

UNIVERSIDADE FEDERAL DE ALFENAS

ANDRÉIA DO CARMO DE OLIVEIRA

**DESEMPENHO DA MISTURA DE DISTRIBUIÇÕES A
PRIORI NA PREDIÇÃO DE VERANICOS DISCRETOS
MÁXIMOS**

ALFENAS/MG

2025

ANDRÉIA DO CARMO DE OLIVEIRA

**DESEMPENHO DA MISTURA DE DISTRIBUIÇÕES A
PRIORI NA PREDIÇÃO DE VERANICOS DISCRETOS
MÁXIMOS**

Dissertação apresentada ao Programa de Pós-Graduação 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: Prof. Dr. Davi Butturi-Gomes.

Coorientador: Prof. Dr. Luiz Alberto Beijo.

ALFENAS/MG

2025

Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central

Oliveira, Andréia do Carmo de.

Desempenho da mistura de distribuições a *priori* na predição de veranicos discretos máximos / Andréia do Carmo de Oliveira. - Alfenas, MG, 2024.

73 f. : il. -

Orientador(a): Davi Butturi-Gomes.

Dissertação (Mestrado em Estatística Aplicada e Biometria) -
Universidade Federal de Alfenas, Alfenas, MG, 2024.

Bibliografia.

1. Distribuição Generalizada de Valores Extremos. 2. Inferência Bayesiana. 3. Veranico. I. Butturi-Gomes, Davi, orient. II. Título.

ANDRÉIA DO CARMO DE OLIVEIRA

Desempenho da Mistura de Distribuições a *Priori* na predição de veranicos discretos máximos

O Presidente da banca examinadora abaixo assina a aprovação da 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: 20 de dezembro de 2024.

Prof. Dr. Davi Butturi-Gomes

Presidente da Banca Examinadora

Instituição: Universidade Federal de São João Del-Rei - UFSJ

Prof. Dr. Everton Batista da Rocha

Instituição: Universidade Federal de Goiás - UFG

Profa. Dr. Gilberto Rodrigues Liska

Instituição: Universidade Federal de São Carlos - UFSCar



Documento assinado eletronicamente por **Davi Butturi Gomes, Usuário Externo**, em 20/12/2024, às 11:33, 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, informando o código verificador **1418621** e o código CRC **9D1D9332**.

Dedico aos meus filhos Augusto e Lavínia, ao meu esposo Reginaldo, aos meus pais José Reginaldo e Ivonete, aos meus irmãos Letícia e Felipe e aos meus sobrinhos Miguel, Lorena e Yasmin.

AGRADECIMENTOS

A Deus pelo dom da vida e por sempre iluminar meu caminho, dando-me forças para vencer todos os obstáculos e tornar esse sonho possível.

Aos meus amados pais José Reinaldo e Ivonete, pelo amor incondicional, pela educação, pela confiança e que apesar de todas as dificuldades ao longo do caminho, não mediram esforços para me apoiar e me ver feliz.

Aos meus irmãos Letícia e Felipe, pela amizade, pelo apoio incondicional e pelas palavras de incentivo nos momentos em que mais precisei.

Aos meus filhos Augusto e Lavínia e meus sobrinhos Miguel, Lorena e Yasmin, por encherem minha vida de alegria, amor e propósito. Cada conquista é também de vocês, e espero que este trabalho inspire a busca pelos seus próprios sonhos. Amo vocês incondicionalmente.

Ao meu amigo Edson Aparecido da Silva Júnior pelo companheirismo e amizade, tanto nas horas de estudo quanto nas horas de lazer. Serei eternamente grata pelos momentos que passamos juntos.

Aos meus amigos, cuja presença e apoio foram fundamentais ao longo desta jornada. Cada conversa, palavra de encorajamento e gesto de amizade foi essencial para que eu conseguisse superar os desafios que encontrei. Em especial, Amanda Melo, Ronaldo Lopes, Daniela, Bruna, Samanta, Paloma, Laís, Tamyres, Simone e Dayani.

Agradeço aos meus companheiros de mestrado Thales, Karoline e Mariana pela amizade e companheirismo no decorrer desta jornada.

Ao meu orientador Prof. Dr. Davi Butturi-Gomes, pela paciência, ensinamentos, orientação, força e amizade durante o desenvolvimento do trabalho.

Ao meu coorientador Prof. Dr. Luiz Alberto Beijo, pelo exemplo de pessoa, pelo profissionalismo, pela inspiração, por todo carinho, incentivo, confiança e conhecimento compartilhado. Com certeza, seu apoio e orientação foram fundamentais para a finalização do trabalho.

À secretária e amiga Martha Priscila Azevedo deixo minha gratidão especial. Sua dedicação e organização foram indispensáveis para o andamento de cada etapa, e sua gentileza e disponibilidade fizeram toda a diferença em momentos de dúvida e desafio. Agradeço por sempre estar pronta para ajudar e pelo apoio constante que me motivou a seguir em frente.

Aos professores e funcionários do programa de Pós-Graduação em Estatística Aplicada e Biometria pela disponibilidade e contribuição para a minha formação. Em especial, ao Prof.

Dr. Denismar Alves Nogueira.

Agradeço aos professores Dr. Gilberto Rodrigues Liska e Dr. Everton Batista da Rocha, membros das bancas de exame de qualificação e defesa de mestrado, pela valiosa participação, pelas sugestões enriquecedoras e pelas contribuições que foram essenciais para o desenvolvimento deste estudo.

Um agradecimento especial eu reservo, ao meu esposo Reginaldo Leandro da Silva, cuja paciência, compreensão e apoio incondicional tornaram esta jornada possível. Agradeço por estar ao meu lado em cada momento, pelo encorajamento constante e por compreender as horas dedicadas a este trabalho. Sua presença foi minha fonte de força e motivação, e sou profundamente grata por compartilhar essa conquista com você.

A todos que de alguma forma foram responsáveis por essa vitória tão esperada, meus sinceros agradecimentos.

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

“O êxito da vida não se mede pelo caminho que você conquistou, mas sim pelas dificuldades que superou no caminho.”

(Washington, 2013)

RESUMO

A ocorrência de períodos consecutivos de dias sem chuva durante a estação chuvosa (veranicos), pode comprometer as atividades do setor agrícola além de ocasionar problemas no abastecimento de água. Neste sentido, a modelagem de veranicos máximos é de grande importância para o planejamento de ações dos órgãos públicos, visando minimizar os impactos causados por eles. Dessa forma, o objetivo deste trabalho é verificar o desempenho da mistura de distribuições informativas *a priori* na predição de veranicos máximos no município alvo Brejinho - PE empregando a distribuição Generalizada de Valores Extremos (GEV) via Inferência Bayesiana. Foram utilizados dados de 12 municípios pertencentes a bacia do Rio Pajeú, localizada no estado de Pernambuco, para gerar informações e para compor a mistura de distribuições informativas *a priori*. Para obtenção dos valores *a posteriori*, utilizou-se o método Monte Carlo via Cadeia de Markov, por meio do *software* R. Foram comparadas sete estruturas de distribuição *a priori* para o ajuste da distribuição Generalizada de Valores Extremos (GEV) aos conjuntos de dados de veranico máximo anual. Com o objetivo de analisar qual o modelo mais eficiente para a previsão de veranico máximo anual para o município de Brejinho - PE, foram verificados o DIC (Critério de Informação da Deviance), a acurácia, o erro médio de predição e a amplitude média para os tempos de retorno de 2, 4, 6 e 8 anos. Foi observado que a estrutura com mistura de *prioris* proporcionou aumento na precisão e acurácia nas predições de veranicos máximos para o município de Brejinho - PE. A partir desses resultados, foram feitas as predições de veranicos máximos em Brejinho - PE, para os tempos de retorno de 5, 10, 20, 30, 50, 100 e 200 anos, que podem ajudar no planejamento, possibilitando evitar catástrofes, minimizar perdas agrícolas e adotar medidas preventivas mais eficazes.

Palavras-chave: distribuição Generalizada de Valores Extremos; inferência bayesiana; veranico.

ABSTRACT

The occurrence of consecutive dry days during the rainy season (dry spells) can compromise agricultural activities and cause water supply issues. In this context, modeling maximum dry spells is crucial for public agencies' planning efforts to mitigate their impacts. Thus, this study aims to evaluate the performance of mixtures of informative prior distributions in predicting maximum dry spells in the target municipality of Brejinho, Pernambuco, using the Generalized Extreme Value (GEV) distribution via Bayesian inference. Data from 12 municipalities within the Pajeú River basin, located in Pernambuco, were used to generate information and compose the mixture of informative prior distributions. The Markov Chain Monte Carlo (MCMC) method, implemented in the R software, was employed to obtain the posterior marginal distributions. Seven prior distribution structures were compared to adjust the GEV distribution to the annual maximum dry spell datasets. To identify the most efficient model for predicting annual maximum dry spells in Brejinho, criteria such as the Deviance Information Criterion (DIC), accuracy, mean prediction error, and average range for return periods of 2, 4, 6, and 8 years were analyzed. Results showed that the prior mixture structure improved prediction precision and accuracy for maximum dry spells in Brejinho. Based on these findings, predictions for maximum dry spells in Brejinho were made for return periods of 5, 10, 20, 30, 50, 100, and 200 years, offering valuable insights for planning. These predictions can help prevent disasters, minimize agricultural losses, and implement more effective preventive measures.

Keywords: Generalized Extreme Value distribution (GEV); bayesian inference; dry spell.

LISTA DE TABELAS

Tabela 1 –	Localização geográfica e tamanho das séries de dados de veranicos máximos na estação chuvosa dos municípios da bacia do Rio Pajeú. . .	33
Tabela 2 –	Distância (km) entre os municípios pertencentes a bacia do Rio Pajeú considerando a latitude e longitude.	36
Tabela 3 –	Estruturas utilizadas no ajuste da distribuição GEV aos dados de veranico máximo de Brejinho - PE, destacando as respectivas abreviações, expressões, configurações de pesos (s) e o parâmetro γ , que representa o peso atribuído entre o tamanho das amostras e as distâncias de cada município.	38
Tabela 4 –	Estatísticas descritivas de veranicos máximos na estação chuvosa dos municípios pertencentes a bacia do Rio Pajeú.	45
Tabela 5 –	Resultados dos testes de hipóteses (valor- p), Ljung-Box e Cox-Stuart ao nível de significância de 5%, referente aos dados de veranicos máximos na estação chuvosa dos municípios pertencentes a bacia do Rio Pajeú.	46
Tabela 6 –	Estimativas dos parâmetros (Desvio Padrão) da distribuição GEV via inferência bayesiana, utilizando <i>priori</i> não informativa e resultados do teste de Kolmogorov-Smirnov (KS) ao nível de significância de 5% . .	47
Tabela 7 –	Estimativas dos parâmetros da distribuição GEV via inferência bayesiana para o município de Brejinho - PE no período de 2000 a 2013, utilizando <i>priori</i> não informativa, informativa e mistura de <i>prioris</i> , com seus respectivos intervalos HPD de 95% e resultados dos critérios de convergência.	49
Tabela 8 –	Resultados do Critério de Informação da Deviance (DIC), Acurácia (%), Amplitude Média e Erro Médio de Predição (%) do veranico máximo predito para os tempos de retorno de 2, 4, 6 e 8 anos (valor observado), para o município de Brejinho-PE.	50
Tabela 9 –	Resultados dos testes (valor- p) de Ljung-Box e Cox-Stuart à 5% de significância, para os dados de veranico máximo do município de Brejinho-PE.	52

Tabela 10 – Estimativas dos parâmetros da distribuição GEV via inferência bayesiana, utilizando mistura de <i>prioris</i> , com seus respectivos intervalos HPD de 95% e resultados dos critérios de convergência.	53
Tabela 11 – Níveis de retorno para os tempos de 5, 10, 20, 30, 50, 100 e 200 anos, seus respectivos intervalos HPD com 95% de probabilidade e resultados dos critérios de convergência.	53

LISTA DE FIGURAS

Figura 1 – Mapa de localização da bacia hidrográfica do rio Pajeú	
.....	31

SUMÁRIO

1	INTRODUÇÃO	12
2	OBJETIVOS	14
2.1	OBJETIVOS ESPECÍFICOS	14
3	REVISÃO DE LITERATURA	15
3.1	VERANICOS	15
3.2	TEORIA DE VALORES EXTREMOS	16
3.2.1	Blocos de Máximos	17
3.3	INFERÊNCIA BAYESIANA	20
3.3.1	Distribuição <i>a priori</i>	21
3.3.2	Distribuição <i>a posteriori</i>	24
3.3.3	Monte Carlo via Cadeias de Markov - MCMC	27
3.3.4	Algoritmo Metropolis-Hastings	28
3.3.5	Verificação da Convergência	28
4	MATERIAL E MÉTODOS	30
4.1	CARACTERIZAÇÃO DA ÁREA DE ESTUDO	30
4.2	DADOS DE PRECIPITAÇÃO	32
4.3	ANÁLISE DESCRITIVA E TESTES PRELIMINARES	33
4.4	ESTIMAÇÃO VIA INFERÊNCIA BAYESIANA	34
4.5	OBTENÇÃO DA DISTRIBUIÇÃO A <i>POSTERIORI</i>	39
4.6	NÍVEL DE RETORNO	40
4.7	AVALIAÇÃO DE DESEMPENHO DOS MODELOS	41
4.8	RECURSOS COMPUTACIONAIS	43
5	RESULTADOS E DISCUSSÃO	44
5.1	ESTATÍSTICA DESCRITIVA	44
5.2	OBTENÇÃO DA <i>PRIORI</i>	45
5.3	PREVISÃO DE VERANICOS MÁXIMOS PARA BREJINHO - PE	48
5.4	ANÁLISE DO VERANICO MÁXIMO DO MUNICÍPIO DE BREJINHO - PE	52
6	CONCLUSÕES	56
	REFERÊNCIAS	57
	APÊNDICE	62

1 INTRODUÇÃO

A frequência e a intensidade de eventos climáticos têm aumentado nas últimas décadas. Como exemplos de eventos climáticos extremos é possível citar a ocorrência de temperaturas máximas ou mínimas extremas, índices extremos de umidade relativa do ar e grande volume de precipitação pluvial, seguido ou não, de longo período de estiagem. A ocorrência de períodos consecutivos de dias sem chuva durante a estação chuvosa é denominada “veranico”. Em geral, quanto maior for período de interrupção da precipitação durante a estação chuvosa, maior será o impacto na agricultura familiar, causando problemas nas condições de vida, como escassez de trabalho e renda, aumento no preço dos alimentos, efeitos emocionais, dentre outros.

