

**UNIVERSIDADE FEDERAL DE ALFENAS**

**MATHEUS DE SOUZA COSTA**

**AVALIAÇÃO DE DIFERENTES PRIORIS NA ESTIMAÇÃO DOS PARÂMETROS DA  
DISTRIBUIÇÃO GEV E NA PREDIÇÃO DE QUANTIS EXTREMOS**

Alfenas/MG

2021

**MATHEUS DE SOUZA COSTA**

**AVALIAÇÃO DE DIFERENTES PRIORIS NA ESTIMAÇÃO DOS PARÂMETROS DA  
DISTRIBUIÇÃO GEV E NA PREDIÇÃO DE QUANTIS EXTREMOS**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas, MG, como parte dos requisitos para a obtenção do título de Mestre. Linha de Pesquisa: Modelagem Estatística e Estatística Computacional

Orientador: Luiz Alberto Beijo

Coorientador: Reinaldo Antônio Gomes Marques

Alfenas/MG

2021

Sistema de Bibliotecas da Universidade Federal de Alfenas  
Biblioteca Central

Costa, Matheus de Souza.

Avaliação de diferentes prioris na estimação dos parâmetros da distribuição GEV e na predição de quantis extremos / Matheus de Souza Costa. - Alfenas, MG, 2021.

56 f. : il. -

Orientador(a): Luiz Alberto Beijo.

Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, MG, 2021.

Bibliografia.

1. Teoria de valores extremos. 2. Estrutura a priori informativa. 3. Nível de retorno. 4. Acurácia. 5. Erro médio de predição. I. Beijo, Luiz Alberto, orient. II. Título.

**AValiação DE DIFERENTES PRIORIS NA ESTIMAÇÃO DOS PARÂMETROS DA DISTRIBUIÇÃO GEV E NA PREDIÇÃO DE QUANTIS EXTREMOS**

A Banca examinadora abaixo-assinada aprova a Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria.

Aprovada em: 15 de dezembro de 2021.

Prof. Dr. Luiz Alberto Beijo  
Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Prof. Dr. Rodrigo dos Santos Targino  
Instituição: Fundação Getúlio Vargas - FGV-RJ

Prof. Dr. Gilberto Rodrigues Liska  
Instituição: Universidade Federal de São Carlos - UFSCar



Documento assinado eletronicamente por **Luiz Alberto Beijo, Professor do Magistério Superior**, em 15/12/2021, às 16:05, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Gilberto Rodrigues Liska, Usuário Externo**, em 15/12/2021, às 16:15, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rodrigo dos Santos Targino, Usuário Externo**, em 15/12/2021, às 17:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://sei.unifal-mg.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0651452** e o código CRC **A9766B46**.

## **AGRADECIMENTOS**

Agradeço a Deus, primeiramente, que me deu força para concluir esta etapa de minha vida.

Aos meus familiares, pelo amor, incentivo, força e apoio incondicional. Em especial, minha mãe Marília, meu pai Aloísio, minha irmã Marcela e minha noiva Geilma.

Aos meus orientadores Prof. Dr. Luiz Alberto Beijo e Prof. Dr. Reinaldo Antônio Gomes Marques, por suas orientações neste trabalho, pela paciência e todos os ensinamentos.

A todos os amigos que direta ou indiretamente participaram da minha formação, o meu eterno agradecimento.

À instituição e ao corpo docente do Programa de Pós-graduação em Estatística Aplicada e Biometria da UNIFAL-MG por sua contribuição na minha formação.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## RESUMO

A abordagem bayesiana tem sido muito aplicada em estudos de valores extremos devido a possibilidade da redução da incerteza nas estimativas. Contudo, a maneira de se elicitar uma estrutura *a priori* informativa, quando se trata de valores extremos ajustados por meio da distribuição generalizada de valores extremos (GEV), ainda não está completamente estabelecida. Deste modo, objetivou-se, neste trabalho, avaliar diferentes estruturas *a priori* em uma abordagem Bayesiana da distribuição GEV ajustada à dados de máximos provenientes de simulação e de dados reais de precipitação máxima. As estruturas *a priori* utilizadas foram: distribuições Normais Trivariadas com matrizes de variância e covariância multiplicadas por 1, 4, distribuição Normal Trivariada não-informativa; distribuição Gama para diferenças de quantis extremos, com diferentes variabilidades. Para o caso com dados simulados, avaliou-se a estimação dos parâmetros e de quantis com o modelo GEV associado a cada uma dessas estruturas *a priori*, com respeito aos respectivos valores reais, através de viés e viés médio relativo percentual em diferentes cenários de simulação e tamanhos amostrais. E para os dados reais, objetivou-se avaliar, através de acurácia e precisão das estimativas dos níveis de retorno, o desempenho das referidas estruturas *a priori* em uma abordagem Bayesiana para ajustar a distribuição GEV a dados de precipitações diárias máximas anuais de São João da Boa Vista-SP, no período de 1971 a 2017. Utilizou-se como *priori* dados de precipitação máxima anual de Lavras-MG e Silvianópolis-MG, comparando-se o critério de informação deviance (DIC), acurácia, erro médio de predição e amplitude média intervalar das predições de precipitações máximas dos modelos. Quanto aos principais resultados do caso simulado, observou-se que distribuições *a priori* com menor variabilidade, produzem estimativas mais precisas de todos os parâmetros da distribuição GEV, considerando pequenos tamanhos amostrais. Nos cenários em que se simulou dados de precipitação o modelo com *priori* Gama apresentou menor viés de parâmetros e níveis de retorno. Para os dados reais, pelo critério DIC, não houve diferenças substanciais entre os modelos. O modelo com estrutura *a priori* normal trivariada com matriz de variâncias e covariâncias multiplicada por 4, com informações *a priori* de Silvianópolis, forneceu estimativas de níveis de retorno mais precisas e acuradas.

Palavras-chave: Teoria de valores extremos; estrutura *a priori* informativa; nível de retorno; acurácia; erro médio de predição.

## ABSTRACT

The Bayesian approach has been widely applied in studies of extreme values due to the possibility of reducing uncertainty in the estimates. However, the way to elicit an informative a priori structure, when dealing with extreme values adjusted through the generalized distribution of extreme values (GEV), is not yet fully established. Thus, the objective of this work was to evaluate different prior structures in a Bayesian approach of the GEV distribution fitted to maximum data from simulation and real data of maximum precipitation. The prior structures used were: Trivariate Normal distributions with variance and covariance matrices multiplied by 1, 4, non-informative Trivariate Normal distribution; Gamma distribution for extreme quantile differences, with different variabilities. For the case with simulated data, the estimation of parameters and quantiles was evaluated with the GEV model associated with each of these prior structures, with respect to the respective real values, through bias and percentage relative mean bias in different simulation scenarios and sample sizes. And for the real data, the objective was to evaluate, through the accuracy and precision of the estimates of the levels of return, the performance of the referred prior structures in a Bayesian approach to the GEV distribution fitted to data of maximum daily precipitations of São João da Boa Vista-SP, from 1971 to 2017. Data of maximum annual rainfall from Lavras-MG and Silvianópolis-MG were used as prior, comparing the deviance information criterion (DIC), accuracy, mean prediction error and interval mean amplitude of the models' maximum precipitation predictions. As for the main results of the simulated case, it was observed that prior structures with less variability produce more accurate estimates of all parameters of the GEV distribution, considering small sample sizes. In the scenarios in which precipitation data were simulated, the model with Gamma prior structure showed lower parameter bias and return levels. For the real data, by the DIC criterion, there were no substantial differences between the models. The model with a trivariate normal prior structure with a matrix of variances and covariances multiplied by 4, with prior information from Silvianópolis, provided more precise and accurate estimates of return levels.

Keywords: Extreme value theory; informative prior structure; return level; accuracy; mean prediction error.

## SUMÁRIO

1	<b>INTRODUÇÃO</b>	8
2	<b>REFERENCIAL TEÓRICO</b>	10
2.1	TEORIA DE VALORES EXTREMOS	10
2.1.1	<b>Distribuição GEV e suas Propriedades</b>	11
2.2	ABORDAGEM BAYESIANA	13
2.2.1	<b>Algoritmo Metropolis-Hastings</b>	15
2.2.1.1	Critérios de convergência	16
2.3	ESTRUTURAS <i>A PRIORI</i>	17
2.3.1	<b>Estrutura Gama para diferenças de quantil</b>	18
2.3.2	<b>Estrutura Normal Trivariada</b>	19
3	<b>MATERIAL E MÉTODOS</b>	20
3.1	DADOS	20
3.1.1	<b>Dados simulados</b>	20
3.1.2	<b>Precipitação pluvial máxima anual de São João da Boa Vista-SP</b>	21
3.1.3	<b>Informações a priori: precipitação pluvial máxima anual de Lavras-MG e Silvianópolis</b>	21
3.1.4	<b>Análise de independência e estacionariedade</b>	22
3.2	ESTRUTURAS <i>A PRIORI</i>	22
3.2.1	<b>Casos simulados</b>	22
3.2.2	<b>Estudo de precipitação</b>	24
3.2.2.1	Priori Fundamentada nas informações de Lavras-MG	24
3.2.2.2	Priori Fundamentada nas informações de Silvianópolis	26
3.2.3	<b>Função de Verossimilhança da Distribuição GEV</b>	27
3.2.4	<b>Posteriori</b>	28
3.3	MÉTRICAS DE ANÁLISE DAS ESTIMATIVAS	29
3.3.1	<b>Para dados simulados</b>	29
3.3.2	<b>Para dados de precipitação</b>	29
4	<b>RESULTADOS E DISCUSSÃO</b>	31
4.1	ANÁLISE DOS CENÁRIOS SIMULADOS	31
4.2	ESTUDO DE PRECIPITAÇÕES MÁXIMAS DE SÃO JOÃO DA BOA VISTA	43
5	CONCLUSÕES	50
	<b>REFERÊNCIAS</b>	51
	<b>APÊNDICES</b>	56



## 1 INTRODUÇÃO

A modelagem de variáveis associadas a eventos extremos, tais como precipitações máximas, inundações, seca, geada, ondas de calor, é um tema de grande interesse atualmente. Eventos como esses podem ter consequências catastróficas para a humanidade, impactando a saúde e causando danos em construções e propriedades. Na agropecuária, por exemplo, a ocorrência de um grande volume de precipitação em um curto espaço de tempo pode gerar prejuízos, por conta de erosão, enchente, alagamentos, e além de danos físicos diretos nas plantas pode provocar até a perda vidas devido a deslizamentos e soterramentos (BEIJO; VIVANCO; MUNIZ, 2009; TUCCI; SILVEIRA, 2004).

Gomes e Guillou (2015) relatam que o desenvolvimento de técnicas para a estimativa e previsão de eventos raros tem contribuído para salvar recursos naturais ameaçados de extinção, com base na modelagem do clima e de outros fenômenos ambientais, como precipitação, temperatura e inundações. Entre essas metodologias, a utilização de Inferência Bayesiana associada a distribuições de probabilidade como a distribuição generalizada de valores extremos (GEV, do inglês *Generalized extreme value distribution*) apresenta vantagens, como a redução das incertezas da estimação dos parâmetros e de quantis destas distribuições (BEIJO; VIVANCO; MUNIZ, 2009).

Conforme apontado por Coles e Powell (1996), a escassez de dados extremos, principalmente os de variáveis hidrológicas, é uma das razões pelas quais se defende a utilização de Inferência Bayesiana na análise, pois esta permite a incorporação de informações *a priori*. Contudo, a maneira de se elicitar uma distribuição *a priori* informativa, quando se trata de valores extremos, ainda não está completamente estabelecida (COLES; TAWN, 1996).

Uma das estratégias é a assumir uma estrutura Normal Trivariada para os parâmetros da GEV. Uma abordagem baseada nessa construção de priori foi proposta por Coles e Powell (1996), com a utilização de uma distribuição *a priori* Normal Trivariada para os parâmetros de posição, escala e forma da distribuição GEV. Stephenson e Ribatet (2006) ressaltam que, embora essa estrutura seja muito flexível, ela requer a determinação de muitos hiperparâmetros.

Outros métodos de elaboração de estruturas *a priori* para distribuições de extremos baseiam-se no espaço do quantil, para probabilidades fixas, e no espaço de probabilidade, para quantis fixos. Nesses casos considera-se distribuições para os quantis extremos, ou para a diferença ou razão entre eles. Distribuições *a priori* considerando o espaço do quantil ou da

probabilidade requerem a determinação de uma quantidade menor de hiperparâmetros, sendo, portanto, mais fáceis de elicitar (STEPHENSON; RIBATET, 2006). Alguns estudos, considerando essas construções de priori, foram realizados por Crowder (1992), Coles e Tawn (1996) e Beijo, Vivanco e Muniz (2009).

Além disso, segundo o paradigma Bayesiano, todo problema científico é singular, e o conhecimento que se tem sobre um certo parâmetro de um modelo varia de um problema para outro e entre investigadores diferentes (PAULINO; TURKMAN; MURTEIRA, 2003). Por conta disso, diferentes abordagens sobre elicitções de prioris em uma modelagem Bayesiana devem ser consideradas. Paulino, Turkman e Murteira (2003) afirmam que isso não compromete a coerência da análise.

Sendo assim, objetiva-se com este trabalho avaliar diferentes estruturas *a priori* em uma abordagem Bayesiana da distribuição GEV ajustada à dados de máximos provenientes de simulação e uma aplicação a dados reais de precipitação máxima. Para o caso simulado, serão analisados vies dos parâmetros em diferentes cenários de simulação e tamanhos amostrais. E, na aplicação a dados reais, serão examinados a acurácia e precisão das estimativas de precipitação máxima anual de São João da Boa Vista através do ajuste da distribuição GEV com diferentes estruturas *a priori*.

## 2 REFERENCIAL TEÓRICO

Nesta seção são apresentadas as fundamentações da teoria de valores extremos, da distribuição generalizada de valores extremos, e da abordagem Bayesiana. São descritas também as estruturas *a priori* que serão utilizadas no trabalho.

### 2.1 TEORIA DE VALORES EXTREMOS

A Teoria de Valores Extremos (TVE) é uma teoria importante em Estatística e Probabilidade que possibilita a análise do comportamento de valores extremos (DEY; YAN, 2016). Embora esses valores tenham uma probabilidade baixa de ocorrer, eles podem ter um impacto no ambiente em que acontecem, por conta dos eventos extremos relacionados a eles. A partir da TVE pode-se extrapolar dados amostrais, encontrando estimadores para quantidades relevantes, como quantis extremos (COLES, 2001). Algumas áreas como engenharia, economia e meteorologia estão entre as que mais aplicam a TVE em suas modelagens.

Conforme Haan e Ferreira (2006), a TVE e o teorema do limite central possuem certa semelhança: o teorema central do limite tem interesse no comportamento assintótico das somas parciais de variáveis  $\{Y_i\}$  independentes e identicamente distribuídas, com  $i = 1, \dots, n$  e  $n \rightarrow \infty$ , já a TVE preocupa-se com distribuições limite para máximo (ou mínimo) de  $\{Y_i\}$ . Neste trabalho, porém, o interesse é em máximos, assim: seja  $F_n(y)$  a função de distribuição de  $M_n = \max(Y_1, \dots, Y_n)$ , com  $\{Y_n\}$ ,  $n \in N$  sendo uma sequência de variáveis i.i.d. e  $n \rightarrow \infty$ , logo:

$$F_n(y) = P(M_n \leq y) = F^n(y),$$

que converge para zero 0 ou para 1. Assim, a fim de obter um distribuição limite não degenerada, é necessária uma normalização (HAAN; FERREIRA, 2006).

Suponha que exista uma sequência de constantes  $a_n > 0$  e  $b_n$  real ( $n = 1, 2, \dots$ ), de tal modo que a partir do Teorema de Tipos Extremais (COLES, 2001) tem-se:

$$\frac{M_n - b_n}{a_n} \rightarrow G(y),$$

quando  $n \rightarrow \infty$ , sendo  $G$  uma função de distribuição não-degenerada, a qual pode corresponder a três famílias de distribuições, as quais são chamadas de distribuição de valores extremos:

$$\text{Gumbel: } F(y) = \exp \left\{ - \exp \left[ - \left( \frac{y-b}{a} \right) \right] \right\}, \quad -\infty < y < \infty \quad (2.1)$$

$$\text{Frechét: } F(y) = \exp \left\{ - \left( \frac{y-b}{a} \right)^{-\alpha} \right\} \quad y > b \quad (2.2)$$

$$\text{Weibull: } F(y) = \exp \left\{ - \left[ - \left( \frac{y-b}{a} \right)^\alpha \right] \right\}, \quad y \leq 0, \quad y < b, \quad (2.3)$$

em que  $a > 0$  e  $\alpha > 0$ .

Embora seja comum utilizar uma dessas distribuições na análise de extremos, Coles (2001) ressalta dois pontos a se considerar:

1. necessita-se uma técnica para escolher qual das três famílias é mais apropriada aos dados que são objeto de estudo;
2. assumindo que uma técnica foi escolhida, as inferências posteriores pressupõem que esta escolha seja correta e não abrangem a incerteza envolvida na seleção, mesmo que seja uma incerteza substancial.

Ainda conforme Coles (2001), uma reformulação do Teorema de Fisher-Tippet pode fornecer uma alternativa, a qual pode ser vista na próxima subsubseção.

### 2.1.1 Distribuição GEV e suas Propriedades

Essa reformulação foi realizada por Jenkinson (1955), pela qual unificou as distribuições Gumbel, Frechét e Weibull em uma forma chamada de distribuição GEV. Pela parametrização de Jenkinson, a função de probabilidade acumulada da distribuição GEV é dada por:

$$F(y) = \exp \left\{ - \left[ 1 + \xi \left( \frac{y-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, \quad (2.4)$$

em que  $\xi$ ,  $\sigma$  e  $\mu$  denotam, respectivamente, os parâmetros de forma, escala e posição. Conforme [Naghetini e Pinto \(2007\)](#), o valor de  $\xi$  determina a forma assintótica de valores extremos: se  $\xi > 0$ , a GEV representa a distribuição Frechét (2.2), definida apenas para  $y > (\mu - \sigma/\xi)$ , se  $\xi < 0$ , a GEV representa a distribuição Weibull (2.3), definida apenas para  $y < (\mu - \sigma/\xi)$ , se  $\xi = 0$ , a GEV representa a distribuição Gumbel (2.1).

De acordo com [Bautista, Zocchi e Angelocci \(2004\)](#), derivando-se (2.4) em relação a  $y$  obtém-se a função densidade da GEV, dada por:

$$f(y) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{y - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp \left\{ - \left[ 1 + \xi \left( \frac{y - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}. \quad (2.5)$$

A distribuição GEV tem sido empregada em diversos estudos envolvendo variáveis como precipitação máxima, temperatura máxima, níveis máximos de poluição, ventos máximos, inflação máxima, entre outras. Pode-se citar trabalhos sobre modelagem:

- de precipitação máxima, como os de [Abreu et al. \(2018\)](#) e [Kumar, Swain e Vinay \(2019\)](#);
- de vazões máximas, como o trabalho de [WU et al., \(2018\)](#);
- de ventos máximos, como os estudos de [Liska et al. \(2013\)](#), [Kang, Ko e Huh \(2015\)](#) e [Diriba et al. \(2017\)](#);
- de níveis de poluição máxima, como os de [Ercelebi e Toros \(2009\)](#) e [Masseran e Safari \(2020\)](#).

