Estimating flood recurrence uncertainty for non-stationary regimes

Estimativa da incerteza de recorrência de cheias para regimes não estacionários

Yan Ranny Machado Gomes[1] , **Laís de Almeida Marques[2]**   **& Christopher Freire Souza[3]**

[1]Universidade Federal de Pernambuco, Recife, PE, Brasil
[2]Universidade de Brasília, Brasília, DF, Brasil
[3]Universidade Federal de Alagoas, Maceió, AL, Brasil
E-mails: yanr.machado@gmail.com (YRMG), marques.lais@gmail.com (LAM), christopher.souza@ctec.ufal.br (CFS)

Abstract

Assuming non-stationarity in flood frequency models is still controversial due to uncertainty in estimates. In this study, a hierarchical Bayesian framework for flood frequency analysis is presented without assuming the stationarity hypothesis. We account data and model uncertainty in all modelling steps and use the Pardo River, Brazil, as study case. Results showed the presence of increasing trends in floods in Pardo River. The stationary model underestimated floods compared to the non-stationary model. Physical-based covariates models performed better than time-based showing the importance of adding physical covariates to explain the trend behavior. The presented model is adaptable to other case. Finally, this study provided guidance for the flood recurrence estimation under non-stationary conditions.

Keywords: Flood Frequency Analysis; Bayesian; Markov Chain Monte Carlo; Extremes; Climate Change.

Resumo

Assumir a não estacionariedade em modelos de frequência de cheia ainda é controverso devido à incerteza nas estimativas. Neste estudo, um modelo bayesiano hierárquico para análise de frequência de cheias é apresentado sem assumir a hipótese de estacionariedade. Levamos em consideração incertezas dos dados e do modelo em todas as etapas de modelagem e utilizamos o rio Pardo, Brasil, como estudo de caso. Os resultados mostraram a presença de tendências crescentes nas cheias no rio Pardo. O modelo estacionário subestimou as cheias em comparação com o modelo não estacionário. Os modelos de covariantes de base física tiveram melhor desempenho do que os baseados em tempo, mostrando a importância de adicionar covariantes físicas para explicar o comportamento da tendência. O modelo apresentado é adaptável a outros estudos de caso. Finalmente, este estudo forneceu orientação para a estimativa de recorrência de cheias em condições não estacionárias.

Palavras-chave: Análise de Frequência de Cheias; Bayesiano; Monte Carlo via Cadeias de Markov; Extremos; Mudanças Climáticas.

INTRODUCTION

An accurate characterization of magnitude, frequency, duration, and timing of floods is key for water management (Merz et al., 2012). The estimation of flood recurrence is useful in many applications, including natural stream design, urban zoning and planning, and design of hydraulic structures (Gordon et al., 2004; Montanari et al., 2009). The Flood frequency analysis (FFA) is a method of fitting a probability distribution to a dataset to estimate the future occurrence of an event of interest, allowing us to determine the probability of flood's magnitude and frequency, and therefore, supporting water management practices and policies (Rao & Hamed, 2000; Brunner et al., 2021; Slater et al., 2021).

Traditional FFA methods require flood data set to be: i) independent and identically distributed, and ii) stationary (i.e., statistical properties do not change over time). When assuming stationarity, we're implying that the chosen return period (RP) in the structure's design will be the same by the end of its planned design life (Salas & Obeysekera, 2014).

However, as many studies have identified significant trends in hydrological series, the stationarity assumption does not always support. Trends in hydrological series were already identified in many studies in different regions. The 100-year flood may become more common in many watersheds in the United States (Vogel et al., 2011), trends in magnitude of snowmelt floods were identified in Canada (Cunderlik & Ouarda, 2009), annual streamflow in Europe (Stahl et al., 2012) and downward annual maxima flood trends in the southeast and southwest regions of Australia (Ishak et al., 2013). In Brazil, Detzel et al. (2011) identified trends in inflow series of hydroelectric plants. However, a more recent study by Berghuijs et al. (2017) found smaller increases in the frequency and magnitude of extreme floods in Brazil compared to Europe and the United States when analyzing multi-continental changes. Bartiko et al. (2019) observed more pronounced trends in the frequency rather than the magnitude of floods on an annual scale in Brazil. Despite this, the authors noted that flood events are becoming more frequent and intense in regions of Brazil with wet conditions and less frequent and intense in drier regions.

Although it is difficult to distinguish the exact cause of trends, researchers have identified several potential drivers, including atmospheric factors such as natural climate variability and human-induced climate change, catchment characteristics such as urbanization and land use changes, and alterations to rivers and streams resulting from in-stream channel engineering (Merz et al., 2012, Bloschl et al., 2015). According to Milly et al. (2008), anthropogenic effects on Earth's climate and its consequences on water cycle makes it unfeasible to consider the stationary assumption.

Conversely, some authors (Montanari & Koutsoyiannis, 2014; Serinaldi & Kilsby, 2015) argue that stationarity should be the default assumption in hydrologic series modeling due to uncertainties added to the nonstationary model, affecting its reliability, as well as the difficulty to demonstrate the main driver in hydrologic series changes. Therefore, these authors suggest there should be strong evidence of changes in extreme events before assuming non-stationarity.

To assess uncertainty in non-stationary models, the Bayesian analysis seems a promising approach. The Bayesian framework considers the addition covariates and, therefore, reduce uncertainties (Serinaldi & Kilsby, 2015). Some of the usual FFA methods, such as the Maximum-Likelihood may have some drawbacks related to instability in the likelihood function of small sample sizes, leading to estimates with high bias and variance (Hosking et al., 1985).

Several non-stationary techniques have been developed for FFA. For instance, Martins & Stedinger (2000) proposed an approach based on the Generalized Maximum Likelihood. Hosking (1990) adapted the L-moments estimation for non-stationary conditions, as further demonstrated by Gado & Nguyen (2016). Another approach involves a two-parameter distribution utilizing Generalized Additive Models for Location, Scale and Shape (GAMLSS), as proposed by Villarini et al. (2009). Some alternatives to narrow uncertainty in FFA include considering peaks-over-threshold (Thiombiano et al., 2017), adding other covariates such as drainage area (Lima et al., 2016; Wu et al., 2019), precipitation (Prosdocimi et al., 2014; Šraj et al., 2016), climate related covariates (Dong et al., 2019; Sharma et al., 2022), and a mixture of climate and physiographic basin characteristics (Viglione et al., 2016).

Numerous studies in FFA (Martins & Stedinger, 2000; El Adlouni et al., 2006; Gado & Nguyen, 2016; Bhat et al., 2019) use model selection methodologies that penalizes models with higher number of parameters (Burnham & Anderson, 2004) instead of considering the direct uncertainty caused by data and model parameterization (Vehtari et al., 2017). Bayesian studies often does not integrate over the posterior of parameters by considering point estimates like the posterior mean (Šraj et al., 2016). Bootstrapping is a common approach to account for prediction uncertainty by calculating confidence intervals (Serinaldi & Kilsby, 2015; Gado & Nguyen, 2016). However, its use is not straightforward for the non-stationary case, when we commonly have hierarchical data structure (*e.g.*, precipitation), whereas the bootstrap was originally designed for data modelled as i.i.d (Field & Welsh, 2007; Ren et al., 2010). Additionally, bootstrap does not consider informative priors, a drawback as no expert knowledge is used to narrow model uncertainty in a world where hydrological data is scarce.