O estudo dos veranicos é de especial importância econômica e ambiental, uma vez que sua ocorrência pode comprometer a viabilidade da produção agrícola em uma determinada região, além de causar problemas com o abastecimento de água. Logo, o conhecimento do comportamento dessa variável é imprescindível para que sejam feitas as devidas intervenções, por parte da população e dos órgãos públicos, no sentido de minimizar o impacto dessa variável climática. A modelagem do comportamento dos veranicos pode ser feita a partir da contagem do número de dias sem chuva em determinado período (em cada mês ou em cada estação ao longo de vários anos) e extrair os blocos de máximos.

O entendimento do comportamento da série de dados de veranico máximo pode fornecer informações importantes para a compreensão desse fenômeno. Essas informações podem ser utilizadas para descrever o comportamento dessa variável e, assim, tornar possível a previsão de valores máximos. A ferramenta estatística que torna possível a modelagem de dados extremos, como veranicos máximos, é a Teoria de Valores Extremos (TVE) (Bautista; Zocchi; Angelocci, 2004).

Uma distribuição percentente a TVE bastante utilizada na modelagem de eventos extremos é a distribuição Generalizada de Valores Extremos (GEV, do inglês *Generalized Extreme Values*). Segundo Coles (2001), há diversos métodos para se obter as estimativas dos parâmetros da GEV, um deles é o método da Máxima Verossimilhança que possui boas propriedades assintóticas. Porém, não há como garantir a verificação das boas propriedades assintóticas para amostras pequenas (Coles; Dixon, 1999). Além disso, os dados necessários para a aplicação da TVE são geralmente escassos, tanto devido ao relativo baixo investimento brasileiro em Ciência e Tecnologia como devido a aspectos teóricos da técnica estatística, ou seja, apenas os dados

extremas são utilizados.

De acordo com Coles e Powell (1996), uma maneira de contornar esta dificuldade é utilizar a inferência bayesiana. Uma vantagem dessa abordagem é a inclusão de informações *a priori* das observações, o que permite diminuir as incertezas com relação aos parâmetros, fazendo com que essa abordagem seja eficaz mesmo em situações que apresentem tamanhos amostrais reduzidos. As informações *a priori* podem ser a respeito de locais próximos ou locais com características semelhantes à área de estudo, possibilitando ao pesquisador a obtenção de maior acurácia nas estimativas dos parâmetros. Assim, a abordagem bayesiana se mostra eficaz para a estimação dos parâmetros e para a predição de veranicos máximos, oferecendo uma solução frente às limitações dos métodos tradicionais.

2 OBJETIVOS

O objetivo geral do presente estudo foi verificar o desempenho da mistura de distribuições informativas *a priori* na predição de veranicos discretos máximos no município de Brejinho - PE, ajustados à distribuição GEV via inferência bayesiana.

2.1 OBJETIVOS ESPECÍFICOS

Especificamente, objetivou-se:

- a) Obter o desempenho preditivo da GEV quando ajustada aos dados de veranicos máximos de Brejinho - PE na ausência de informação *a priori*;
- b) Obter o desempenho preditivo da GEV quando ajustada aos dados de veranicos máximos de Brejinho - PE, utilizando, como informação *a priori*, os resultados obtidos pelo ajuste da GEV via inferência bayesiana aos dados de veranicos máximos do município mais próximo disponível;
- c) Obter o desempenho preditivo da GEV quando ajustada aos dados de veranicos máximos de Brejinho - PE, utilizando, como informação *a priori*, os resultados, ponderados pela distância, obtidos pelo ajuste da GEV via inferência bayesiana aos dados de veranicos máximos de todos os municípios disponíveis;
- d) Comparar o DIC, a acurácia, a precisão e amplitude média das predições obtidas.

3 REVISÃO DE LITERATURA

Nesta seção, é apresentado o referencial teórico sobre os conceitos de veranicos, teoria dos valores extremos e inferência bayesiana, utilizados na execução deste estudo.

3.1 VERANICOS

A variação climática é estabelecida por uma série de ocorrências atmosféricas e oceânicas, sobretudo suas interatividades. Pode-se enfatizar que episódios de climas secos e úmidos, por sua vez, influenciam significativamente a sociedade. Durante a estação chuvosa, observa-se períodos curtos de *déficit* hídrico que podem ocasionar prejuízos às culturas e problemas no abastecimento de água potável ou para afins de geração de energia hidroelétrica (Cupolillo, 2008). Esse período é caracterizado como veranico.

Fernandes (2014), define veranico como dias consecutivos sem chuva ou com precipitação menor ou igual a 2 mm/dia. Para Neto e Villela (1986), períodos de precipitação menores que 3 mm/dia já são considerados como veranicos. Minuzzi *et al.* (2005) considera como veranico o período de pelo menos três dias sem chuva ou com precipitação inferior a 1 mm. E para o Sleiman (2008), um veranico pode ser definido com duração mínima de cinco dias.

Segundo Cupolillo (2008), os veranicos podem ser classificados de acordo com a intensidade, frequência e tempo de duração. Quanto maior a intensidade, a frequência e a duração deste fenômeno, maiores serão as consequências nas atividades do setor agrícola. A influência do veranico sobre a produtividade da cultura pode ser acentuada, principalmente quando coincide com a fase na qual a planta é mais sensível a deficiência hídrica (Carvalho *et al.*, 2000).

A atividade do setor agrícola é influenciada por fatores climáticos, como radiação solar, umidade relativa do ar, velocidade do vento e, principalmente, pela temperatura e precipitação. Entretanto, o déficit hídrico afeta vários aspectos do crescimento vegetal; os efeitos mais óbvios do estresse hídrico se referem à redução do tamanho das plantas, de sua área foliar e da produtividade da cultura (Farias *et al.*, 2008).

Quando se trata da disponibilidade hídrica da região Nordeste, em especial do semiárido, Assis (2012) afirma que essa região é caracterizada pela grande variabilidade da precipitação no tempo e no espaço, em que ocorrem sucessivos períodos secos durante a estação chuvosa,

os quais são denominados como “veranico”. A influência do veranico sobre a produtividade da cultura pode ser acentuada, principalmente quando coincide com a fase na qual a planta é mais sensível à deficiência hídrica (Carvalho *et al.*, 2000).

Menezes, Brito e Lima (2010), em um estudo para verificar a relação entre o período de veranico e as produções de cana-de-açúcar, arroz, abacaxi, algodão, sisal, milho e feijão, chegaram a conclusão de que as produções de milho e feijão são dependentes da duração dos veranicos. Carvalho *et al.* (2000), em um estudo que teve como objetivo estimar as perdas relativas na produção de milho, influenciadas pela ocorrência de veranico na fase de floração concluíram que quanto maior o período de veranico, maior é a perda na produção de milho.

Nesse sentido, é importante prever a ocorrência dos veranicos para que os prejuízos causados por esse fenômeno possam ser minimizados. Especialmente, a duração extremal de veranicos máximos pode ser estudada utilizando a Teoria de Valores Extremos.

3.2 TEORIA DE VALORES EXTREMOS

A Teoria de Valores Extremos (TVE) é uma importante área da estatística que permite o estudo de eventos extremos que possuem baixa frequência relativa, ou seja, são eventos que acontecem com rara frequência, mas que podem obter em certas situações, resultados catastróficos. A TVE é, segundo Beijo e Avelar (2011), uma metodologia utilizada para analisar o comportamento das observações extremas, ou seja, dos pontos mais afastados da média e tem sido importante para o planejamento de atividades sujeitas aos efeitos de eventos com essas características.

Conforme Mendes (2004), Gumbel afirma que a preocupação com a modelagem de eventos extremos começou no século XVII com estudos em astronomia. Ainda de acordo com a autora, alguns pesquisadores colocam o artigo de Bortkiewicz (1922) como uma referência no desenvolvimento formal da TVE, em que ele analisou a amplitude de uma amostra da distribuição normal. Alguns anos depois, em 1927, Fréchet publicou um trabalho introduzindo a distribuição assintótica do máximo de uma amostra. Em 1928, Fisher (1890-1962) e Tippet (1902-1985) estudaram o mesmo problema de Fréchet, porém de forma independente. Eles mostraram que existem três tipos possíveis para a distribuição do máximo, a tipo I de Gumbel, a tipo II de Fréchet e a tipo III de Weibull (Bautista, 2002).

Segundo Mendes (2004), valor extremo é o máximo (ou mínimo) de uma série de dados num determinado período, que é denominado como Blocos de Máximo (ou Mínimo) ou valores que ultrapassam um limiar suficientemente alto (ou baixo), que é chamado de POT (Peaks Over Threshold). Neste trabalho, os dados serão separados por blocos de máximos.

3.2.1 Blocos de Máximos

Seja uma amostra com Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com função de distribuição $F(y)$. O máximo das n variáveis aleatórias é definido da seguinte forma,

$$M_n = \max\{Y_1, Y_2, \dots, Y_n\}.$$

A função de distribuição de M_n é dada por,

$$\begin{aligned} F_{M_n}(y) &= P(M_n \leq y) \\ &= P(Y_1 \leq y \cap Y_2 \leq y \cap \dots \cap Y_n \leq y) \\ &= P(Y_1 \leq y) \times P(Y_2 \leq y) \times \dots \times P(Y_n \leq y) \\ &= [F(y)]^n, \end{aligned}$$

para $y \in \mathbb{R}$, $n \in \mathbb{N}$.

A medida que n cresce, ou seja, quando $n \rightarrow \infty$ a função M_n é degenerada. Contudo, este problema pode ser contornado pelo Teorema de Fisher e Tippett (1928).

Teorema 3.1 (Fisher; Tippett, 1928) *Sejam Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes e identicamente distribuídas. Defina $M_n = \max\{Y_i\}$. Para constantes $a_n > 0$ e $b_n \in \mathbb{R}$, Fisher e Tippett (1928) mostraram que,*

$$P\left(\frac{M_n - b_n}{a_n} \leq y\right) \xrightarrow[n \rightarrow \infty]{D} G(y),$$

em que $G(\cdot)$ é alguma função de distribuição não-degenerada e $\xrightarrow[n \rightarrow \infty]{D}$ denota a convergência em distribuição.

Segundo Fisher e Tippett (1928) e Gnedenko (1943), essa função de distribuição não degenerada $G(\cdot)$ converge para um dos três tipos de função de distribuição:

Distribuição Gumbel (Tipo I):

$$G(y) = \exp \left\{ - \exp \left[- \left(\frac{y - \mu}{\sigma} \right) \right] \right\} \quad \text{para } -\infty < y < \infty;$$

Distribuição Fréchet (Tipo II)

$$G(y) = \begin{cases} \exp \left\{ - \left(\frac{y - \mu}{\sigma} \right)^{-\xi} \right\} & \text{para } y > \mu, \\ 0 & \text{para } y \leq \mu; \end{cases}$$

Distribuição Weibull (Tipo III)

$$G(y) = \begin{cases} \exp \left\{ - \left[- \left(\frac{y - \mu}{\sigma} \right)^\xi \right] \right\} & \text{para } y < \mu, \\ 1 & \text{para } y \geq \mu; \end{cases}$$

em que $\sigma > 0$, $-\infty < \mu < \infty$ e $\xi > 0$, sendo os parâmetros de escala, de posição e de forma.

Bautista (2002) afirma que um problema prático é a escolha da distribuição de valores extremos mais adequada para uma amostra de dados. Para isso, Jenkinson (1955), propôs uma composição das três distribuições de valores extremos, a distribuição generalizada de valores extremos, ou GEV (*Generalized Extreme Values*).