Como a distribuição GEV (2.4) não é simétrica, o parâmetro de posição ( $\mu$ ) não é a média da distribuição GEV. Além disso, o parâmetro de escala ( $\sigma$ ) não é o desvio padrão.

Contudo, o foco na análise de máximos com a distribuição GEV, não é em seus parâmetros, mas sim em quantis de particular interesse, como, níveis de retorno, por exemplo. O nível retorno,  $y(t)$ , para a distribuição GEV, associado com um período de retorno  $t = \frac{1}{p_t}$ . Para obter este quantil basta igualar a função de distribuição da GEV, dada em (2.4), à  $1 - p_t$ , isto é,

$$F(y(t)) = 1 - p_t \quad (2.6)$$

e aplicar a função inversa  $F^{-1}$  na equação 2.6 e isolar  $y(t)$ , resultando em:

$$y(t) = q_p = \mu + \frac{\sigma}{\xi} [z_p^{-\xi} - 1] \quad (2.7)$$

em que  $z_p = -\ln(1 - p_t)$ .

O nível de retorno  $y(t)$  é, portanto, um quantil correspondente à probabilidade  $p_t$  da cauda superior da distribuição GEV.

## 2.2 ABORDAGEM BAYESIANA

A abordagem Bayesiana pode ser aplicada em diferentes situações; e conforme destacam [Feinberg e van der Linden \(2005\)](#), além de permitir a estimação de parâmetros de um modelo, contribui para maior flexibilidade: na avaliação do ajuste do modelo; na comparação de modelos; no tratamento de dados ausentes; e no “ajuste” de um modelo de maneiras que não podem ser feitas usando rotinas pré-existentes em *softwares* (por exemplo, liberando ou impondo restrições).

Especificamente, a análise Bayesiana pode ser uma maneira de realizar inferências sobre os parâmetros ( $\mu$ ,  $\sigma$  e  $\xi$ ) e sobre os quantis da distribuição GEV. Esta técnica tem sido utilizada com sucesso em estudos sobre valores extremos. Porém, antes de estabelecer as bases teóricas que permitem aplicar a metodologia Bayesiana à teoria de valores extremos, é importante ressaltar algumas considerações sobre este tipo de análise.

De acordo com [Gelman \*et al.\* \(2013\)](#), uma característica essencial dos métodos Bayesianos é o uso explícito de probabilidade para quantificar incerteza nas inferências baseadas na análise estatística dos dados. Conforme esses autores explicam, a análise Bayesiana pode ser idealizada considerando três etapas, a saber: estabelecimento de um modelo de probabilidade para todas as quantidades observáveis e não observáveis do problema, cálculo e interpretação da distribuição a posteriori e, finalmente, avaliação do ajuste e relevância do modelo.

A primeira etapa origina uma questão: como construir especificações de probabilidade apropriadas para quantidades não-observáveis, como os parâmetros da GEV? A segunda etapa muitas vezes exige grande esforço computacional dependendo da complexidade das integrais a serem computadas e a terceira requer um equilíbrio de técnica e análise associadas ao contexto do problema ([GELMAN \*et al.\*, 2013](#)).

[Turkman, Paulino e Müller \(2019\)](#) afirmam que o objetivo da inferência Bayesiana é representar a incerteza *a priori*, antes de observar os dados, sobre os parâmetros do modelo com

uma distribuição de probabilidade e atualizar essa incerteza *a priori* com os dados observados para produzir uma distribuição de probabilidade *a posteriori* para o(s) parâmetro(s) que contém menos incerteza. Do ponto de vista Bayesiano, qualquer quantidade cujo valor verdadeiro seja incerto, como os parâmetros de um modelo, pode ser representada através de distribuições de probabilidade.

Assumindo que se tem valores observados  $\mathbf{y}$  de uma variável  $Y$ , uma função  $f(\mathbf{y}|\boldsymbol{\theta})$  e uma distribuição  $\pi(\boldsymbol{\theta})$ , então, segundo [Turkman, Paulino e Müller \(2019\)](#), o Teorema de Bayes implica que:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (2.8)$$

em que  $\pi(\boldsymbol{\theta}|\mathbf{y})$  é a distribuição *a posteriori* de  $\boldsymbol{\theta}$  dada a observação de  $\mathbf{y}$ . A informação inicial é caracterizada pela distribuição *a priori*  $\pi(\boldsymbol{\theta})$ , a qual é modificada pelos dados observados sendo atualizado para  $\pi(\boldsymbol{\theta}|\mathbf{y})$ . O denominador em (2.8), denotado por  $f(\mathbf{y})$ , é a distribuição marginal de  $Y$ , qualquer que seja o valor de  $\boldsymbol{\theta}$ .

Assim, considerando uma amostra de  $Y$ , a expressão 2.8 pode ser escrita como

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto L(\boldsymbol{\theta}; \mathbf{y})\pi(\boldsymbol{\theta}) \quad (2.9)$$

em que,  $\propto$  representa proporcionalidade.  $L(\boldsymbol{\theta}; \mathbf{y})$  é a função de verossimilhança, a qual pode ser interpretada como uma quantificação da informação sobre  $\boldsymbol{\theta}$  que é provida pelos valores observados de  $Y$ .  $\pi(\boldsymbol{\theta})$  é também conhecida como *priori*, ou distribuição *a priori*. Os parâmetros de uma distribuição *a priori* são conhecidos como hiperparâmetros.

Quando  $\boldsymbol{\theta}$  é um vetor de parâmetros, como no caso na distribuição GEV ( $\boldsymbol{\theta} = \mu, \sigma, \xi$ ), pode ser desejável realizar inferências sobre um dos componentes de  $\boldsymbol{\theta}$ , por exemplo,  $\mu$ . Nesse caso, deve-se encontrar a marginal desse parâmetro a partir da distribuição *a posteriori* conjunta  $\pi(\boldsymbol{\theta}|\mathbf{y})$ ,

$$\pi(\mu|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{\boldsymbol{\theta} \neq \mu}.$$

A integração da distribuição conjunta *a posteriori* geralmente não pode ser obtida analiticamente, deve-se, portanto, utilizar técnicas iterativas como o algoritmo Metropolis-Hastings ([TURKMAN; PAULINO; MÜLLER, 2019](#)). Este algoritmo realiza uma simulação de Monte Carlo via Cadeias de Markov (MCMC), que gera amostras de parâmetros

do modelo a partir de sua distribuição conjunta a posteriori por amostragem em cadeia de Markov. A partir dessa técnica, há, portanto, um passeio no espaço paramétrico em estudo.

De acordo [Gelman et al. \(2013\)](#), sob uma abordagem Bayesiana, pode-se construir intervalos de credibilidade (ou regiões) para os parâmetros. Dado  $y$ , e uma vez determinada uma distribuição a *posteriori*, encontra-se um intervalo credível para um parâmetro de  $\theta$ . O intervalo é formado por dois valores em  $\theta$ , por exemplo,  $[\theta_A, \theta_B]$ , de modo que

$$P(\theta_A < \theta < \theta_B | y) = \int_{\theta_A}^{\theta_B} \pi(\theta | y) d\theta = 1 - \alpha \quad (2.10)$$

em que  $1 - \alpha$  é o nível de credibilidade desejado.

No entanto, o intervalo dado pela expressão 2.10 não é único. Assim, costuma-se utilizar um intervalo de alta densidade a *posteriori* HPD (do inglês *Highest Probability Density*), o qual possui os valores mais credíveis a posteriori de  $\theta$  com o nível  $1 - \alpha$  de credibilidade.

### 2.2.1 Algoritmo Metropolis-Hastings

Os métodos MCMC consistem em simular uma cadeia de Markov,  $\theta_1, \theta_2, \dots$ , cuja distribuição estacionária é  $\pi(\theta | y)$ . Um dos métodos MCMC é o algoritmo Metropolis-Hastings, o qual é uma generalização de Hastings (1970) do método proposto por Metropolis. [Morales e Causil \(2019\)](#) descrevem o algoritmo assim:

1. Inicia-se com um valor  $\theta_0$  e, digamos que o valor atual é  $\theta_t$ .
2. Simular um valor candidato  $\hat{\theta}$ , de uma densidade proposta  $q(\hat{\theta} | \theta_t)$ .
3. Calcular a probabilidade de aceitar o valor gerado:

$$\alpha(\theta, \hat{\theta}) = \min \left\{ 1, \frac{\pi(\hat{\theta} | y) q(\theta_t | \hat{\theta})}{\pi(\theta_t | y) q(\hat{\theta} | \theta_t)} \right\}.$$

4. Simular  $u$  de uma distribuição uniforme  $U[0,1]$ .
5. Se  $u < \alpha$ , aceite  $\theta_{t+1} = \hat{\theta}$ . Em outro caso, rejeite  $\hat{\theta}$  e tome  $\theta_{t+1} = \theta_t$  e volte ao passo 2.

A densidade à qual objetiva-se obter ( $\pi(\theta | y)$ ) é utilizada no processo através do quociente



$\pi(\hat{\theta}|\mathbf{y})/\pi(\theta_t|\mathbf{y})$ , por conseguinte, não é necessário conhecer a constante de normalização para implementar o algoritmo.

As densidades  $q(\hat{\theta}|\theta_t)$  propostas para cada parâmetro da distribuição GEV, no algoritmo implementado no pacote *evdbayes* do *software R*, segundo [Stephenson e Ribatet \(2006\)](#) são:

- $\hat{\mu} \sim N(\mu_t, s_\mu^2)$ .
- $\hat{\sigma} \sim LN(\ln \sigma_t, s_\sigma^2)$ .
- $\hat{\xi} \sim N(\xi_t, s_\xi^2)$ .

Segundo [Gamerman e Lopes \(2006\)](#), se o algoritmo Metropolis-Hastings é implementado de maneira correta, a convergência da cadeia a posteriori é alcançada, independentemente dos valores iniciais estabelecidos. Precisa-se de um número elevado de iterações para aproximar-se à distribuição estacionária. Também é necessário estabelecer um número de iterações a serem descartadas, para eliminar aquelas que não estão em estado estacionário.

### 2.2.1.1 Critérios de convergência

Os métodos MCMC são simulações iterativas, portanto os valores consecutivos estão altamente correlacionados ([MORALES; CAUSIL, 2019](#)). Desta forma, necessita-se métodos (critérios) para monitorar a convergência das cadeias *a posteriori*, com o fim de determinar se a cadeia alcançou a distribuição limite ou de equilíbrio; dado que utilizar cadeias que não atinjam a convergência, pode levar a realizar inferências equivocadas sobre os parâmetros ou hipóteses de interesse no estudo. [Nogueira, Sáfadi e Ferreira \(2004\)](#), ressaltam que é importante verificar a convergência das cadeias para evitar que iterações, aquém e além das necessárias, sejam executadas.

De acordo com [Gamerman \(1997\)](#) existem métodos de convergência formais e informais, este último consiste na análise gráfica da trajetória da cadeia ao longo das iterações, a qual permite observar se a cadeia apresenta algum indício de não estacionariedade. Quanto aos métodos formais, destaca-se os métodos propostos por [Raftery e Lewis \(1992\)](#), [Geweke \(1991\)](#) e [Heidelberger e Welch \(1983\)](#).

Segundo [Morales e Causil \(2019\)](#), o critério de Raftery e Lewis, realiza um diagnóstico do número de iterações, baseando-se no critério da exatidão da estimação do quantil  $q$ , com certa

probabilidade  $p$ . Este critério também proporciona o fator de dependência (FD) das iterações e, sugere-se concluir que as iterações são independentes quando o FD é menor do que 5.

O critério de Geweke, consiste em realizar um teste de igualdade de médias através da estatística  $Z$ , que segue uma distribuição normal padrão. Assim, realiza-se o teste utilizando uma amostra com 10% dos valores iniciais da cadeia e uma outra amostra com 50% dos valores finais da cadeia. Se a estatística  $|Z| > 1,96$ , considera-se que as médias são diferentes. Este procedimento também permite observar qual deve ser o número de iterações iniciais a serem descartadas (*burn-in*).

Por último, o critério de Heidelberg e Welch, faz uso da estatística Cramer-Von Mises para verificar a estacionariedade da cadeia. O teste consiste em descartar 10% das primeiras iterações para verificar a estacionariedade. Se esta não é alcançada, descartam-se mais cadeias até descartar 50% destas. Conclui-se que não há indícios de não estacionariedade da cadeia quando o valor- $p$  é maior do que um nível de significância estabelecido pelo pesquisador, geralmente de 5%.

### 2.3 ESTRUTURAS A PRIORI

Kotz e Nadarajah (2000) explicam que o uso prático da abordagem Bayesiana em valores extremos está frequentemente associado a dificuldades para escolher informações a *priori* e distribuição a *priori* para os parâmetros. Apesar disso, muitos estudos com estes temas tem sido desenvolvidos como os de Coles e Powell (1996), Coles e Tawn (1996), Coles e Tawn (2005), Beijo, Vivanco e Muniz (2009), Diriba *et al.* (2017), entre outros. As distribuições a priori propostas nestes estudos baseiam-se em informações a *priori* sobre quantis extremos, diferenças de quantis ou mesmo informações sobre os próprios parâmetros

Algumas das estruturas a *priori* mais utilizadas para distribuição GEV e que serão utilizadas neste trabalho são descritas nas subsubseções a seguir.

### 2.3.1 Estrutura Gama para diferenças de quantil

Esta abordagem foi proposta por [Coles e Tawn \(1996\)](#), que sugerem a construção da priori baseada em quantis com probabilidade fixa e pequena.

Considerando que o quantil da distribuição GEV ( $q_p$ ) é dado pela expressão (2.7), a distribuição *a priori* pode ser construída em termos dos quantis ( $q_{p_1}, q_{p_2}, q_{p_3}$ ) para probabilidades especificadas  $p_1 > p_2 > p_3$ . Uma vez que  $q_{p_1} < q_{p_2} < q_{p_3}$ , pode-se trabalhar com as diferenças ( $\tilde{q}_{p_1}, \tilde{q}_{p_2}, \tilde{q}_{p_3}$ ), de modo que,  $\tilde{q}_{p_i} = q_{p_i} - q_{p_{i-1}}$  para  $i = 1, 2, 3$ , em que  $q_{p_0}$  é o ponto final inferior físico da variável do processo. Para precipitação máxima, considera-se  $q_{p_0} = 0$ . Assume-se que as distribuições *a priori* para diferenças de quantis são consideradas independentes, com

$$\tilde{q}_{p_i} \sim \text{Gama}(\alpha_i; \beta_i), \quad (2.11)$$

em que  $\alpha_i, \beta_i > 0$ , para  $i = 1, 2, 3$ . As diferenças ( $\tilde{q}_{p_2}, \tilde{q}_{p_3}$ ) dependem apenas dos parâmetros de escala e forma ( $\sigma, \xi$ ). A informação *a priori* sobre o parâmetro de posição  $\mu$  se dá apenas através de  $\tilde{q}_{p_1}$ . Os hiperparâmetros  $(\alpha_1, \alpha_2, \alpha_3)$  e  $(\beta_1, \beta_2, \beta_3)$ , e as probabilidades  $p_1 > p_2 > p_3$ , devem ser especificadas.

Esta construção leva a função de densidade *a priori*

$$\pi(\boldsymbol{\theta}) \propto \mathbf{J} \prod_{i=1}^3 \tilde{q}_{p_i}^{\alpha_i-1} \exp\{-\tilde{q}_{p_i}/\beta_i\}, \quad (2.12)$$

desde que  $q_{p_1} < q_{p_2} < q_{p_3}$ .  $\mathbf{J}$  é o Jacobiano da transformação de  $(q_{p_1}, q_{p_2}, q_{p_3})$  para  $\boldsymbol{\theta} = (\mu, \sigma, \xi)$ , dado por

$$\mathbf{J} = \sigma/\xi^2 \left| \sum_{i,j \in \{1,2,3\}} (-1)^{i+j} (x_i x_j)^{-\xi} \ln(x_j/x_i) \right|, \quad \text{com } i < j,$$

em que  $x_i = -\ln(1 - p_i)$  para  $i = 1, 2, 3$ .

Quando  $\xi = 0$ , a distribuição *a priori* é definida pela continuidade, usando

$$\lim_{\xi \rightarrow 0} q_{p_i} = \mu - \sigma \ln x_i, \quad i = 1, 2, 3$$

e

$$\lim_{\xi \rightarrow 0} \mathbf{J} = \sigma/2 \left| \sum_{i,j \in \{1,2,3\}} (-1)^{i+j} \ln x_i \ln x_j \ln(x_j/x_i) \right|, \quad \text{com } i < j.$$

### 2.3.2 Estrutura Normal Trivariada

Seguindo a sugestão de [Coles e Powell \(1996\)](#), assumindo que a distribuição Normal Trivariada para o vetor de parâmetros da GEV,  $\boldsymbol{\theta}' = (\mu, \ln(\sigma), \xi)$ , é dada por:

$$\pi(\boldsymbol{\theta}') \propto |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta}' - \boldsymbol{\Phi}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\theta}' - \boldsymbol{\Phi}_0) \right\}, \quad (2.13)$$

$\boldsymbol{\Phi}_0$  é o vetor de médias dos hiperparâmetros,  $\boldsymbol{\Phi}_0 = (\mu_0, \ln(\sigma_0), \xi_0)$  e  $\boldsymbol{\Sigma}_0$  é a matriz de variâncias e covariâncias, que tem a seguinte forma:

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \text{Var}(\mu_0) & \text{Cov}(\mu_0, \sigma_0) & \text{Cov}(\mu_0, \xi_0) \\ \text{Cov}(\mu_0, \sigma_0) & \text{Var}(\sigma_0) & \text{Cov}(\sigma_0, \xi_0) \\ \text{Cov}(\mu_0, \xi_0) & \text{Cov}(\sigma_0, \xi_0) & \text{Var}(\xi_0) \end{bmatrix}$$

Mais detalhes sobre essa estrutura *a priori* podem ser encontrados em [Coles e Powell \(1996\)](#).

### 3 MATERIAL E MÉTODOS

#### 3.1 DADOS

##### 3.1.1 Dados simulados

Com o intuito de se ter cenários de simulação mais próximos da realidade, utilizou-se estimativas dos parâmetros posição, escala e forma da distribuição GEV, ajustada em diferentes casos, para gerar séries de máximos, por meio de um procedimento de geração de números pseudo-aleatórios. Os valores-base dos parâmetros  $\mu$ ,  $\sigma$  e  $\xi$ , fundamentos em diferentes estudos, para realizar a simulação, são descritos na Tabela 1.

Tabela 1 – Valores-base dos parâmetros da GEV obtidos com informação de estudos de extremos compondo cada cenário-base (CB).