Therefore, due to the necessity to better understand uncertainty in non-stationary flood recurrence credibility intervals, we present a hierarchical Bayesian model for flood risk estimation of non-stationarity rivers using GEV distribution. First, we provide guidelines to FFA under a fully Bayesian approach. We use the GEV distribution as likelihood in our Bayesian model and set its parameters to vary depending on covariates. Secondly, using the hydrological data of Pardo River watershed in southeast Brazil, we assess the proposed method by analyzing the flood recurrence intervals under the non-stationary assumption. We incorporate the use of precipitation information as explainable variable of non-stationarity and compare with the time covariant (in years) case.

MATERIAL AND METHODS

We used Bayesian inference for flood estimation (Figure 1). We used annual maximum streamflow and GEV as likelihood for all models. We set time and precipitation as covariates (99th quantile and total annual precipitation) and compared results

to stationary model. After sampling from Markov Chain Monte Carlo (MCMC), we assessed model convergence and quantified the goodness of fit for all tested models. We selected the Pardo River as study case to assess model performance.

Bayesian theory and MCMC

The Bayesian theorem involves calculating the probability of a set of parameters (θ) given a set of observations (X). Unlike frequentist approaches, which treat the variable of interest as a constant value, Bayesian theory considers it as a random variable to account for the uncertainty in its exact value. This uncertainty is represented by a probability distribution that can be calculated from the observations.

Some advantages of Bayesian method are its ability to provide the parameters' full posterior distribution, instead of only a point estimate, which allows the evaluation of uncertainties in the model, known as credible intervals.

To employ the Bayesian approach, we integrate the prior distribution, P(θ), which represents our prior knowledge about the parameter θ, with the likelihood of the dataset, P(X|θ). In cases where we lack robust prior information about the parameters, it is advisable to choose an uninformative prior distribution, which can be uniform or flat. Conversely, if we possess sufficient background information, we can set informative priors to better inform the analysis. The Bayes' theorem (Equation 1) is expressed as follows:

$$P(\theta|X) = \frac{P(\theta)P(X|\theta)}{P(X)} \tag{1}$$

Where $P(\theta|X)$ is the posterior distribution, or the probability of the hypothesis given the data or evidence X; $P(\theta)$ is the prior distribution; $P(X|\theta)$ is the likelihood function, or the probability of the data under the hypothesis, and it is usually given by a known distribution (for instance, GEV distribution); $P(X)$ makes the posterior distribution of probabilities for each hypothesis sum to one, and for this reason, is called normalizing constant.

For simple problems $P(X)$ can be obtained analytically. However, in non-trivial cases where the parameter vector is large, computational methods are required such as the Markov Chain Monte Carlo (MCMC; Gelman et al., 2013).

MCMC construct a Markov Chain whose stationary distribution is the target distribution (*i.e.*, Posterior) after a large number of steps. Among the several MCMC techniques, the Metropolis Hastings (MH) (Metropolis et al., 1953; Hastings, 1970) has been widely used in hydrology studies (Han & Coulibaly, 2017). It consists of sampling from a target distribution by making a random proposal for new parameter values and deciding if those values are incorporated to the chain or if they are discarded based on defined criteria. This procedure is repeated until the desired number of iterations is achieved, attaining the samples from the target distribution in the accepted parameters values.
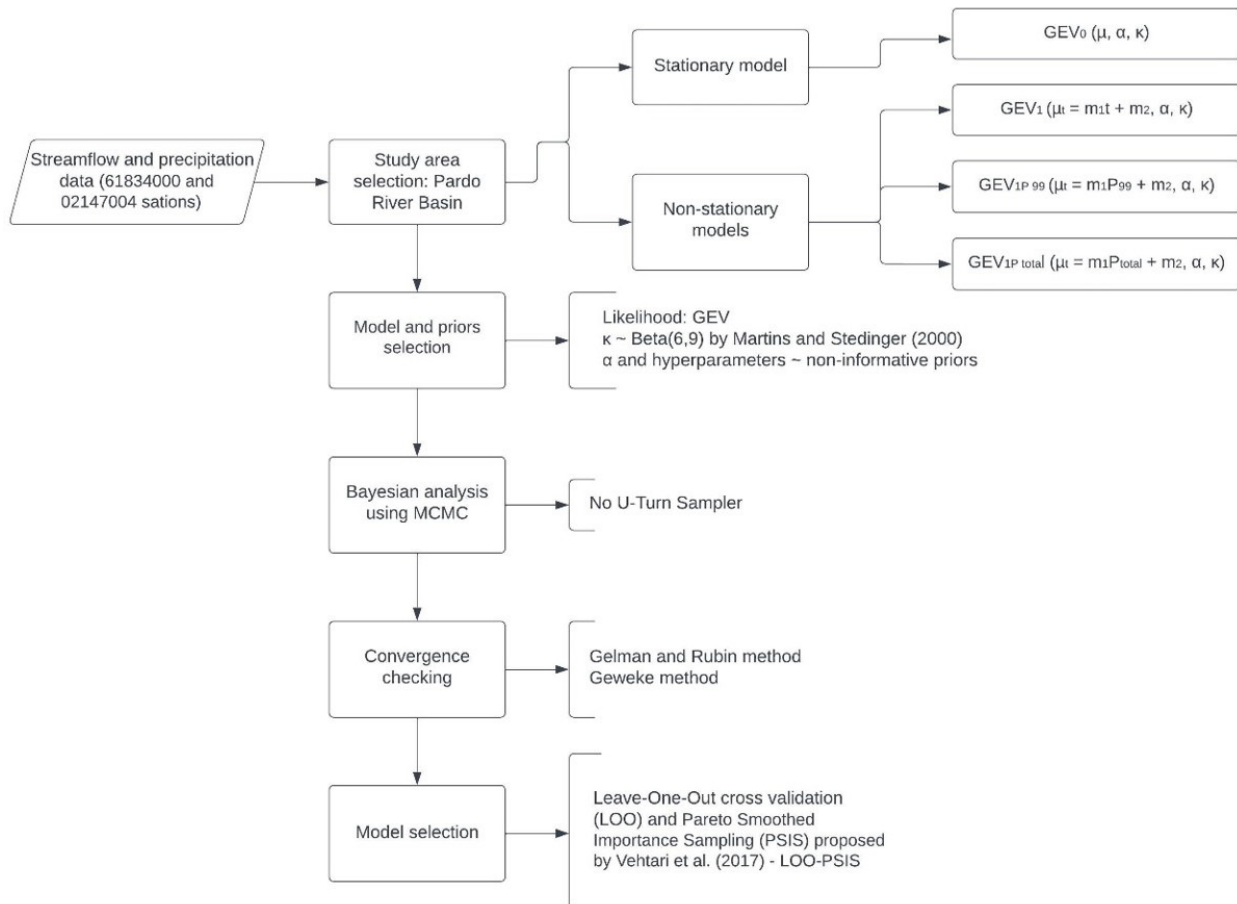

Figure 1. Workflow and summary of used techniques.