A distribuição GEV, proposta por Jenkinson (1955), possui a seguinte função de distribuição:

$$F(y) = \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (3.1)$$

definida em $-\infty < y < \mu - \frac{\sigma}{\xi}$ para $\xi < 0$, $\mu - \frac{\sigma}{\xi} < y < \infty$ para $\xi > 0$ e, $-\infty < y < \infty$ quando $\xi \rightarrow 0$, sendo μ, σ e ξ os parâmetros de posição, de escala e de forma, respectivamente, com $\sigma > 0$.

Derivando-se (3.1) em relação a y , obtém-se a função de densidade de probabilidade da distribuição GEV,

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

em que $-\infty < x < \mu - \frac{\sigma}{\xi}$ para $\xi < 0$ e $\mu - \frac{\sigma}{\xi} < x < \infty$ para $\xi > 0$.

Há vários métodos disponíveis para estimar os parâmetros da distribuição GEV, entre os quais destaca-se a inferência bayesiana. Segundo Coles e Powell (1996), uma das vantagens da inferência bayesiana é a possibilidade de acomodar informações *a priori* por meio de uma distribuição, o que ajuda a diminuir as incertezas na estimativa dos parâmetros e dos quantis da distribuição GEV. A informação obtida a partir dos dados coletados, contribui para a formação do modelo por meio da função de verossimilhança $L(\boldsymbol{\theta}|\mathbf{y})$, em que $\boldsymbol{\theta}$ é o vetor de parâmetros e y representa a amostra.

A função de verossimilhança da distribuição GEV pode ser expressa por:

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

Logo, a função suporte da distribuição GEV é expressa por:

$$l(\mu, \sigma, \xi|\mathbf{y}) = \sum_{i=1}^n \left\{ -\ln \sigma - \left(\frac{1+\xi}{\xi} \right) \ln \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right] - \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}. \quad (3.2)$$

Diversos autores exploraram a modelagem da distribuição GEV em conjunto com a inferência bayesiana, aplicando essas metodologias em diferentes contextos e áreas de estudo.

Ferreira *et al.* (2024a) modelaram o comportamento da temperatura máxima trimestral da cidade de Divinópolis-MG, ajustando a distribuição GEV às séries históricas de temperaturas máximas via método da máxima verossimilhança e inferência bayesiana. Os autores verificaram que, para todos os trimestres, a inferência bayesiana propiciou melhores estimativas da temperatura máxima em comparação com o método de máxima verossimilhança.

Costa, Beijo e Avelar (2022) utilizaram a distribuição GEV com estimação de parâmetros via inferência bayesiana para prever a vazão máxima do reservatório de Furnas. Os autores concluíram que a utilização das informações *a priori* proporcionou aumento na precisão e acurácia das estimativas de vazão máxima.

Ferreira *et al.* (2024b) compararam a qualidade das previsões da temperatura máxima em Lavras-MG, com base nas normas climatológicas do período de 1981 a 2010, e nas previsões obtidas pelo ajuste da distribuição GEV via inferência bayesiana, levando em conta diferentes estruturas de distribuições *a priori* informativas e não informativas. Os autores concluíram que

a distribuição GEV ajustada por meio da inferência bayesiana proporcionou melhores previsões para a temperatura máxima em Lavras-MG e que, na maioria dos meses, as *prioris* informativas apresentaram desempenho superior em relação às *prioris* não informativas.

3.3 INFERÊNCIA BAYESIANA

A teoria da inferência bayesiana surgiu no meio científico por volta de 1930 com base em alguns estudos teóricos como o de Jeffreys (1939). Entretanto, estes estudos envolviam resoluções de integrais complexas, o que fez com que o uso dos métodos bayesianos ficassem limitados até a década de 90. Este problema foi solucionado por meio do trabalho de Gelfand *et al.* (1990), onde foi utilizado um recurso de simulação dinâmica, o algoritmo Amostrador de Gibbs, que emprega as Cadeias de Markov elaborada por Geman e Geman (1984).

Segundo Rossi (2011), a principal diferença entre as abordagens bayesiana e clássica está na interpretação do parâmetro. Enquanto, na abordagem clássica, o parâmetro é geralmente um escalar ou vetor desconhecido e, na maioria das vezes, fixo, na abordagem bayesiana ele é sempre considerado como um vetor ou escalar aleatório, também desconhecido.

A metodologia bayesiana consiste na obtenção da distribuição *a posteriori*, $p(\boldsymbol{\theta}|\mathbf{y})$, dos parâmetros. Para tanto, é necessário combinar uma distribuição de probabilidade para os parâmetros desconhecidos, denominada de distribuição *a priori*, $p(\boldsymbol{\theta})$, e a informação sobre os dados, que é descrita por meio da função de verossimilhança, $L(\boldsymbol{\theta}|\mathbf{y})$. Essa combinação é feita por meio do Teorema de Bayes (Box; Tiao, 1992). A distribuição *a posteriori* de um parâmetro $\boldsymbol{\theta}$ contém toda a informação probabilística a respeito deste parâmetro e sua função de densidade *a posteriori* é a melhor descrição do processo de inferência (Beijo, 2006).

Teorema 3.2 (Box; Tiao, 1992) *Suponha que $\mathbf{y} = (y_1, \dots, y_n)$ é um vetor de n observações cuja distribuição de probabilidade $p(\mathbf{y}|\boldsymbol{\theta})$ depende dos valores dos parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Suponha também que $\boldsymbol{\theta}$ tem uma distribuição de probabilidade $p(\boldsymbol{\theta})$. Logo,*

$$p(\mathbf{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) = p(\mathbf{y}, \boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y}) \cdot p(\mathbf{y}).$$

Dado os valores observados \mathbf{y} , a distribuição condicional de $\boldsymbol{\theta}$ é:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}, \boldsymbol{\theta})}{p(\mathbf{y})}, \quad (3.3)$$

em que $p(\mathbf{y})$ pode ser escrita como:

$$p(\mathbf{y}) = \mathbb{E}p(\mathbf{y}|\boldsymbol{\theta}) = c^{-1} = \begin{cases} \int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} & \text{se } \boldsymbol{\theta} \text{ contínuo,} \\ \sum p(\mathbf{y}|p(\boldsymbol{\theta})) & \text{se } \boldsymbol{\theta} \text{ discreto,} \end{cases}$$

em que a soma ou a integral é assumida no espaço paramétrico de $\boldsymbol{\theta}$, e $\mathbb{E}[f(\boldsymbol{\theta})]$ é a esperança matemática de $f(\boldsymbol{\theta})$ em relação a $p(\boldsymbol{\theta})$. Logo, a expressão (3.3) pode ser escrita alternativamente como,

$$p(\boldsymbol{\theta}|\mathbf{y}) = cp(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (3.4)$$

A expressão (3.4), ou seu equivalente, é referida como Teorema de Bayes. Nesta expressão, $p(\boldsymbol{\theta})$ descreve informações de $\boldsymbol{\theta}$ sem o conhecimento dos dados e é denominada distribuição *a priori*. Correspondentemente, $p(\boldsymbol{\theta}|\mathbf{y})$ acomoda a atualização sobre $\boldsymbol{\theta}$ após observar os dados e é denominada distribuição *a posteriori*. A quantidade c é uma constante de normalização necessária para que a distribuição *a posteriori* integre ou some 1.

Agora, dado os valores \mathbf{y} , $p(\mathbf{y}|\boldsymbol{\theta})$ em (3.4) é considerada como uma função de $\boldsymbol{\theta}$. Sendo assim considerada, é denominada função de verossimilhança de $\boldsymbol{\theta}$ dado \mathbf{y} , ou seja, $L(\boldsymbol{\theta}|\mathbf{y})$. Portanto, a fórmula de Bayes pode ser escrita como:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto L(\boldsymbol{\theta}|\mathbf{y})p(\boldsymbol{\theta}),$$

ou seja, Distribuição *a posteriori* \propto Verossimilhança x Distribuição *a priori*, em que \propto representa proporcionalidade.

3.3.1 Distribuição *a priori*

Segundo Gelman *et al.* (2014), a principal característica da abordagem bayesiana é a informação *a priori* que se tem sobre uma quantidade de interesse $\boldsymbol{\theta}$. Essa informação representa probabilisticamente o conhecimento a respeito de $\boldsymbol{\theta}$ antes da obtenção dos dados.

A distribuição *a priori* pode ser classificada de acordo com o grau de informação de $\boldsymbol{\theta}$, sendo ela *priori* informativa ou *priori* não informativa. A *priori* informativa é utilizada nos casos em que o pesquisador possui algum conhecimento prévio do que está sendo estudado. Por

outro lado, *a priori* não informativa é utilizada quando o pesquisador possui pouca ou nenhuma informação para incluir na análise.

Berger (1985), define os tipos de *priori* quanto ao grau de informação da seguinte forma:

1. *Priori* informativa: é uma distribuição que favorece um ou mais valores de θ no espaço paramétrico.
2. *Priori* não informativa: é uma distribuição que não favorece nenhum valor de θ no espaço paramétrico.

Esses tipos de distribuições, conforme observado por Paulino, Turkman e Murteira (2003), podem servir como referência, sendo úteis quando é preciso determinar uma distribuição *a posteriori* em situações em que o conhecimento disponível é insuficiente para propor uma distribuição subjetiva. Além disso, são úteis para comparar os resultados com a teoria clássica e, por fim, para avaliar a influência de *prioris* subjetivas nas inferências quando comparadas com *prioris* de referência.

Segundo Paulino, Turkman e Murteira (2003), existem várias formas de especificar uma *priori* não informativa, como o método de Bayes-Laplace, o método de Jeffreys, o método de Box-Tiao, *Priori Vaga (Flat)*, entre outros.

- **Método de Bayes-Laplace:** Princípio da Razão Insuficiente. Na ausência da razão insuficiente para favorecer uma probabilidade em relação a outras, devido à ausência de informações *a priori*, deve-se assumir que todas as probabilidades são igualmente prováveis.
- **Método de Jeffreys (1961):** A *priori* de Jeffreys é obtida com base na matriz de informação de Fisher $I(\theta)$. Considerando o caso uniparamétrico, a *priori* de Jeffreys é expressa por,

$$P(\theta) = \sqrt{I(\theta)},$$

em que,

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \ln f(y|\theta) \right)^2 \middle| \theta \right].$$

Agora, considerando o caso multiparamétrico, a matriz de informação de Fisher é expressa por $I(\boldsymbol{\theta})$, em que o ij -ésimo elemento da matriz é,

$$I_{ij}(\boldsymbol{\theta}) = E \left(\frac{\partial \ln f(y|\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln f(y|\boldsymbol{\theta})}{\partial \theta_j} \middle| \boldsymbol{\theta} \right).$$

A *priori* de Jeffreys é proporcional à raiz quadrada do determinante da matriz de informação esperada de Fisher,

$$p(\boldsymbol{\theta}) \propto \sqrt{|I_{ij}(\boldsymbol{\theta})|}.$$

- **Método de Box e Tiao (1992):** A principal ideia foi encontrar uma nova parametrização $\phi = \phi(\boldsymbol{\theta})$ para o modelo $f(y|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$ tal que a função de verossimilhança resultante fosse simplesmente uma translação baseada nos dados, isto é,

$$L(\boldsymbol{\theta}|y) \propto g[\phi(\boldsymbol{\theta}) - m(y)],$$

em que g é uma função cuja forma não depende de y , e $m(y)$ é a função que descreve como a verossimilhança L se desloca com a variação de y (ou seja, o parâmetro de localização da verossimilhança L).

Após identificar uma transformação ϕ adequada, seleciona-se uma distribuição para ela que garanta que a distribuição a *posteriori* seja, na prática, a verossimilhança normalizada. Em termos aproximados, isso significa que,

$$\pi(\phi|y) \propto \frac{L(\phi|y)}{\int L(\phi|y)}.$$

A distribuição *a priori* não informativa de Box-Tiao para o parâmetro original $\boldsymbol{\theta}$ é, aproximadamente, proporcional ao determinante do jacobiano da transformação utilizada, que é dado por

$$\pi(\boldsymbol{\theta}) \propto \left| \frac{d\phi}{d\boldsymbol{\theta}} \right|.$$

- **Priori Vaga (Flat):** Uma distribuição plana (também chamada de vaga ou *flat*) representa uma situação de pouca ou nenhuma informação prévia sobre o parâmetro em questão,

de modo que exerce uma influência mínima sobre a distribuição *a posteriori*. Nessa configuração, a informação contida nos dados, expressa pela função de verossimilhança, torna-se o fator dominante na formação da distribuição *a posteriori* do parâmetro.

Mais detalhes sobre os métodos para definir uma *priori* não informativa podem ser encontrados em Paulino, Turkman e Murteira (2003). Daqui para frente neste trabalho, toda menção a uma distribuição *a priori* “não informativa” deverá ser entendida efetivamente como uma *priori* vaga.

Para Box e Tiao (1992), a distribuição *a priori* é a forma de quantificar a incerteza sobre o parâmetro, porém, a *priori* não informativa não representa necessariamente o desconhecimento do pesquisador sobre θ . Esse tipo de *priori* deve ser usada como forma de viabilizar a inferência *a posteriori*.

