Revista Brasileira de Geografia Física

ISSN:1984-2295

Homepage:https://periodicos.ufpe.br/revistas/rbgfe

Modelagem de precipitação máxima na bacia do rio São Francisco: uma comparação entre diferentes métodos de estimativa de parâmetros da função generalizada de valores extremos

Márcio Adalberto[1] Andrade e Luiz Felippe Salemi[2]

[1]Doutorando em Ciências Ambientais na Universidade de Brasília – UNB, Brasília, DF. Companhia de Desenvolvimento dos Vales do São Francisco e do Parnaiba - DF. SGAN 601 Módulo I Asa Norte. CEP: 70830019 - Brasília, DF – Brasil. Email: marcio.adalberto@gmail.com
[2]Professora doutor na Universidade de Brasília – UNB, Brasília, DF. Universidade de Brasília, Campus Planaltina. Campus da UnB. Área Universitária (Planaltina). CEP 73345010 - Brasília, DF - Brasil Email: lfsalemi@unb.br

Artigo recebido em 30/04/2024 e aceito em 23/05/2024

R E S U M O

Estudos de precipitação máxima têm grande relevância para o gerenciamento de recursos hídricos. A função de densidade de probabilidade generalizada de valores extremos (GEV) tem auxiliado na modelação de eventos extremos naturais. Existem vários métodos para inferir os parâmetros da função GEV, como o de máxima verossimilhança (MV) e momentos-L (ML). A inferência Bayesiana (IB) constitui uma alternativa à abordagem frequentista na medida em que permite combinar a informação conferida pelas amostras aleatórias de precipitações máximas com informação a priori sobre os parâmetros na estimação de modelos. Entretanto, poucos trabalhos utilizam a abordagem Bayesiana. Objetivou-se com o presente trabalho, apresentar a modelação da função GEV, com parâmetros estimados pelos seguintes métodos: MV, ML e IB. Com testes de adequação, definiu-se modelo probabilístico mais adequado, e realizou a predição de valores de precipitação máxima diária para os tempos de retorno 10, 25, 50, 100, 200, 500 e 1000 anos. A área de estudo foi a bacia hidrográfica do rio São Francisco e os de dados utilizados neste estudo foram obtidos a partir do banco de dados da Agência Nacional de Águas e Saneamento Básico (ANA). Utilizando o teste de AD, verificou-se que os métodos ML e IB foram considerados os mais adequados para estudos de probabilidade de precipitação máxima diária anual. Verificou-se que as estações com séries menos assimétricas, o método IB obteve melhores resultados enquanto as mais assimétricas o método ML foi melhor. Pode-se inferir que o método IB demostrou maiores incertezas para predição de chuvas com tempo de retorno acima de 100 anos.

Palavras Chaves**:** eventos extremos de precipitação, estimativa de parâmetros, inferência bayesiana.

Modelling of extreme rainfall in the São Francisco river basin: a comparison between different methods of estimating parameters of the generalized extreme value fuction

A B S T R A C T

Studies of extreme rainfall have great relevance for water resource management. The Generalized Extreme Value (GEV) density function has assisted in the modeling of extreme natural events. There are several methods for inferring the parameters of the GEV function, such as the Maximum Likelihood Estimation (MV) and L-Moments (ML) methods. Inference Bayesian (IB) constitutes an alternative to the frequentist approach in that it allows combining the information given by random samples of maximum precipitation with prior information about the parameters in the estimation of models. However, few works use the Bayesian approach. The aim of this work was to present the modeling of the GEV function, with parameters estimated by the following methods: MV, ML and IB. With goodness-of-fit tests, the most suitable probabilistic model was defined, and prediction of daily maximum precipitation values for return levels 10, 25, 50, 100, 200, 500, and 1000 years was performed. The study area was the São Francisco river basin, and the data used in this study were obtained from the National Water and Sanitation Agency database. Using the AD test, it was verified that the and IB methods were considered the most suitable for studies of annual daily maximum precipitation probability. It was found that the stations with less asymmetrical series, the IB method obtained better results while the most asymmetrical the ML method was better. It can be inferred that the IB method showed greater uncertainties for predicting rains with return time above 100 years.

Keywords: extreme precipitation events, parameter estimation, Bayesian inference.

Introdução

Os estudos de eventos extremos de precipitação são de grande importância na adaptação e avaliação da vulnerabilidade do homem ao impacto das mudanças climáticas. Compreender essa relação em face ao aquecimento global envolve aspectos hidrometeorológicos, entre eles a compreensão acerca dos eventos de fortes precipitações que possuem implicações ambientais, sociais e econômicas, como inundações, deslizamento de encostas, erosão de terras férteis, entre outros (Biswas e Bhattacharya, 2023; Kiefer et al., 2021).

A precipitação máxima diária é um gatilho para inundações e deslizamentos de terra (Wang et al., 2021; Chaithong, 2022). A Organização Mundial da Saúde (2023) relata que eventos extremos de precipitação e inundações estão aumentando em frequência e intensidade, e ainda mais ao longo deste século devido às mudanças climáticas.

Tendo em vista a expressiva concentração das populações em grandes áreas urbanas, a questão da vulnerabilidade das cidades em face aos eventos de precipitações máximas precisa ser trabalhada tanto em função da possível indução de ocorrência de inundações pela urbanização como também de mudanças climáticas globais. Portanto, se faz necessário quantificar tais eventos. Essas informações podem ser usadas como subsídio ao desenvolvimento de ferramentas e estratégias visando o incremento da resiliência das cidades. Descrever o comportamento estatístico de eventos extremos é crucial para limitar o risco de falha de obras e, consequentemente, limitar a perda de vidas humanas (Moccia et al., 2021).

Sigrist et al. (2012) afirmaram que a precipitação é um fenômeno complexo, que varia no espaço e no tempo. Dessa forma, a adaptação e consequente minimização dos impactos de eventos extremos sobre a sociedade passam primariamente pelo estudo e conhecimento científico da magnitude e frequência de ocorrência desses eventos ao longo do tempo, em particular visando a modelagem da ocorrência dos mesmos (Lima, 2011).

Uma estratégia para estudar eventos extremos de precipitação é usar uma dada distribuição de probabilidade para modelar a frequência dos dados observados de chuvas. O uso dessas distribuições é crucial para a previsão e, portanto, para a prevenção de desastres envolvendo chuvas máximas extremas (Yuan e Hu, 2020). Diversas distribuições teóricas são usadas para estudar eventos extremos máximos, como a distribuição Log-Normal de 2 e 3 parâmetros e a Gumbel (Back et al., 2011; Mello e Viola, 2013; Beskow et al., 2015), Gamma de 2 parâmetros (Franco et al., 2014), Generalizada de Valores Extremos – GEV (Bracken et al., 2018).