Nonetheless, a characteristic and drawback of MH is its own randomness, called the "random walk" since it tends to waste computational time or explore narrow regions of the parameters space in complex problems. Hence, Hamiltonian Monte Carlo (HMC) algorithm is a more efficient way to explore the parameter space avoiding the random walk and sensitivity to correlated parameters (Hartmann & Ehlers, 2017). Traditional HMC requires the specification of two parameters, and it is highly sensitive to them: step size and number of steps. In this matter, the No U-Turn Sampler (NUTS) use a recursive algorithm that eliminates the need to set a number of steps. NUTS performs equally or more efficiently than traditional HMC (Hoffman & Gelman, 2014). In this work, we used the the Probabilistic Programming framework written in Python, PyMC3 (Salvatier et al., 2016).

Stationary and non-stationary models

The Generalized Extreme Value (GEV) distribution is commonly used to model annual streamflow maxima (Katz et al., 2002). It includes three distribution families according to the shape parameter (κ): Gumbel ($\kappa = 0$), Fréchet ($\kappa < 0$) and Weibull ($\kappa > 0$). GEV distribution in its cumulative form (Equation 2) is represented by:

$$F(y) = \exp\left\{-\left[1 - \kappa\left(\frac{y-\mu}{\alpha}\right)^{1/\kappa}\right]\right\} \qquad (2)$$

where $\left(\mu + \frac{\alpha}{\kappa}\right) \leq y < \infty$ when $\kappa < 0$ (Fréchet), $-\infty < y < \infty$ when $\kappa = 0$ (Gumbel) and $-\infty < y \leq \left(\mu + \frac{\alpha}{\kappa}\right)$ when $\kappa > 0$ (Weibull). μ, α, κ are the location, scale and shape parameters, respectively. To calculate quantiles with non-exceedance probability P, the inverse of the cumulative distribution function is used.

Reasonable shape parameters lie between -0.5 and 0.5 due to GEV restrictions regarding skew and variance. Due to these mathematical restrictions and based on hydrological experience, it is recommended a $\kappa \sim Beta(p = 6, q = 9)$ prior centered in -0.1 (Martins & Stedinger, 2000). Moreover, for heavy tailed cases, most common in hydrological practice, the choice of parameters 6 and 9 gives better quantile estimates (Park, 2005). For location and scale, when no information is available, uninformative parameters are chosen.

Convergence checking

As in any other optimization method, the convergence of the MCMC process must be analyzed for a proper inference, as we are assuming that the samples are derived from the true posterior distribution. We can determine the minimum number of samples required to ensure a reasonable approximation. There are several known methods to perform convergence diagnostics. Once no method works in every case, it is recommended a combination of strategies to evaluate MCMC sampler convergence (Cowles & Carlin, 1996; El Adlouni et al., 2006). In this study, we chose Gelman and Rubin (Gelman & Rubin, 1992) and Geweke (Geweke, 1991) tests due to their ease interpretation and wide use in the literature (Najafi & Moradkhani, 2013; Nguyen et al., 2014; Seidou et al., 2012).

The Gelman and Rubin test is based on the comparison of the variation within the chains in relation to the total variation across the chains for all the iterations. Several Markov chains are run in parallel and the square root of the ratio between average within-chain and average between-chains variance of the likelihood are computed during the iterations. This ratio should ideally be equal to 1 if convergence is achieved.

The Geweke test splits the MCMC resulting chain in two parts and compare their mean. If the chain has converged, their mean should be similar. With this test is also possible to determine the burn-in period of the simulation (or how many initial iterations to discard), which would be equivalent to the smallest early portion of the chain that passes the diagnostic (El Adlouni et al., 2006).

Model selection

Cross-validation (CV) is primarily used to estimate model performance when limited data is available. In the standard approach, called k-fold CV, we split the dataset in k smaller sets. For each subset we sampled a group as a holdout dataset and took the remaining groups as a training dataset. We fitted the model using the training set and then, evaluated the model using the discarded set. In the Leave-One-Out cross-validation (LOO), we iterated through each data point to test model performance. To implement K-fold cross-validation we need to partition the data repeatedly and fit the model on every partition, which is computationally costly. To approximate LOO, importance sampling we used the Pareto Smoothed Importance Sampling (PSIS) (Vehtari et al., 2017). One advantage of PSIS-LOO is that the method is more robust than the Watanabe–Akaike information criterion (WAIC) in the finite case with weak priors or influential observations. As PSIS-LOO accounts for the whole posterior it is a fully Bayesian method, unlike AIC or DIC. Therefore, we used PSIS-LOO to assess prediction accuracy. Models with lower scores are preferred than the others. One concern regarding using LOO is the fact that we are using "future" data to validate past data, as we hold-out one data at a time for validation (Bürkner et al., 2020). However, LOO is able to select a model which can describe well the structure in the time series. The main assumption we have to make is that it is likely that such model we choose as the "best" would also make good predictions for future data as it did for the calibration data.

CASE STUDY

The Pardo River Basin was selected to assess the presented model in this study. Streamflow and precipitation records were obtained from Hidroweb from the Brazilian National Water Agency (ANA). Based on non-stationarity evidence in literature (Santos et al., 2016) we selected streamflow data from the Clube de Regatas station (code 61834000), located in the city of Ribeirão Preto, São Paulo Brazil. The streamflow timeseries correspond to 1941-2014 period. Data from nearby precipitation gauge station (code 02147004), was used. We considered a 5% maximum gap to include each year into the used dataset. After this pre analysis, and to pair both streamflow and precipitation data, we considered the period of 1945-2012.

Annual flood time series, annual precipitation percentile 99th, and total annual precipitation were grouped from the time series dataset. All three variables indicate an increasing trend (Figure 2).

All models in this study use GEV distribution as likelihood in the Bayesian framework. We chose to vary the location parameter in function of time and precipitation. The shape parameter was assumed as constant in time for the principle of parsimony. The models are summarized as:

- $GEV_0(\mu,\alpha,\kappa)$: A stationary case, where all parameters are constant;

- $GEV_1(m_1 t + m_2, \alpha, \kappa)$: A non-stationary case, where the location parameter (μ) is linearly dependent on time as covariate;

- $GEV_{1P99}(m_1 P_{99} + m_2, \alpha, \kappa)$: A non-stationary case, where the location parameter (μ) is linearly dependent on the 99th percentile of daily rainfall values during a hydrological year;

- $GEV_{1PTotal}(m_1 P_{Total} + m_2, \alpha, \kappa)$: A non-stationary case, where the location parameter μ) is linearly dependent on the total annual precipitation.