Quando se trata de uma *priori* informativa, faz-se referência à acomodação de informações prévias na análise, fornecidas por alguém com conhecimento especializado sobre o problema em questão, como um pesquisador, estatístico ou outro especialista. Essas informações, embora subjetivas em sua essência, muitas vezes têm como base dados objetivos, como séries históricas relacionadas ao problema em estudo ou a problemas análogos (Paulino; Turkman; Murteira, 2003).

A escolha da distribuição *a priori* desempenha um papel fundamental na análise bayesiana. A utilização de uma *priori* conjugada pode ser uma estratégia eficaz para integrar a informação prévia ao modelo. De acordo com Paulino, Turkman e Murteira (2003), uma distribuição *a priori* é considerada conjugada quando, ao ser combinada com a função de verossimilhança, resulta em uma distribuição *a posteriori* que pertence à mesma família da *priori*.

Uma outra vantagem da utilização de *prioris* conjugadas é que elas simplificam os cálculos, resultando em uma forma analítica que é fácil de manipular para a distribuição *a posteriori*. Entretanto, uma desvantagem, é que a distribuição escolhida nem sempre representa adequadamente a incerteza *a priori* associada aos parâmetros.

3.3.2 Distribuição *a posteriori*

A distribuição *a posteriori* do parâmetro θ é proporcional ao produto entre a distribuição *a priori* e a função de verossimilhança dos dados. Sendo assim, toda inferência relacionada ao

parâmetro é realizada por meio da distribuição *a posteriori*, pois ela contém toda informação probabilística de interesse a respeito do parâmetro (Martins-Filho *et al.*, 2008).

O processo de inferência sobre um parâmetro é realizado por meio da obtenção da sua distribuição marginal *a posteriori*, integrando a distribuição conjunta *a posteriori* em relação aos outros parâmetros. Dessa forma, é possível calcular a estimativa dos parâmetros, que são os valores centrais da distribuição *a posteriori*. Uma forma de expressar a incerteza associada a uma estimativa pontual é por meio de intervalos, conhecidos na inferência bayesiana como intervalos de credibilidade (ICr), os quais podem ser calculados a partir da distribuição *a posteriori*. De acordo com Paulino, Turkman e Murteira (2003), a região de credibilidade é definida da seguinte forma:

$R(Y)$ é uma região de credibilidade λ para θ se,

$$P(\theta \in R(Y)|Y) = \int_{R(Y)} p(\theta|Y)d\theta \quad (3.5)$$

Assim, será escolhido o subconjunto de θ que possui a maior credibilidade *a posteriori*, ou seja, aquele que apresenta a mesma credibilidade λ e inclui todos os valores de θ .

Conforme Paulino, Turkman e Murteira (2003), existe uma infinidade de regiões de credibilidade com o mesmo grau de credibilidade, mas apenas um apresenta alta densidade *a posteriori*, o qual é denominado de intervalo HPD (*Highest Probability Density*).

Segundo Paulino, Turkman e Murteira (2003), a região $R(Y)$ de credibilidade λ para θ , é uma região com densidade (probabilidade) *a posteriori* máxima (HPD) se,

$$\sup_{\theta \notin R(Y)} p(\theta|Y) \leq c_\lambda \leq \inf_{\theta \in R(Y)} p(\theta|Y) \quad (3.6)$$

para algum c_λ tal que $P(\theta \in R(Y)|Y) \geq \lambda$, ou equivalente, se

$$R(Y) = \{\theta : p(\theta|Y) \geq c_\lambda\} \quad (3.7)$$

com $c_\lambda > 0$ a maior constante tal que,

$$\int_{R(Y)} p(\theta|Y)d\theta \geq \lambda. \quad (3.8)$$

A região HPD inclui os valores de θ que possuem a maior densidade *a posteriori*. Todos os pontos dentro da região HPD possuem uma densidade de probabilidade $p(\theta|Y)$ maior ou

igual à densidade dos valores que estão fora dessa região. Isso significa que a região HPD é a mais compacta e densa possível, garantindo que todos os valores de θ pertencentes a ela têm uma densidade de probabilidade maior ou igual à densidade dos valores fora da região.

Segundo Gamerman e Lopes (2006), em alguns casos, não é possível derivar analiticamente as distribuições marginais *a posteriori*, tornando necessário o uso de técnicas iterativas, como os algoritmos de Amostrador de Gibbs e Metropolis-Hastings, que se baseiam na técnica de Monte Carlo via Cadeias de Markov (MCMC).

Conforme mencionado por Gamerman e Lopes (2006), o amostrador de Gibbs é um processo caracterizado como Markoviano, pois possui uma memória curta, em que o valor atual é determinado apenas pela variável anterior. Esse algoritmo utiliza a simulação de Monte Carlo para gerar os valores dos parâmetros com base em distribuições de probabilidades conhecidas, fundamentando-se nos princípios das Cadeias de Markov (Gamerman; Lopes, 2006).

O amostrador de Gibbs tem como objetivo gerar amostras a partir de uma distribuição conjunta complexa, iterando sobre as variáveis da distribuição. Durante esse processo, o algoritmo realiza a amostragem a partir da distribuição condicional de cada variável, mantendo as outras fixas, repetindo essa etapa até que as amostras converjam para a distribuição desejada. Essa abordagem é especialmente eficiente quando as distribuições condicionais são bem conhecidas e simples de implementar.

Por outro lado, o algoritmo de Metropolis-Hastings é utilizado para amostrar a partir de uma distribuição alvo quando a distribuição condicional dos parâmetros não é conhecida. Ele propõe um novo estado com base em uma distribuição de proposta, aceitando essa nova amostra com uma probabilidade que depende da razão das densidades entre a distribuição alvo e a proposta. Se a nova amostra não for aceita, a amostra anterior é mantida.

Esse algoritmo é fundamentado no princípio dos métodos de rejeição, em que um valor é gerado a partir de uma distribuição auxiliar e é aceito com uma probabilidade específica. Esse mecanismo de ajuste garante que a cadeia de Markov alcance a distribuição de equilíbrio ao longo do tempo (Gamerman; Lopes, 2006).

3.3.3 Monte Carlo via Cadeias de Markov - MCMC

Os métodos de Monte Carlo e as Cadeias de Markov são técnicas fundamentais em diversas áreas da ciência e engenharia, graças à sua eficiência na solução de problemas complexos por meio de simulações. Embora sejam metodologias distintas, juntas formam a base do Método de Monte Carlo via Cadeias de Markov (MCMC).

A técnica de Monte Carlo, envolve gerar múltiplas amostras sucessivas de uma mesma população, calculando as estimativas dos parâmetros de interesse em cada uma delas. Em alguns cenários, isso implica a repetição de simulações de observações baseadas em uma distribuição específica. Quando se obtém um grande número de amostras, a média das estimativas pode fornecer uma boa aproximação do parâmetro real que se está investigando (Hammersley; Handscomb, 1964).

Por outro lado, segundo Gamerman e Lopes (2006), uma cadeia de Markov é um processo estocástico $\theta_0, \theta_1, \dots$, se a distribuição de θ_j , dado todos os valores anteriores $\theta_1, \theta_2, \dots, \theta_{j-1}$ depende apenas de θ_{j-1} . Assim,

$$P(\theta_j \in A | \theta_1, \theta_2, \dots, \theta_{j-1}) = P(\theta_j \in A | \theta_{j-1})$$

para qualquer subconjunto A , em que j é o número de estados.

Quando combinados, o método de Monte Carlo e as Cadeias de Markov oferecem uma metodologia para amostragem e estimação, conhecida como MCMC (Monte Carlo via Cadeias de Markov). Esse método é um procedimento estocástico que gera repetidamente amostras pseudo-aleatórias que caracterizam a distribuição dos parâmetros de interesse (Hamra; MacLehose; Richardson, 2013).

Os métodos de Monte Carlo via cadeias de Markov (MCMC) são simulações iterativas que se baseiam em cadeias de Markov, o que implica que os valores gerados não serão independentes entre si. O objetivo é simular um passeio aleatório no espaço de θ que converge para uma distribuição estacionária, a qual representa a distribuição de interesse no problema (Gamerman; Lopes, 2006).

3.3.4 Algoritmo Metropolis-Hastings

A estratégia na amostragem por MCMC consiste em estabelecer uma cadeia de Markov irreduzível e aperiódica, para a qual a distribuição estacionária é equivalente a distribuição *a posteriori* de interesse. Um método genérico para a construção desse tipo de cadeia de Markov é o algoritmo de Metrópolis-Hastings. Os passos a seguir descrevem o algoritmo:

- inicie com um valor θ^0 , e suponha que o valor atual seja θ^t ;
- simule um valor-candidato θ^* a partir de uma densidade proposta, $\pi_0(\theta^*|\theta^t)$;
- obtenha a razão

$$R = \frac{p(\theta^*|\mathbf{y})\pi_0(\theta^t|\theta^*)}{p(\theta^t|\mathbf{y})\pi_0(\theta^*|\theta^t)};$$

- obtenha a probabilidade de aceitação $P = \min\{R, 1\}$;
- simule um número aleatório u de uma distribuição uniforme $U(0,1)$;
- se $u < P$, aceite $\theta^{t+1} = \theta^*$ como o próximo ponto da cadeia. Caso contrário, rejeite θ^* e tome $\theta^{t+1} = \theta^t$

Mais detalhes sobre o algoritmo de Metropolis-Hastings e demais aspectos da implementação de métodos e MCMC podem ser encontrados nos trabalhos de Raftery e Lewis (1995) e Tierney (1994).

3.3.5 Verificação da Convergência

Segundo Gamerman e Lopes (2006), a avaliação de convergência em métodos MCMC pode ser realizada por meio de métodos formais e informais, sendo este último baseado na análise gráfica da trajetória da cadeia ao longo das iterações. Quanto aos métodos formais, destaca-se os métodos propostos por Geweke (1992) e Heidelberger e Welch (1983) que usam resultados baseados em análise espectral e Raftery e Lewis (1992) que permite calcular quantas iterações são necessárias para uma cadeia atingir a distribuição estacionária por meio da estimação de quantis *a posteriori* com uma precisão previamente fixada.

De acordo com Nogueira, Safadi e Ferreira (2004), o método de Geweke (1992) fornece um diagnóstico para a ausência de convergência. Este método consiste em dividir a cadeia de Markov em duas partes e testar se as médias de ambas as partes são iguais. Geralmente, a primeira parte equivale aos primeiros 10% da cadeia após o “*burn in*” (valores iniciais descartados) e a segunda parte equivale aos 50% finais da cadeia. A cadeia convergirá quando a hipótese de igualdade não for rejeitada.

O critério de Heidelberger e Welch (1983) utiliza testes estatísticos para avaliar a hipótese nula de estacionariedade da cadeia. Se a hipótese nula for rejeitada, descarta-se 10% das iterações iniciais e o teste é repetido. Caso a hipótese seja rejeitada novamente, descarta-se mais 10% das iterações iniciais, até que sejam descartados 50% das iterações iniciais. Se a hipótese nula for rejeitada ou o teste não conseguir ser realizado, isto indica falha de estacionariedade, implicando que é necessário um número maior de iterações. Caso contrário, o número inicial de iterações é indicado como o tamanho do “*burn in*” (Nogueira; Safadi; Ferreira, 2004).

O critério de Raftery e Lewis (1992) é um método que estima quantas iterações são necessárias para que o algoritmo apresente convergência, em que fornece estimativas para “*burn in*” (total de iterações que devem ser descartadas) e para o “*thin*” (distância mínima entre as iterações). Isso se dá considerando-se uma precisão definida, pelas estimativas dos quantis de uma função (Nogueira; Safadi; Ferreira, 2004).