CB	Parâmetros da dist. GEV			Variável do estudo	Referência
	$\mu$	$\sigma$	$\xi$		
a	70	30	0,4	Precipitação máxima (mm)	<a href="#">Diriba et al. (2014)</a>
b	70	25	0,05	Precipitação máxima (mm)	<a href="#">Martins et al. (2018)</a>
c	34	2	-0,4	Temperatura máxima (° C)	<a href="#">Aguirre, Nogueira e Beijo (2020)</a>
d	6,78	0,561	-0,2	Intensidade máxima de terremotos	<a href="#">Maruyama et al. (2020)</a>
e	1200	340	0,1	Vazão Máxima ( $m^3s^{-1}$ )	<a href="#">Kousar et al. (2020)</a>

Fonte: Do autor.

Cada um dos cenários de simulação foi constituído combinando-se os valores-base dos parâmetros da distribuição GEV (Tabela 1) e um dos três de tamanhos amostrais de 30, 60 e 120, totalizado-se 15 cenários distintos. Replicou-se 1000 vezes cada um desses cenários. Para realizar a simulação desses máximos utilizou-se a função *rgev()* que está implementada no pacote *evd* ([STEPHENSON, 2002](#)) do *software* R ([R CORE TEAM, 2018](#)). Esta função possui os seguintes argumentos *n*, *loc*, *scale*, *shape*, que correspondem a tamanho da amostra, parâmetro de posição, parâmetro de escala e parâmetro de forma, respectivamente.

A rotina para realização da simulação descrita nessa subsecção está no apêndice A.

### 3.1.2 Precipitação pluvial máxima anual de São João da Boa Vista-SP

Para o estudo de caso, utilizou-se dados a partir dos registros pluviográficos, que correspondem a precipitações pluviais diárias (mm) de 1971 a 2020 do município de São João da Vista - SP, fornecidos pelo Instituto Nacional de Meteorologia (INMET). As coordenadas geográficas do município são:  $21^{\circ}58'8''$  ao Sul (latitude) e  $46^{\circ}47'52''$  à Oeste (longitude) (IBGE, 2011).

O tipo de clima dominante em São João da Boa Vista - SP é o Cwb, que, pela classificação de Köppen, é caracterizado pelo clima subtropical de altitude, com chuvas no verão e seca no inverno (ALVARES *et al.*, 2013).

Conforme a metodologia de blocos de máximos (COLES, 2001), em cada ano foi extraída a precipitação pluvial diária máxima. O conjunto de dados foi dividido em dois subconjuntos. O primeiro subconjunto compreende o período de 1971 a 2000 e foi reservado para estimação dos parâmetros do modelo. E o segundo subconjunto compreende o período de 2001 a 2020 e corresponde à fase de teste, para comparar a precisão e acurácia das previsões do modelo.

### 3.1.3 Informações a priori: precipitação pluvial máxima anual de Lavras-MG e Silvianópolis

Uma das opções foi utilizar dados de Lavras para obtenção de informações *a priori* devido a semelhança climática entre essa região e São João da Boa Vista - SP. Dados de precipitações pluviais diárias (mm) de Lavras foram obtidos junto ao INMET, os registros pluviométricos são da Estação Climatológica Principal de Lavras-MG, localizada no campus da Universidade Federal de Lavras, Minas Gerais. A cidade de Lavras possui coordenadas geográficas de  $24^{\circ}28'15''$  ao Sul (latitude) e  $50^{\circ}29'19''$  à oeste (longitude) (IBGE, 2011). Segundo a classificação de Köppen (ALVARES *et al.*, 2013), o clima da região é do tipo Cwa, com inverno seco e verão quente e chuvoso. Assim, extraiu-se a precipitação pluvial diária máxima, em cada ano, de 1971 a 2000, formando um bloco de máximos com 30 observações.

Assim, extraiu-se a precipitação pluvial diária máxima, em cada ano, de 1971 a 2000, formando um bloco de máximos com 30 observações.

A outra fonte de informação *a priori* foi baseado em um estudo de [Martins et al. \(2018\)](#), os quais analisaram os dados de precipitação máxima de Silvianópolis.

### 3.1.4 Análise de independência e estacionariedade

Utilizou-se o teste de Ljung-Box para analisar a independência da série de máximos construída ([LJUNG; BOX, 1978](#)). A hipótese nula do teste de Ljung-Box é que as autocorrelações nos dados são iguais a zero.

Para analisar se há tendência (monotônica) na série, aplicou-se o teste de Mann-Kendall ([MCLEOD, 2011](#)). A hipótese nula do teste é de que não existe tendência na série. Para todos os testes foram adotados o nível de significância de 5%.

Estes testes foram aplicados apenas na análise dos dados reais, uma vez que a função *rgev()* gera valores independentes de uma distribuição GEV estacionária, logo as séries de máximos, geradas por esse procedimento, satisfazem as pressuposições de independência e de estacionariedade.

## 3.2 ESTRUTURAS A PRIORI

Nesta subseção serão apresentados os hiperparâmetros das distribuições a priori para os casos simulados e para o caso real (precipitação máxima de São João da Boa Vista - SP).

### 3.2.1 Casos simulados

Considerou-se distribuições *a priori*: distribuição normal trivariada não informativa, distribuição normal trivariada informativa com matriz de variância multiplicada por 1 ou por 4; e duas distribuições a priori gama para a diferença de quantis, com variabilidades distintas.

Na Tabela 2 são apresentados os hiperparâmetros das distribuições *a priori* normais trivariadas (expressão 2.13). Associou-se a distribuição GEV (função de verossimilhança 3.8) com tais distribuições *a priori*, obtendo-se: modelo não informativo ( $M_{NTNI}$ ), modelo com

variâncias dos parâmetros multiplicada por 1 ( $M_{NTV1}$ ) e modelo com variâncias dos parâmetros multiplicada por 4 ( $M_{NTV4}$ ). Nesses casos, utilizou-se, como hiperparâmetros, as estimativas dos parâmetros da distribuição GEV, assim como suas variâncias, de cada um dos cenários de simulação; esses valores podem ser conferidos nos próprios artigos citados na Tabela 1.

Tabela 2 – Hiperparâmetros (médias e de variâncias) das distribuições normais *a priori* para os parâmetros  $\mu$ ,  $\sigma$  e  $\xi$  da distribuição GEV para cada cenário-base (CB).

CB	Modelo	$\mu_0$	$\sigma_0$	$\xi_0$	$Var(\mu_0)$	$Var(\sigma_0)$	$Var(\xi_0)$
a	$M_{NTNI}$	0	0	0	10000,0	10000,0	100,0
	$M_{NTV1}$	70,0	30,0	0,4	18,08	15,29	0,01
	$M_{NTV4}$	70,0	30,0	0,4	72,32	61,15	0,05
b	$M_{NTNI}$	0	0	0	10000,0	10000,0	100,0
	$M_{NTV1}$	70,0	25,0	0,05	6,300	12,816	0,006
	$M_{NTV4}$	70,0	25,0	0,05	25,200	51,266	0,026
c	$M_{NTNI}$	0	0	0	10000,0	10000,0	100,0
	$M_{NTV1}$	34,0	2,0	-0,4	0,155	0,079	0,016
	$M_{NTV4}$	34,0	2,0	-0,4	0,621	0,316	0,063
d	$M_{NTNI}$	0	0	0	10000,0	10000,0	10000,0
	$M_{NTV1}$	6,78	0,561	-0,2	0,004	0,002	0,004
	$M_{NTV4}$	6,78	0,561	-0,2	0,015	0,008	0,015
e	$M_{NTNI}$	0	0	0	10000000,0	1000000,0	10000,0
	$M_{NTV1}$	1200,0	340,0	0,1	1521,0	1296,0	0,008
	$M_{NTV4}$	1200,0	340,0	0,1	6084,0	5184,0	0,032

Fonte: Do autor.

Na Tabela 3 são apresentados os hiperparâmetros das distribuições gama (expressão 2.12). Associou-se a distribuição GEV (função de verossimilhança 3.8) com tais distribuições *a priori*, obtendo-se: modelo com menor variabilidade ( $M_{Gama1}$ ), modelo com maior variabilidade ( $M_{Gama2}$ ). Para determinação de tais hiperparâmetros, utilizou-se estimativas níveis de retorno presentes também nos artigos citados na Tabela 1. Através da função “gamma.parms.from.quantiles()” no *software R*, é possível obter os parâmetros da distribuição gama, os quais permitem a obtenção das diferenças de quantis desejada. Esta função utiliza como princípio, outras funções básicas do *R* relacionadas à distribuição gama como “rgamma()”, “dgamma()” e “lgamma()” e um método numérico (Newton-Raphson).



Tabela 3 – Hiperparâmetros, forma e escala ( $\alpha_i$  e  $\beta_i$ ), das distribuições gama *a priori* de ( $\tilde{q}_{pi}$ ), considerando limites inferiores (LI) e superiores (LS) da estimativa intervalar das diferenças de quantis, em cada cenário-base (CB).

CB	$p_i$	$M_{Gama1}$				$M_{Gama2}$			
		$\tilde{q}_{pi}$		Hiperparâmetros		$\tilde{q}_{pi}$		Hiperparâmetros	
		LI	LS	$\alpha_i$	$\beta_i$	LI	LS	$\alpha_i$	$\beta_i$
a	0,1	146	226	80,98	2,27	146	226	80,98	2,27
	0,01	148	583	8,65	38,1	0	437	3,42	84,06
	0,001	262	1695	4,88	168,46	0	1547	0,57	818,03
b	0,1	114,14	147,68	226,03	0,58	114,14	147,68	226,03	0,58
	0,02	33,54	74,1	24,05	2,14	12,23	38,09	1,09	27,86
	0,01	0	107,64	13,84	1,73	0	112,19	1,08	29,37
c	0,1	36,59	37,73	16324,61	0,0022	36,59	37,73	16324,61	0,0022
	0,02	0,9	1,43	72,16	0,02	0	2,57	17,73	0,096
	0,01	0,21	0,46	1,46	0,099	0	2,13	27,36	0,0552
d	0,1	7,65	8,22	2975,76	0,0027	7,65	8,22	2975,76	0,0027
	0,02	0,44	0,5	940,79	0,0005	0	1,07	3357,17	0,0003
	0,002	0,31	0,86	15,24	0,0361	0	1,49	1,72	62,145
e	0,1	1700	2050	438,91	4,26	1700	2050	438,91	4,26
	0,01	450	920	30,53	21,76	100	1270	2,83	182,59
	0,001	250	1130	7,23	84,49	0	1950	0,58	718,84

Fonte: Do autor.

### 3.2.2 Estudo de precipitação

#### 3.2.2.1 Priori Fundamentada nas informações de Lavras-MG

Para os modelos com priori Gama para diferenças de quantis da GEV, utilizou-se a estimativa intervalar do nível de retorno para obter informações *a priori*, através de um intervalo de confiança de  $(1 - \alpha)\%$

$$IC_{95\%}[y(t)] = \widehat{y(t)} \pm z_{\alpha/2} \sqrt{Var(\widehat{y(t)})}, \quad (3.1)$$

em que,  $\widehat{y(t)}$  é uma estimativa de precipitação máxima obtida via método da máxima verossimilhança para o tempo de retorno  $p_t = \frac{1}{t}$ ,  $z_{\alpha/2}$  o quantil da normal padrão tal que  $P(|Z| < z_{\alpha/2})$  com um nível de significância  $\alpha$  e  $Var(\widehat{y(t)})$  é a estimativa da variância do nível de retorno  $y(t)$ . No *software R*, a função *confint()* permite obter os intervalos de confiança. Assim, ajustou-se a distribuição GEV à série de precipitação máxima de Lavras-MG com o método da máxima verossimilhança, obtendo-se estimativas intervalares para níveis de retorno

de 10, 100 e 200 anos, os quais são apresentados na Tabela 4.

Tabela 4 – Estimativas intervalares de precipitação máxima para tempos de retorno (TR) de 10, 100 e 200 anos série de máximos de Lavras-MG, via máxima verossimilhança.

Estim.	Níveis de Retorno		
	TR-10	TR-100	TR-200
LI	95,33	110,06	96,57
LS	154,10	249,48	312,63

Fonte: Do autor.

LI e LS são os limites inferior e superior, respectivamente, do intervalo de confiança de 95%.

Estes valores *a priori* serviram como referência para a determinar os hiperparâmetros da distribuição Gama. Para melhor visualização, os valores *a priori* de  $\tilde{q}_{p_i}$ , assim como os respectivos hiperparâmetros  $\alpha_i$  e  $\beta_i$  das distribuições Gama são apresentados na Tabela 6. Considerou-se como 0 (zero) o limite inferior, nos casos em que a diferença entre os quantis foi negativa, uma vez que a distribuição Gama não assume valores negativos.

Para melhor visualização, os valores *a priori* de  $\tilde{q}_{p_i}$ , assim como os respectivos hiperparâmetros  $\alpha_i$  e  $\beta_i$  das distribuições Gama são apresentados novamente na Tabela 5.

Tabela 5 – Hiperparâmetros, forma e escala ( $\alpha_i$  e  $\beta_i$ ), das distribuições gama *a priori* de ( $\tilde{q}_{p_i}$ ), considerando limites inferiores (LI) e superiores (LS) da estimativa intervalar das diferenças de quantis de precipitações máximas de Lavras.

$M_{\text{Gama1-L}}$					$M_{\text{Gama2-L}}$				
$p_i$	$\tilde{q}_{p_i}$		Hiperparâmetros		$p_i$	$\tilde{q}_{p_i}$		Hiperparâmetros	
	LI	LS	$\alpha_i$	$\beta_i$		LI	LS	$\alpha_i$	$\beta_i$
0,1	95,33	154,1	67,09	1,83	0,1	95,33	154,1	67,09	1,83
0,01	14,73	95,38	4,87	9,48	0,01	0	154,15	0,9845	42,17
0,005	0	63,15	1,3	14,7	0,005	0	202,57	0,9157	57,73

Fonte: Do autor.

Para obtenção das informações sobre hiperparâmetros da distribuição Normal Trivariada informativa, o vetor de médias dos hiperparâmetros  $\Phi_0$  e a matriz de variância  $\Sigma_0$ , considerou-se o ajuste da distribuição da GEV via método da máxima verossimilhança aos

dados de precipitação máxima diária anual de Lavras-MG. Os parâmetros foram estimados pelo método da máxima verossimilhança (aproximação numérica Nelder-Mead) através da função  $fgev()$  do pacote *evd*; as variâncias e covariâncias dos estimadores de máxima verossimilhança são dados pelos elementos da inversa da matriz de informação de Fisher, através da função  $var.cov()$  do pacote *evd* (STEPHENSON, 2002) do *software* R (R CORE TEAM, 2018).

Logo, tem-se os seguintes hiperparâmetros:

Normal Trivariada com variância multiplicada por 1:

$$\Phi_0 = (68,98; 20,27; 0,19) \quad e \quad \Sigma_0 = \begin{bmatrix} 23,61 & 12,99 & -0,68 \\ 12,99 & 15,97 & -0,48 \\ -0,68 & -0,48 & 0,07 \end{bmatrix}. \quad (3.2)$$

Normal Trivariada com variância multiplicada por 4:

$$\Phi_0 = (68,98; 20,27; 0,19) \quad e \quad \Sigma_0 = \begin{bmatrix} 94,45 & 51,95 & -2,71 \\ 51,95 & 63,89 & -1,92 \\ -2,71 & -1,92 & 0,26 \end{bmatrix}. \quad (3.3)$$

E para a estrutura *a priori* Normal Trivariada não informativa:

$$\Phi_0 = (0; 0; 0) \quad e \quad \Sigma_0 = \begin{bmatrix} 10000 & 0 & 0 \\ 0 & 10000 & 0 \\ 0 & 0 & 100 \end{bmatrix}. \quad (3.4)$$

Denominou-se os modelos:  $M_{NTV1-L}$ ,  $M_{NTV4-L}$ ,  $M_{NTNI}$ ; os quais são a associação da distribuição GEV com as estruturas *a priori* (3.2), (3.3) e (3.4), respectivamente.

### 3.2.2.2 Priori Fundamentada nas informações de Silvianópolis

Os hiperparâmetros das prioris foram obtidos com base no artigo de Martins *et al.* (2018) e correspondem aos mesmos utilizados no estudo de simulação do cenário-base **b**, os quais também foram descritos nas Tabelas 2 e 3.

Os hiperparâmetros das distribuições *a priori* normais trivariadas são dados a seguir.

Tabela 6 – Hiperparâmetros, forma e escala ( $\alpha_i$  e  $\beta_i$ ), das distribuições gama *a priori* de ( $\tilde{q}_{pi}$ ), considerando limites inferiores (LI) e superiores (LS) da estimativa intervalar das diferenças de quantis de precipitações máximas de Silvianópolis.

$M_{\text{Gama1-S}}$					$M_{\text{Gama2-S}}$				
$p_i$	LI	$\tilde{q}_{pi}$	LS	Hiperparâmetros $\alpha_i$ $\beta_i$	$p_i$	LI	$\tilde{q}_{pi}$	LS	Hiperparâmetros $\alpha_i$ $\beta_i$
0,1	114,14	147,68	226,03	0,58	0,1	114,14	147,68	226,03	0,58
0,02	33,54	74,1	24,05	2,14	0,02	12,23	38,09	1,09	27,86
0,01	0	107,64	13,84	1,73	0,01	0	112,19	1,08	29,37

Fonte: Do autor.

Normal Trivariada com variância multiplicada por 1:

$$\Phi_0 = (70,0; 25,0; 0,05) \quad e \quad \Sigma_0 = \begin{bmatrix} 6,3 & 0 & 0 \\ 0 & 12,816 & 0 \\ 0 & 0 & 0,006 \end{bmatrix}. \quad (3.5)$$

Normal Trivariada com variância multiplicada por 4:

$$\Phi_0 = (70,0; 25,0; 0,05) \quad e \quad \Sigma_0 = \begin{bmatrix} 25,2 & 0 & 0 \\ 0 & 51,266 & 0 \\ 0 & 0 & 0,026 \end{bmatrix}. \quad (3.6)$$

Denominou-se os modelos:  $M_{\text{NTV1-S}}$  e  $M_{\text{NTV4-S}}$ ; os quais são a associação da distribuição GEV com as estruturas *a priori* (3.5) e (3.6), respectivamente.

### 3.2.3 Função de Verossimilhança da Distribuição GEV

Considerando  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ ,  $n$  observações independentes e identicamente distribuídas, com distribuição GEV, a função de verossimilhança, com  $\xi \neq 0$ , é dada por

$$L((\mu, \sigma, \xi); \mathbf{y}) = \frac{1}{\sigma^n} \prod_{i=1}^n \left\{ \left[ 1 + \xi \left( \frac{y_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \right\} \exp \left\{ \sum_{i=1}^n \left\{ - \left[ 1 + \xi \left( \frac{y_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi}\right)} \right\} \right\}. \quad (3.7)$$

Por ser mais prático trabalhar com somas do que com produtos, é comum utilizar o logaritmo da função de verossimilhança. Logo, o logaritmo (neperiano) da distribuição GEV é

dado por (WHANG et al., 2017):

$$\begin{aligned} l((\mu, \sigma, \xi); \mathbf{y}) &= \ln[L(\boldsymbol{\theta}; \mathbf{y})] \\ &= -n \ln \sigma - \left( \frac{1 + \xi}{\xi} \right) \sum_{i=1}^n \ln \left[ 1 + \xi \left( \frac{y_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^n \left[ 1 + \xi \left( \frac{y_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \end{aligned} \quad (3.8)$$

Se  $\xi = 0$  (distribuição Gumbel), então o logaritmo da função de verossimilhança é

$$l((\mu, \sigma); \mathbf{y}) = -n \ln \sigma + \sum_{i=1}^n \left[ \left( -\frac{y_i - \mu}{\sigma} \right) - \exp \left( -\frac{y_i - \mu}{\sigma} \right) \right] \quad (3.9)$$

### 3.2.4 Posteriores

A distribuição *a posteriori* é resultante do produto função de verossimilhança (3.2.3) e a densidade *a priori*  $\pi(\boldsymbol{\theta})$ , a qual pode ser dada por uma das expressões (2.11) e (2.13). Esse produto, no entanto, não possui solução analítica, sendo assim utilizou-se técnicas de MCMC através do *software R* (R CORE TEAM, 2018) para obtenção de uma solução numérica.