Time (t) was counted from the beginning of the record for GEV_1.

As indicated in section 2.2, Beta prior for the shape parameter was used for all models. For all other parameters and hyper-parameters uninformative priors were set. Details on distributions and boundaries are presented in Supplementary Material S1.

To make predictions we extrapolate covariates. As for GEV_1 covariate is time, extrapolation is evident, for the precipitation-based models we performed a linear regression of recorded precipitations and time to forecast the near-future precipitation data using the assumption floods will change in the same rate as in the recorded period.

RESULTS AND DISCUSSIONS

Model convergence

Figure 3 presents the trace plots for the GEV_1 model parameters obtained to Pardo River timeseries. It shows that all parameters follow a central tendency, meaning they are stable and there is low dependence from the initial state (i.e., they have low autocorrelation). All traceplots from GEV_1 present a zig-zag motion indicating a good exploration of the posterior distribution. To guarantee the chain has converged to its equilibrium value, we drawn multiple long chains and compared marginal parameter densities when estimating FFA. To choose the "long chain" quantity we checked different (computationally feasible) chain lengths from 1,500 to 10,000 and respective impacts on posterior summaries such as Highest Posterior Densities (HPD; most likely values of the parameters with a fixed probability, *e.g.*, 95%), mean and median values. There is no rule to choose chain lengths, but there was no evidence of longer chains needed in our sensitivity analysis. In our study, all trace plots converged (Figure 3). Parameter estimates do not move quickly from the central tendency after a few iterations from the initial value. This pattern means we have a good starting value (other trace plots not shown here).
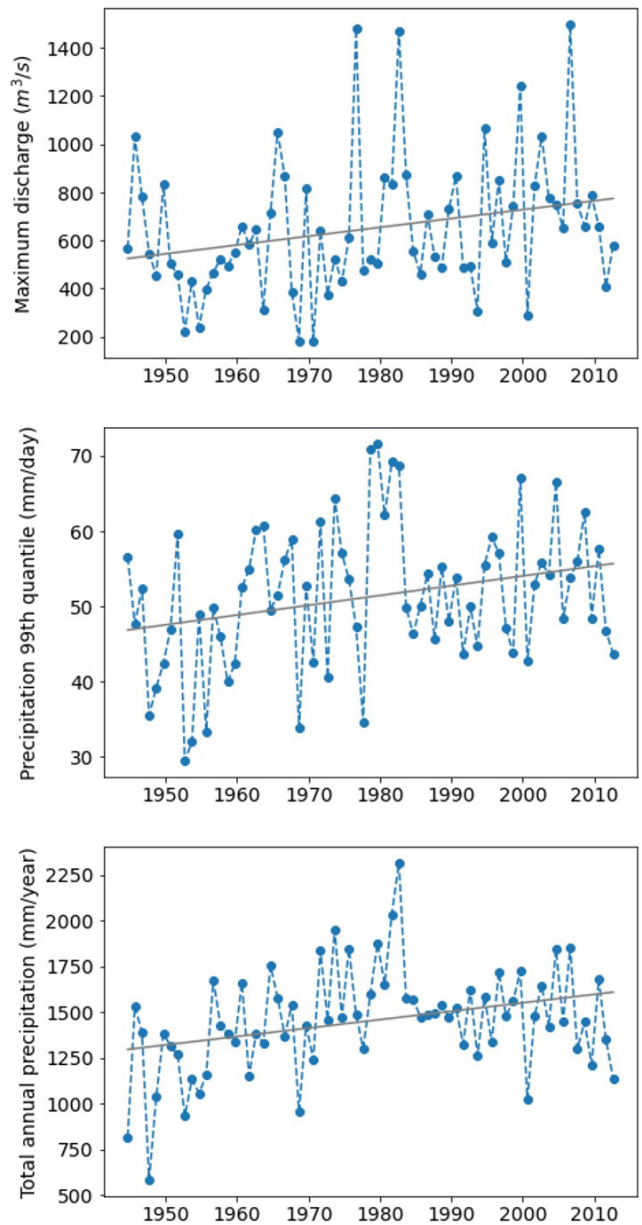

Figure 2. Time series in the Pardo River of discharge, annual 99th quantile precipitation and total annual precipitation. In gray, linear regression for time series.

Chain length values can also be chosen depending on burn in period identified by model converge diagnostics. Geweke and Gelman and Rubin tests showed convergence for a burn-in-period of 500 iterations and chain length of 2500 (parameters not shown here). Multiple chains were generated and then retested using convergence diagnostics. They did not show considerable variation in terms of convergence parameters. Due to the results of Geweke and Gelman and Rubin tests, we assume that the model achieved convergence. It is important to note that if some of the above methods fail convergence detection, the alternative is to try a different parametrization strategy, such as different priors and respective hyper parameters, or resetting model structure. Changing to more informative priors, rebuilding model setup, and even gathering more data might improve model convergence.
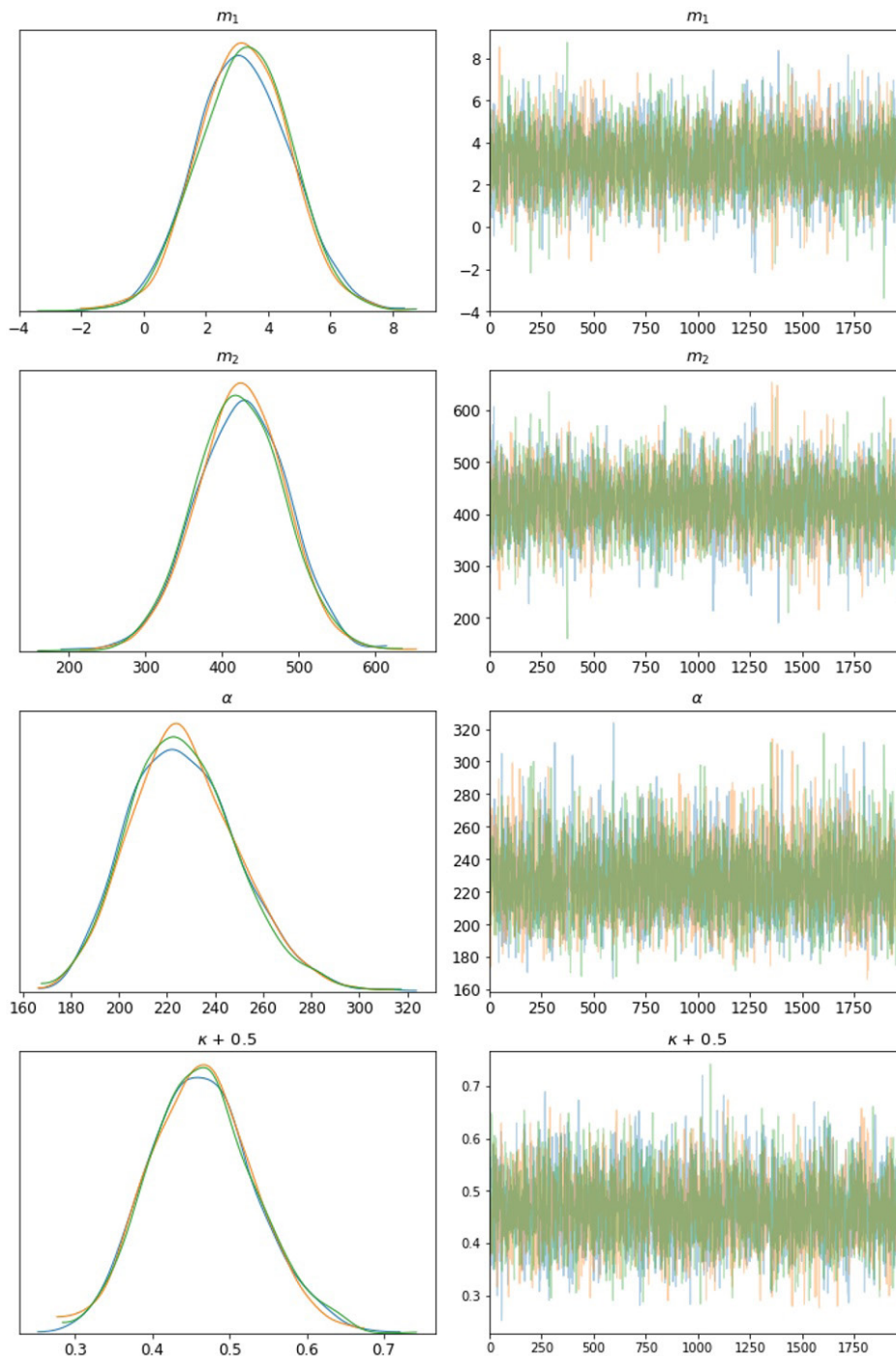
Figure 3. Trace plots and posterior densities of GEV_1 parameters in Pardo River. Different colors represent multiple simulated chains.