Em uma série temporal diária se o valor máximo for selecionado de um bloco predeterminado de tempo, por exemplo, anualmente, então a série temporal resultante seguirá uma distribuição de GEV de três parâmetros (Westra et al., 2013). Esta distribuição descreverá uma ampla gama de comportamentos de cauda e é denominada de método de blocos de máximo. Na análise de frequência máxima de bloco, uma distribuição GEV estacionária é ajustada a um conjunto de dados em um único local e os tempos de retorno são calculados para determinados intervalos (Ferreira & de Haan, 2015). Uma distribuição GEV estacionária é aquela em que os parâmetros dessa distribuição são constantes no tempo.

A metodologia de inferência estatística exerce influência sobre a qualidade do ajuste de um modelo probabilístico. Existem vários métodos de inferência estatística como o de máxima verossimilhança e momentos L. Naghettini e Pinto (2007) relataram que o método de máxima verossimilhança produz estimadores de baixa resistência. No entanto, envolve equações geralmente não lineares e implícitas. De acordo com esses autores, o método dos momentos L resulta em estimadores de parâmetros comparáveis em qualidade aos produzidos pelo método da máxima verossimilhança, sendo frequentemente mais precisos para pequenas amostras comumente usadas em estudos hidrológicos (Naghettini e Pinto, 2007).

A inferência Bayesiana constitui uma alternativa à abordagem frequentista na medida em que permite combinar a informação conferida pelas amostras aleatórias de precipitações máximas com informação a priori sobre os parâmetros na estimação de modelos de frequência. Existem vários estudos em que os métodos Bayesianos são utilizados combinando diferentes tipos de informação a priori tendo em vista a obtenção de um modelo unificado de probabilidades de precipitações máximas a posteriori, tais como Diriba e Debusho (2020), Ahmad et al. (2019), Bracken et al. (2018), Cotton (2008).

Uma das vantagens da Inferência Bayesiana está no fato de permitir a incorporação de informações a priori, melhorando, dessa forma, o processo de inferência. Contudo, em casos em que não existam informações, é comum o uso de *priores* que tenha uma alta variação, refletindo a ausência de informações prévias genuínas (Garcia-Aristizabal et al., 2015). As deficiências da

Inferência Bayesiana podem incluir perdas de informações (uma vez que apenas o máximo de cada ano é considerado) e se o tamanho das amostras forem pequenas (Moccia et al., 2021)

Nerantzaki e Papalexiou (2022), em uma revisão bibliográfica sobre o tema, apontaram que os estudos recentes sobre eventos extremos de precipitação são melhores descritos por distribuições de cauda pesada, tipo GEV, e que o Método dos Momentos é normalmente a primeira escolha na estimativa de parâmetros de distribuição, mas é superado em qualidade por métodos como o dos momentos L, máxima verossimilhança, mínimos quadrados e Bayesianos. De acordo com os esses autores, a distribuição GEV é a escolha mais popular para análise com aplicações para eventos extremos em muitos países tais como Áustria, Sibéria, Suíça, Itália, Etiópia, Escócia, Inglaterra, EUA.

Em face a relevância que o assunto requer, as definições dos melhores métodos para obtenção dos parâmetros da função de distribuição de probabilidades podem impactar o dimensionamento de obras hidráulicas como as de drenagem urbana, barragens, pontes, entre outras. Buscar ter o melhor conhecimento é fundamental no sentido de minimizar os prejuízos financeiros e perdas de vidas humanas.

Diante do exposto, os objetivos deste estudo foram: (i) comparar os parâmetros da Distribuição de Probabilidades Generalizada de Valores Extremos (GEV) estimados pelos os Métodos de Máxima Verossimilhança (MV), dos Momento - L (ML) e Inferência Bayesiana (IB); (ii) definir o modelo probabilístico mais adequado, com base na análise dos testes de adequação, para a variável aleatória precipitação máxima diária anual; (iii) com base no melhor modelo, predizer os valores de precipitação máxima diária para os tempos de retorno 10, 25, 50, 100, 200, 500 e 1000 anos.

Metodologia

Área de Estudo

A bacia hidrográfica do rio São Francisco (BSF) possui uma população de cerca de 18 milhões de habitantes em uma área de 636.216,80 km² que representa cerca de 7,5% da área total do Brasil e se estende pelos Estados de Minas Gerais, Bahia, Goiás, Pernambuco, Sergipe e Alagoas, além do Distrito Federal. Conforme o Plano de Recursos Hídricos da Bacia do Rio São Francisco (2017), a água que flui da bacia serve para múltiplos usos incluindo abastecimento doméstico, abastecimento industrial, irrigação, pesca, energia hidrelétrica e transporte. Mais de 330.000 hectares da bacia são dedicados à agricultura intensiva (11% da agricultura irrigada total do Brasil). A energia hidrelétrica que é gerada ali tem um potencial superior a 10.500 MW (CBHSF, 2017).

A bacia apresenta grande diversidade quanto às áreas irrigáveis, cobertura vegetal e fauna aquática. No tocante aos solos, predominam na bacia hidrográfica os latossolos (35,3%), os neossolos (26,5%) e os cambissolos (15,8%). A maior parte das terras é irrigável com aptidão restrita ou moderada (50%). Em relação à cobertura vegetal, a bacia contempla fragmentos de diversos biomas, salientando-se a Floresta Atlântica em suas cabeceiras, o Cerrado (Alto e Médio São Francisco) e a Caatinga (Médio e Submédio São Francisco). Com relação à fauna aquática, observa-se que o rio São Francisco apresenta a maior biomassa e diversidade de peixes de água doce da região Nordeste (CBHSF, 2017).

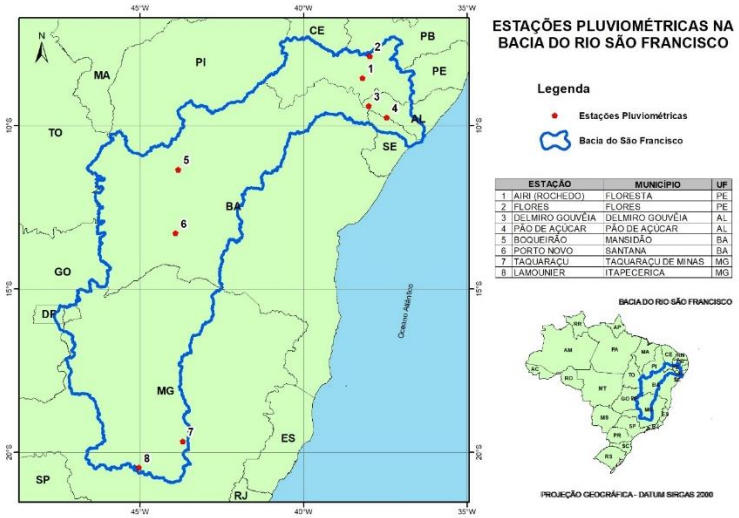

Figura 1. Distribuição das estações pluviométricas estudadas neste estudo, na bacia hidrográfica do rio São Francisco.