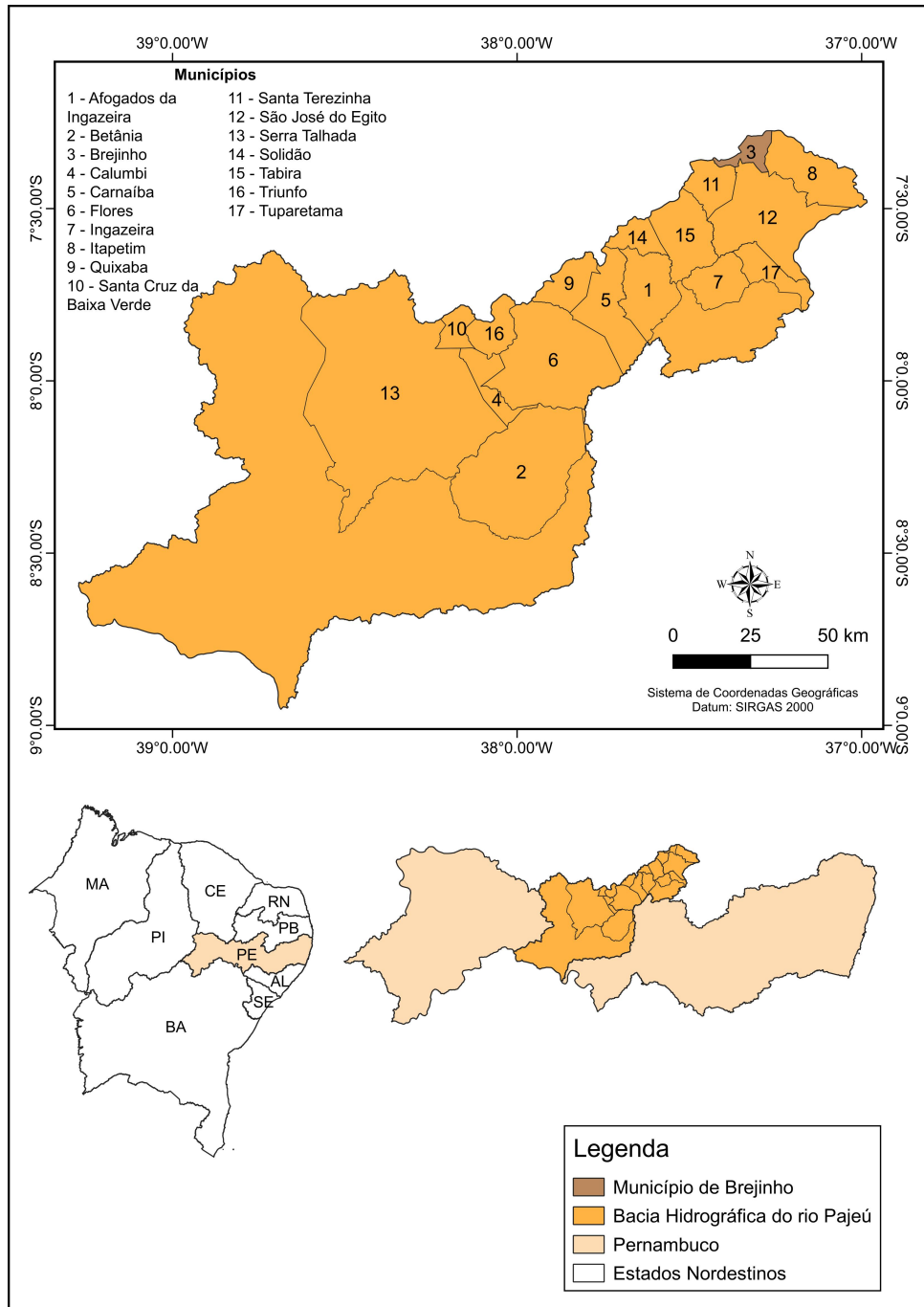
4 MATERIAL E MÉTODOS

Nesta seção, serão apresentadas as características da área de estudo, os detalhes da obtenção dos dados utilizados e as metodologias que serviram de base para o desenvolvimento deste trabalho.

4.1 CARACTERIZAÇÃO DA ÁREA DE ESTUDO

A área de estudo compreende os municípios pertencentes a bacia do Rio Pajeú, localizada no estado de Pernambuco. Os municípios inseridos nessa bacia são: Afogados da Ingazeira, Betânia, Brejinho, Calumbi, Carnaíba, Flores, Ingazeira, Itapetim, Quixabá, Santa Cruz da Baixa Verde, Santa Terezinha, São José do Egito, Serra Talhada, Solidão, Tabira, Triunfo e Tuparetama. Na Figura 1, é apresentado um mapa indicando a localização da bacia hidrográfica do rio Pajeú.

Figura 1 – Mapa de localização da bacia hidrográfica do rio Pajeú



Fonte: Elaborado pela autora.

Legenda: A figura apresenta dois painéis. No painel superior há um mapa com os municípios pertencentes à Bacia do Rio Pajeú. No painel inferior, há dois mapas, à esquerda o mapa do Brasil destacando-se o Estado de Pernambuco e, à direita, o mapa de Pernambuco destacando-se a localização Bacia do Rio Pajeú.

Sendo a maior bacia do estado de Pernambuco, a bacia do Rio Pajeú possui uma área de 16.685,63 km², o que equivale a 16,9% da área total do estado. A bacia encontra-se nas seguintes coordenadas geográficas: 07°16'20" e 08°56'01" de latitude sul e 36°59'00" e 38°57'45"

de longitude oeste. Segundo Assis (2012), o clima da mesorregião do sertão pernambucano, onde está inserida a bacia do Rio Pajeú é do tipo BSw'h' e BSwh', quente e seco, conforme a classificação climática de Koppen.

4.2 DADOS DE PRECIPITAÇÃO

Os dados de precipitação diária (uma estrutura tipicamente de séries temporais), expressos em altura de lâmina d'água (mm), foram obtidos junto aos postos pluviométricos dos municípios em estudo. Esses dados estão disponibilizados para acesso no Banco de Dados da Agência Pernambucana de Águas e Clima (APAC). Os dados de precipitação diária foram organizados a fim de que a variável aleatória do estudo seja o veranico máximo, ou seja, um intervalo de dias secos dentro do período chuvoso. Para ser considerado um veranico, foi adotado como um período mínimo sem chuva o intervalo igual ou superior a cinco dias (Sleiman, 2008) e com precipitação igual ou inferior a 2 mm (Fernandes, 2014). A Tabela 1 apresenta a localização geográfica e tamanho da série de dados de veranicos máximos dos municípios em estudo:

Tabela 1 – Localização geográfica e tamanho das séries de dados de veranicos máximos na estação chuvosa dos municípios da bacia do Rio Pajeú.

Posto Pluviométrico	Latitude	Longitude	Área (Km ²)	Tamanho (anos)
Afogados da Ingazeira	-7,75	-37,64	377,69	42
Betânia	-8,27	-38,03	1.244,07	58
Brejinho	-7,34	-37,28	106,27	21
Calumbi	-7,93	-38,15	179,31	29
Carnaíba	-7,80	-37,79	427,80	51
Flores	-7,86	-37,96	995,55	21
Ingazeira	-7,68	-37,47	243,66	11
Itapetim	-7,37	-37,20	404,84	56
Quixabá	-7,74	-37,85	210,70	48
Santa Cruz da Baixa Verde	-7,82	-38,15	114,93	16
Santa Terezinha	-7,37	-37,48	195,58	22
São José do Egito	-7,44	-37,28	798,87	54
Serra Talhada	-8,03	-38,12	2.979,99	40
Solidão	-7,60	-37,65	138,39	20
Tabira	-7,58	-37,53	388,00	46
Triunfo	-7,83	-38,10	191,51	50
Tuparetama	-7,60	-37,31	179,56	28

Fonte: Elaborada pela autora.

Com a finalidade de selecionar a melhor estrutura de *priori* para predição de veranicos máximos no município de Brejinho - PE, o conjunto de dados foi dividido em duas partes. O período de 2000 a 2013 foi reservado para estimação dos parâmetros da distribuição GEV. O período de 2014 a 2021 corresponde à fase teste, por meio da qual é possível verificar o desempenho das estruturas de *priori*.

4.3 ANÁLISE DESCRITIVA E TESTES PRELIMINARES

A análise descritiva dos dados foi realizada por meio da estimação da média, dos valores mínimo e máximo, dos primeiros e terceiros quartis, assim como do desvio padrão e do coeficiente de variação.

A pressuposição de independência e a presença de tendência (não estacionariedade) na série de veranicos máximos, foram avaliadas, respectivamente, pelos testes de Ljung-Box (Ljung; Box, 1978) e Cox-Stuart (Cox; Stuart, 1955), com nível de significância de 5%. Para verificar o ajuste da distribuição GEV à série de veranicos máximos, foi utilizado o teste de Kolmogorov-Smirnov (Kolmogorov, 1933) e (Smirnov, 1939), também com nível de significância de 5%.

4.4 ESTIMAÇÃO VIA INFERÊNCIA BAYESIANA

Para estimação dos parâmetros da distribuição GEV via inferência bayesiana, utilizou-se a distribuição *a priori* Normal Trivariada sugerida por Coles e Powell (1996):

$$p(\boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\lambda})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta} - \boldsymbol{\lambda}) \right\}, \quad (4.1)$$

em que $\boldsymbol{\theta} = (\mu, \ln(\sigma), \xi)$ é o vetor de parâmetros da distribuição GEV, $\boldsymbol{\lambda}$ é o vetor de médias dos hiperparâmetros e $\boldsymbol{\Sigma}$ é a matriz de variâncias e covariâncias dada por

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

Para elicitare informações para a construção das distribuições *a priori*, a distribuição GEV foi ajustada aos dados de todos os municípios em estudo, exceto Brejinho - PE, utilizando inferência bayesiana com *priori* não informativa vaga. Esse ajuste gerou estimativas dos parâmetros (médias *a posteriori*) que serviram de base para a construção das *prioris* informativas.

Com o objetivo de selecionar a estrutura que apresentasse melhor desempenho nas previsões, foram ajustadas sete configurações de *priori* da distribuição GEV aos dados de veranico máximo do município de Brejinho - PE. A primeira estrutura ajustada foi com *priori* não informativa, utilizando hiperparâmetros das médias iguais a zero e variâncias altas:

$$\boldsymbol{\lambda}_1 = (0; 0; 0) \quad e \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 10^4 & 0 & 0 \\ 0 & 10^4 & 0 \\ 0 & 0 & 10^2 \end{bmatrix}.$$

Na segunda estrutura, inicialmente foram calculadas as distâncias euclidianas entre os postos pluviométricos dos municípios em estudo, visando determinar o posto pluviométrico mais próximo ao de Brejinho - PE. Pela Tabela 2, verificou-se que o posto pluviométrico mais próximo ao de Brejinho - PE é o do município de Itapetim, a uma distância de 9,4 Km.

Tabela 2 – Distância (km) entre os municípios pertencentes a bacia do Rio Pajeú considerando a latitude e longitude.

Posto Pluviométrico	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) Afogados da Ingazeira	0																
(2) Betânia	71,8	0															
(3) Brejinho	60,3	131,9	0														
(4) Calumbi	59,7	39,9	116,1	0													
(5) Carnaíba	17,4	58,3	75,9	42,2	0												
(6) Flores	37,3	45,9	94,5	22,3	19,9	0											
(7) Ingazeira	20,3	89,8	43,1	79,9	37,7	57,6	0										
(8) Itapetim	64,2	135,2	9,4	121,7	80,6	99,8	45,4	0									
(9) Quixabá	23,2	61,9	76,9	39,2	9,4	17,9	42,4	82,6	0								
(10) Santa Cruz da Baixa Verde	56,8	51,5	109,7	12,2	39,8	21,4	76,6	116,1	34,2	0							
(11) Santa Terezinha	45,6	116,6	22,3	96,4	58,6	75,8	34,3	30,9	57,8	89,1	0						
(12) São José do Egito	52,5	123,6	11,1	110,2	68,9	88,2	33,8	11,7	71,1	104,8	23,4	0					
(13) Serra Talhada	61,3	28,3	120,1	11,5	44,3	25,8	81,5	125,1	43,8	23,4	101,5	113,3	0				
(14) Solidão	16,6	85,1	49,9	66,1	26,9	44,7	21,7	55,8	26,9	60,3	31,6	44,5	70,3	0			
(15) Tabira	22,4	94,1	38,3	78,6	37,6	56,6	12,9	43,2	39,5	73,4	23,9	31,6	81,9	13,4	0		
(16) Triunfo	51,5	49,3	105,5	12,3	34,3	15,8	71,4	111,6	29,3	5,6	85,2	100,2	22,2	55,8	68,7	0	
(17) Tuparetama	40,0	108,6	28,9	99,6	57,4	77,2	19,8	28,2	61,5	95,9	31,6	18,0	101,2	37,5	24,4	90,8	0

Fonte: Elaborada pela autora.

Sendo assim, as informações para os hiperparâmetros da distribuição normal trivariada foram extraídas da distribuição GEV ajustada via inferência bayesiana aos dados de veranicos máximos do município de Itapetim - PE. Para o vetor de médias da distribuição normal trivariada (λ), foram utilizados os valores obtidos das estimativas dos parâmetros (médias a *posteriori*) e o logaritmo natural dessa estimativa para o parâmetro σ da distribuição GEV. Com isso, a *priori* informativa do parâmetro de escala da distribuição GEV foi definido da seguinte forma:

$\sigma \sim \text{Lognormal} \left(a, \frac{1}{b^2} \right)$, em que a e b são obtidos a partir da solução do sistema não-linear:

$$\begin{cases} \exp(a + \frac{1}{2}b^2) - \hat{\sigma} = 0 \\ \exp(2a + b^2) \cdot [\exp(b^2) - 1] - \text{Var}(\hat{\sigma}) = 0, \end{cases} \quad (4.2)$$

em que $\hat{\sigma}$ e $\text{Var}(\hat{\sigma})$ são, respectivamente a média a *posteriori* e a variância da *posteriori* do parâmetro escala da distribuição GEV obtidos no ajuste dos dados de veranico máximo de cada município, exceto Brejinho - PE. A solução do sistema não-linear são os parâmetros μ e σ da Lognormal para que a distribuição tenha média igual a estimativa do parâmetro de escala da *priori* e desvio padrão igual ao erro padrão do parâmetro de escala.