Parameter estimation

Table 1 shows the parameter estimation results for the studied models considering time and rainfall-based covariates. We found similar shape estimations (negative and close to zero) regardless of the studied model. For all the non-stationary models we found the angular coefficient for the location parameter m_1) to be positive not only for the posterior mean, but for the whole considered HPD. This confirms our suspicions of an increasing trend in Pardo River. In GEV_1 the hyperparameter m_1 can give us
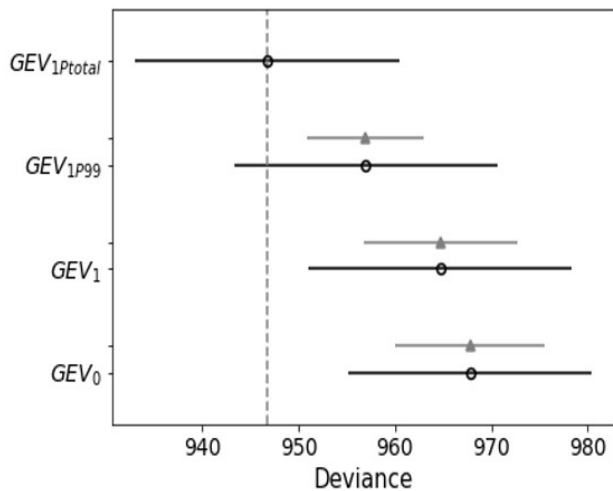
an indication of how much flow discharge increases per year. Compared to data mean (650 m³/s) it represents approximately 0.9% of the flow discharge per year. For hydraulic structures with long design lifespan this increase could be impactful under stationarity for the whole design period.

Comparison and prediction

GEV_{1P99} and $GEV_{1PTotal}$ performed better than GEV_0 and GEV_1 (Figure 4). LOO-score is represented in terms of Deviance, which

Table 1. Summary of the posterior estimates for all four models using time and rainfall-based covariates. Estimated parameters (posterior mean) are followed by 94% credible intervals (in brackets).

Model	μ	m_1	m_2	α	κ
GEV_0	528.38 [469.64; 586.65]	-	-	233.41 [195.04; 279.57]	-0.034 [-0.152; 0.096]
GEV_1	-	3.13 [0.43;5.84]	425.58 [313.56;529.41]	226.48 [184.82;269.86]	-0.032 [-0.149; 0.099]
GEV_{1P99}	-	8.64 [3.92;13.41]	92.55 [-134.14;355.49]	211.67 [175.78;255.43]	-0.016 [-0.139; 0.123]
$GEV_{1PTotal}$	-	0.424 [0.26;0.59]	-74.318 [-321.29;160.74]	194.27 [158.81;232.51]	-0.006 [-0.130; 0.149]

Figure 4. Models' comparison in terms of PSIS-LOO and WAIC. Empty circle represents the values of LOO and respective black error bars associated with them are the values of the standard deviation of LOO. Gray plots correspond to Watanabe–Akaike Information Criterion (WAIC).

is the LOO log-score multiplied by -2 (lower Deviance is better). The model that has total annual precipitation as covariate resulted as the best model in our study considering the cross-validation analysis.

Comparing GEV_0 and GEV_1 in terms of the PSIS-LOO score, GEV_1 performed best (964.6 over 967.9), but errors overlap. Taking in to account the uncertainty in their estimation, the PSIS-LOO seems quite flat across these two models, meaning uncertainty mostly overlaps across models. This indicates that even though GEV_1 performed best, it does not differ much from the stationary model in terms of PSIS-LOO. In this case, model complexity could not outweigh its benefit in absolute fit than the simpler model (even when there is an established increase trend). PSIS-LOO should guide our decision about the best model. However, it should not be the only metric to decide which model to effectively choose. Also, one limitation from PSIS-LOO or others cross-validation metrics is that they can only provide a relative test of model quality. In other words, PSIS-LOO cannot answer if the model is good in absolute sense (Vehtari et al., 2017). This means that we are comparing which model performs best with these specific chosen parameters. To find more informative priors and test different covariates that may improve model accuracy is one of the main focus on Bayesian Flood Frequency Analysis (Han & Coulibaly, 2017).

Figure 5 shows flood prediction from years 2022-2050 to the 2 and 100 years return period for our models. 2 and 100 years were chosen to represent a common and a rare (and potentially catastrophic) event. GEV_1 overestimated more than other non-stationary models and presents higher uncertainty even for low return period values such as the 2-year presented. As LOO accounts for the whole posterior, the higher uncertainty in GEV_1 agrees with higher LOO values for this model compared to the precipitation-based ones. GEV_1 also shows higher differences in trend attenuation (the slope hyperparameter) then the other non-stationary, which shows the value of adding physical-based covariates to non-stationary models for narrowing uncertainty. $GEV_{1Ptotal}$ and GEV_{1P99} present similar results on prediction. However, GEV_{1P99} slightly underestimate compared to $GEV_{1Ptotal}$. The 99th quantile for the maximum predicted precipitation for year 2050 is 60.6 mm, which means a discharge of 1552 m^3/s for the expected 100 years flood considering the posterior mean of our model. This value is 2.1% higher than the predicted value using $GEV_{1Ptotal}$ for 2022-year, 1520 m^3/s.