Dados Pluviométricos

Os conjuntos de dados utilizados neste estudo foram obtidos a partir do banco de dados hidrometeorológicos da Agência Nacional de Águas (ANA), através da plataforma HidroWeb - Sistema de Informações Hidrológicas, e correspondem a séries temporais de precipitação diária na BSF. Foram utilizados registros diários de chuva das estações pluviométricas operadas pela CPRH (Serviço Geológico do Brasil) sob responsabilidade da ANA.

Tabela 1 Estações pluviométricas e precipitação média anual para o período analisado.

Estações Pluviométricas									
Região Fisiográfica	Nome	Código	Município / UF	UF	Latitude	Longitude	Altitude (m)	Clima	Pluviosidade Média Anual
Alto São Francisco	LAMOUNIER	2045005	ITAPECERICA	MG	-20.4722	-45.0361	738	Tropical semi-úmido	1462
	TAQUARAÇU	1943023	TAQUARAÇU DE MINAS	MG	-19.6639	-43.6881	710	Tropical semi-úmido	1332
Médio São Francisco	PORTO NOVO	1343008	SANTANA	BA	-13.2925	-43.9092	442	Semi-árido	897
	BOQUEIRÃO	1143010	MANSIDÃO	BA	-11.3553	-43.8456	401	Semi-árido	901
Sub Médio São Francisco	AIRI (ROCHEDO)	838000	FLORESTA	PE	-8.5392	-38.1928	361	Semi-árido	495
	FLORES	737027	FLORES	PE	-7.8678	-37.9725	460	Semi-árido	728
Baixo São Francisco	PIRANHAS	937023	PIRANHAS	AL	-9.6261	-37.7561	110	Tropical semi-úmido	491
	DELMIRO GOUVÉIA	937013	DELMIRO GOUVÉIA	AL	-9.3928	-37.9942	256	Tropical semi-úmido	509

Diversos estudos realizados sobre eventos extremos de chuvas no Brasil fizeram o uso de séries históricas seguindo um limite mínimo de 10 a 15 anos de observações (Caldeira et al., 2015; Beskow et al. 2015; Aragão et al., 2013).

Para o presente estudo foram utilizados aqueles que satisfizeram os seguintes critérios: (a) as duas séries por região fisiográfica com maiores registros, neste estudo as séries possuem no mínimo 75 anos de dados, (b) porcentagem de valores ausentes por ano menor que 20% e (c) falhas não registradas no período chuvoso de outubro a março. Foi adotado como ano hidrológico, o período de outubro a setembro, conforme descrito no Plano Decenal de Recursos Hídricos da Bacia Hidrográfica do Rio São Francisco para o Controle de Cheias (CBHSF, 2004).

Tabela 2 Sumário Estatístico das séries de precipitação máxima das estações pluviométricas do estudo.

Região Fisiográfica	Estação Pluviométrica	Tamanho da Amostra	Mínimo (mm)	Máximo (mm)	Média (mm)	Desvio Padrão	Coeficiente de Variação	Coeficiente de Assimetria	Coeficiente de Curtose
Alto São Francisco	Lamounier	76	53,00	175,00	89,40	28,54	0,32	1,08	4,00
	Taquaraçu	76	47,40	147,20	83,73	22,47	0,27	1,02	3,74
Médio São Francisco	Porto Novo	75	40,60	230,10	79,99	30,21	0,38	2,62	12,50
	Boqueirão	82	35,00	213,80	80,02	27,69	0,35	1,85	8,94
Sub Médio São Francisco	Airi (Rochedo)	82	26,00	160,00	62,52	24,39	0,39	1,21	5,25
	Flores	105	30,50	137,80	74,81	19,57	0,26	0,24	3,86
Baixo São Francisco	Piranhas	77	16,40	129,00	55,64	26,04	0,47	0,90	3,42
	Delmiro Gouvéia	79	22,70	148,00	60,32	25,03	0,42	1,07	4,32

No estudo foi realizada uma etapa de análise preliminar de tratamento dos dados, sendo aplicados testes para avaliar a adequação dos dados a postulados básicos para a análise de frequência. Os testes utilizados no presente estudo são melhor detalhados em Naghettini e Pinto (2007). Foi utilizado o teste de Aleatoriedade proposto por pelo Instituto de Hidrologia da Grã-Bretanha (1975), onde a rejeição da hipótese de aleatoriedade de uma série hidrológica é realizada por meio do teste não-paramétrico do número de inflexões. O teste de homogeneidade proposto por Mann-Whitney (1947), que avalia se uma dada amostra provém de uma única e idêntica população. O teste de independência de Wald-Wolfowitz (1943) que essencialmente avalia que nenhuma observação presente na amostra pode influenciar a ocorrência de qualquer outra observação seguinte. E o teste de Spearman (NERC, 1975) para avaliar a estacionariedade da série, este refere-se ao fato que, excluídas as flutuações aleatórias, as observações amostrais são invariantes, com relação à cronologia de suas ocorrências.

Função Densidade de Probabilidade

A distribuição GEV é uma família de distribuição de probabilidade que incorpora a distribuição de Gumbel, de Fréchet e de Weibull. Todas estas funções possuem distribuições de máximas e foram introduzidas por Jenkison, em 1955 (Naghettini & Pinto, 2007). A função GEV possui três parâmetros: forma, escala e posição. O parâmetro de forma representa onde a maioria dos dados está concentrada. O parâmetro de escala descreve quão espalhado a distribuição e na medida em que o valor do parâmetro cresce, a função se torna mais espalhada. O parâmetro de posição

descreve a posição da distribuição em uma dada direção do eixo horizontal (Beijo et al., 2009).

A função de densidade de probabilidade da distribuição GEV, de acordo com Beskow et al. (2015) possui a seguinte descrição:

$$f(x) = \frac{1}{\sigma} . \left\{ (1-k).\ \left[(1-k).\left(\frac{x-\mu}{\alpha}\right) \right]^{\left(\frac{1}{k-1}\right)} exp\left\{ -\left[(1+\ k).\left(\frac{x-\beta}{\alpha}\right) \right]^{-\left(\frac{1}{k}\right)} \right\} \right\} \quad (1)$$

Em que:
f(x) = função da densidade de probabilidade (x);
x = variável aleatória (precipitação);
α = parâmetro de escala (que varia entre 0 e +∞);
μ = parâmetro de posição (varia de –∞ e + ∞);
k =parâmetro de forma (varia de –∞ e + ∞).

Para k> 0, a distribuição GEV é delimitada acima de acordo com a Eq. (2) e para k <0, o limite inferior de GEV é dado pela Eq. (3). Com −∞ <x <+ ∞ e k = 0, o GEV refere-se à distribuição assintótica do tipo I de valores extremos ou à distribuição de Gumbel, ε≤ x ≤ + ∞ e k <0, à distribuição assintótica do tipo II de valores extremos, ou Fréchet e −∞ <x <ω e k> 0, a distribuição assintótica do tipo III de valor extremo ou Weibull.

$$\omega = \beta + \frac{\alpha}{k} \quad (2)$$

$$\varepsilon = \beta + \frac{\alpha}{k} \quad (3)$$

Para o presente estudo, foram empregados, na estimativa dos parâmetros das distribuições de probabilidade, os métodos dos momentos-L (ML), Máxima Verossimilhança (MV) e Inferência Bayesiana (IB). Para detalhes destas metodologias recomenda-se a leitura em Naghettini & Pinto (2007).