Neste trabalho foi utilizados o pacote *evdbayes*, que possui função própria para a obtenção de distribuições *a posteriori* através do algoritmo Metropolis-Hastings. As funções *prior.norm* e *prior.quant* constroem, respectivamente, as distribuições *a priori*: Normal Trivariada, expressão (2.13); Gama, expressão (2.11). A função *posterior* gera uma cadeia de Markov  $\theta_0, \dots, \theta_n$  com a distribuição alvo  $\pi(\boldsymbol{\theta}|\mathbf{y})$ .

Para obtenção de cada marginal *a posteriori* dos parâmetros, foram geradas cadeias MCMC com 650000 iterações, sendo excluídos os 250000 primeiros e realizado salto a cada 60 iterações. Obtendo-se cadeias finais de aproximadamente 5000 valores. A convergência das cadeias foi monitorada utilizando-se o pacote *coda - Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output* (PLUMMER et al., 2006), disponível no *software R* (R CORE TEAM, 2018).

Os critérios para verificação da convergência das cadeias geradas por MCMC foram: o critério de Raftery e Lewis (1992), rejeitando-se as cadeias em que o fator de dependência foi maior ou igual a 5; a estatística  $Z$  de Geweke (1991), sendo rejeitadas aquelas cadeias em que  $|Z| \geq 1,96$ ; e o critério de Heidelberger e Welch (1983), adotando o nível de significância de 5%.

### 3.3 MÉTRICAS DE ANÁLISE DAS ESTIMATIVAS

Tendo em vista que se propõe realizar análises dos modelos em dois contextos distintos. O primeiro é através de dados provenientes de simulação, o que requer técnicas de avaliação que considerem o fato dos parâmetros serem conhecidos. Logo para esse caso, utilizou-se viés e viés médio relativo percentual das estimativas dos parâmetros. Para o outro caso, de dados reais, utilizou-se critério de Informação Deviance (DIC), e medidas de precisão e acurácia para analisar os níveis de retorno. Estas métricas são definidas nesta subseção.

#### 3.3.1 Para dados simulados

Com o intuito de avaliar o desempenho dos modelos a partir de dados simulados utilizou-se o Viés e o Viés absoluto médio relativo percentual (VAMRP) das estimativas dos parâmetros e dos níveis de retorno, os quais são dados a seguir:

$$\text{Viés}(\omega) = \hat{\omega}_l - \omega \quad (3.10)$$

e

$$\text{VAMRP} = \left| \frac{\sum_{l=1}^s \hat{\omega}_l - \omega}{\omega} \right| \times 100\% \quad (3.11)$$

em que  $\omega$  é um dos parâmetros  $\mu$ ,  $\sigma$  ou  $\xi$ ; ou um dos níveis de retorno associados aos tempos de 10, 50, 100 e 200 anos;  $\hat{\omega}_l$  é o valor estimado do parâmetro na  $l$ -ésima simulação  $s$  corresponde ao número total de simulações (1000).

#### 3.3.2 Para dados de precipitação

Para avaliar o desempenho da distribuição GEV sob diferentes estruturas de prioris, foi calculado o critério de informação do desvio (DIC, do inglês *Deviance Information Criterion*), a acurácia, a amplitude média das estimativas intervalares (HPD) de 95% de probabilidade e o

erro médio de predição.

O DIC de um modelo é uma medida de qualidade do ajuste semelhante ao BIC e que é utilizada quando se tem amostras de uma cadeia *a posteriori* que foi obtida por simulação MCMC (BAROSSO-FILHO; ACHCAR; SOUZA, 2010). Conforme definição de Spiegelhalter *et al.* (2002), o DIC é dado por:

$$DIC = -2\langle \ln(\pi(\mathbf{y}|\boldsymbol{\theta})) \rangle + 2(\ln(\pi(\mathbf{y}|\langle \boldsymbol{\theta} \rangle)) - \langle \ln(\pi(\mathbf{y}|\boldsymbol{\theta})) \rangle), \quad (3.12)$$

em que  $\langle \dots \rangle$  denota uma média sobre a distribuição *a posteriori*. Se a diferença entre os DICs de dois modelos for maior ou igual 5, descarta-se o modelo de maior DIC.

Analisou-se a acurácia e a precisão das predições dos níveis de retorno para 5, 10, 15, 20 anos. A acurácia (o número total de acertos) foi obtida analisando quantos valores observados, para um determinado tempo de retorno, pertencem à respectiva estimativa intervalar (HPD) de 95% de probabilidade. A precisão foi analisada através da amplitude média intervalar (AMI) e do erro médio de predição (EMP), que são dados respectivamente por:

$$AMI = \frac{1}{N} \sum_{i=1}^N LS_i - LI_i \quad (3.13)$$

e

$$EMP = \frac{1}{N} \sum_{i=1}^N \left| \frac{y(t)_i - \hat{y}(t)_i}{y(t)_i} \right| \times 100\%,$$

em que,  $N$  é o número de predições;  $LS_i$  e  $LI_i$  são os limites superior e inferior, respectivamente, da  $i$ -ésima estimativa de  $HPD_{95\%}$ ;  $y(t)_i$  é precipitação máxima observada e  $\hat{y}(t)_i$  é a precipitação máxima estimada para o  $i$ -ésimo tempo de retorno  $t$ .

Para calcular a acurácia, verifica-se se  $y_i \in (HPD_i)_{(1-\alpha)100\%}$ , em que  $x_i$  é o valor observado no  $i$ -ésimo tempo de retorno e  $(HPD_i)_{(1-\alpha)100\%}$  é o intervalo de alta densidade *a posteriori* com  $1 - \alpha$  de probabilidade. Será considerada a contagem do número de valores observados que pertenciam ao intervalo, considerando essa contagem como a acurácia.

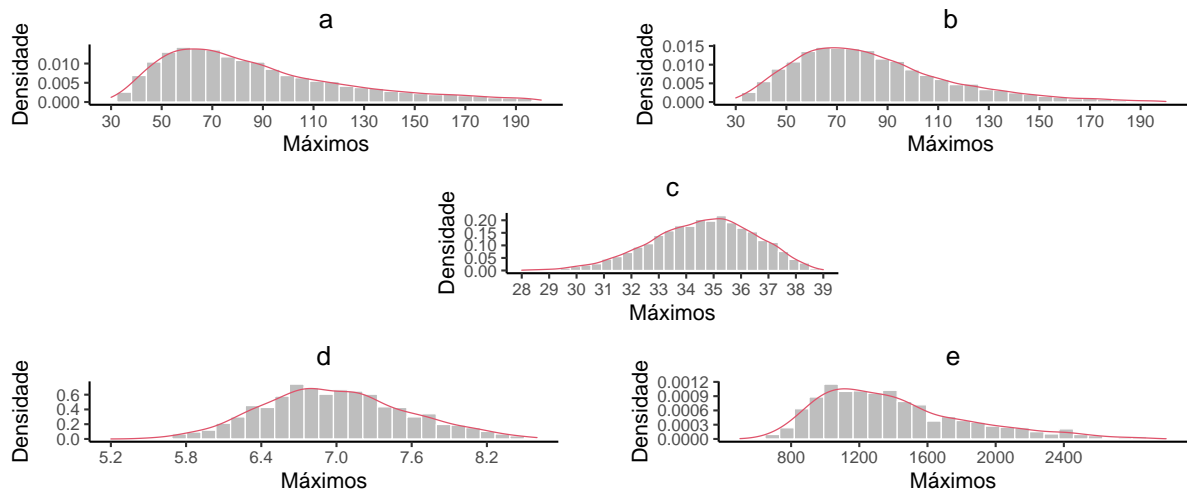
Todas as análises estatísticas foram feitas com auxílio do *software* R (R CORE TEAM, 2021) e seus respectivos pacotes: *coda* (PLUMMER *et al.*, 2006), *evd* (STEPHENSON, 2002), *evdbayes* (STEPHENSON; RIBATET, 2014).

## 4 RESULTADOS E DISCUSSÃO

### 4.1 ANÁLISE DOS CENÁRIOS SIMULADOS

Na Figura 1 são apresentados os histogramas dos valores simulados em cada cenário-base da distribuição GEV.

Figura 1 – Histogramas nos cenários da distribuição GEV com parâmetros:  $\mu = 70, \sigma = 30, \xi = 0,4$  (**a**);  $\mu = 70, \sigma = 25, \xi = 0,05$  (**b**);  $\mu = 34, \sigma = 2, \xi = -0,4$  (**c**);  $\mu = 6,78, \sigma = 0,561, \xi = -0,2$  (**d**); e  $\mu = 1200, \sigma = 340, \xi = 0,1$  (**e**).



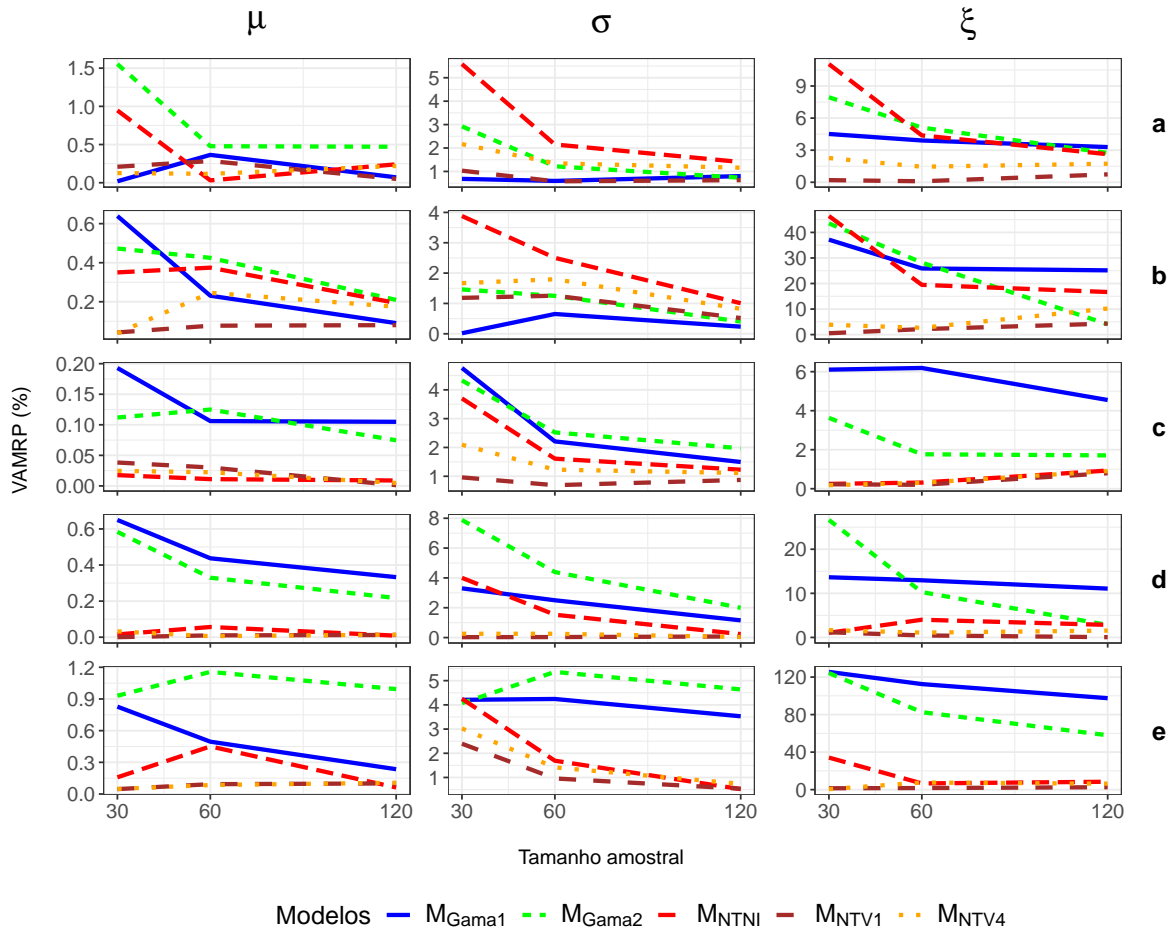
Fonte: Do autor.

Pode-se observar que, nos cenários da distribuição GEV com parâmetro  $\xi$  positivo (Cenários **a**, **b** e **e**, Figura 1), a cauda superior da distribuição é longa, conforme esperado dado os valores dos parâmetros na simulação. Já nos cenários com  $\xi$  negativo (cenários-base **c** e **d**), a cauda é finita e a distribuição apresenta menor grau de assimetria.

Na Figura 2 é apresentado o viés absoluto médio relativo percentual (VAMRP) dos parâmetros posição, escala e forma; nas Figuras 3, 4 e 5 são apresentados os Boxplots do viés dos parâmetros (considerando 1000 simulações).



Figura 2 – Viés absoluto médio relativo percentual (VAMRP) das estimativas dos parâmetros  $\mu$ ,  $\sigma$  e  $\xi$  para cada modelo (dependente da escolha da priori) nos cenários com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 70, \sigma = 30, \xi = 0,4$  (a);  $\mu = 70, \sigma = 25, \xi = 0,05$  (b);  $\mu = 34, \sigma = 2, \xi = -0,4$  (c);  $\mu = 6,78, \sigma = 0,561, \xi = -0,2$  (d); e  $\mu = 1200, \sigma = 340, \xi = 0,1$  (e).

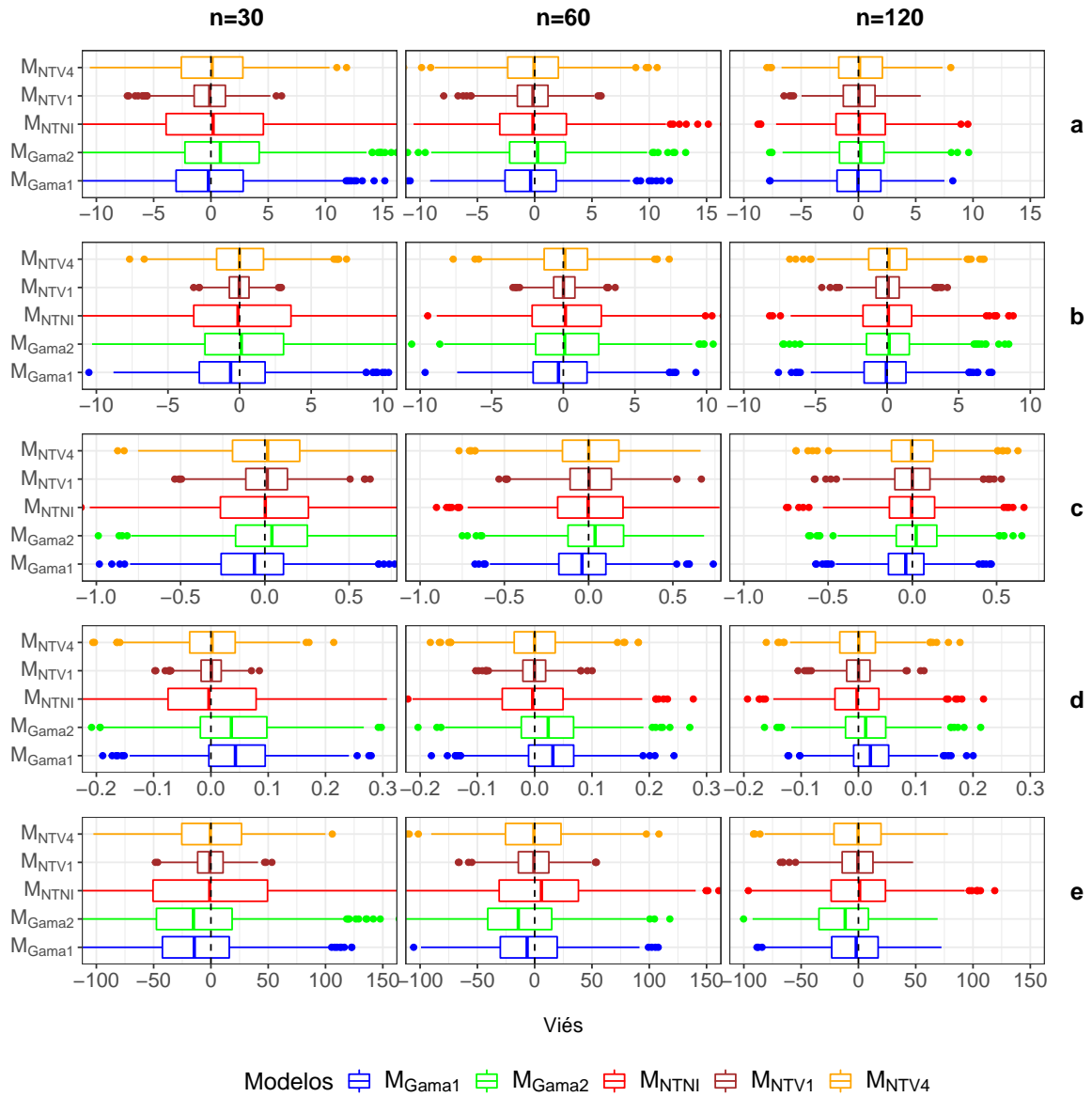


Fonte: Do autor.

Observa-se que, de maneira geral, distribuições a priori mais informativas, como os modelos  $M_{NTV1}$  e  $M_{NTV4}$ , produzem estimativas mais precisas de todos os parâmetros da distribuição GEV, considerando pequenos tamanhos amostrais (Figura 2). Com o aumento do tamanho da amostra, tem-se a redução do viés do parâmetro. Observa-se também que os modelos com priori Gama apresentam maior VAMRP nos cenários **c**, **d** e **e**.

Com tamanho amostral grande (120), identifica-se dois padrões: primeiro, nos cenários em que simulou-se parâmetro forma positivo e parâmetro de posição igual a 70 (Cenário **a** e Cenário **b**, Figura 2), não há diferença clara entre resultados dos modelos; segundo, no restante dos cenários, os modelos com priori normal trivariada produzem estimativas mais precisas dos parâmetros.

Figura 3 – Boxplot do viés do parâmetro de posição  $\mu$  para cada modelo (dependente da escolha da priori) nos cenários com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 70, \sigma = 30, \xi = 0,4$  (a);  $\mu = 70, \sigma = 25, \xi = 0,05$  (b);  $\mu = 34, \sigma = 2, \xi = -0,4$  (c);  $\mu = 6,78, \sigma = 0,561, \xi = -0,2$  (d); e  $\mu = 1200, \sigma = 340, \xi = 0,1$  (e).