Logo,

$$\lambda_2 = (24,723; 2,395; 0,248) \quad e \quad \Sigma_2 = \begin{bmatrix} 2,884 & 0 & 0 \\ 0 & 0,017 & 0 \\ 0 & 0 & 0,016 \end{bmatrix}. \quad (4.3)$$

Por fim, as outras estruturas ajustadas foram a mistura de *prioris*, utilizando, como informação prévia, os resultados, ponderados pela distância, obtidos pelo ajuste da distribuição GEV via inferência bayesiana aos dados de veranicos máximos de todos os municípios em estudo, exceto Brejinho - PE.

Para a obtenção da mistura de *prioris*, foram realizados os seguintes passos:

1. as informações para os hiperparâmetros foram extraídas da distribuição GEV, via inferência bayesiana, para os veranicos máximos anuais de cada município em estudo, exceto, Brejinho - PE;
2. Com essas informações, as distribuições *a priori* normais trivariadas foram construídas

para cada município, mantendo-se a utilização da distribuição Lognormal para modelar o parâmetro de escala σ da GEV.

3. foi calculado o peso para cada Normal Trivariada, utilizando-se a seguinte equação:

$$\omega_j^{(s)} = \begin{cases} \gamma^{(s)} \left(\frac{n_j}{\sum n_j} \right) + (1 - \gamma^{(s)}) \left(\frac{\sum d_j}{d_j} \right), & \text{se } s = 1, 2, 3, 4. \\ \left(\frac{n_j}{\sum n_j} \right) \left(\frac{\sum d_j}{d_j} \right), & \text{se } s = 5. \end{cases} \quad (4.4)$$

em que n_j é o número de anos utilizado no ajuste via inferência bayesiana do j -ésimo município, $\sum_{j=1}^m n_j$ é o total de observações dos municípios não-alvo, isto é, excluindo-se Brejinho - PE, d_j é a distância entre o j -ésimo município não-alvo e Brejinho - PE, $\sum_{j=1}^m d_j$ é o total das distâncias, γ é o peso dado entre as distâncias e os tamanhos das amostras de cada município. A fim de assegurar que os pesos resultantes somem 1, procedeu-se a uma normalização, em que cada peso $\omega_j^{(s)}$ foi dividido pela soma total dos pesos obtidos. Além disso, foram atribuídos diferentes pesos às observações e às distâncias. Assim, as diferentes estruturas de *prioris* foram organizadas da seguinte forma:

Tabela 3 – Estruturas utilizadas no ajuste da distribuição GEV aos dados de veranico máximo de Brejinho - PE, destacando as respectivas abreviações, expressões, configurações de pesos (s) e o parâmetro γ , que representa o peso atribuído entre o tamanho das amostras e as distâncias de cada município.

<i>Prioris</i>	Abreviação	Expressão	s	γ
Não Informativa	NI	Equação 4.1	-	-
Informativa	INF	Equação 4.1	-	-
Mistura de <i>Prioris</i> 1	MP ₁	Equação 4.5	1	0
Mistura de <i>Prioris</i> 2	MP ₂	Equação 4.5	2	$\frac{1}{4}$
Mistura de <i>Prioris</i> 3	MP ₃	Equação 4.5	3	$\frac{1}{3}$
Mistura de <i>Prioris</i> 4	MP ₄	Equação 4.5	4	$\frac{1}{2}$
Mistura de <i>Prioris</i> 5	MP ₅	Equação 4.5	5	-

Fonte: Elaborada pela autora.

4. foi realizada a soma das distribuições a *priori* normais trivariadas dando o respectivo peso para cada uma delas. A mistura de *prioris* pode ser expressa pela seguinte equação,

$$p(\boldsymbol{\theta}) \propto \sum_{j=1}^k \omega_j^{(s)} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\lambda}_j)^T \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{\theta} - \boldsymbol{\lambda}_j) \right\}, \quad (4.5)$$

em que $\omega_j^{(s)}$ são os pesos calculados na expressão 4.4, $\boldsymbol{\lambda}_j$ são os vetores de média e $\boldsymbol{\Sigma}_j$ são as matrizes de variância e covariância.

4.5 OBTENÇÃO DA DISTRIBUIÇÃO A *POSTERIORI*

De forma geral, a distribuição conjunta *a posteriori* para a distribuição GEV é dada pelo produto da função de verossimilhança (3.2) e a distribuição *a priori* (4.1):

$$p(\boldsymbol{\theta}|y) \propto \sum_{i=1}^n \left\{ -\ln \sigma - \left(\frac{1+\xi}{\xi} \right) \ln \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right] - \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \times \\ \times \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\lambda})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\lambda}) \right\}.$$

Portanto, a estrutura 1 utilizou $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1$ e $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$, a estrutura 2 utilizou $\boldsymbol{\lambda} = \boldsymbol{\lambda}_2$ e $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$, e, por fim, a distribuição conjunta *a posteriori* das demais estruturas foi obtida pela seguinte expressão:

$$p(\boldsymbol{\theta}|y) \propto \sum_{i=1}^n \left\{ -\ln \sigma - \left(\frac{1+\xi}{\xi} \right) \ln \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right] - \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \times \\ \times \sum_{j=1}^k w_j \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\lambda}_j)^T \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{\theta} - \boldsymbol{\lambda}_j) \right\}.$$

No entanto, a integração da distribuição conjunta *a posteriori*, para obter as distribuições marginais para cada parâmetro, não possui solução analítica. Sendo assim, foi utilizada técnicas de MCMC por meio do algoritmo Metrópolis-Hastings para gerar amostras das distribuições marginais *a posteriori*.

Para obter cadeias, considerando *priori* não informativa (estrutura 1), foram realizadas 440.000 iterações, excluindo-se as 20.000 primeiras (*burn-in*) e realizando-se saltos a cada 60 iterações (*thin*), obtendo-se uma cadeia final com 7.000 iterações. Para as estruturas de *pri-*

ori informativa (INF), MP_1 , MP_2 e MP_3 , foram realizadas 110.000 iterações, excluindo-se as 10.000 primeiras e realizando-se saltos a cada 20 iterações, obtendo-se também cadeias finais com 5.000 iterações. Para as estruturas MP_4 e MP_5 foram geradas cadeias com 220.000 iterações, excluindo-se as 10.000 primeiras e realizando-se saltos a cada 40 iterações, obtendo-se cadeias finais com 5.000 iterações. É importante observar que foi gerada apenas uma cadeia final para cada uma das estruturas.

Para analisar a convergência das cadeias, foram utilizados três critérios. O primeiro, proposto por Geweke (1992), verifica se a diferença padronizada entre as médias é significativa. Caso não seja, há indicação de convergência das cadeias. O segundo critério, desenvolvido por Raftery e Lewis (1992), estima o número de iterações necessárias para que o algoritmo alcance a convergência, fornecendo estimativas para o *burn-in*, que representa o total de iterações a serem descartadas, e para o *thin*, que corresponde à distância mínima entre as iterações. Por fim, o critério de Heidelberger e Welch (1983) emprega testes de significância para avaliar a hipótese nula de estacionariedade dos valores amostrados.

De acordo com Nogueira, Safadi e Ferreira (2004), a convergência das cadeias pode ser verificada quando no critério de Raftery e Lewis, valores do fator de dependência próximos de 1 indicam independência entre as iterações, sugerindo convergência. Pelo critério de Geweke, a convergência é indicada quando $|Z_G| < z_{\frac{\alpha}{2}}$. Considerando $\alpha = 0,05$, esse limiar corresponde a $z_{\frac{\alpha}{2}} \approx 1,96$, que representa o quantil de 97,5% da distribuição normal padrão. Já no critério de Heidelberger e Welch, a estacionariedade da série é verificada quando $(p > 0,05)$

4.6 NÍVEL DE RETORNO

O nível de retorno $y(T)$ é o valor que se espera que seja superado, pelo menos uma vez, no período de tempo T . De acordo com Coles (2001), $y(T)$ é obtido a partir dos quantis da distribuição GEV e calculados por,

$$y(T) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left\{ 1 - \left[-\ln \left(1 - \frac{1}{T} \right) \right]^{-\hat{\xi}} \right\}, \quad (4.6)$$

em que $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$ são as estimativas dos parâmetros da distribuição GEV e T é o tempo de retorno dado em anos.

4.7 AVALIAÇÃO DE DESEMPENHO DOS MODELOS

A fim de avaliar o desempenho das estruturas foram analisados o DIC (Critério de Informação da Deviance), a acurácia, erro médio de predição e amplitude média dos intervalos HPD de 95%. Os níveis de retorno foram calculados, com base na equação apresentada na subseção 4.6, para os tempos de 2, 4, 6 e 8 anos.

Segundo Spiegelhalter *et al.* (2002), o DIC é uma aproximação assintótica para grandes amostras e é válido quando a distribuição *a posteriori* se aproxima de uma distribuição normal multivariada. Assim, o desvio (Deviance) é definido como,

$$D(\boldsymbol{\theta}) = -2 \ln L(\boldsymbol{\theta}) + C,$$

em que $\boldsymbol{\theta}$ é o vetor de parâmetros do modelos, $L(\boldsymbol{\theta})$ representa a função de verossimilhança e C é uma constante que se anula, portanto não é necessário conhecê-la ao comparar dois modelos (no presente estudo, duas estruturas de *priori*). Assim, o critério DIC, proposto por Spiegelhalter *et al.* (2002), pode ser escrito como,

$$DIC = D(\hat{\boldsymbol{\theta}}) + 2p_D,$$

em que $D(\hat{\boldsymbol{\theta}})$ representa o desvio avaliado na média *a posteriori* e p_D é o número efetivo de parâmetros no modelo, dado por,

$$p_D = \bar{D} - D(\hat{\boldsymbol{\theta}}),$$

sendo $\bar{D} = E[D(\hat{\boldsymbol{\theta}})]$ o desvio médio *a posteriori*, o qual avalia a qualidade do ajuste do modelo aos dados. Os modelos mais adequados são os que apresentam um valor menor para o DIC.

Spiegelhalter *et al.* (2002), propõem o seguinte critério para calcular a diferença absoluta entre os valores de DIC de dois modelos, A e B,

$$\Delta_{DIC} = |DIC_A - DIC_B|.$$

Sendo assim, se $\Delta_{DIC} < 5$, a diferença entre os modelos não é substancial; se $5 \leq \Delta_{DIC} \leq 10$, a diferença entre os modelos é substancial; e se $\Delta_{DIC} > 10$, descarta-se definitivamente o modelo com maior valor de DIC (Spiegelhalter *et al.*, 2002).

A acurácia foi avaliada verificando se a estimativa intervalar engloba ou não o valor observado para os tempos de retorno de 2, 4, 6 e 8 anos. O erro médio de predição (EMP) pode ser expresso por

$$EMP = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \quad (4.7)$$

em que, y_i é o veranico máximo observado para o i -ésimo tempo de retorno, \hat{y}_i é o veranico máximo previsto para o i -ésimo tempo de retorno e $N = 4$ é o número de tempos de retorno estudados (2, 4, 6 e 8 anos).

A amplitude média foi calculada com base nas amplitudes dos intervalos HPD de 95%. Sendo a amplitude dada por:

$$Amp = LS - LI, \quad (4.8)$$

em que LS e LI representam, respectivamente, o limite superior e o limite inferior da estimativa intervalar de 95% para os níveis de retorno. Dessa forma, a amplitude média é dada por,

$$Amp_m = \frac{\sum_{i=1}^4 Amp_i}{4}, \quad (4.9)$$

em que $i = 1, 2, 3, 4$, correspondendo aos níveis de retorno para 2, 4, 6 e 8 anos, respectivamente.

Para selecionar a melhor estrutura seguiu-se a seguinte ordem: primeiro foram ordenadas as estruturas em relação aos valores de DIC, descartando-se as estruturas com maior valor de DIC, em que $\Delta_{DIC} > 10$. No segundo momento avaliou-se a acurácia das estruturas que possuem $\Delta_{DIC} < 10$. Permanecendo o empate, avaliou-se o erro médio de predição e a amplitude média como critérios adicionais.

Com base na estrutura selecionada, foram estimados os níveis de retorno para os períodos de 5, 10, 20, 30, 50, 100 e 200 anos, utilizando-se a série completa de veranico máximo do município de Brejinho - PE.

4.8 RECURSOS COMPUTACIONAIS

Para a realização das análises estatísticas foi utilizado o *software* computacional estatístico R (R CORE TEAM, 2021) e seus respectivos pacotes: *coda* (Plummer *et al.*, 2006), *evd* (Stephenson, 2002) e *evdbayes* (Stephenson; Ribatet., 2023). O pacote *evdbayes* (Stephenson; Ribatet., 2023) foi utilizado para obter as cadeias *a posteriori* das estruturas 1 e 2 (NI e INF) via MCMC utilizando Metrópolis-Hastings. O algoritmo de Metrópolis-Hastings para obter as cadeias *a posteriori* da estrutura 3 (MP₁ a MP₅), foram implementadas em R (R CORE TEAM, 2021), especialmente para os objetivos deste estudo.