Abordagem Bayesiana

A análise bayesiana permite a incorporação do conhecimento prévio na probabilidade visando melhorar a confiabilidade da estimativa.

Distribuição a Priori

De acordo com Coles & Powell (1996), o modelo GEV não admite uma distribuição conjugada a Priori, com isto são propostas obter famílias conjugadas para as sub-classes do modelo proposto, ou seja, distribuição conjungada para cada um dos parâmetros do modelo, constituindo assim uma Priori Hierárquica. Estes autores comentam que mesmo se a forma para cada parâmetro estiver disponível, isto pode não significar qual é claramente a priori conjungada apropriada, destacando-se a dificuldade de se identificar certas dependências entre os parâmetros de escala σ e forma k .Neste trabalho, os autores construíram uma priori usando a distribuição normal (μ, log σ, k), baseando as estimativas destes hiperparâmetros em outras localizações, onde o processo de interesse foi observado. Devido ao fato que $\sigma > 0$ é necessário aplicar uma transformação logarítmica no parâmetro de escala, de modo que o mesmo permaneça positivo, passando ser então adotado log σ (Cotton, 2008).

Conhecimento prévio sobre os parâmetros, que podem vir de outros conjuntos de dados ou da experiência e intuição física de um modelador, podem ser incluídos no modelo através da construção de distribuição prévia apropriada. Em alguns casos pode ser válido para utilizar este dispositivo para incluir conhecimento genuíno sobre o processo sob modelagem. Mais comumente, é difícil obter tal conhecimento e distribuições vagas que têm grande variação são então adotados para refletir esta ignorância prévia. A ferramenta bayesiana que adopta as distribuições prévias vagas pode ser dito que é essencialmente formal, mas leva a uma inferência em que parâmetro a incerteza é devidamente formalizada e para quais inferências as previsões são naturalmente tratadas (Coles et al. 2003).

Sendo assim, propõe-se neste trabalho uma priori baseada na abordagem utilizada por Coles & Powel (1996), que trata com uso da normal para cada parâmetro. De acordo com Smith (2005), a priori possuirá a seguinte forma:

$$Z \sim GEV\ (\alpha, \mu, k) \quad (4)$$

Em que:

$$k \sim N\ (0; 10) \quad (5)$$

$$\mu \sim N\ (0; 100) \quad (6)$$
$$\log \alpha \sim N\ (0; 1000) \quad (7)$$

Inferência Bayesiana

A Inferência Bayesiana constitui uma alternativa à inferência estatística baseada na função de verosimilhança. Sob o paradigma Bayesiano, os parâmetros são entendidos como variáveis aleatórias com distribuições a priori que devem ser definidas sem levar em consideração os dados amostrais. A inferência Bayesiana consiste na utilização do teorema de Bayes para obter uma distribuição a posteriori dos parâmetros que combina a informação fornecida pela amostra com a informação adicional antecedente e independente dos dados. O teorema de Bayes refere que:

$$\pi\ (\theta|x) = \frac{L\ (x|\theta)\ \pi\ (\theta)}{\int L\ (\theta|x)\pi\ (\theta)\ d\theta} \quad (8)$$

Em que $\pi(\theta \mid x)$ e $\pi\ (\theta)$ são, respectivamente as densidades a posteriori e a priori dos parâmetros, e $L\ (x|\theta)$ é a função de verosimilhança ou FDP conjunta da amostra aleatória x, embora expressa numa notação diferente no contexto Bayesiano. Na equação (8), o integral no denominador refere-se a uma constante de normalização da densidade a posteriori e, frequentemente, a sua solução analítica é muito complexa ou mesmo impossível.

Dado que o denominador é uma constante, é comum representar o teorema de Bayes por uma forma simplificada, tal como mostrada na equação (9):

$$\pi(\theta \mid x) \propto L(x \mid \theta)\pi(\theta) \quad (9)$$

Para implementação dos métodos relacionados no presente estudo, foi utilizado o Software R, e o algoritmo desenvolvido por Stephenson (2002).

Análise da qualidade dos métodos de estimação dos parâmetros da GEV

Segundo Mello e Silva (2013), para a aplicação de uma distribuição de probabilidades é indispensável analisar se a mesma representa adequadamente bem, a relação funcional entre os valores do evento e as respectivas frequências de ocorrências. Para isso, há necessidade de se comprovar previamente se a distribuição é adequada para a série histórica a ser trabalhada, analisando o comportamento das frequências teóricas em função das observadas.

A adequação dos modelos probabilísticos às séries anuais máximas diárias de precipitação foram verificadas pelos testes Kolmogorov – Smirnov (KS), Qui-quadrado (χ^2) e Anderson – Darling (AD), todos sob a hipótese H_0 de que as amostras dos dados sigam a distribuição de probabilidade testada em um nível de significância de 5%. Tais testes têm sido amplamente utilizados para verificar ajustes de distribuições de probabilidade no campo da hidrologia (Caldeira et al., 2015; Sansigolo, 2008; Ben-Zvi, 2009, Back et al., 2011, Franco et al., 2014,), especialmente os dois últimos uma vez que eles conferem mais importância ao comportamento da função da distribuição de probabilidade nas caudas de distribuição.

O KS é baseado na maior diferença, em valor absoluto, entre as probabilidades cumulativas teórica e empírica. O erro absoluto máximo entre as frequências teórica e observada é comparado a um valor crítico tabelado que depende do tamanho da amostra e do nível de relevância. Se o erro absoluto máximo for inferior ao valor crítico tabelado, isso significa que Hipótese H_0 é aceita (Back, 2001).

Apesar de ser bastante semelhante ao teste de KS, o teste de AD (Beskow et al. 2015) confere maior importância às caudas das distribuições, apresentando potencial na verificação do ajuste série assintótica. Neste teste, o valor calculado é multiplicado por um fator de correção, que é uma função da distribuição de probabilidade. Por fim, esse valor calculado é comparado a um valor estatisticamente nulo, que é uma função da distribuição de probabilidade e do nível de significância.

A estatística χ^2 é derivada da soma das diferenças quadráticas entre as frequências observada e teórica e seu valor é comparado a um valor tabulado, que varia em função dos graus de liberdade e do nível de significância. O χ^2 calculado pode ser considerado como um indicador da precisão do ajuste da função de probabilidade, pois corresponde ao erro quadrático médio. Se o χ^2 calculado for menor que o χ^2 tabulado, a hipótese H_0 será aceita (Franco et al., 2014).

Tempo de retorno

A atenção à análise de eventos ambientais extremos às vezes não se baseia somente nas estimativas de parâmetros de distribuição de valores extremos, e sim se aplica o modelo ajustado para calcular as outras grandezas. As estimativas do tempo de retorno desempenham um papel dinâmico na modelagem da precipitação e para calcular o risco futuro associado com períodos de

retorno em conformidade com um modelo ajustado (Ahmad et al., 2022).