All non-stationary models that have time as covariate, are extrapolations of the trend pattern and infer about events that have not been observed. However, there is no guarantee that the increasing pattern of floods in Pardo River will hold indefinitely. Future data may indicate change in the non-stationarity intensity. Therefore, it is important to identify what are the main drivers of non-stationarity to predictions (Viglione et al., 2016).

Considerations on the time frame predictions must be taken parsimoniously, while prediction data include a new layer of uncertainty. Even though a linear regression is fair for comparing different models, it severely underestimates the variance of the series and is unsuitable for a long-time ahead extrapolation. An alternative is using Regional Climate Models (Giorgi, 2019), which can provide more reliable information of climate-related variables in a changing climate (Gebrechorkos et al., 2019). It is also important to note that to design bigger hydraulic structures such as dams, we need to estimate decamillennial floods. In the proposed framework, it would require covariates to be extrapolated 10,000 years in the future as well, which would clearly incorporate more uncertainty to the problem.

Our results showed significant differences in quantile estimates between stationary and non-stationary models, with lower discharge values for the stationary case. Unjustified stationarity assumptions could lead to an increase in risk failure of hydraulic structures. Our results suggest that precipitation is a good covariate to explain the non-stationarity characteristic in Pardo River corroborating with other studies (Prosdocimi et al., 2014; Šraj et al., 2016; Viglione et al., 2016).
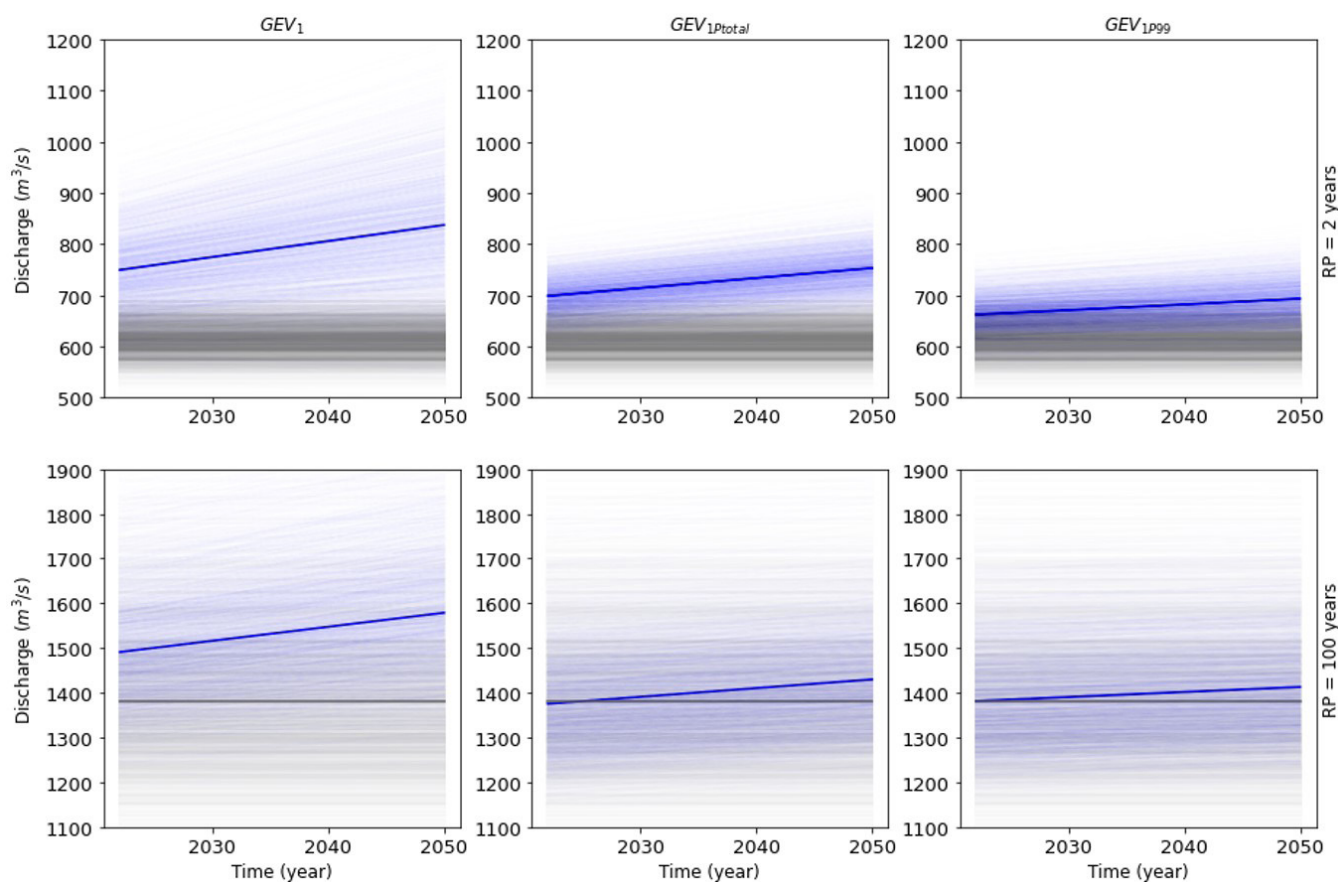
Figure 5. Discharge prediction for GEV_1, $GEV_{1Ptotal}$ and GEV_{1P99} models (in blue). In gray, stationary prediction.

In fact, other climate indicators should be explored for Pardo River such as done by other authors (Zhang et al., 2019; Dong et al., 2019; Liu et al., 2014). Using multiple explanatory variables can be an attractive option because it may help to better explain the variability in the data. However, it is important to study these variables carefully since including more variables does not necessarily lead to a reduction in uncertainty (Zhang et al., 2019).

As the integration of physical variables into non-stationary FFA methods seems promising, they still rely on extrapolation as in the time-based GEV_1 and have their own forecast uncertainty. The Bayesian method can deal with this uncertainty by adding a new level in model hierarchy. Naturally, this adds complexity to the model and their implications on how to model uncertainty should yet be explored which is out of the scope of this study. The representativeness of river discharge time series also must be considered when performing FFA. Small size and insufficient datasets are associated with even higher uncertainties and consequently unreliable estimations.

CONCLUSIONS

The Bayesian framework for flood forecast presented in this study is flexible and adaptable to all cases. By presenting uncertainty in prediction return periods decision makers have the option to be even more cautious on flow return levels when designing hydraulic structures. The addition of covariates and

expert knowledge to update priors were extremely valuable to increase model accuracy and make better predictions.

The limitation of the research work is still the uncertainty in extrapolating trends, as examining the assumption of the stationarity. Several aspects from the future work can be tackled from this study: consideration of regional data to better estimate priors, evaluation of Bayesian model flexibility and estimation uncertainty regarding the combination of multiple physical covariates, understand the role of other physical variables such as radar data, climate data and remotely sensed data.