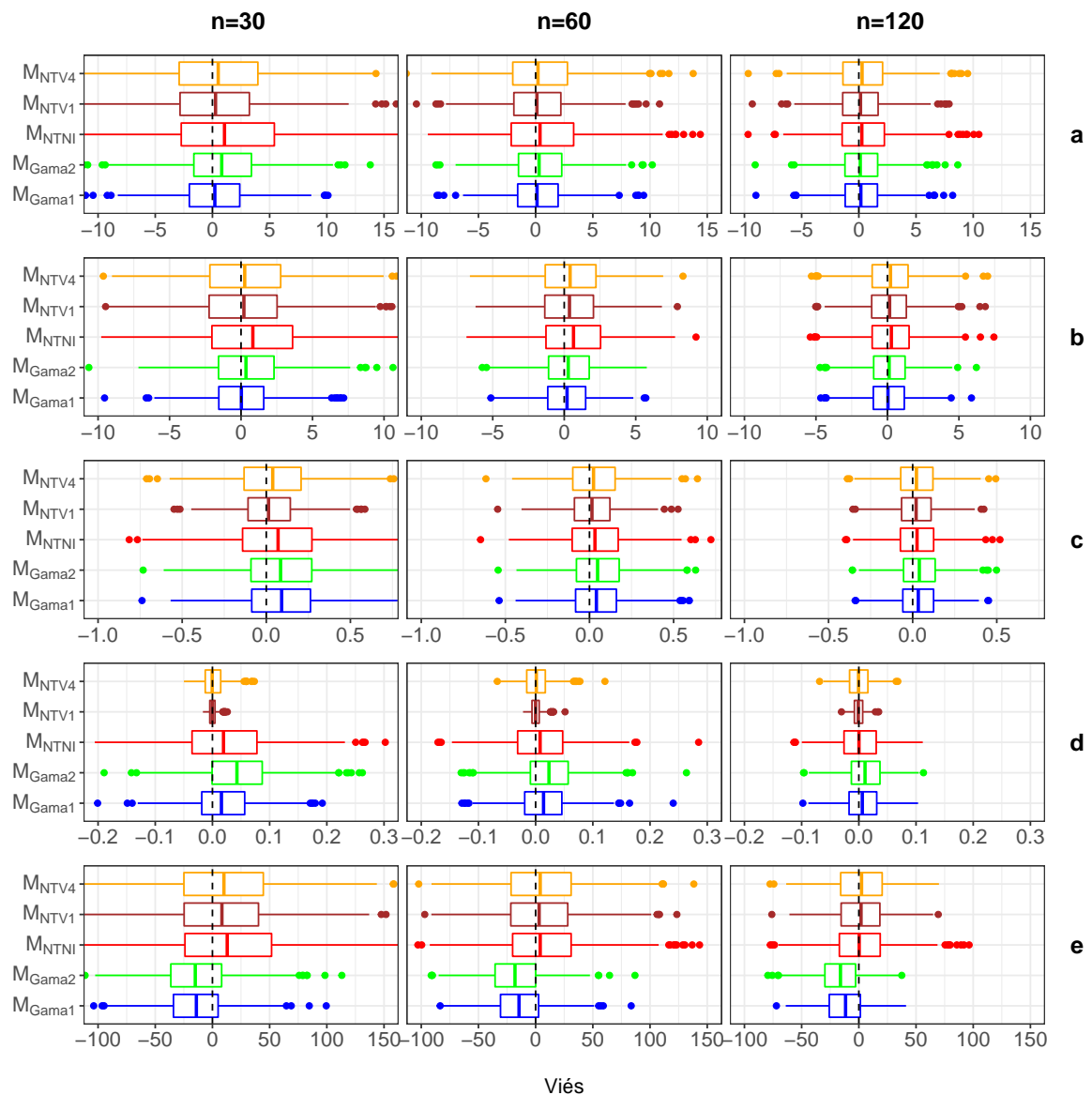


Fonte: Do autor.

Verifica-se que, independentemente do cenário, os vieses de  $\mu$  (Boxplots da Figura 3) dos modelos com priori normal trivariada, informativos e não informativos, estão centrados em 0, isto é, não há uma direção clara para o viés desse parâmetro com estes modelos. Observa-se a amplitude dos Boxplots do modelo  $M_{NTV1}$  foi menor (mais preciso) que os outros, quando se considera o tamanho amostral  $n = 30$  e  $n = 60$ . Já com o tamanho amostral  $n = 120$ , os

resultados são similares.

Figura 4 – Boxplot do viés do parâmetro de escala,  $\sigma$ , para cada modelo (dependente da escolha da priori) nos cenários com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 70, \sigma = 30, \xi = 0,4$  (a);  $\mu = 70, \sigma = 25, \xi = 0,05$  (b);  $\mu = 34, \sigma = 2, \xi = -0,4$  (c);  $\mu = 6,78, \sigma = 0,561, \xi = -0,2$  (d); e  $\mu = 1200, \sigma = 340, \xi = 0,1$  (e).



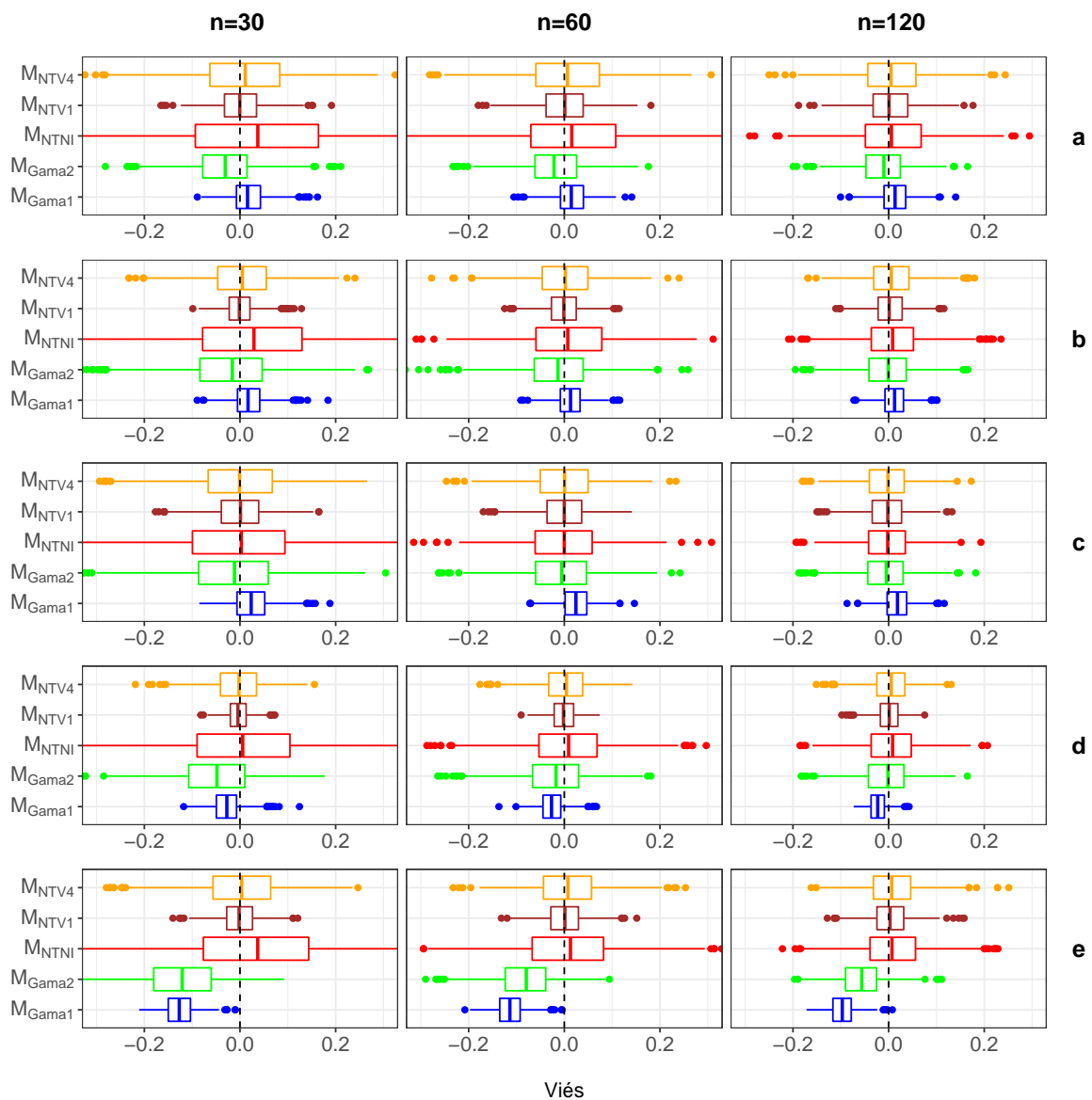
Fonte: Do autor.

Não houve diferenças consideráveis entre resultados de viés para  $\sigma$  dos cenários **a**, **b** e **c** (Boxplots da Figura 4). O boxplot de viés com o modelo  $M_{NTV1}$  apresentou, novamente, menor amplitude nos cenários **d** e **e**.

O viés de  $\xi$  em todos os modelos com priori normal trivariada centrou-se em 0 (zero)

(Figura 5). O modelo  $M_{NTV1}$  apresentou menor amplitude do Boxplot em todos os cenários, sendo, mais uma vez, mais preciso. Os modelos  $M_{Gama1}$  e  $M_{Gama2}$  produziram vieses negativos para  $\xi$ , subestimando este parâmetro em todos os cenários.

Figura 5 – Boxplot do viés do parâmetro de forma,  $\xi$ , para cada modelo (dependente da escolha da priori) nos cenários com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 70, \sigma = 30, \xi = 0,4$  (a);  $\mu = 70, \sigma = 25, \xi = 0,05$  (b);  $\mu = 34, \sigma = 2, \xi = -0,4$  (c);  $\mu = 6,78, \sigma = 0,561, \xi = -0,2$  (d); e  $\mu = 1200, \sigma = 340, \xi = 0,1$  (e).



Fonte: Do autor.

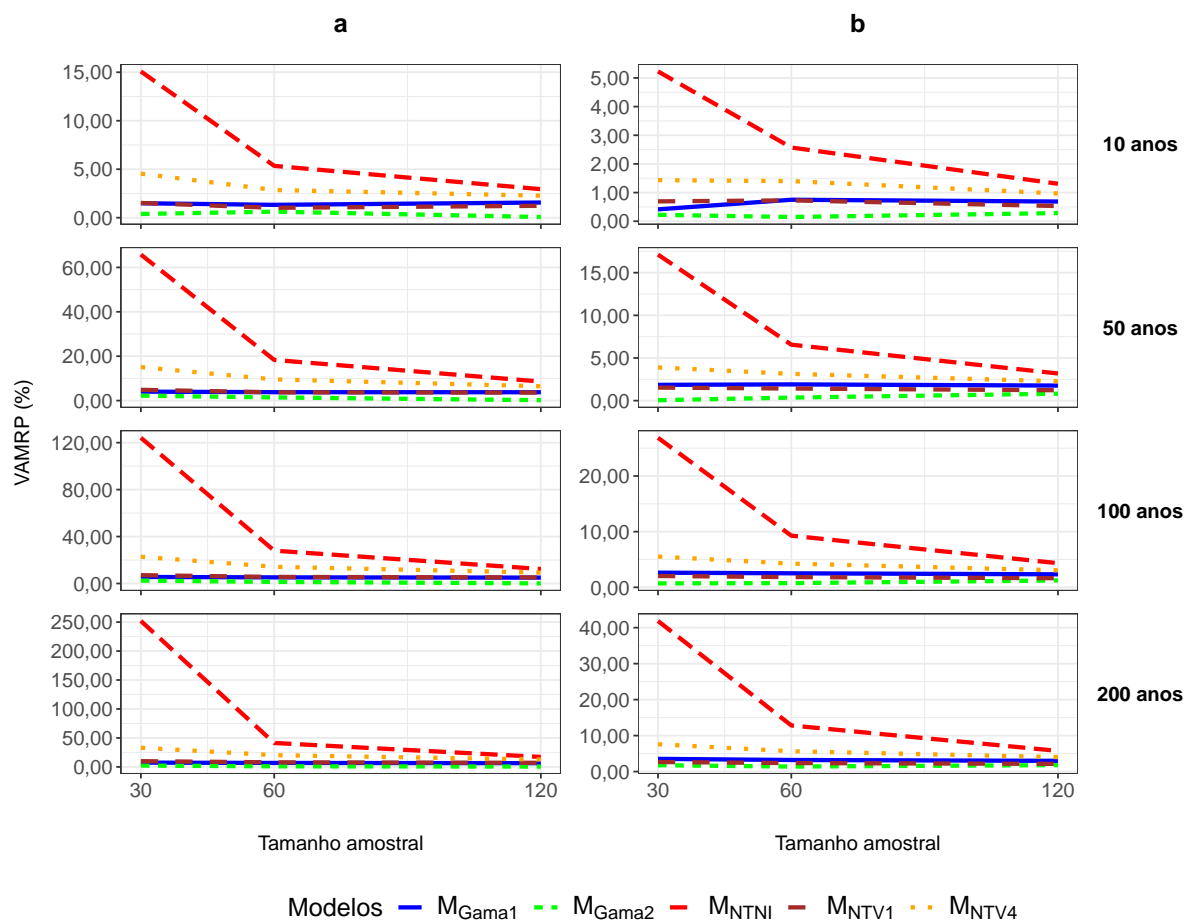
O modelo não informativo,  $M_{NTNI}$ , apresentou menor precisão (maior variabilidade do viés) para todos cenários e parâmetros (Boxplots Figuras 3, 4 e 5). Os resultados deste modelo

aproximam-se dos demais com o aumento do tamanho amostral, e ainda mais dos obtidos pelos modelos com priori normal trivariada informativos.

Pode-se citar diversos trabalhos, nos quais constatou-se que distribuições *a priori* informativas, proporcionaram estimativas mais precisas e acuradas dos parâmetros da distribuição GEV, como: [Beijo, Vivanco e Muniz \(2009\)](#), [Diriba et al. \(2014\)](#), [Martins et al. \(2018\)](#). Estes trabalhos corroboram os encontrados aqui, ressaltando-se o ganho que se tem na estimativa de parâmetros quando se tem informações *a priori* e estas são incorporadas de maneira adequada na abordagem Bayesiana do problema.

Nas Figuras 6 e 7 é apresentado o viés absoluto médio relativo percentual (VAMRP) dos níveis de retorno de 10, 50, 100 e 200 anos.

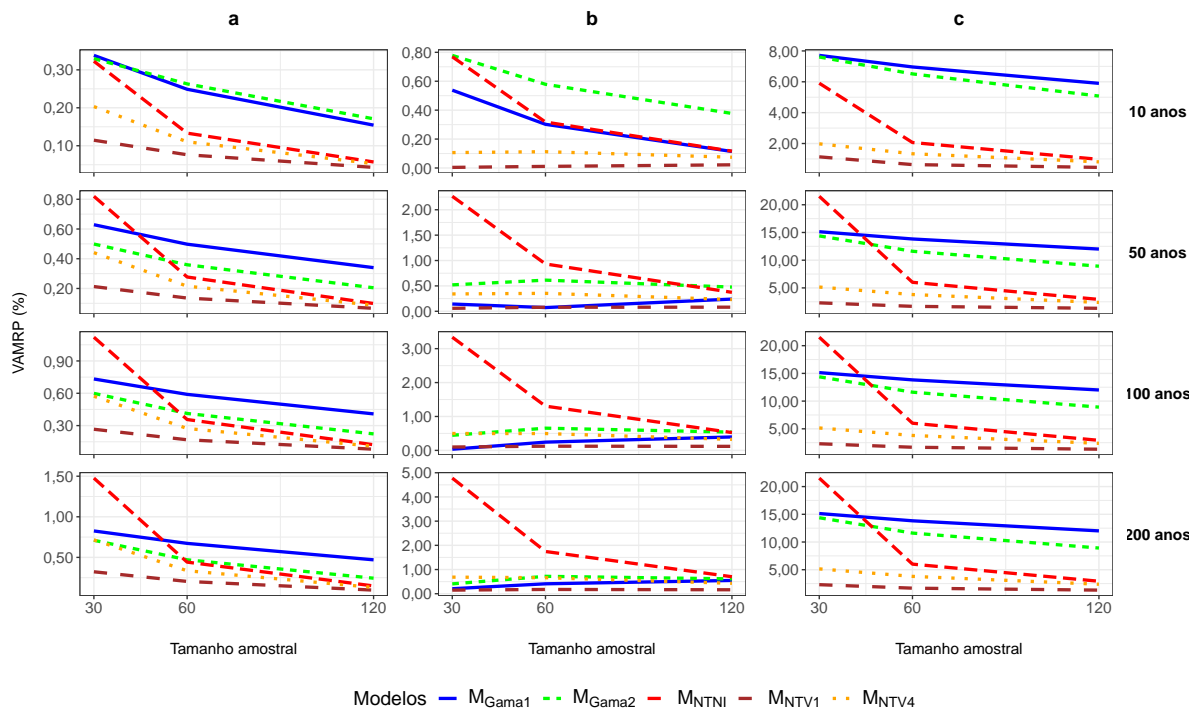
Figura 6 – Viés absoluto médio relativo percentual (VAMRP, %) das estimativas dos níveis de retorno para os tempos de 10, 50, 100 e 200 anos, para cada modelo (dependente da escolha da priori) nos cenários com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 70, \sigma = 30, \xi = 0,4$  (a);  $\mu = 70, \sigma = 25, \xi = 0,05$  (b).



Fonte: Do autor.

Pode-se constatar que o modelo  $M_{Gama2}$  foi levemente mais preciso nas predições de níveis de retorno em todos os tempos analisados nos cenários-base **a** e **b** (Figura 6). Além disso, no cenário-base **a** e para tempos de retorno maiores ou iguais a 50 anos, a precisão das predições desse modelo tende a aumentar com o incremento do tamanho amostral. No cenário-base **b**, com tamanhos de amostras maiores, ocorre diminuição da distância entre os vieses dos modelos.

Figura 7 – Viés absoluto médio relativo percentual (VAMRP, %) das estimativas dos níveis de retorno para os tempos de 10, 50, 100 e 200 anos, para cada modelo (dependente da escolha da priori) nos cenários com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 34, \sigma = 2, \xi = -0,4$  (**c**);  $\mu = 6,78, \sigma = 0,561, \xi = -0,2$  (**d**); e  $\mu = 1200, \sigma = 340, \xi = 0,1$  (**e**).



Fonte: Do autor.

É possível perceber que o modelo  $M_{NTV1}$  apresentou menor viés para todos os níveis de retorno nos cenários **c**, **d** e **e** (Figura 7). A redução do viés nas estimativas é observada com tamanhos de amostra maiores. O modelo  $M_{Gama1}$  sofre redução de precisão na predição de níveis de retorno maiores, com o aumento do tamanho de amostra no cenário **d**.