De posse dos valores numéricos dos parâmetros e considerando-se a frequência de excedência, a estimativa da precipitação máxima associada a um tempo de retorno é dada por:

$$X_{TR} = \mu + \frac{\sigma}{k} * \left\{1 - \left[-LN\left(1 - \frac{1}{TR}\right)\right]^{k}\right\} \quad (10)$$

Em que XTR é o valor da variável hidrológica associada a um tempo de retorno em anos.

Baseado no exposto, também foi investigado neste estudo a influência do modelo teórico de probabilidades na precipitação diária máxima anual estimados eventos associados a intervalos de recorrência de 10, 25, 50, 100, 500 e 1000 anos, com intervalo de confiança no nível de 95%.

Resultados e discussão

Na etapa de análise preliminar de dados, todas as séries de dados das estações pluviométricas selecionadas foram submetidas a testes para avaliar a adequação a postulados básicos para a análise de frequência e as hipóteses de aleatoriedade, homogeneidade, independência e estacionariedade, sendo consideradas adequadas aquelas dentro de um nível de significância de 5%.

A Tabela 3 sintetiza os parâmetros obtidos para a função de densidade de probabilidade GEV pelos três diferentes métodos nas estações pluviométricas abordadas no estudo.

Tabela 3 Parâmetros da função de densidade de probabilidade GEV estimados pelos diferentes métodos.

Região Fisiográfica	Estação Pluviométrica	Método	Paramêtros		
			Posição (μ)	Escala (σ)	Forma (k)
Alto São Francisco	Lamounier	ML	75,66	20,82	0,078
		MV	74,86	18,98	0,178
		IB	75,66	20,82	0,078
	Taquaraçu	ML	72,92	16,49	0,074
		MV	73,04	16,18	0,077
		IB	72,85	16,92	0,092
Médio São Francisco	Porto Novo	ML	66,37	15,21	0,246
		MV	67,22	17,43	0,128
		IB	66,90	18,14	0,149
	Boqueirão	ML	67,23	17,92	0,122
		MV	67,89	19,35	0,051
		IB	67,50	20,01	0,070
Sub Médio São Francisco	Airi (Rochedo)	ML	51,14	18,21	0,046
		MV	51,17	17,80	0,057
		IB	50,98	18,48	0,083
	Flores	ML	67,81	18,72	-0,251
		MV	67,24	18,98	-0,200
		IB	66,79	19,41	-0,181
Baixo São Francisco	Piranhas	ML	43,30	19,94	0,041
		MV	43,41	19,50	0,046
		IB	42,92	20,24	0,071
	Delmiro Gouvéia	ML	48,54	18,85	0,047
		MV	48,59	18,39	0,057
		IB	48,18	19,05	0,084

Na figura 2 são apresentados os ajustes dos modelos probabilísticos à distribuição GEV para a precipitação máxima diária anual das estações pluviométricas da bacia hidrográfica do rio São Francisco.

Ao verificar o comportamento gráfico das FDP da GEV aos dados observados pode-se inferir que os métodos aplicados não mostraram aderência aos dados observados nas estações de Porto Novo, Médio São Francisco e Flores, no Sub Médio São Francisco. Observa-se também que, no caso das estações Taquaraçú (Alto São Francisco) e Piranhas (no Baixo São Francisco), os dados observados são pouco assimétricos, o que permitiu uma maior aderência da FDP da GEV.

A Tabela 4 apresenta os resultados dos testes de aderência para cada método trabalhado para distribuição GEV.

De acordo com Franco et al. (2014), há fragilidade desses testes para a análise de séries históricas assintóticas e devem ser evitados para reduzir erros do tipo I, ou seja, aceitar a hipótese H_o de que a distribuição é adequada sendo que a mesma pode não representar adequadamente a população de dados. Mesmo diante dessa limitação. Todos os métodos aplicados no estudo se mostraram adequados de acordo com os testes de Qui-quadrado e Kolmogorov-Smirnov.

A exemplo do apresentado nos resultados deste trabalho, o teste de KS mostrou-se menos rigoroso em diversos estudos (Franco et al., 2014; Beskow et al., 2015). Por se tratar de um teste qualitativo. Caldeira et al. (2015) sugere-se que não é possível utilizá-lo como parâmetro comparativo do ajuste entre diferentes distribuições de probabilidade. Apesar disso, devido à sua simplicidade, o teste de KS tem larga utilização na literatura.

O teste AD é interessante por analisar o comportamento dos ajustes atribuindo maior peso aos valores das caudas das distribuições, o que é extremamente relevante para dados com característica assintótica, tais como os de precipitação máxima diária anual (Franco et al., 2014).