REFERENCES

Bartiko, D., Oliveira, D. Y., Bonumá, N. B., & Chaffe, P. L. B. (2019). Spatial and seasonal patterns of flood change across Brazil. *Hydrological Sciences Journal*, *64*(9), 1071-1079. http://dx.doi.org/10.1080/02626667.2019.1619081.

Berghuijs, W. R., Aalbers, E. E., Larsen, J. R., Trancoso, R., & Woods, R. A. (2017). Recent changes in extreme floods across multiple continents. *Environmental Research Letters*, *12*(11), 114035. http://dx.doi.org/10.1088/1748-9326/aa8847.

Bhat, M. S., Alam, A., Ahmad, B., Kotlia, B. S., Farooq, H., Taloor, A. K., & Ahmad, S. (2019). Flood frequency analysis of river

Jhelum in Kashmir basin. *Quaternary International*, *507*, 288-294. http://dx.doi.org/10.1016/j.quaint.2018.09.039.

Bloschl, G., Gaál, L., Hall, J., Kiss, A., Komma, J., Nester, T., Parajka, J., Perdigão, R. A. P., Plavcová, L., Rogger, M., Salinas, J. L., & Viglione, A. (2015). Increasing river floods: fiction or reality? *WIREs. Water*, *2*(4), 329-344. http://dx.doi.org/10.1002/wat2.1079.

Brunner, M. I., Slater, L., Tallaksen, L. M., & Clark, M. (2021). Challenges in modeling and predicting floods and droughts: a review. *Wiley Interdisciplinary Reviews: Water*, *8*(3), e1520. http://dx.doi.org/10.1002/wat2.1520.

Bürkner, P. C., Gabry, J., & Vehtari, A. (2020). Approximate leave-future-out cross-validation for Bayesian time series models. *Journal of Statistical Computation and Simulation*, *90*(14), 2499-2523. http://dx.doi.org/10.1080/00949655.2020.1783262.

Burnham, K. P., & Anderson, D. R. (2004). Model selection: understanding AIC and multimodel inference, with contrasts to BIC. *Sociological Methods & Research*, *33*, 261-304. http://dx.doi.org/10.1177/0049124104268644.

Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, *91*(434), 883-904. http://dx.doi.org/10.1080/01621459.1996.10476956.

Cunderlik, J. M., & Ouarda, T. B. (2009). Trends in the timing and magnitude of floods in Canada. *Journal of Hydrology (Amsterdam)*, *375*(3-4), 471-480. http://dx.doi.org/10.1016/j.jhydrol.2009.06.050.

Detzel, D. H. M., Bessa, M. R., Vallejos, C. A., Santos, A. B., Thomsen, L. S., Mine, M. R., Bloot, M., & Estrocio, J. (2011). Estacionariedade das afluências às usinas hidrelétricas brasileiras. *RBRH*, *16*(3), 95-111. http://dx.doi.org/10.21168/rbrh.v16n3.p95-111.

Dong, Q., Zhang, X., Lall, U., Sang, Y. F., & Xie, P. (2019). An improved nonstationary model for flood frequency analysis and its implication for the Three Gorges Dam, China. *Hydrological Sciences Journal*, *64*(7), 845-855. http://dx.doi.org/10.1080/02626667.2019.1596274.

El Adlouni, S., Favre, A. C., & Bobée, B. (2006). Comparison of methodologies to assess the convergence of Markov chain Monte Carlo methods. *Computational Statistics & Data Analysis*, *50*(10), 2685-2701. http://dx.doi.org/10.1016/j.csda.2005.04.018.

Field, C. A., & Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *69*(3), 369-390. http://dx.doi.org/10.1111/j.1467-9868.2007.00593.x.

Gado, T. A., & Nguyen, V. (2016). An at-site flood estimation method in the context of nonstationarity II. Statistical analysis of floods in Quebec. *Journal of Hydrology (Amsterdam)*, *535*, 722-736. http://dx.doi.org/10.1016/j.jhydrol.2015.12.064.

Gebrechorkos, S. H., Hülsmann, S., & Bernhofer, C. (2019). Regional climate projections for impact assessment studies in East Africa. *Environmental Research Letters*, *14*(4), 044031. http://dx.doi.org/10.1088/1748-9326/ab055a.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457-472.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton: CRC Press.

Geweke, J. F. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments* (No. 148). Minneapolis: Federal Reserve Bank of Minneapolis.

Giorgi, F. (2019). Thirty years of regional climate modeling: where are we and where are we going next? *Journal of Geophysical Research, D, Atmospheres*, *124*(11), 5696-5723. http://dx.doi.org/10.1029/2018JD030094.

Gordon, N. D., McMahon, T. A., Finlayson, B. L., Gippel, C. J., & Nathan, R. J. (2004). *Stream hydrology: an introduction for ecologists* (2nd ed.). Hoboken: John Wiley & Sons, LTD.

Han, S., & Coulibaly, P. (2017). Bayesian flood forecasting methods: a review. *Journal of Hydrology (Amsterdam)*, *551*, 340-351. http://dx.doi.org/10.1016/j.jhydrol.2017.06.004.

Hartmann, M., & Ehlers, R. S. (2017). Bayesian inference for generalized extreme value distributions via Hamiltonian Monte Carlo. *Communications in Statistics. Simulation and Computation*, *46*(7), 5285-5302. http://dx.doi.org/10.1080/03610918.2016.1152365.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97-109. https://doi.org/10.1093/biomet/57.1.97.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(1), 1593-1623.

Hosking, J. R. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B. Methodological*, *52*(1), 105-124. http://dx.doi.org/10.1111/j.2517-6161.1990.tb01775.x.

Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, *27*(3), 251-261. http://dx.doi.org/10.1080/00401706.1985.10488049.

Ishak, E. H., Rahman, A., Westra, S., Sharma, A., & Kuczera, G. (2013). Evaluating the non-stationarity of Australian annual maximum flood. *Journal of Hydrology (Amsterdam)*, *494*, 134-145. http://dx.doi.org/10.1016/j.jhydrol.2013.04.021.

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8-12), 1287-1304. http://dx.doi.org/10.1016/S0309-1708(02)00056-8.

Lima, C. H., Lall, U., Troy, T., & Devineni, N. (2016). A hierarchical Bayesian GEV model for improving local and regional flood quantile estimates. *Journal of Hydrology (Amsterdam)*, *541*, 816-823. http://dx.doi.org/10.1016/j.jhydrol.2016.07.042.

Liu, D., Guo, S., Lian, Y., Xiong, L., & Chen, X. (2014). Climate-informed low-flow frequency analysis using nonstationary modelling. *Hydrological Processes*, *29*, 2112-2124. http://dx.doi.org/10.1002/hyp.10360.

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. http://dx.doi.org/10.1029/1999WR900330.

