flow uncertainty estimation. *Journal of Hydrologic Engineering*, 25(5), 04020012.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). New York: Chapman and Hall/CRC.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3), 423-453.
- Joint Committee for Guides in Metrology – JCGM. (2010). *Evaluation of measurement data—Guide to the expression of uncertainty in measurement*. Sèvres: JCGM.
- Juston, J., Jansson, P., & Gustafsson, D. (2014). Rating curve uncertainty and change detection in discharge time series: case study with 44-year historic data from the Nyangores river, Kenya. *Hydrological Processes*, 28(4), 2509-2523.
- Kastali, A., Zeroual, A., Remaoun, M., Serrano-Notivolli, R., & Moramarco, T. (2021). Design flood and flood-prone areas under rating curve uncertainty: area of Vieux-Ténès, Algeria. *Journal of Hydrologic Engineering*, 26(3), 1-12.
- Kiang, J. E., Gazorian, C., McMillan, H., Coxon, G., Coz, J., Westerberg, I. K., Belleville, A., Sevrez, D., Sikorska, A. E., Petersen-Øverleir, A., Reitan, T., Freer, J., Renard, B., Mansanarez, V., & Mason, R. (2018). A comparison of methods for streamflow uncertainty estimation. *Water Resources Research*, 54(10), 7149-7176.
- Koutsoyiannis, D. (2021). *Stochastics of hydroclimatic extremes: a cool look at risk*. Athens: Kallipos.
- Lang, M., Pobanz, K., Renard, B., Renouf, E., & Sauquet, E. (2010). Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis. *Hydrological Sciences Journal*, 55(6), 883-898.
- Le Coz, J. (2012). *A literature review of methods for estimating the uncertainty associated with stage-discharge relations*. World Meteorological Organization. Retrieved from <http://citeserx.ist.psu.edu/viewdoc/download?doi=10.1.1.400.8656&rep=rep1&type=pdf>
- Le Coz, J., Renard, B., Bonnifait, L., Branger, F., & Boursicaud, R. (2014). Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: a Bayesian approach. *Journal of Hydrology*, 509, 573-587.
- Lima, C. H. R., Lall, U., Troy, T., & Devineni, N. (2016). A hierarchical Bayesian GEV model for improving local and regional flood quantile estimates. *Journal of Hydrology*, 541, 816-823.

- Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3), 737-744.
- McClymont, K., Morrison, D., Beevers, L., & Carmen, E. (2020). Flood resilience: a systematic review. *Journal of Environmental Planning and Management*, 63(7), 1151-1176.
- McMillan, H. K., & Westerberg, I. K. (2015). Rating curve estimation under epistemic uncertainty. *Hydrological Processes*, 29(7), 1873-1882.
- McMillan, H., Freer, J., Pappenberger, F., Krueger, T., & Clark, M. (2010). Impacts of uncertain river flow data on rainfall – runoff model calibration and discharge predictions. *Hydrological Processes*, 24(10), 1270-1284.
- McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrological Processes*, 26(26), 4078-4111.
- McMillan, H., Seibert, J., Petersen-Overleir, A., Lang, M., White, P., Snelder, T., Rutherford, K., Krueger, T., Mason, R., & Kiang, J. (2017). How uncertainty analysis of streamflow data can reduce costs and promote robust decisions in water management applications. *Water Resources Research*, 53(7), 5220-5228.
- Moges, E., Demissie, Y., Larsen, L., & Yassin, F. (2021). Review: sources of hydrological model uncertainties and advances in their analysis. *Water*, 13(28), 1-23.
- Moyeed, R. A., & Clarke, R. T. (2005). The use of Bayesian methods for fitting rating curves, with case studies. *Advances in Water Resources*, 28, 807-818.
- Naggettini, M. (2017). *Fundamentals of statistical hydrology*. Cham: Springer.
- Ocio, D., Vine, N., Westerberg, I., Pappenberger, F., & Buytaert, W. (2017). The role of rating curve uncertainty in real-time flood forecasting. *Water Resources Research*, 53(5), 4197-4213.
- Osorio, A. L. N. A., & Reis Junior, D. S. (2016, May). A Bayesian approach for the evaluation of rating curve uncertainties in flood frequency analyses. In *World Environmental and Water Resources Congress*, 482–491. West Palm Beach, United States of America. Retrieved in 2022, April 30, from <https://ascelibrary.org/doi/epdf/10.1061/9780784479858.050>
- Paes, R. (2011). *Análise da translação da onda de cheia efluente do reservatório da UHE Manso na bacia hidrográfica do rio Cuiabá, Mato Grosso* (Master dissertation). Retrieved in 2021, September 5, from: <https://www.teses.usp.br/teses/disponiveis/18/18138/tde-02062011-131557/pt-br.php>
- Reitan, T., & Petersen-Overleir, A. (2009). Bayesian methods for estimating multi-segment discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, 23(5), 627-642.
- Renard, B. (2011). A Bayesian hierarchical approach to regional frequency analysis. *Water Resources Research*, 47(11), 1-21.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. *Water Resources Research*, 46(5), 1-22.
- Sampaio, J., & Costa, V. (2021). Bayesian regional flood frequency analysis with GEV hierarchical models under spatial dependency structures. *Hydrological Sciences Journal*, 66(3), 422-433.
- Serviço Geológico do Brasil. (2011). *Atlas pluviométrico do Brasil*. Retrieved in 2022, February 3 from <http://www.cprm.gov.br/publique///Mapas-e-Publicacoes/Atlas-Pluviometrico-do-Brasil-1351.html>
- Sikorska, A. E., & Renard, B. (2017). Calibrating a hydrological model in stage space to account for rating curves uncertainties: general framework and key challenges. *Advances in Water Resources*, 105, 51-66.
- Stan Development Team. (2022). *RStan: the R interface to Stan*. R package version 2.26.9. Retrieved in 2022, May 31. Stan Development Team, from <https://mc-stan.org/>.
- Steinbakk, G. H., Thorarinsdottir, T. L., Reitan, T., Schlichting, L., Hølleland, S., & Engeland, K. (2016). Propagation of rating curve in design flood estimation. *Water Resources Research*, 52(9), 6897-6915.
- Thorarinsdottir, T. L., Steinbakk, G. H., Schlichting, L., & Engeland, K. (2018). Bayesian regional flood frequency analysis for large catchments. *Water Resources Research*, 54(9), 6929-6947.
- United Nations Office for Disaster Risk Reduction – UNDRR. (2019). *Global assessment report on disaster risk reduction*. Geneva: UNDRR.
- US Army Corps of Engineers. Institute for Water Resources. Hydrologic Engineering Center. (1986). *Accuracy of computed water surface profiles*. Davis: Hydrologic Engineering Center.
- Westerberg, I. K., Sikorska-Senoner, A. E., Viviroli, D., Vis, M., & Seibert, J. (2020). Hydrological model calibration with uncertain discharge data. *Hydrological Sciences Journal*. Ahead of print.
- World Meteorological Organization (2017). *Guidelines for the assessment of uncertainty for hydrometric measurement*. Geneva: World Meteorological Organization.

### Authors contributions

Luan Marcos da Silva Vieira: Methodology definition, results analysis and article writing.

Júlio César Lôbo Sampaio: Methodology definition, results analysis and article writing.

Veber Afonso Figueiredo Costa: Methodology definition, results analysis and article writing.

**Editor-in-Chief:** Adilson Pinheiro

Julian Cardoso Eleutério: results analysis and writing of the article.

**Associated Editor:** Carlos Henrique Ribeiro Lima

## SUPPLEMENTARY MATERIAL

Supplementary material accompanies this paper.

**Figure S1.** Stage records time series, flow measurement campaigns and Annual Maximum Series (AMS) of the stage time series considered for the flood frequency analysis. Red lines on the bottom of the chart indicate missing daily stage records.

**Figure S2.** Estimated rating curve at Acorizal gauging station under scenarios A, B and C. The points and error bars show the measured stage-discharge at the field and its uncertainty (two standard deviation). The annual maximum stage records are shown in the lower part of the figures in dark-blue lines. The stage records used in the Flood Frequency Analysis of scenarios B and C are shown in shorter yellow lines. In scenario A, the whole dataset is used to infer the rating curve's parameters and prediction interval. In scenarios B and C, the stage-discharge dataset is reduced from 410 records to 40 records by keeping only one measurement per year and removing records above the estimated bankfull stage.

**Figure S3.** Posterior estimates of the GEV parameters  $\mu$ ,  $\sigma$  and  $\xi$  under scenarios A, B and C. Light grey lines show the individual posterior distribution of each 100 rating curves that comprise model M1 mixture distribution.

**Table S1.** Point estimates for the 100-year quantile and its 95% credible interval (in the parenthesis) for models M1 to M3 in scenarios A, B and C.

This material is available as part of the online article from <https://www.scielo.br/j/rbrh>