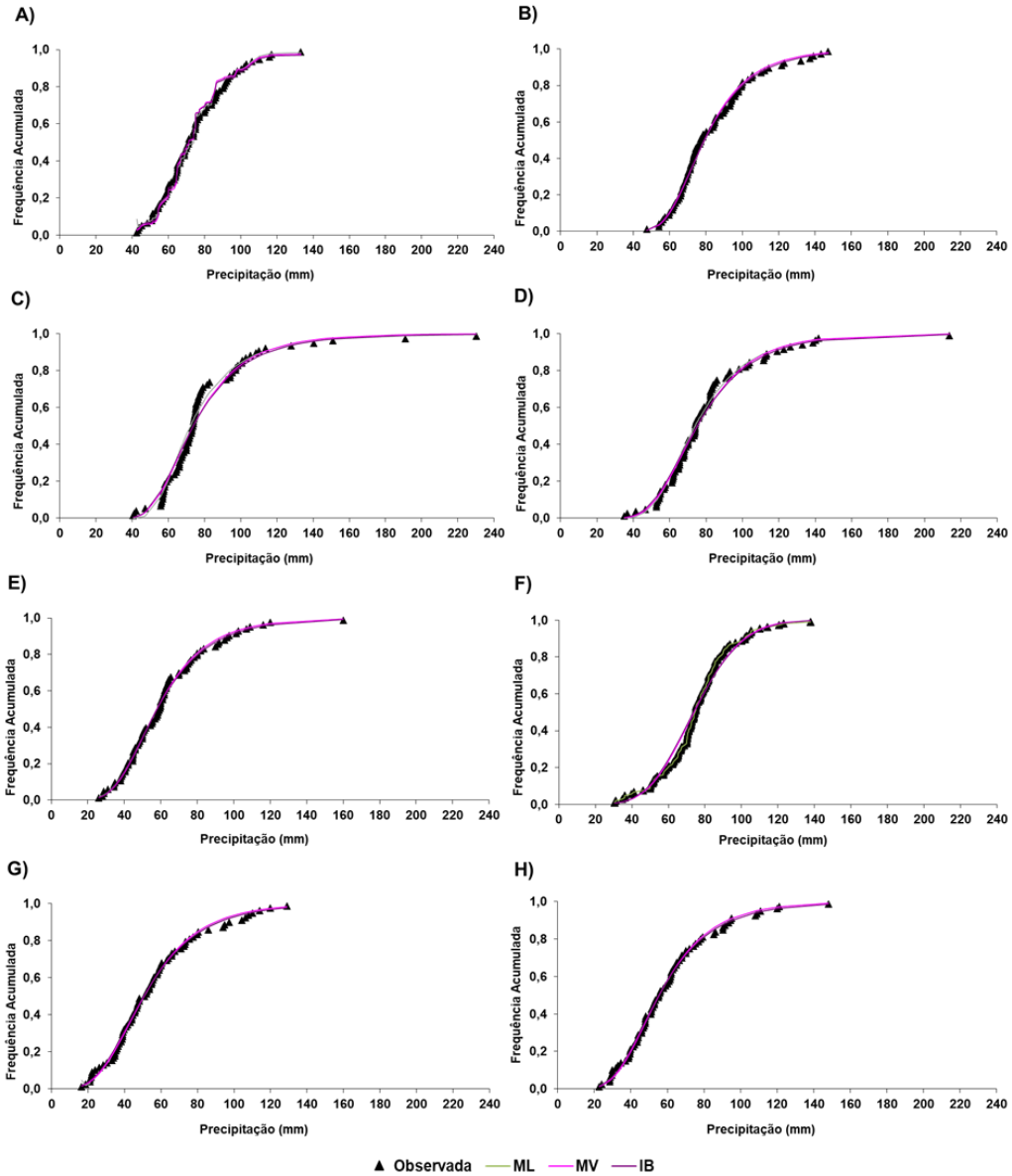

Figura 2. Ajuste dos modelos probabilísticos à precipitação máxima diária anual de estações pluviométricas da bacia hidrográfica do rio São Francisco, sendo: A) Lamounier; B) Taquaraçú; C) Porto Novo; D) Boqueirão; E) Airi; f) Flores; G) Piranhas e H) Delmiro Gouvêia.

Tabela 4. Adequabilidade das distribuições de probabilidades para as séries históricas de precipitação máxima diária anual, para a Bacia São Francisco, Minas Gerais.

Região Fisiográfica	Estação Pluviométrica	Método	Teste de Aderência		
			Anderson-Darling	Kolmorov-Smirnov	χ^2
Alto São Francisco	Lamounier	Tabelado	**0,757**	**0,154**	**11,070**
		ML	0,445	0,063	4,331
		MV	0,450	0,072	4,141
		IB	0,366	0,062	4,032
	Taquaraçu	Tabelado	**0,757**	**0,154**	**11,070**
		ML	0,231	0,067	3,490
		MV	0,241	0,056	3,707
		IB	0,220	0,053	3,263
Médio São Francisco	Porto Novo	Tabelado	**0,757**	**0,155**	**11,070**
		ML	1,246	0,095	5,828
		MV	1,088	0,115	4,398
		IB	1,218	0,117	5,016
	Boqueirão	Tabelado	**0,757**	**0,148**	**12,592**
		ML	0,515	0,061	5,551
		MV	0,550	0,080	7,280
		IB	0,648	0,083	8,325
Sub Médio São Francisco	Airi (Rochedo)	Tabelado	**0,757**	**0,148**	**12,592**
		ML	0,183	0,051	0,741
		MV	0,193	0,047	0,799
		IB	0,190	0,053	0,764
	Flores	Tabelado	**0,757**	**0,140**	**14,067**
		ML	0,835	0,072	7,764
		MV	0,955	0,092	5,532
		IB	1,117	0,090	9,687
Baixo São Francisco	Piranhas	Tabelado	**0,757**	**0,157**	**11,070**
		ML	0,270	0,043	4,722
		MV	0,250	0,046	5,076
		IB	0,230	0,042	4,326
	Delmiro Gouvéia	Tabelado	**0,757**	**0,151**	**11,070**
		ML	0,151	0,037	3,236
		MV	0,165	0,041	3,375
		IB	0,158	0,035	3,308

Observação: Os valores em vermelho indicam que o valor calculado foi superior ao tabelado, indicando que o método não foi adequado para representar a série.

Dentre os três testes de aderência empregados neste estudo, foi possível observar uma restrição maior gerada pelo teste de AD. Na Tabela 2, o valor tabelado para referido teste, sob a hipótese H_0 de que as amostras dos dados sigam a distribuição de probabilidade testada em um nível de significância de 5% para todas as estações foi de 0,757. Em duas FDP este valor foi superado sendo então rejeitando assim seus resultados.

Utilizando o teste de AD, verificou-se que, em duas estações, os ajustes aplicados às suas séries históricas foram considerados como inadequados em todos os métodos aplicados. Trata-se das estações de Porto Novo e Flores, que pela figura 2c e 2f, pode se observar dificuldade nos ajustes dos dados observados aos modelos teóricos.

A característica positiva deste teste é possiblidade de inferir quantitativamente o melhor método em comparação com os demais, isto se dá com o menor resultado do teste refletindo no melhor método empregado. Conforme exposto, utilizando como critério o teste AD, pode-se afirmar em três estações o método mais adequado foi IB (Estações de Lamounier, Taquaraçú e Piranhas) e nas outras três estações (Estações Boqueirão, Airi e Delmiro Gouvêia), o método ML foi o mais adequado. Conforme relatado em Diriba e Debusho (2020) e Ahmad et al. (2022), as estimativas de parâmetros por IB frequentemente supera ML, contudo é possível verificar a superioridade das mesmas sobre os demais métodos.

Com relação ao desempenho da IB, pode-se afirmar que o método se mostrou mais adequado para as séries em estações que possuíam a menor assimetria dos dados observados, comportando-se melhor em séries sem presença de observações extremas.

Nas séries com que com maior assimetria, o método que apresentou melhores resultados foi ML. Junqueira Júnior et al. (2015) testaram a aderência da distribuição a função GEV em estações na bacia hidrográfica do rio Doce, pelo teste de χ^2, para diferentes métodos de ajuste das distribuições (ML e MV). Os autores verificaram que o teste $\chi 2$ foi mais rigoroso para os ajustes das distribuições pelos métodos dos momentos-L. Beskow et al. (2015) utilizaram os mesmos testes utilizados neste estudo e verificaram que o teste de AD foi o mais rigoroso em relação aos testes de $\chi 2$ e KS. Esses resultados demonstram que, a exemplo do observado no presente trabalho, o teste AD tende a ser mais rigorosos que os testes de $\chi 2$ e KS.

Com relação às predições de eventos extremos para diferentes tempos de retorno, observou-se que a modelagem usando a IB apresentou precipitações com valores mais elevados, principalmente tratando de tempos de retorno acima de 100 anos, exceção feita a estação Boqueirão (Figura 3). Em relação ao intervalo de confiança, os resultados obtidos pela IB apresentaram incertezas maiores como na Estação Lamounier, cujos resultados variaram de 197 mm até 1.676 mm, com valor esperado de 506 mm. Os valores de precipitação máxima provável são essenciais para o dimensionamento de obras hídricas tais como pontes, barragens, redes de drenagem urbana entre outros, podendo superdimensionamento, uso de recursos financeiros mais ou sub-dimensionamento, que pode provocar enchentes, perda de vidas humanas e prejuízos econômicos.