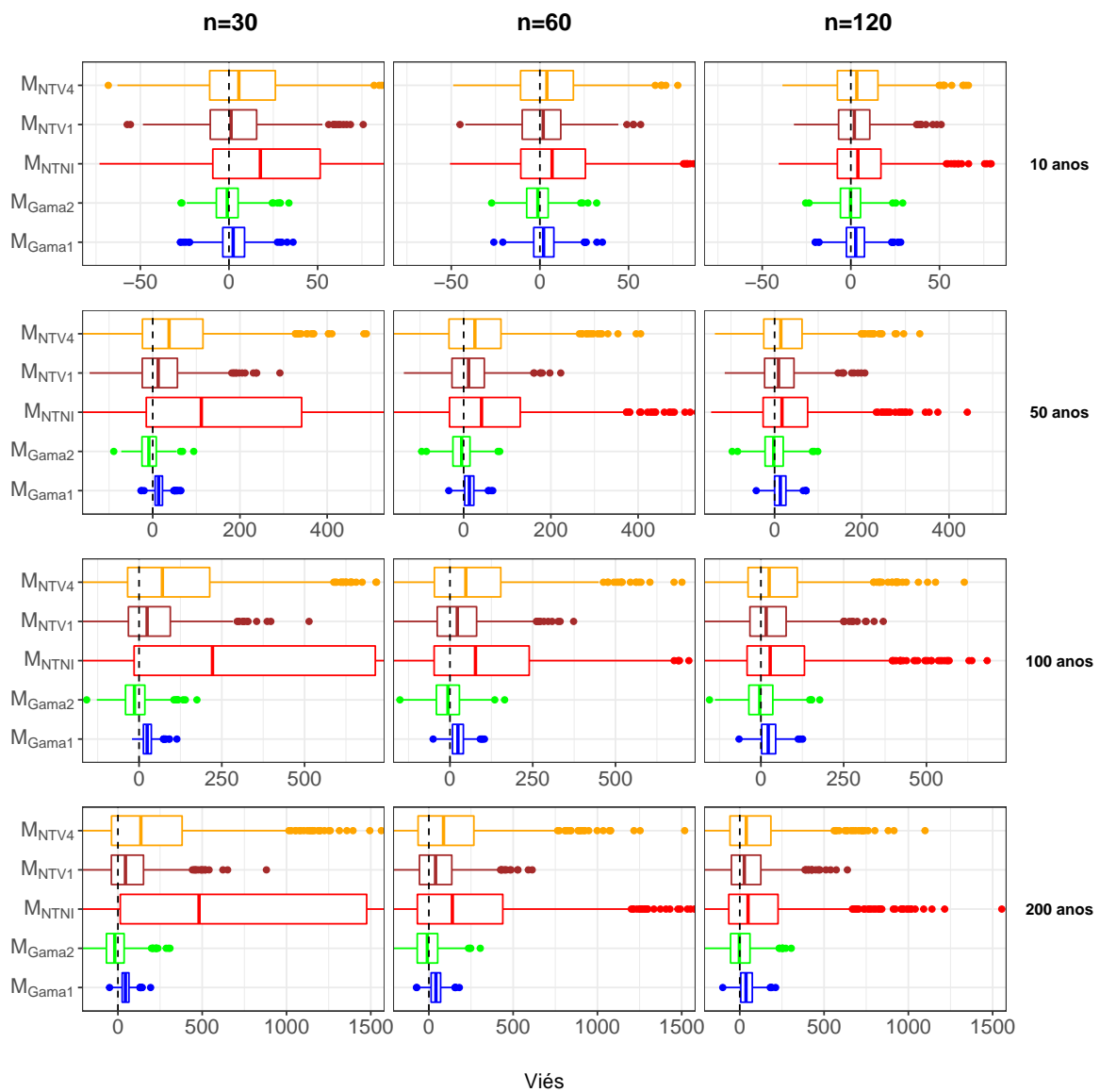
De maneira geral, com tempos de retorno mais altos obteve-se vieses mais elevados, independentemente da estrutura a priori utilizada em cada modelo (Figuras 6 e 7).

Nas Figuras 8, 9, 10, 11 e 12 estão apresentados os Boxplots do viés dos níveis de

retorno em cada cenário-base.

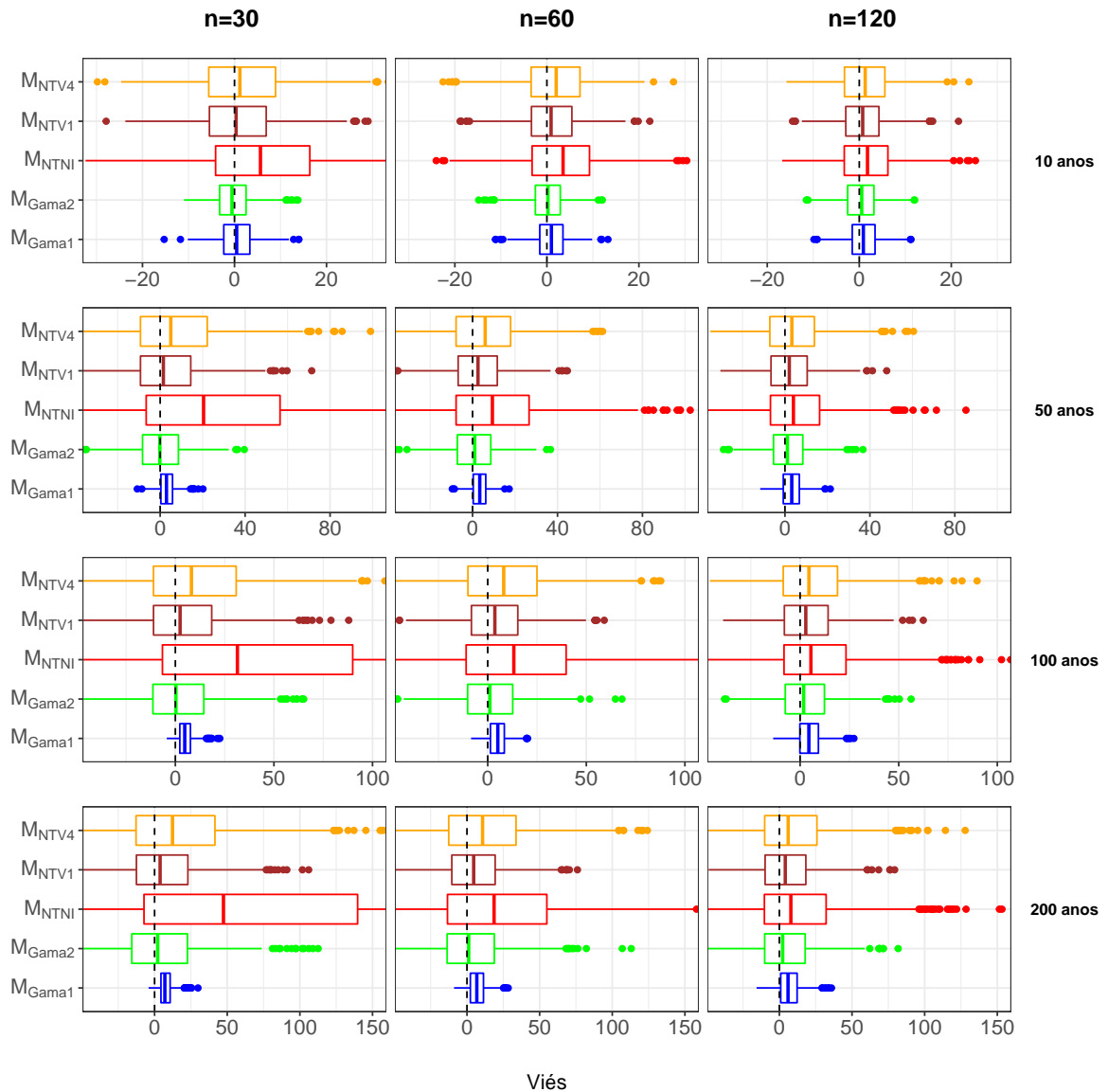
Conforme pode-se observar nas Figuras 8 e 9 (cenário-base **a** e **b**), os Boxplots tiveram o mesmo padrão, com os modelos apresentando viés positivo e este é crescente a medida em que se aumenta os tempos de retorno. Além disso, nestes dois cenários-base e também no cenário-base **c** (Figura 10), o modelo  $M_{Gama1}$  apresenta menor amplitude da Boxplot dos níveis de retorno, indicando maior precisão da estimativa.

Figura 8 – Boxplot do viés de níveis de retorno para os tempos de 10, 50, 100 e 200 anos para cada modelo (dependente da escolha da priori) no Cenário-base (**a**) com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 70, \sigma = 30, \xi = 0,4$ .



Fonte: Do autor.

Figura 9 – Boxplot do viés de níveis de retorno para os tempos de 10, 50, 100 e 200 anos para cada modelo (dependente da escolha da priori) no Cenário-base (b) com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 70, \sigma = 25, \xi = 0,05$ .

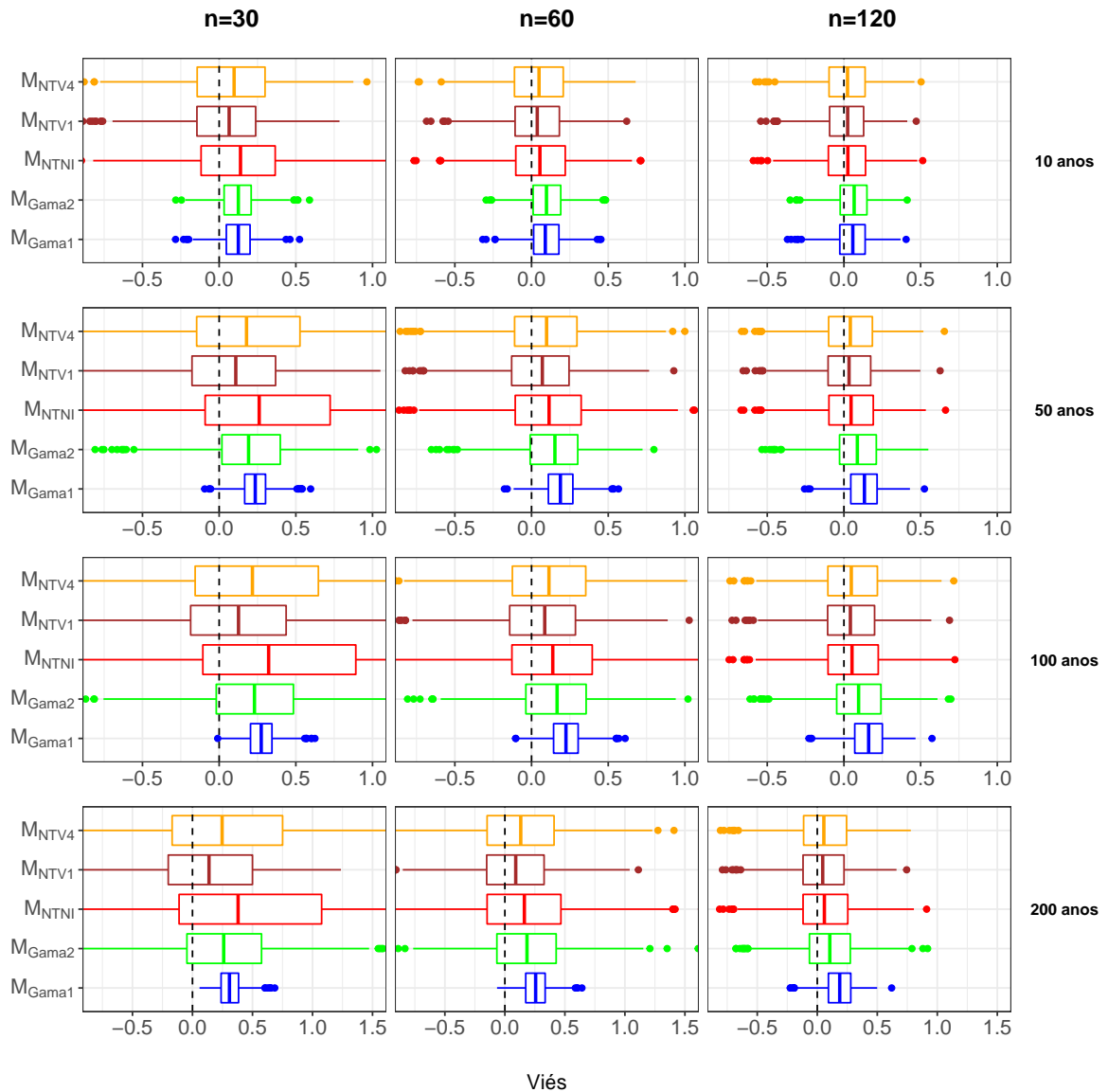


Fonte: Do autor.

No cenário-base **d** (Figura 11), os modelos com estruturas *a priori* mais informativas ( $M_{Gama1}$  e  $M_{NTV1}$ ) foram mais precisos. Já no cenário-base **e** (Figura 12), os modelos com estruturas *a priori* apresentaram viés mais centralizado, isto é, nem subestimaram ou superestimaram os níveis de retorno de maneira tão significativa. Constatou-se que apenas com os modelos com priori Gama, no cenário **e**, houve viés negativo, isto é, houve subestimação dos níveis de retorno.



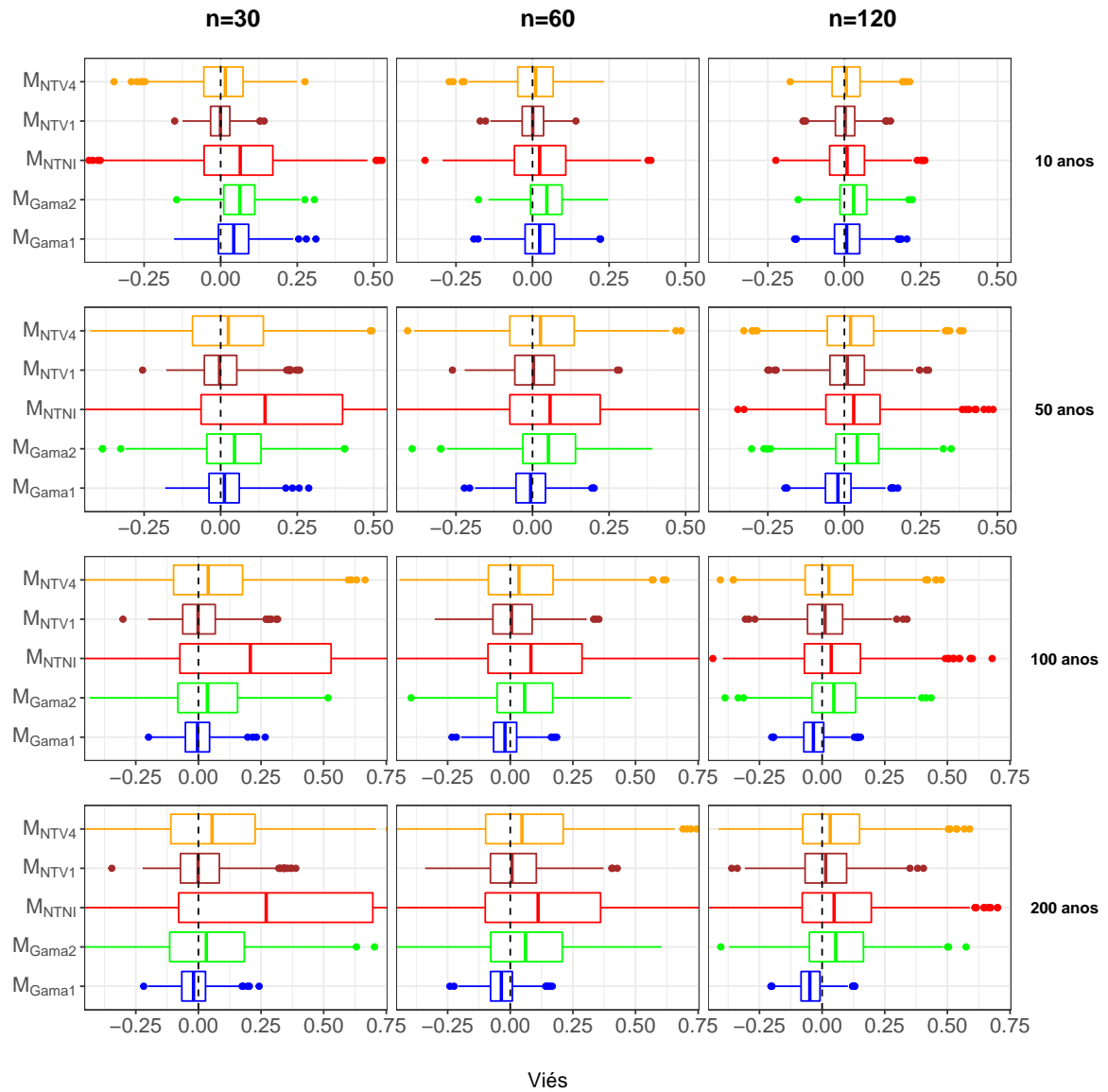
Figura 10 – Boxplot do viés de níveis de retorno para os tempos de 10, 50, 100 e 200 anos para cada modelo (dependente da escolha da priori) no Cenário-base (c) com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 34, \sigma = 2, \xi = -0,4$ .



Fonte: Do autor.

Com períodos de retorno mais curtos, os vieses de predição tendem a estar mais perto de zero em todos os cenários independente do tipo de priori do modelo, este resultado corrobora os encontrados por [Butturi-Gomes, Beijo e Avelar \(2019\)](#), os quais estudaram frequência de veranicos por meio da distribuição GEV e uma abordagem Bayesiana com diferentes estruturas *a priori* informativas e não informativas. E quanto mais longo o período de retorno, mais assimétrica é a distribuição do viés.