5 RESULTADOS E DISCUSSÃO

5.1 ESTATÍSTICA DESCRITIVA

As estatísticas descritivas para a característica veranico máximo, expressa em anos dos municípios pertencentes a bacia do rio Pajeú, localizada no Sertão Pernambucano estão submetidas na Tabela 4.

Os resultados revelaram uma variação considerável nas médias dos veranicos máximos, variando de 12,7 dias, no município de Flores, a 43,3 dias, em Tuparetama. Nota-se que em locais como Flores (CV de 70,0%) e Santa Terezinha (CV de 73,7%), a alta variabilidade aponta para veranicos menos previsíveis, com variações expressivas de um ano para outro. Por outro lado, Tabira apresenta o menor coeficiente de variação (35,2%), o que sugere uma menor variabilidade nos valores de veranicos máximos ao longo dos anos e uma variação anual mais estável.

Observa-se na Tabela 4 que municípios como Betânia, Carnaíba, Santa Terezinha e Tuparetama registram valores máximos de 120 dias, sugerindo que esses locais podem enfrentar períodos de estiagem prolongada, mesmo durante a estação mais úmida. Já em municípios como Triunfo, com valor máximo de 43 dias, a amplitude é menor, possivelmente refletindo uma menor variabilidade nos eventos de veranico máximo.

Em Brejinho - PE, a média dos veranicos máximos durante a estação chuvosa é de 28,9 dias, o que indica a ocorrência de períodos de estiagem relativamente prolongados, mesmo na estação mais úmida. Quanto aos valores mínimos e máximos, observa-se que Brejinho - PE pode enfrentar desde veranicos mais curtos, com um mínimo de 15 dias, até veranicos longos, com máximos de até 58 dias.

Tabela 4 – Estatísticas descritivas de veranicos máximos na estação chuvosa dos municípios pertencentes a bacia do Rio Pajeú.

Posto Pluviométrico	N	Média	DP	CV (%)	MIN	MAX	Q1	Md	Q3
Afogados da Ingazeira	42	25,9	11,5	44,5	11	61	18,0	22,0	30,5
Betânia	58	34,9	21,0	60,1	10	120	22,2	30,0	43,0
Brejinho	21	28,9	11,7	40,6	15	58	19,0	27,0	37,0
Calumbi	29	30,5	17,1	56,1	14	99	20,0	26,0	37,0
Carnaíba	51	30,4	17,8	45,4	9	120	21,0	28,0	34,5
Flores	21	26,9	11,3	42,0	13	64	22,0	23,0	29,0
Ingazeira	11	27,4	16,5	60,0	7	63	18,5	20,0	33,5
Itapetim	56	33,6	17,9	54,1	11	98	20,0	28,5	41,7
Quixabá	48	23,0	9,4	40,7	6	55	18,7	21,0	27,0
Santa Cruz da Baixa Verde	16	26,3	9,9	37,6	13	45	17,7	26,0	33,2
Santa Terezinha	22	40,8	30,0	73,7	15	120	21,2	30,0	49,0
São José do Egito	54	32,2	16,9	52,7	11	94	20,2	27,5	37,5
Serra Talhada	40	27,0	12,1	44,8	8	64	18,7	24,0	34,2
Solidão	20	27,7	12,1	43,7	13	59	20,5	23,0	33,2
Tabira	46	28,5	10,0	35,2	11	60	22,0	26,5	34,7
Triunfo	50	20,0	7,6	38,1	7	43	15,0	18,5	25,0
Tuparetama	28	43,3	19,6	45,4	15	120	32,5	41,0	49,0

Fonte: Elaborada pela autora.

Nota: N: Número de Observações; DP: Desvio Padrão; CV: Coeficiente de Variação (%); MIN: Mínimo; MAX: Máximo; Q1: 1º Quartil; Md: Mediana; Q3: 3º Quartil.

5.2 OBTENÇÃO DA *PRIORI*

Na Tabela 5, são apresentados os resultados dos testes de Ljung Box e Cox-Stuart dos municípios pertencentes a bacia do rio Pajeú, localizada no Sertão Pernambucano.

Tabela 5 – Resultados dos testes de hipóteses (valor- p), Ljung-Box e Cox-Stuart ao nível de significância de 5%, referente aos dados de veranicos máximos na estação chuvosa dos municípios pertencentes a bacia do Rio Pajeú.

Posto Pluviométrico	Ljung-Box	Cox-Stuart
Brejinho	0,403	0,687
Afogados da Ingazeira	0,649	0,824
Betânia	0,433	0,036
Calumbi	0,229	0,013
Carnaíba	0,991	0,307
Flores	0,724	0,753
Ingazeira	0,373	1,00
Itapetim	0,114	0,557
Quixabá	0,588	0,405
Santa Cruz da Baixa Verde	0,837	0,289
Santa Terezinha	0,071	0,065
São José do Egito	0,246	0,143
Serra Talhada	0,639	0,503
Solidão	0,028	0,508
Tabira	0,516	0,664
Triunfo	0,815	0,043
Tuparetama	0,843	0,180

Fonte: Elaborada pela autora.

Os resultados do teste de Ljung-Box, a um nível de significância de 5%, indicam que todas as séries de veranicos máximos são independentes, exceto para o município de Solidão, que apresentou um valor- $p = 0,028$.

Em relação ao teste de Cox-Stuart, que também adotando um nível de significância de 5%, observou-se que as séries anuais de veranico máximo nos municípios de Betânia, Calumbi e Triunfo não apresentam estacionariedade, indicando a existência de uma tendência significativa.

Para os municípios em que as pressuposições de independência e estacionariedade foram satisfeitas, a distribuição GEV foi ajustada via inferência bayesiana utilizando *priori* não informativa. Os resultados das estimativas dos parâmetros da distribuição GEV estão apresentados na Tabela 6.

Tabela 6 – Estimativas dos parâmetros (Desvio Padrão) da distribuição GEV via inferência bayesiana, utilizando *priori* não informativa e resultados do teste de Kolmogorov-Smirnov (KS) ao nível de significância de 5%

Posto Pluviométrico	Parâmetros			Valor-p (Teste KS)
	μ	σ	ξ	
Afogados da Ingazeira	20,220 (1,360)	7,711 (1,160)	0,197 (0,139)	0,634
Carnaíba	22,639 (1,666)	10,579 (1,377)	0,189 (0,113)	0,452
Flores	21,795 (1,822)	7,511 (1,620)	0,198 (0,174)	0,380
Ingazeira	19,891 (4,779)	13,838 (5,051)	0,212 (0,332)	0,367
Itapetim	24,723 (1,698)	11,068 (1,455)	0,248 (0,126)	0,489
Quixabá	18,863 (1,240)	7,864 (0,938)	-0,014 (0,095)	0,384
Santa Cruz da Baixa Verde	21,692 (2,709)	9,241 (2,557)	0,051 (0,351)	0,481
Santa Terezinha	24,563 (3,041)	12,078 (3,706)	0,727 (0,292)	0,621
São José do Egito	24,163 (1,534)	10,011 (1,285)	0,0234 (0,113)	0,637
Serra Talhada	21,404 (1,654)	9,261 (1,302)	0,082 (0,125)	0,755
Tabira	23,913 (1,358)	8,235 (1,037)	0,017 (0,114)	0,507
Tuparetama	34,973 (2,838)	14,015 (2,312)	0,099 (0,124)	0,297

Fonte: Elaborada pela autora.

Pelos resultados do teste de Kolmogorov-Smirnov (Tabela 6), pôde-se observar que a

distribuição GEV ajustada via inferência bayesiana e *priori* não informativa, ajustou às séries de veranicos máximos anuais dos municípios pertencentes a bacia do Rio Pajeú.

5.3 PREVISÃO DE VERANICOS MÁXIMOS PARA BREJINHO - PE

A Tabela 7 apresenta os resultados das estimativas dos parâmetros da distribuição GEV via inferência bayesiana para o município de Brejinho - PE no período de 2000 a 2013, com seus respectivos intervalos HPD de 95% e dos critérios de convergência.

Tabela 7 – Estimativas dos parâmetros da distribuição GEV via inferência bayesiana para o município de Brejinho - PE no período de 2000 a 2013, utilizando *priori* não informativa, informativa e mistura de *prioris*, com seus respectivos intervalos HPD de 95% e resultados dos critérios de convergência.

<i>Prioris</i>	Parâmetros	Média a <i>posteriori</i>	<i>HPD</i> _{95%}	Critérios		
				ZG	R.L	H.W
NI	μ	23,984	[17,984; 30,579]	-1,066	1,090	0,368
	σ	10,067	[4,917; 16,532]	-0,353	1,090	0,702
	ξ	0,224	[-0,379; 0,889]	-0,252	1,080	0,242
INF	μ	24,757	[21,852; 27,416]	-0,352	1,070	0,399
	σ	10,532	[8,299; 13,035]	-1,215	1,020	0,142
	ξ	0,228	[-0,026; 0,457]	-0,399	0,989	0,765
MP ₁	μ	24,068	[20,521; 27,512]	-1,153	1,050	0,619
	σ	9,697	[6,916; 12,837]	1,084	1,090	0,172
	ξ	0,198	[-0,091; 0,474]	-0,830	1,050	0,268
MP ₂	μ	23,980	[20,325; 27,275]	1,422	0,982	0,185
	σ	9,688	[6,702; 12,618]	0,979	1,000	0,557
	ξ	0,202	[-0,097; 0,470]	0,712	1,020	0,676
MP ₃	μ	24,005	[20,315; 27,363]	1,479	1,030	0,189
	σ	9,694	[6,637; 12,699]	1,431	1,050	0,399
	ξ	0,203	[-0,094; 0,484]	0,468	0,982	0,619
MP ₄	μ	23,978	[20,199; 27,287]	-0,538	1,000	0,708
	σ	9,672	[6,701; 12,726]	-0,607	1,050	0,538
	ξ	0,203	[-0,104; 0,487]	-1,367	1,000	0,074
MP ₅	μ	24,108	[20,467; 27,078]	1,246	1,020	0,669
	σ	9,738	[7,009; 12,577]	-0,054	0,971	0,436
	ξ	0,196	[-0,081; 0,444]	-1,852	1,020	0,542

Fonte: Elaborada pela autora.

Nota: ZG: Estatística do critério de Geweke; R.L: Fator do critério de Raftery e Lewis;

H.W: Valor-p do critério de Heidelberger-Welch.

Conforme apresentado na Tabela 7, o zero está dentro do intervalo de credibilidade do parâmetro de forma. Entretanto, segundo Coles e Pericchi (2003), a redução da distribuição

GEV para a distribuição Gumbel deve ser feita com cautela, mesmo quando testes padrões a indicam, pois a distribuição GEV estima melhor os quantis extremos.

Os resultados apresentados na Tabela 7 indicam que para todos os parâmetros das *prioris* não informativa, informativa e mistura de *prioris* os fatores de dependência do critério de Raftery e Lewis foram próximos de 1, o que indica independência entre as iterações. Pelo critério de Geweke tem-se que $|ZG| < 1,96$, indicando que não há indícios de ausência de convergência. Os valores-p do critério de Heidelberg e Welch foram não significativos ($p > 0,05$), verificando que a série é estacionária. Sendo assim, não há evidências de ausência de não convergência das cadeias a *posteriori*. Uma vez que não há evidências que indicam a não convergência das cadeias a *posteriori*, os níveis de retorno preditos para os tempos de retorno de 2, 4, 6 e 8 anos foram obtidos via inferência bayesiana.

Para determinar qual das estruturas de *priori* é mais adequada para a localidade em estudo, foi realizado uma análise do Critério de Informação Deviance (DIC), juntamente com a avaliação da acurácia, erro médio de predição (EMP) e amplitude média na previsão dos níveis de retorno.

Tabela 8 – Resultados do Critério de Informação da Deviance (DIC), Acurácia (%), Amplitude Média e Erro Médio de Predição (%) do veranico máximo predito para os tempos de retorno de 2, 4, 6 e 8 anos (valor observado), para o município de Brejinho-PE.

<i>Prioris</i>	Veranico máximo predito e $HPD_{95\%}$				DIC	AC	EMP	AMP _m
	2 anos (40 dias)	4 anos (40 dias)	6 anos (40 dias)	8 anos (45 dias)				
NI	27,8 [20,6; 35,3]	38,9 [26,4; 52,7]	46,3 [30,1; 68,4]	52,4 [32,4; 82,6]	104,2	75%	17,4%	32,4
INF	28,8 [25,6; 31,8]	40,0 [34,9; 45,6]	46,8 [38,8; 54,6]	52,0 [42,5; 62,8]	101,1	75%	15,2%	13,2
MP ₁	27,8 [23,4; 31,6]	37,9 [30,4; 45,1]	44,1 [33,7; 54,1]	48,6 [35,8; 61,2]	100,9	75%	14,0%	17,2
MP ₂	27,7 [23,2; 31,3]	37,8 [30,5; 44,9]	44,0 [33,3; 53,5]	48,6 [35,9; 61,3]	100,9	75%	14,0%	17,1
MP ₃	27,7 [23,1; 31,7]	37,9 [30,3; 45,1]	44,1 [33,7; 54,3]	48,7 [35,6; 61,3]	100,9	75%	14,0%	17,4
MP ₄	27,7 [23,3; 31,6]	37,8 [30,4; 44,9]	44,0 [33,3; 53,9]	48,7 [35,5; 61,2]	100,9	75%	14,0%	17,3
MP ₅	27,8 [23,9; 31,6]	37,9 [30,6; 44,1]	44,0 [34,2; 52,7]	48,6 [37,1; 60,3]	100,7	75%	13,0%	15,7

Fonte: Elaborada pela autora.