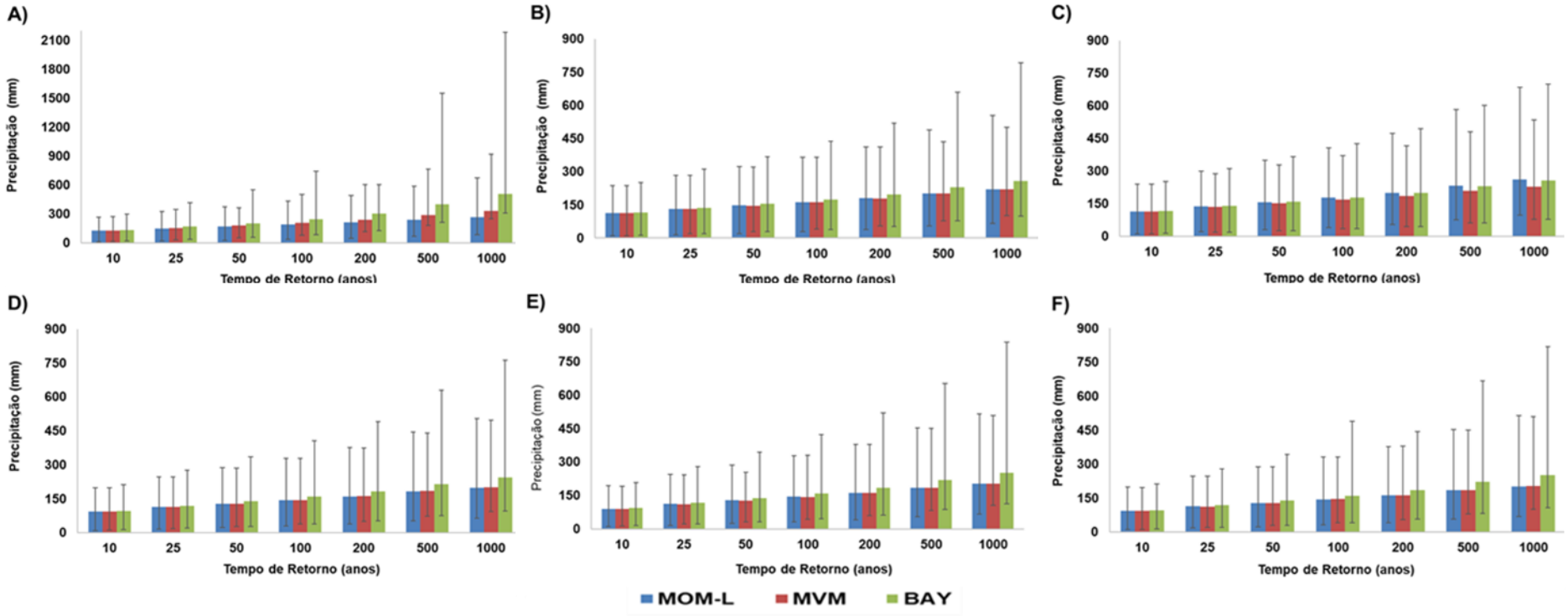

Figura 3 –Valores de precipitação máxima esperada (mm) nas Estações: A) Lamounier; B) Taquaraçú; C) Boqueirão; D) Airi; E) Piranhas e F) Delmiro Gouvêia. Barras de desvio representam os respectivos intervalos de confiança no nível de 95%.

Conclusões

1. Na comparação de métodos para estimativa dos parâmetros da função GEV, utilizando o teste de AD, verificou-se que os métodos ML e IB foram considerados os mais adequados para estudos de probabilidade de precipitação máxima diária anual.
2. O método IB teve melhor performance em estações com menor assimetria dos dados observados ao passo que o ML, obteve melhor resultados em séries de dados cujos dados possuíam valores mais assimétricos.
3. O método IB demostrou maiores incertezas para predição de chuvas com tempo de retorno acima de 100 anos.
4. O estudo mostra a importância do Projetista adotar diferentes métodos para estimativas dos parâmetros da GEV visando minimizar os riscos decorrentes de precipitações máximas em obras hidráulicas.
5. Observa-se, também que devidos a grande variabilidade espacial das precipitações na BSF, os métodos para estimativa estimativas dos parâmetros da GEV se mostraram inadequados para representar duas séries em estações fluviométricas que se encontram semiárido.

Agradecimentos

Os autores deste trabalho agradecem a Codevasf pelo apoio aos estudos e a Universidade de Brasília.

Referências

Ahmad, T., Ahmad, I., Arshad, I. A., Bianco, N. 2022. A comprehensive study on the Bayesian modelling of extreme rainfall: A case study from Pakistan. International Journal of Climatology 42, 208-224. https://doi.org/10.1002/joc.7240.

Ahmad, I., Ahmad, T., Almanjahie, I.M. 2019. Modeling of extreme rainfall in Punjab, Pakistan using Bayesian and frequentist approach. Applied Ecology and Environmental Research 17, 13729-13748. http://dx.doi.org/10.15666/aeer/1706_13729 13748.

Aragão, R. 2013. Chuvas intensas para o estado de Sergipe com base em dados desagregados de chuva diária. Revista Brasileira de Engenharia Agrícola e Ambiental 17, 243-252. https://doi.org/10.1590/S1415-43662013000300001.

Back, A.J., Hen, A., Oliveira, J.L.R. 2011. Heavy rainfall equations for Santa Catarina, Brazil. Revista Brasileira de Ciências do Solo 35, 2127-2134. https://doi.org/10.1590/S0100-06832011000600027.

Back, A.J. 2001. Seleção de distribuição de probabilidade para chuvas diárias extremas do estado de Santa Catarina. Revista Brasileira Meteorologia 16, 211-222, 2001.

Beijo, L. A., Vivanco, M. J. F., Muniz, J. A. 2009. Análise Bayesiana no estudo do tempo de retorno das precipitações máximas em Jaboticabal (SP). Ciência e Agrotecnologia 33, 261-270. https://doi.org/10.1590/S1413-70542009000100036.

Beskow, S., Caldeira T. L., Mello, C.R.; Faria, L.C., Guedes, H.A. 2015. Multiparameter probability distributions for heavy rainfall modeling in extreme southern Brazil, Journal of Hydrology: Regional Studies 4, 123-133. https://doi.org/10.1016/j.ejrh.2015.06.007.

Biswas, J., Bhattacharya, S. 2023. Investigation of Nonstationary Association of Monsoon Temperature and Precipitation Extremes through Past and Future over East-Central India. Pure Application Geophysical 180, 1143-1171. https://doi.org/10.1007/s00024-023-03242-w.