Merz, B., Kundzewicz, Z. W., Delgado, J., Hundecha, Y., & Kreibich, H. (2012). Detection and attribution of changes in flood hazard and risk. In Z. W. Kundzewicz (Ed.), *Changes in flood risk in Europe* (pp. 435-458). Boca Raton: CRC Press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087-1092. http://dx.doi.org/10.1063/1.1699114.

Milly, P. C., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: whither water management? *Science*, *319*(5863), 573-574. http://dx.doi.org/10.1126/science.1151915.

Montanari, A., Shoemaker, C. A., & Van de Giesen, N. (2009). Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology: an overview of issues and challenges. *Water Resources Research*, *45*(12), 1-2. http://dx.doi.org/10.1029/2009WR008471.

Montanari, A., & Koutsoyiannis, D. (2014). Modeling and mitigating natural hazards: stationarity is immortal! *Water Resources Research*, *50*(12), 9748-9756. http://dx.doi.org/10.1002/2014WR016092.

Najafi, M. R., & Moradkhani, H. (2013). Analysis of runoff extremes using spatial hierarchical Bayesian modeling. *Water Resources Research*, *49*(10), 6656-6670. http://dx.doi.org/10.1002/wrcr.20381.

Nguyen, C. C., Gaume, E., & Payrastre, O. (2014). Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. *Journal of Hydrology (Amsterdam)*, *508*, 385-396. http://dx.doi.org/10.1016/j.jhydrol.2013.09.058.

Park, J. S. (2005). A simulation-based hyperparameter selection for quantile estimation of the generalized extreme value distribution. *Mathematics and Computers in Simulation*, *70*(4), 227-234. http://dx.doi.org/10.1016/j.matcom.2005.09.003.

Prosdocimi, I., Kjeldsen, T. R., & Svensson, C. (2014). Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. *Natural Hazards and Earth System Sciences*, *14*(5), 1125-1144. http://dx.doi.org/10.5194/nhess-14-1125-2014.

Rao, A. R., & Hamed, K. G. (2000). *Flood frequency analysis*. Boca Raton: CRC Press.

Ren, S., Lai, H., Tong, W., Aminzadeh, M., Hou, X., & Lai, S. (2010). Nonparametric bootstrapping for hierarchical data. *Journal of Applied Statistics*, *37*(9), 1487-1498. http://dx.doi.org/10.1080/02664760903046102.

Salas, J. D., & Obeysekera, J. (2014). Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, *19*(3), 554-568. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000820.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ. Computer Science*, *2*, e55. http://dx.doi.org/10.7717/peerj-cs.55.

Santos, C. A., Lima, A. M. M., Farias, M. H. C. S., Aires, U. R. V., & Serrão, E. A. O. (2016). Análise estatística da não estacionariedade de séries temporais de vazão máxima anual diária na bacia hidrográfica do Rio Pardo. *Holos*, *32*(7), 179-193. http://dx.doi.org/10.15628/holos.2016.4892.

Seidou, O., Ramsay, A., & Nistor, I. (2012). Climate change impacts on extreme floods I: combining imperfect deterministic simulations and non-stationary frequency analysis. *Natural Hazards*, *61*, 647-659. http://dx.doi.org/10.1007/s11069-011-0052-x.

Serinaldi, F., & Kilsby, C. G. (2015). Stationary is undead: uncertainty dominates the distribution of extremes. *Advances in Water Resources*, *77*, 17-36. http://dx.doi.org/10.1016/j.advwatres.2014.12.013.

Sharma, S., Ghimire, G. R., Talchabhadel, R., Panthi, J., Lee, B. S., Sun, F., et al (2022). Bayesian characterization of uncertainties surrounding fluvial flood hazard estimates. *Hydrological Sciences Journal*, *67*(2), 277-286. http://dx.doi.org/10.1080/02626667.2021.1999959.

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., Kelder, T., Kowal, K., Lees, T., Matthews, T., Murphy, C., & Wilby, R. L. (2021). Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. *Hydrology and Earth System Sciences*, *25*(7), 3897-3935. http://dx.doi.org/10.5194/hess-25-3897-2021.

Šraj, M., Viglione, A., Parajka, J., & Blöschl, G. (2016). The influence of non-stationarity in extreme hydrological events on flood frequency estimation. *Journal of Hydrology and Hydromechanics*, *64*(4), 426-437. http://dx.doi.org/10.1515/johh-2016-0032.

Stahl, K., Tallaksen, L. M., Hannaford, J., & Van Lanen, H. A. J. (2012). Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble. *Hydrology and Earth System Sciences*, *16*(7), 2035-2047. http://dx.doi.org/10.5194/hess-16-2035-2012.

Thiombiano, A. N., El Adlouni, S., St-Hilaire, A., Ouarda, T. B., & El-Jabi, N. (2017). Nonstationary frequency analysis of extreme daily precipitation amounts in Southeastern Canada using a peaks-over-threshold approach. *Theoretical and Applied Climatology*, *129*, 413-426. http://dx.doi.org/10.1007/s00704-016-1789-7.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413-1432. http://dx.doi.org/10.1007/s11222-016-9696-4.

Viglione, A., Merz, B., Viet Dung, N., Parajka, J., Nester, T., & Blöschl, G. (2016). Attribution of regional flood changes based on scaling fingerprints. *Water Resources Research*, *52*(7), 5322-5340. http://dx.doi.org/10.1002/2016WR019036.

Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., & Krajewski, W. F. (2009). Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources*, *32*(8), 1255-1266. http://dx.doi.org/10.1016/j.advwatres.2009.05.003.

Vogel, R. M., Yaindl, C., & Walter, M. (2011). Nonstationarity: flood magnification and recurrence reduction factors in the United States 1. *Journal of the American Water Resources Association*, *47*(3), 464-474. http://dx.doi.org/10.1111/j.1752-1688.2011.00541.x.

Wu, Y. B., Xue, L. Q., & Liu, Y. H. (2019). Local and regional flood frequency analysis based on hierarchical Bayesian model in Dongting Lake Basin, China. *Water Science and Engineering*, *12*(4), 253-262. http://dx.doi.org/10.1016/j.wse.2019.12.001.

Zhang, X., Duan, K., & Dong, Q. (2019). Comparison of nonstationary models in analyzing bivariate flood frequency at the Three Gorges Dam. *Journal of Hydrology (Amsterdam)*, *579*, 124208. http://dx.doi.org/10.1016/j.jhydrol.2019.124208.

Authors contributions

Yan Ranny Machado Gomes: Conceived the presented idea, performed the programming and calculations, analyzed the results, wrote the manuscript, wrote the final draft of the manuscript.

Laís de Almeida Marques: Analyzed the results, wrote the manuscript.

Christopher Freire Souza: Conceived the presented idea, analyzed the results, reviewed the manuscript.

Editor in-Chief: Adilson Pinheiro
Associated Editor: Adilson Pinheiro

SUPPLEMENTARY MATERIAL

Supplementary Material S1: Details on chosen priors.

This material is available as part of the online article from https://doi.org/10.1590/2318-0331.282320230031