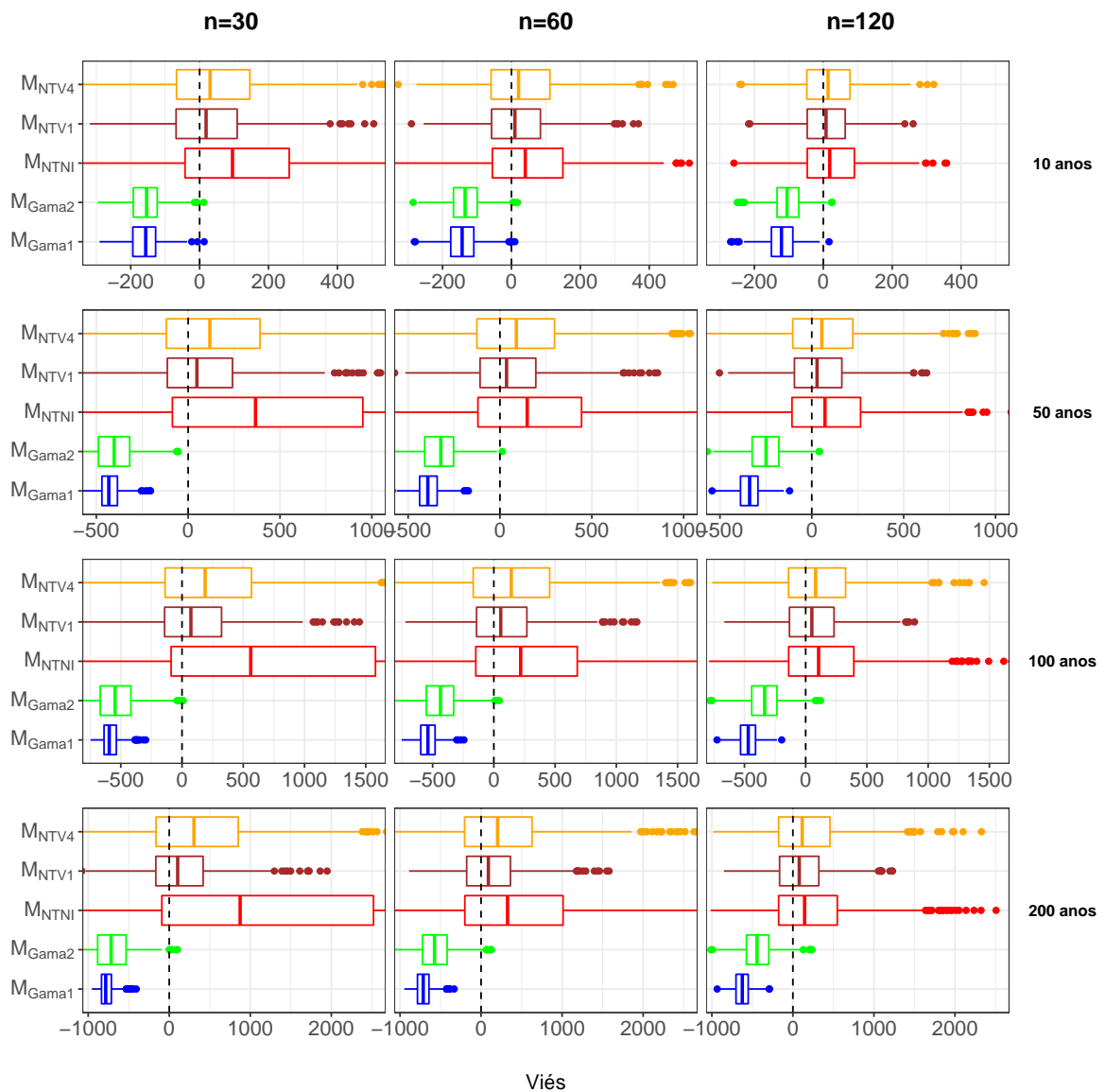
Figura 11 – Boxplot do viés de níveis de retorno para os tempos de 10, 50, 100 e 200 anos para cada modelo (dependente da escolha da priori) no Cenário-base (**d**) com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 6,78$ ,  $\sigma = 0,561$ ,  $\xi = -0,2$ .



Fonte: Do autor.

O modelo com priori normal não informativo apresentou menor precisão na predição dos níveis de retorno, assim como ocorreu com as predições dos parâmetros.

Figura 12 – Boxplot do viés de níveis de retorno para os tempos de 10, 50, 100 e 200 anos para cada modelo (dependente da escolha da priori) no Cenário-base (e) com amostras de tamanhos 30, 60 e 120 da distribuição GEV com parâmetros:  $\mu = 1200, \sigma = 340, \xi = 0,1$ .

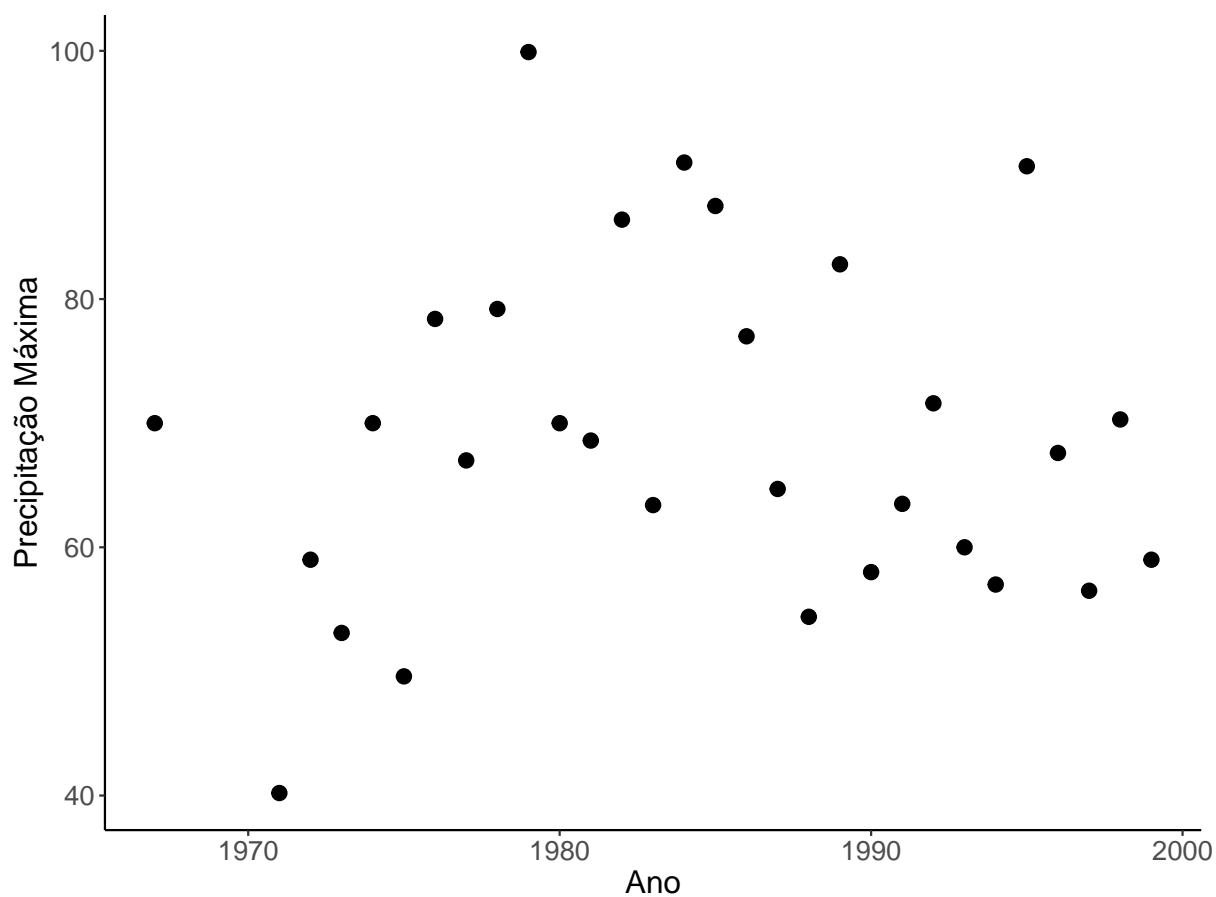


Fonte: Do autor.

## 4.2 ESTUDO DE PRECIPITAÇÕES MÁXIMAS DE SÃO JOÃO DA BOA VISTA

Na Figura 13 são apresentadas as precipitações pluviárias diárias máximas, por ano, de São João da Boa Vista, as quais estão entre 40 mm e 100 mm.

Figura 13 – Precipitações pluviárias diárias máximas, por ano, de São João da Boa Vista - SP de 1971 a 2000.



Fonte: Do autor.

Algumas medidas descritivas da série de precipitação máxima anual são apresentadas na Tabela 7, assim como os resultados dos testes de independência (Ljung-Box) e tendência (Mann-Kendall).

Tabela 7 – Medidas descritivas da série precipitações máximas de São João da Boa Vista - SP e valores-p dos testes de independência (Ljung-Box) e tendência (Mann-Kendall) da série.

Mínimo	Mediana	Média	Máximo	CV (%)	Ljung-Box	Mann-Kendall
40,20	68,10	68,88	99,90	19,94%	0,722	1,00

Fonte: Do autor.

O resultado do teste de Ljung-Box indica que as observações de precipitação máxima são independentes (valor-p = 0,722) (TABELA 7). Portanto, a função de verossimilhança da distribuição GEV pode ser utilizada na análise. Além disso, não há evidência de tendência monótona na série de precipitação pelo teste Mann-Kendall (valor-p = 1,00), isto é, não há indícios de não estacionariedade das observações de precipitação máxima de São João da Boa Vista - SP. Os resultados dos critérios de convergência das cadeias MCMC, bem como as estimativas dos parâmetros com os respectivos intervalos  $HPD_{95\%}$  são apresentados na Tabela 8.

A partir dos resultados da Tabela 8, conclui-se que há evidências de convergência das cadeias usando métodos de MCMC. Por exemplo, pelo critério de Raftery e Lewis, considerando o fator de dependência (RL), nota-se que os valores foram próximos de 1 indicando independência entre iterações. Além disso, pelo critério de Geweke, os valores absolutos encontrados foram menores que o 1,96 apontando que não há sinais de não convergência. Finalmente, analisando-se o critério de Heidelberger e Welch, constatou-se que a série é estacionária (valor-p > 0,05).

Tendo em vista que não há evidências contra a convergência das cadeias, pode-se prosseguir com a análise das estimativas *a posteriori*. Analisando os resultados da Tabela 8, as estimativas de média *a posteriori* do parâmetro de forma ( $\xi$ ) dos modelos com priori Normal Trivariada foram negativas, assim como o modelo  $M_{\text{Gamma}2}$ .

Tabela 8 – Estimativas de média e desvio padrão (DP) a posteriori dos parâmetros da distribuição GEV com diferentes distribuições *a priori*, com os respectivos intervalos  $HPD_{95\%}$ , assim como os critérios de convergência das cadeias MCMC: estatística  $|Z|$  de Geweke; fator de dependência de Raftery e Lewis (RL); valor-p do critério de Heidelberger e Welch (HW).

Modelo	Par.	Estim.	DP	HPD		$ Z $	RL	HW
				LI	LS			
$M_{Gama1-L}$	$\mu$	63,81	2,81	58,39	69,07	0,31	4,26	0,69
	$\sigma$	14,99	2,12	11,00	19,15	0,29	4,02	0,78
	$\xi$	0,01	0,11	-0,20	0,20	0,08	3,50	0,35
$M_{Gama2-L}$	$\mu$	64,06	63,95	58,53	69,49	0,78	1,46	0,60
	$\sigma$	14,79	14,57	10,76	18,99	0,37	1,78	0,27
	$\xi$	-0,05	-0,04	-0,32	0,21	0,25	1,59	0,64
$M_{NTV1-L}$	$\mu$	65,56	2,25	61,10	69,63	0,73	4,62	0,69
	$\sigma$	13,78	1,95	10,11	17,59	0,38	3,73	0,79
	$\xi$	-0,08	0,13	-0,31	0,20	0,85	4,26	0,57
$M_{NTV4-L}$	$\mu$	64,42	2,51	59,69	69,59	-0,25	4,50	0,28
	$\sigma$	13,59	1,97	10,25	17,57	1,40	3,80	0,63
	$\xi$	-0,16	0,13	-0,39	0,12	-1,35	4,02	0,12
$M_{Gama1-S}$	$\mu$	66,05	2,89	60,40	71,81	0,59	2,17	0,66
	$\sigma$	18,37	2,17	14,53	22,76	0,11	2,80	0,29
	$\xi$	0,17	0,05	0,06	0,27	-0,20	2,39	0,27
$M_{Gama2-S}$	$\mu$	65,71	2,99	59,91	71,56	0,27	1,49	0,57
	$\sigma$	18,27	2,13	14,39	22,53	0,16	1,33	0,56
	$\xi$	0,17	0,12	-0,07	0,38	0,23	1,36	0,66
$M_{NTV1-S}$	$\mu$	66,37	1,90	62,92	70,06	-1,74	2,05	0,13
	$\sigma$	14,13	2,04	10,53	18,16	-1,17	2,18	0,28
	$\xi$	-0,02	0,07	-0,16	0,13	-0,75	1,71	0,73
$M_{NTV4-S}$	$\mu$	64,43	2,34	59,93	68,99	-0,24	1,57	0,86
	$\sigma$	13,33	1,91	9,98	17,14	0,77	1,44	0,87
	$\xi$	-0,10	0,11	-0,31	0,12	-0,18	1,54	0,42
$M_{NTNI}$	$\mu$	63,44	2,71	58,47	69,13	-1,04	4,15	0,80
	$\sigma$	13,63	2,00	9,92	17,50	-1,82	3,99	0,15
	$\xi$	-0,18	0,13	-0,41	0,07	1,90	4,49	0,81

Fonte: Do autor.

Quanto aos desvios padrão das estimativas a posteriori de  $\mu$  e  $\sigma$ , pode-se afirmar que foram baixos em relação a média em todas as estruturas *a priori*. Além disso, nos modelos com priors normais trivariadas, os desvios padrão *a posteriori* de  $\mu$  são um pouco mais baixos do que com *a priori* não informativa. Esta redução no desvio padrão pode refletir a diminuição da incerteza devido ao uso de informações adicionais. Em um estudo de ventos máximos de Cape South, África do Sul, [Diriba et al. \(2017\)](#) obtiveram resultado semelhante através de abordagem bayesiana com priori informativa, reduzindo o desvio padrão *a posteriori* do parâmetro  $\mu$  com a utilização de informações *a priori* de regiões próximas ao local de interesse.

Na Tabela 9 são apresentados a média, desvio padrão e intervalos  $HPD_{95\%}$  das estimativas *a posteriori* dos níveis de retorno para os tempos de 5, 10, 15 e 20 anos do modelo GEV considerando todas as estruturas *a priori*, assim como os resultados dos critérios de convergência das cadeias MCMC.

Tendo em vista os resultados dos critérios de Geweke, Raftery e Lewis e Heidelberger e Welch (TABELA 9), não há evidências contra a convergência das cadeias *a posteriori* dos níveis de retorno. Pode-se interpretar os níveis de retorno da seguinte maneira: considerando, por exemplo, o tempo de retorno de 10 anos e o modelo  $M_{NTV1}$  (TABELA 9), espera-se que ocorra, em um tempo médio de 10 anos, uma precipitação maior ou igual a 114,1 mm, com um intervalo de alta densidade *a posteriori* de 95% entre 92,4 mm e 146,06 mm.

Observando-se os resultados dos níveis de retorno na Tabela 9, é possível notar que o modelo  $M_{NTNI}$  proporcionaram níveis de retorno mais baixos (média *a posteriori*), e aumentam-se, mais lentamente, com tempos de retorno mais altos. Isto está diretamente relacionado com a estimativa negativa do parâmetro  $\xi$  (Tabela 8).

Tabela 9 – Estimativas de médias *a posteriori* para níveis de retorno (NR) correspondentes aos tempos de retorno (TR) de 5, 10, 15 e 20 anos e seus respectivos desvios padrão (DP) e intervalos  $HPD_{95\%}$  sob diferentes estruturas *a priori*, assim como os critérios de convergência das cadeias MCMC: estatística  $|Z|$  de Geweke; fator de dependência de Raftery e Lewis (RL); valores-p do critério de Heidelberger e Welch (HW).

Modelo	TR	NR	DP	HPD		$ Z $	RL	HW
				LI	LS			
$M_{Gama1-L}$	5	86,52	4,71	77,54	95,65	0,43	2,29	0,74
	10	98,17	6,62	85,55	111,02	0,44	2,17	0,73
	15	104,86	8,09	89,88	120,75	0,44	2,29	0,68
	20	109,61	9,31	92,09	127,46	0,43	2,29	0,64
$M_{Gama2-L}$	5	95,88	94,95	84,22	109,97	0,90	1,26	0,12
	10	118,47	115,91	94,00	149,20	0,75	1,22	0,18
	15	128,13	124,50	96,58	170,02	0,71	1,30	0,22
	20	137,92	132,58	97,30	191,88	0,67	1,28	0,25
$M_{NTV1-L}$	5	94,45	6,91	82,85	107,96	1,13	2,29	0,63
	10	114,10	16,31	92,40	146,06	1,11	1,91	0,64
	15	122,32	22,09	95,04	165,49	1,10	1,88	0,63
	20	130,57	29,12	96,66	186,42	1,10	1,91	0,62
$M_{NTV4-L}$	5	90,36	5,29	81,33	100,62	0,42	1,98	0,98
	10	105,24	11,38	90,72	129,07	-0,20	1,98	0,50
	15	110,87	14,99	92,66	141,71	-0,34	1,98	0,41
	20	116,19	19,19	94,63	155,87	-0,43	2,09	0,37
$M_{Gama1-S}$	5	97,27	4,67	88,18	106,42	0,26	2,93	0,44
	10	116,07	5,87	105,15	128,10	0,16	1,84	0,49
	15	127,77	6,73	114,97	141,14	0,08	1,74	0,53
	20	136,48	7,46	122,58	151,72	0,02	1,55	0,57
$M_{Gama2-L}$	5	97,04	4,77	87,36	106,12	0,48	1,20	0,28
	10	116,30	7,57	100,78	130,83	0,52	1,18	0,29
	15	128,47	10,23	109,04	149,00	0,57	1,16	0,33
	20	137,64	12,65	112,44	161,89	0,63	1,18	0,37
$M_{NTV1-S}$	5	83,04	4,05	75,64	90,96	0,39	1,20	0,98
	10	91,47	5,80	81,81	103,14	0,44	1,22	0,97
	15	96,03	7,08	84,70	110,35	0,44	1,16	0,95
	20	99,13	8,10	86,93	115,97	0,44	1,08	0,92
$M_{NTV4-S}$	5	87,32	4,55	78,93	96,17	-1,49	1,81	0,18
	10	97,67	6,68	85,92	111,02	-1,48	1,70	0,20
	15	103,48	8,11	89,21	119,18	-1,46	1,51	0,21
	20	107,55	9,21	91,59	125,43	-1,46	1,53	0,22
$M_{NTNI}$	5	81,41	3,58	74,33	88,22	-1,06	3,29	0,45
	10	88,88	4,75	80,58	98,24	-0,27	3,19	0,27
	15	92,74	5,73	83,98	104,77	0,14	2,09	0,25
	20	95,32	6,55	85,09	108,33	0,37	2,05	0,26

Fonte: Do autor.



Para analisar o desempenho do modelo GEV, com diferente estruturas de priori na predição de precipitações máximas considerando as diferentes estruturas *a priori* são apresentados na Tabela 10: critério de informação do desvio (DIC), acurácia, amplitude média intervalar (AMI) e erro médio de predição (EMP, em %).

Tabela 10 – DIC, acurácia, amplitude média intervalar (AMI) e Erro médio de predição (EMP) referentes a predições do modelo GEV, ajustado à série de precipitações máximas anuais de São João da Boa Vista - SP de 1971 a 2000, com estimativas via abordagem Bayesianas considerando diferentes distribuições *a priori*.

Modelo	DIC	Acurácia	AMI	EMP(%)
$M_{NTV4-S}$	245,90	4	15,20	2,73
$M_{NTV1-S}$	248,02	4	17,80	6,50
$M_{Gama1-L}$	248,50	4	17,91	7,68
$M_{NTV4-L}$	245,98	4	19,29	12,24
$M_{NTNI}$	246,26	3	14,38	4,55
$M_{NTV1-L}$	247,05	3	25,11	22,43
$M_{Gama2-L}$	248,56	2	29,26	27,40
$M_{Gama1-S}$	254,58	1	17,31	26,76
$M_{Gama2-S}$	255,10	1	19,39	27,39

Fonte: Do autor.