Bracken, C., Holman, K. D., Rajagopalan, B., Moradkhani, H. A. 2018. Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis. Water Resources Research 54, 243-255. https://doi.org/10.1002/2017WR020403.

Caldeira, T.L., Beskow, S., Mello, C.R., Faria, L.C., Souza, M.R., Guedes, H.A.S. 2015. Modelagem probabilística de eventos de precipitação extrema no estado do Rio Grande do Sul. Revista Brasileira de Engenharia Agrícola e Ambiental 19, 197-203. https://doi.org/10.1590/1807-1929/agriambi.v19n3p197-203.

Chaithong, T. 2022. Influence of changes in extreme daily rainfall distribution on the stability of residual soil slopes. Big Earth

Data 7, 101-125. https://doi.org/10.1080/20964471.2022.2046306.

Coles, S., Pericchi, L.R., Sisson, S. 2003. A fully probabilistic approach to extreme rainfall modeling. Journal of Hydrology 273, 35-50. https://doi.org/10.1016/S0022-1694(02)00353-0.

Coles, S.G., Powell, E.A. 1996. Bayesian Methods in Extreme Value Modelling: A Review and New Developments. International Statistical Review / Revue Internationale de Statistique 64, 119-136. https://doi.org/10.2307/1403426.

Comitê de Bacia Hidrográfica do Rio São Francisco. 2016. Plano de Recursos Hídricos da Bacia do Rio São Francisco 2016 – 2025. p. 520.

Comitê de Bacia Hidrográfica do Rio São Francisco. 2004. Plano Decenal de Recursos Hídricos da Bacia Hidrográfica do Rio São Francisco - PBHSF (2004-2013), Estudo Técnico de Apoio ao PBHSF – Nº 10 Controle de Cheias. 65 p.

Cotton, J. 2008. Bayesian Priors and Estimation of Extreme Rainfall in Camborne. http://empslocal.ex.ac.uk/people/staff/ferro/Projects/cotton. 32 p.

Diriba, T.A., Debusho, L.K. 2020. Modeling dependency effect to extreme value distributions with application to extreme Wind speed at Port Elizabeth, South Africa: a frequentist and Bayesian approaches. Computational Statistics 35, 243–267. https://doi.org/10.1007/s00180-019-00947-2.

Ferreira, A., De Haan, L. 2015. On the block maxima method in extreme value theory: PWM estimators. The Annals of Statistics 43, 276-298. https://doi:10.1214/14-AOS1280.

Franco, C.S. Marques, R.F.P.V., Oliveira, A.S. Oliveira, L.F.C. 2014. Distribuição de probabilidades para precipitação máxima diária na Bacia Hidrográfica do Rio Verde, Minas Gerais. Revista Brasileira de Engenharia Agrícola e Ambiental 18, 735-741. https://doi.org/10.1590/S1415-43662014000700010.

Garcia-Aristizabal, A., Bucchignani, E., Palazzi, E. 2015. Analysis of non-stationary climate-related extreme events considering climate change scenarios: an application for multi-hazard assessment in the Dar es Salaam region, Tanzania. Natural Hazards 75, 289-320. https://doi.org/10.1007/s11069-014-1324-z.

Institute of Hydrology (Grã-Bretanha). (1975). Flood studies report. Natural Environment Research Council.

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

Junqueira, J. A., Mello, C. R., Alves, Geovane J. 2015. Eventos extremos de precipitação no Alto Rio Grande, MG: Análise probabilística. Revista Brasileira de Engenharia Agrícola e Ambiental, v. 19, 301-308. https://doi.org/10.1590/1807-1929/agriambi.v19n4p301-308.

Lima, C. H. R., Kwon, H. H., Kim, Y. T. 2018. A local-regional scaling-invariant Bayesian GEV model for estimating rainfall IDF curves in a future climate. Journal of Hydrology, 566, 73-88. https://doi.org/10.1016/j.jhydrol.2018.08.075.

Kiefer, A., Costa, R., Petsch, C., Sccoti, A. 2021. Panorama das Alterações nos Padrões de Precipitação e Erosão diante de Mudanças Climáticas: Revisão de Literatura. Revista Brasileira de Geografia Física 14, 1724-1747. https://doi.org/10.26848/rbgf.v14.3.p1724-1747.

Mann, H.B., Whitney, D.R. 1947. On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other. Annals of Mathematical Statistics 18, 50-60. https://doi.org/10.1214/aoms/1177730491.

Mello, C.R., Silva, A. M. 2013. Hidrologia: princípios e aplicações em sistemas agrícolas. Editora UFLA, Lavras, 455 p.

Moccia, B., Papalexiou, S.M., Russo, F., Napolitano, F. 2021. Spatial variability of precipitation extremes over Italy using a fine-resolution gridded product. Journal Hydrology: Regional Studies 37, 100906. https://doi.org/10.1016/j.ejrh.2021.100906.

Naghettini, M., Pinto, E. J. A. 2007 Hidrologia Estatística. Editora CPRM, Belo Horizonte, 552 p.

Nerantzaki, S. D.; Papalexiou, S. M. 2022. Assessing extremes in hydroclimatology: A review on probabilistic methods, Journal of Hydrology 605, 127302. https://doi.org/10.1016/j.jhydrol.2021.127302.

Organização Mundial da Saúde (OMS). 2023. Disponível em: https://www.who.int/health-topics/floods#tab=tab_1

Queiroz, M. M. F., Chaudhry, F. H. 2006. Análise de eventos hidrológicos extremos, usando-se a distribuição GEV e momentos LH. Revista Brasileira de Engenharia Agrícola e Ambiental 10, 381-389. https://doi.org/10.1590/S1415-43662006000200020.

Sansigolo, C. A. 2008. Distribuições de extremos de precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, (1917-2006). Revista Brasileira de Meteorologia 23, 341-346. https://doi.org/10.1590/S0102-77862008000300009

Sigrist, F., Künsch, H.R., Stahel, W.A. 2015. Stochastic partial differential equation based modelling of large space–time data sets. Journal of the Royal Statistical Society: Series B. Statistical Methodology 77, 3-33. https://doi.org/10.48550/arXiv.1204.6118

Smith, E. 2005. Bayesian Modelling of Extreme Rainfall. University of Newcastle. 195 p.

Stephenson, A. G. 2002. evd: Extreme Value Distributions. R News 2, 31-32.

Yuan, W., Hu, X. 2020. Comparison of Heavy Rainfall Events Originated from Different Directions of Beijing City. Journal of Meteorological Research 34, 1299-1308. https://doi.org/10.1007/s13351-020-0051-5.

Wang, H., Xuan, Y. 2021. Spatial variation of extreme rainfall observed from two century-long datasets. Geophysical Research Letter 48, 1-12. https://doi.org/10.1029/2020gl091933.

Westra, S., Alexander, L. V., Zwiers, F. W. 2013. Global Increasing Trends in Annual Maximum Daily Precipitation. Journal of Climate 26, 3904-3918. https://doi.org./ 10.1175/JCLI-D-12-00502.1.