Conforme os resultados apresentados na Tabela 10, tem-se que os modelos  $M_{Gama1-S}$  e  $M_{Gama2-S}$  apresentaram DIC maior que os outros modelos com uma diferença maior que 5 unidades, logo pode-se descartar estes modelos. Quanto à acurácia,  $M_{Gama1-L}$ ,  $M_{NTV4-L}$ ,  $M_{NTV1-S}$  e  $M_{NTV4-S}$  acertaram 4 predições. Dentre estes,  $M_{NTV4-S}$  apresentou menor amplitude intervalar e menor erro médio de predição. Assim, pode-se concluir que as estimativas mais acuradas e precisas foram obtidas com o modelo  $M_{NTV4-S}$ .

Os resultados encontrados aqui corroboram os encontrados nos cenários de simulação, quando se considera a precisão e acurácia dos modelos com prioris normais informativos.

A informação *a priori* de Silvianópolis produziu, ao mesmo tempo, os melhores resultados considerando-se a utilização de prioris normais, e os piores resultados quando utilizou-se a Gama. Assim, deve-se considerar não somente a distribuição ou estrutura *a priori* mais adequada para o problema específico, como também a fonte ou local de onde se obtém a informação.

A melhora na acurácia dos resultados de previsão de precipitação máxima, quando se utiliza a distribuição GEV com uma abordagem Bayesiana, foi observada por [Beijo e Avelar \(2011\)](#). Neste estudo, os autores também utilizaram uma distribuição *a priori* (Gumbel) baseada

em quantis extremos.

[Martins \*et al.\* \(2018\)](#) realizaram previsões de precipitações máximas de Silvianópolis via abordagem bayesiana, com erros de previsão em torno 18% e 27%. Portanto, os erros médios de previsão de precipitação máxima pelo modelo GEV encontrados nesse trabalho são coerentes que os reportados na literatura.

## 5 CONCLUSÕES

Foi investigado o comportamento das estruturas *a priori*, em diferentes cenários, isto é, com diferentes valores de parâmetros da distribuição GEV e diferentes tamanhos amostrais, através de simulações, e também analisando-se o comportamento dessas estruturas considerando a predição de quantis extremos. A abordagem Bayesiana mostrou-se eficiente na análise de extremos máximos provenientes de simulação, tendo em vista que os vieses foram próximos de 0 (zero).

Esta abordagem também foi considerada adequada para análise precipitações máximas de São João da Vista - SP, cujas séries de máximos foram consideradas independentes e sem indícios de tendência. As informações *a priori* de Lavras - MG e Silvianópolis permitiram uma construção adequada das estruturas *a priori*. Houve diferenças substanciais entre as predições da distribuição GEV com as estruturas *a priori* utilizadas. As informações *a priori* juntamente com a estrutura *a priori* normal trivariada, com matriz de variância e covariância multiplicada por 4, proporcionou estimativas mais precisas e acuradas dos níveis de retorno, corroborando os resultados obtidos na análise dos dados simulados para cenários-base representados por parâmetros que foram provenientes de estudos de precipitação máxima.

Uma limitação da abordagem adotada neste trabalho foi que a única fonte de informação *a priori* foram os registros históricos de precipitação de outra localidade. Poderia-se, por exemplo, utilizar informações *a priori* a partir de um componente espacial (como topografia) para minimizar eventuais diferenças entre os comportamentos de precipitação máxima das estações meteorológicas.

Cabe ressaltar que, neste trabalho, a abordagem Bayesiana foi aplicada na análise de extremos considerando-se apenas uma parte das estruturas *a priori* disponíveis para os parâmetros da distribuição GEV. Assim como o que foi constatado por [Coles e Powell \(1996\)](#), percebe-se a possibilidade de utilizar a abordagem Bayesiana em análise de valores extremos com uma sofisticação, ou complexidade, maior do que a que foi considerada aqui. Pode-se, por exemplo, considerar como priori um modelo multivariado assimétrico.

## REFERÊNCIAS

- ABREU, M. C. *et al.* Critérios para escolha de distribuições de probabilidade em estudos de eventos extremos de precipitação. **Revista Brasileira de Meteorologia**, São Paulo, SP, v. 33, n. 4, p. 601–613, 2018. Disponível em: <https://doi.org/10.1590/0102-7786334004>. Acesso em: 20 nov. 2020.
- AGUIRRE, A. F. L.; NOGUEIRA, D. A.; BEIJO, L. A. Análise da temperatura máxima de piracicaba (sp) via distribuição gev não estacionária: uma abordagem bayesiana. **Revista Brasileira de Climatologia**, Curitiba, PR, v. 27, 2020.
- ALVARES, C. A. *et al.* Köppen's climate classification map for brazil. **Meteorologische Zeitschrift**, Stuttgart, DE, v. 22, n. 6, p. 711–728, 2013. Disponível em: <http://dx.doi.org/10.1127/0941-2948/2013/0507>. Acesso em: 5 ago. 2020.
- BAROSSO-FILHO, M.; ACHCAR, J. A.; SOUZA, R. M. d. Modelos de volatilidade estocástica em séries financeiras: uma aplicação para o ibovespa. **Economia Aplicada**, Ribeirão Preto, SP, v. 14, n. 1, p. 25–40, 2010. Disponível em: <https://doi.org/10.1590/S1413-80502010000100002>. Acesso em: 11 jul. 2020.
- BAUTISTA, E. A.; ZOCCHI, S. S.; ANGELOCCI, L. R. A distribuição generalizada de valores extremos aplicada ao ajuste dos dados de velocidade máxima do vento em piracicaba, são paulo, brasil. **Revista de Matemática e Estatística**, São Paulo, SP, v. 22, n. 1, p. 95–111, 2004. Disponível em: [http://jaguar.fcav.unesp.br/RME/fasciculos/v22/v22\\_n1/A7\\_SSZocchi.pdf](http://jaguar.fcav.unesp.br/RME/fasciculos/v22/v22_n1/A7_SSZocchi.pdf). Acesso em: 25 dez. 2020.
- BEIJO, L. A.; AVELAR, F. G. Distribuição generalizada de valores extremos no estudo de dados climáticos uma breve revisão e aplicação. **Revista da Estatística da Universidade Federal de Ouro Preto**, Ouro Preto, MG, v. 1, n. 1, p. p. 10–16, 2011. Disponível em: <https://periodicos.ufop.br/rest/article/view/3043/2314>. Acesso em: 10 abr. 2021.
- BEIJO, L. A.; VIVANCO, M. J. F.; MUNIZ, J. A. Análise bayesiana no estudo do tempo de retorno das precipitações pluviométricas máximas em jaboticabal (sp). **Ciência e Agrotecnologia**, Lavras, MG, v. 33, n. 1, p. 261–270, 2009. Disponível em: <https://doi.org/10.1590/S1413-70542009000100036>. Acesso em: 8 maio 2021.
- BUTTURI-GOMES, D.; BEIJO, L. A.; AVELAR, F. G. On modeling the maximum duration of dry spells:: a simulation study under a bayesian approach. **Theoretical and Applied Climatology**, Springer, [S.l.], v. 137, n. 1, p. 1337–1346, 2019.
- COLES, S.; TAWN, J. Bayesian modelling of extreme surges on the uk east coast. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, London, GB, v. 363, n. 1831, p. 1387–1406, 2005. Disponível em: <https://doi.org/10.1098/rsta.2005.1574>. Acesso em: 10 jul. 2020.
- COLES, S. G. **An Introduction to Statistical Modeling of Extreme Values**. Berlin, DE: Springer, 2001. Disponível em: <https://link.springer.com/content/pdf/10.1007/>

[978-1-4471-3675-0.pdf](#). Acesso em: 16 ago. 2020.

COLES, S. G.; POWELL, E. A. Bayesian methods in extreme value modelling: a review and new developments. **International Statistical Review/Revue Internationale de Statistique**, Voorburg, NL, p. 119–136, 1996. Disponível em: [https://www.jstor.org/stable/1403426?seq=1#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/1403426?seq=1#metadata_info_tab_contents). Acesso em: 21 out. 2020.

COLES, S. G.; TAWN, J. A. A Bayesian analysis of extreme rainfall data. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, London, GB, v. 45, n. 4, p. 463–478, 1996. Disponível em: <https://doi.org/10.2307/2986068>. Acesso em: 5 jan. 2021.

CROWDER, M. Bayesian priors based on a parameter transformation using the distribution function. **Annals of the Institute of Statistical Mathematics**, [S.l.], v. 44, n. 3, p. 405–416, 1992. Disponível em: <https://link.springer.com/content/pdf/10.1007/BF00050695.pdf>. Acesso em: 12 fev. 2021.

DEY, D. K.; YAN, J. **Extreme Value Modeling and Risk Analysis Methods and Applications**. New York: CRC PRESS, 2016. Disponível em: <https://www.routledge.com/Extreme-Value-Modeling-and-Risk-Analysis-Methods-and-Applications/Dey-Yan/p/book/9780367737399>. Acesso em: 24 fev. 2021.

DIRIBA, T. A. *et al.* Analysis of extreme rainfall at east london, south africa. In: SOUTH AFRICAN STATISTICAL ASSOCIATION (SASA). **Annual Proceedings of the South African Statistical Association Conference**. [S.l.], 2014. v. 2014, n. con-1, p. 25–32.

\_\_\_\_\_. Bayesian modelling of extreme wind speed at cape town, south africa. **Environmental and Ecological Statistics**, [S.l.], v. 24, n. 2, p. 243–267, 2017. Disponível em: <https://doi.org/10.1007/s10651-017-0369-z>. Acesso em: 2 ago. 2020.

ERCELEBI, S. G.; TOROS, H. Extreme value analysis of istanbul air pollution data. **CLEAN–Soil, Air, Water**, Weinheim, DE, v. 37, n. 2, p. 122–131, 2009. Disponível em: <https://doi.org/10.1002/clen.200800041>. Acesso em: 25 jul. 2020.

FEINBERG, S. E.; VAN DER LINDEN, W. J. **Statistics for social and behavioral sciences**. New York: Springer, 2005.

GAMERMAN, D. Sampling from the posterior distribution in generalized linear mixed models. **Statistics and Computing**, [S.l.], v. 7, n. 1, p. 57–68, 1997.

GAMERMAN, D.; LOPES, H. F. **Markov chain Monte Carlo: stochastic simulation for Bayesian inference**. Boca Raton, USA: CRC Press, 2006. Disponível em: <https://www.routledge.com/Markov-Chain-Monte-Carlo-Stochastic-Simulation-for-Bayesian-Inference/Gamerman-Lopes/p/book/9781584885870>. Acesso em: 12 jan. 2021.

GELMAN, A. *et al.* **Bayesian data analysis**. 3. ed. Boca Raton, USA: Chapman and Hall/CRC, 2013.

GEWEKE, J. F. **Evaluating the accuracy of sampling-based approaches to the calculation**

**of posterior moments.** Minneapolis, USA: Federal Reserve Bank of Minneapolis, 1991. Disponível em: <https://doi.org/10.21034/sr.148>. Acesso em: 12 jan. 2021.

GOMES, M. I.; GUILLOU, A. Extreme value theory and statistics of univariate extremes: a review. **International Statistical Review**, Voorburg, NL, v. 83, n. 2, p. 263–292, 2015. Disponível em: <https://doi.org/10.1111/insr.12058>. Acesso em: 9 jul. 2020.

HAAN, L. de; FERREIRA, A. **Extreme value theory: an introduction.** New York, USA: Springer Science & Business Media, 2006. Disponível em: <https://doi.org/10.1007/0-387-34471-3>. Acesso em: 17 mar. 2020.

HEIDELBERGER, E.; WELCH, D. Simulation run length control in the presence of an initial transient. **Operations Research**, Catonsville, USA, v. 31, n. 6, p. 1109–1144, 1983. Disponível em: <https://www.jstor.org/stable/170841>. Acesso em: 18 set. 2020.

IBGE. **Índice de Nomes Geográficos. Instituto Brasileiro de Geografia e Estatística.** [s.n.], 2011. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv56282.pdf>. Acesso em: 14 fev. 2021.

JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quartely Journal of the Royal Meteorological Society**, [S.l.], v. 81, p. 158–171, April 1955. Disponível em: <https://doi.org/10.1002/qj.49708134804>. Acesso em: 11 out. 2020.

KANG, D.; KO, K.; HUH, J. Determination of extreme wind values using the gev distribution. **Energy**, Oxford, GB, v. 86, n. 15, p. 51–58, 2015. Disponível em: <https://doi.org/10.1016/j.energy.2015.03.126>. Acesso em: 3 jun. 2020.

KOTZ, S.; NADARAJAH, S. **Extreme value distributions: theory and applications.** Singapura: World Scientific, 2000. Disponível em: <https://doi.org/10.1142/p191>. Acesso em: 17 ago. 2020.

KOUSAR, S. *et al.* Some best-fit probability distributions for at-site flood frequency analysis of the ume river. **Journal of Flood Risk Management**, Wiley Online Library, [S. l.], v. 13, n. 3, p. e12640, 2020.

KUMAR, K.; SWAIN, D. K.; VINAY, T. Characterization of rainfall through probability distributions for yadgir district in karnataka, india. **International Journal of Current Microbiology and Applied Sciences**, Tamilnadu, India, v. 8, n. 4, p. 1099–1113, 2019. Disponível em: <https://doi.org/10.20546/ijcmas.2019.804.127>. Acesso em: 25 jan. 2021.

LISKA, G. R. *et al.* Estimativas de velocidade máxima de vento em piracicaba-sp via séries temporais e teoria de valores extremos. **Revista Brasileira de Biometria, São Paulo**, v. 31, n. 2, p. 295–309, 2013. Disponível em: [http://jaguar.fcav.unesp.br/RME/fasciculos/v31/v31\\_n2/A8\\_GilbertoLiska.pdf](http://jaguar.fcav.unesp.br/RME/fasciculos/v31/v31_n2/A8_GilbertoLiska.pdf). Acesso em: 16 jul. 2020.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, Oxford, GB, v. 65, n. 2, p. 297–303, 1978. Disponível em: <https://doi.org/10.2307/2344000>.

[//doi.org/10.1093/biomet/65.2.297](https://doi.org/10.1093/biomet/65.2.297). Acesso em: 4 maio 2020.

MARTINS, T. B. *et al.* Predição da precipitação máxima no município de silvianópolis-mg: Abordagens clássica e bayesiana. **IRRIGA**, Botucatu, SP, v. 23, n. 3, p. 467–479, 2018. Disponível em: <https://doi.org/10.15809/irriga.2018v23n3p467-479>. Acesso em: 10 mar. 2021.

MARUYAMA, F. *et al.* Analyzing the annual maximum magnitude of earthquakes in japan by extreme value theory. **Open Journal of Applied Sciences**, Scientific Research Publishing, [S. l.], v. 10, n. 12, p. 817, 2020.

MASSERAN, N.; SAFARI, M. A. M. S. Intensity–duration–frequency approach for risk assessment of air pollution events. **Journal of environmental management**, Port Harcourt, NG, v. 264, 2020. Disponível em: <https://doi.org/10.1016/j.jenvman.2020.110429>. Acesso em: 21 fev. 2021.

MCLEOD, A. **Kendall: Kendall rank correlation and Mann-Kendall trend test**. [S.l.], 2011. R package version 2.2. Disponível em: <https://CRAN.R-project.org/package=Kendall>. Acesso em: 26 set. 2020.

MORALES, J. C. C.; CAUSIL, C. J. B. **Introducción a la estadística Bayesiana**. Medellín, CO: INSTITUTO TECNOLÓGICO METROPOLITANO–ITM, 2019. Disponível em: <https://dx.doi.org/10.22430/9789585414242>. Acesso em: 13 out. 2020.

NAGHETTINI, M.; PINTO, E. J. **Hidrologia Estatística**. Belo Horizonte, MG: CPRM-Serviço Geológico do Brasil, 2007.

NOGUEIRA, D. A.; SÁFADI, T.; FERREIRA, D. F. Avaliação de critérios de convergência univariados para o método de monte carlo via cadeias de markov. **Revista Brasileira de Estatística**, Rio de Janeiro, RJ, v. 65, n. 224, p. 59–88, 2004. Disponível em: [https://www.researchgate.net/publication/339149423\\_Avaliacao\\_de\\_criterios\\_de\\_convergencia\\_univariados\\_para\\_o\\_metodo\\_de\\_Monte\\_Carlo\\_via\\_cadeias\\_de\\_Markov](https://www.researchgate.net/publication/339149423_Avaliacao_de_criterios_de_convergencia_univariados_para_o_metodo_de_Monte_Carlo_via_cadeias_de_Markov). Acesso em: 8 jan. 2020.

PAULINO, D. C.; TURKMAN, M. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa, PT: FCG, 2003.

PLUMMER, M. *et al.* Coda: Convergence diagnosis and output analysis for mcmc. **R News**, [S. l.], v. 6, n. 1, p. 7–11, 2006. Disponível em: <https://journal.r-project.org/archive/>. Acesso em: 6 maio 2021.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. Disponível em: <https://www.R-project.org/>. Acesso em: 12 jan. 2021.

RAFTERY, A. E.; LEWIS, S. Comment: One long run with diagnostics: implementation strategies for markov chain monte carlo. **Statistical Science**, USA, v. 7, n. 4, p. 493–497, 1992. Disponível em: <https://doi.org/10.1214/ss/1177011143>. Acesso em: 2 out. 2020.

SPIEGELHALTER, D. J. *et al.* Bayesian measures of model complexity and fit. **Journal of**

**the royal statistical society: Series b (statistical methodology)**, London, GB, v. 64, n. 4, p. 583–639, 2002. Disponível em: <https://doi.org/10.1111/1467-9868.00353>. Acesso em: 7 jun. 2020.

STEPHENSON, A.; RIBATET, M. A user's guide to the evdbayes package (version 1.1). In: . [S.l.: s.n.], 2006.

\_\_\_\_\_. **evdbayes**: Bayesian analysis in extreme value theory. [S.l.], 2014. R package version 1.1-1. Disponível em: <https://CRAN.R-project.org/package=evdbayes>. Acesso em: 17 jun. 2020.

STEPHENSON, A. G. evd: Extreme value distributions. **R News**, [S. l.], v. 2, n. 2, p. 0, June 2002. Disponível em: <https://CRAN.R-project.org/doc/Rnews/>. Acesso em: 3 set. 2020.

TUCCI, C. E.; SILVEIRA, A. L. L. **Hidrologia**: Ciência e aplicação. Porto Alegre, RS: UFRGS/ABRH, 2004.

TURKMAN, M. A. A.; PAULINO, C. D.; MÜLLER, P. **Computational Bayesian Statistics**: An introduction. Cambridge, GB: Cambridge University Press, 2019. v. 11.

WU, Y. *et al.* Local and regional flood frequency analysis based on hierarchical Bayesian model: application to annual maximum streamflow for the huaihe river basi. **Hydrology and Earth System Sciences Discussions**, DE, v. 10, n. 5194, p. 2018–22, 2018. Disponível em: <https://doi.org/10.1016/j.wse.2019.12.001>. Acesso em: 8 ago. 2021.



## APÊNDICES

### APÊNDICE B - Rotinas para estimação dos parâmetros $\mu, \sigma, \xi$ da distribuição GEV

Todas as rotinas para realização da simulação e da análise do dados simulados e reais podem ser encontradas em

<https://github.com/matheusmsgit/dissertation.git>