AC: Acurácia; AMP_m: Amplitude Média; EMP: Erro Médio de Predição.

A partir dos resultados apresentados na Tabela 8, foi possível selecionar a melhor estrutura para a predição do veranico máximo no município de Brejinho - PE. Inicialmente, foram

analisados os valores do Critério de Informação da Deviance (DIC). Tendo em vista que não houve diferença substancial nos valores de DIC (Tabela 8), outros critérios pra seleção de modelos podem ser aplicados, como parcimônia e medidas de desempenho.

Dessa forma, verificou-se qual das estruturas obteve maior acurácia, ou seja, maior número de acertos. Conforme os resultados apresentados na Tabela 8, todas as estruturas alcançaram uma acurácia de 75%. Em seguida, foram analisados o Erro Médio de Predição e a amplitude média. A estrutura MP_5 destacou-se por apresentar o menor erro médio de predição (13,0%), o que sugere maior capacidade de fornecer estimativas consistentes e confiáveis, especialmente para os eventos extremos de veranico máximo.

Com base nesses resultados, conclui-se que a estrutura MP_5 é a mais apropriada para a predição do veranico máximo em Brejinho - PE.

A escolha da mistura de *prioris* como estrutura final para as predições do veranico máximo justifica-se por diversos fatores. Primeiramente, essa estrutura tem a capacidade de combinar informações de diferentes *prioris*, cada uma contribuindo com aspectos distintos e relevantes para os dados de veranicos máximos.

Contudo, é importante reconhecer que a utilização de uma mistura de *prioris* pode introduzir um grau adicional de variabilidade nas predições. Essa variabilidade reflete a incerteza ao processo de combinação das *prioris*, podendo, em alguns casos, resultar em leve perda de precisão nas estimativas. No entanto, essa aparente limitação é compensada pelo ganho em acurácia observado, reforçando a relevância prática e teórica da mistura de *prioris*.

Observa-se que, em todos os casos analisados, o uso de uma *priori* informativa proporcionou resultados superiores quanto a precisão do veranico máximo em Brejinho-PE, quando comparado à utilização de uma *priori* não informativa.

Esses resultados corroboram com os apresentados por Reis, Beijo e Butturi-Gomes (2024), que ajustou a distribuição GEV aos conjuntos de dados de veranico máximo anual em diferentes regiões de Minas Gerais. A autora concluiu que a *priori* informativa proporcionou maior precisão na análise de veranicos máximos na maioria dos municípios avaliados.

Esses resultados também estão alinhados com os apresentados por Silva *et al.* (2023), que ajustaram a distribuição GEV às séries históricas de precipitação máxima de Petrópolis - RJ e Poços de Caldas - MG e avaliaram diferentes estruturas de distribuições *a priori*. Os autores observaram que a *priori* informativa obteve melhor desempenho para prever as precipitações máximas de Poços de Caldas - MG e Petrópolis - RJ.

De maneira semelhante, os achados de Branco, Oliveira e Beijo (2022) reforçam esse padrão, ao modelar a série de precipitações máximas anuais de Manhuaçu - MG por meio da GEV utilizando inferência bayesiana. Nesse estudo, a *priori* informativa também demonstrou superioridade na estimativa dos níveis de retorno da precipitação máxima.

5.4 ANÁLISE DO VERANICO MÁXIMO DO MUNICÍPIO DE BREJINHO - PE

A Tabela 9 apresenta o resultado dos testes de Ljung-Box e Cox-Stuart, utilizando a série completa de veranico máximo, abrangendo o período de 2000 a 2021. A partir do teste de Ljung-Box, verificou-se que a série é independente, enquanto o teste de Cox-Stuart indicou que a série é estacionária, ou seja, não apresenta tendência.

Tabela 9 – Resultados dos testes (valor-*p*) de Ljung-Box e Cox-Stuart à 5% de significância, para os dados de veranico máximo do município de Brejinho-PE.

Ljung-Box	Cox-Stuart
0,403	1,000

Fonte: Elaborada pela autora.

Na Tabela 10, são apresentadas as estimativas dos parâmetros com respectivos intervalos HPD de 95% de probabilidade e os resultados dos critérios de convergência. Observa-se que os fatores de dependência de Raftery e Lewis estão próximos de 1, a estatística de Geweke $|ZG| < 1,96$, o valores-*p* do teste de Heidelberg-Welch foram não significativos ($p > 0,05$), indicando a convergência das cadeias *a posteriori*.

Tabela 10 – Estimativas dos parâmetros da distribuição GEV via inferência bayesiana, utilizando mistura de *prioris*, com seus respectivos intervalos HPD de 95% e resultados dos critérios de convergência.

Parâmetros	Média a <i>posteriori</i>	$HPD_{95\%}$	Critérios		
			ZG	R.L	H.W
μ	23,653	[20,455; 26,530]	0,544	1,020	0,728
σ	9,507	[6,972; 12,078]	1,417	1,030	0,663
ξ	0,196	[-0,065; 0,459]	-0,047	0,999	0,580

Fonte: Elaborada pela autora.

Nota: ZG: Critério de Geweke; R.L: Critério de Raftery e Lewis; H.W: Critério de Heidelberger-Welch.

Uma vez obtidas as estimativas dos parâmetros, foram realizadas a predição dos tempos de retorno para 5, 10, 20, 30, 50, 100 e 200 anos. A Tabela 11 apresenta os níveis de retorno, seus respectivos intervalos HPD com 95% de probabilidade e critérios de convergência.

Tabela 11 – Níveis de retorno para os tempos de 5, 10, 20, 30, 50, 100 e 200 anos, seus respectivos intervalos HPD com 95% de probabilidade e resultados dos critérios de convergência.

Tempos de retorno	Níveis de retorno	$HPD_{95\%}$	Critérios		
			ZG	R.L	H.W
5	40,4	[32,9; 47,8]	1,086	0,966	0,877
10	51,1	[38,5; 63,8]	0,748	0,966	0,862
20	63,4	[43,5; 84,7]	0,351	0,982	0,765
30	71,5	[46,7; 101,1]	0,144	0,982	0,724
50	83,1	[47,9; 123,8]	-0,110	0,982	0,613
100	101,6	[52,5; 165,9]	-0,455	0,966	0,470
200	124,4	[54,9; 220,1]	-0,801	0,966	0,357

Fonte: Elaborada pela autora.

Nota: ZG: Critério de Geweke; R.L: Critério de Raftery e Lewis; H.W: Critério de Heidelberger-Welch.

Conforme os resultados apresentados na Tabela 11, é visto que não há evidência que indicam a falta de convergência das cadeias, ou seja, os fatores de dependência de Raftery e Lewis estão próximos de 1, a estatística de Geweke apresenta $|ZG| < 1,96$, e os valores $-p$

do teste de Heidelberg-Welch foram não significativos ($p > 0,05$). Dessa forma, seguiu-se a análise dos níveis de retorno.

Uma interpretação prática para os níveis de retorno de Brejinho - PE, considerando a Tabela 11 e o tempo de retorno do 30 anos é a seguinte: espera-se que em um tempo médio de 30 anos, durante a estação chuvosa, ocorra pelo menos um veranico maior ou igual a 71 dias, além disso, acredita-se com 95% de probabilidade que ocorra um veranico maior ou igual a um valor que esteja entre 46,7 e 101,1 dias.

Esses resultados podem ser comparados com os apresentados por Reis, Beijo e Butturi-Gomes (2024), que avaliou a probabilidade de ocorrência de veranicos em algumas regiões do estado de Minas Gerais. A autora relatou uma probabilidade elevada de veranicos com duração superior a 30 dias na estação chuvosa para áreas centrais e ao norte do estado, com destaque para o município de Sete Lagoas, que apresentou uma probabilidade de ocorrência de veranicos com duração acima de 70 dias. Embora existam diferenças regionais, os resultados apontam para fatores climáticos comuns que influenciam os veranicos em ambas as regiões. Reis, Beijo e Butturi-Gomes (2024) associou o atraso da estação chuvosa na região norte de Minas Gerais à Zona de Convergência do Atlântico Sul (ZCAS). De forma semelhante, os padrões observados em Brejinho - PE podem ser parcialmente explicados pela irregularidade pluviométrica do Sertão Pernambucano, intensificada por fenômenos como o *El Niño*. Como destacado por Marengo *et al.* (2008), a alta variabilidade espacial e temporal das chuvas no semiárido nordestino torna as secas uma característica inevitável da região.

Além dos aspectos climáticos regionais, os resultados também revelam uma relação direta entre o tempo de retorno e os níveis de intensidade dos veranicos. À medida que o tempo de retorno aumenta, os níveis de retorno também aumentam, o que reflete uma tendência esperada para eventos extremos: eventos mais intensos estão associados a períodos mais longos. Esses resultados corroboram os achados de Assis, Sobral e Souza (2012), que analisaram as tendências dos índices de detecção de mudanças climáticas nos rios Brígida e Pajeú, ambos localizados no Sertão Pernambucano. As autoras encontraram uma diminuição no regime de chuvas ao longo dos anos, além de um aumento no número de dias secos consecutivos.

Soares e Nóbrega (2010), ao conduzir um estudo climatológico sobre o comportamento e a frequência dos veranicos no Sertão de Pernambuco, verificaram que, em alguns municípios, os períodos sem chuva podem alcançar uma média de até 120 dias consecutivos. Os autores apontam que esse padrão ressalta a forte concentração temporal das precipitações no semiárido

pernambucano, tanto em escala anual – com chuvas limitadas a poucos meses – quanto dentro dos próprios períodos chuvosos, em que as precipitações se concentram historicamente em um pequeno número de dias.

Os resultados obtidos neste estudo mostram que, para tempos de retorno superiores a 10 anos, os níveis de retorno dos veranicos ultrapassam os 50 dias, alcançando valores extremos de 124 dias para tempos de 200 anos. Essas estimativas, especialmente para tempos mais longos, indicam a presença de eventos extremos com potencial significativo para impactar atividades agrícolas na região. Silva (2014), ao estudar a pluviometria da bacia do rio Paranaíba, destacou que veranicos superiores a 13 dias consecutivos podem causar perdas substanciais às lavouras, devido à combinação da ausência de precipitação e ao aumento das taxas de evapotranspiração. Nesse contexto, os níveis de retorno estimados neste estudo indicam que os veranicos em períodos mais longos representam um risco elevado para a agricultura de sequeiro, que depende exclusivamente da chuva e da capacidade de retenção hídrica do solo, cujas características são diretamente influenciadas pelo clima.

Esses resultados reforçam a necessidade de estratégias de adaptação específicas, como o manejo integrado da água e o uso de culturas mais resistentes à seca. A tendência de aumento na duração dos veranicos ao longo do tempo, observada tanto neste estudo quanto nos de Assis, Sobral e Souza (2012) e Soares e Nóbrega (2010), evidencia a importância de planejar ações que minimizem os impactos dos veranicos em cadeias produtivas agrícolas.

6 CONCLUSÕES

Com base nos resultados obtidos, conclui-se que a série de dados de veranico máximo do município de Brejinho - PE não apresentou evidências de tendência. A distribuição GEV revelou-se apropriada para modelar esses dados, enquanto a estrutura com mistura de *prioris* demonstrou melhor desempenho na previsão dos veranicos máximos para o município. Conclui-se também que em todos os casos, o uso de uma *priori* informativa proporcionou resultados superiores na predição do veranico máximo em Brejinho-PE, quando comparado à utilização de uma *priori* não informativa.

Sendo assim, os resultados obtidos neste trabalho podem ser utilizados por agricultores e gestores, interessados em compreender os impactos dos veranicos sobre a agricultura na região de Brejinho - PE. Essas informações podem auxiliar na escolha de culturas mais resistentes à estiagem, no planejamento das épocas de plantio e colheita, bem como no desenvolvimento de estratégias para minimizar os efeitos da falta de chuva, garantindo maior estabilidade na produção agrícola e na segurança alimentar local.

Uma possível direção para estudos futuros seria estender a aplicação da mistura de *prioris* a diferentes regiões do semiárido e além, com o objetivo de verificar se os padrões encontrados se mantêm consistentes quando a metodologia é aplicada a áreas com características climáticas distintas. Além disso, seria interessante testar diferentes esquemas de pesos para as *prioris*, investigando como a escolha desses pesos pode influenciar os resultados, especialmente em regiões com características climáticas variadas.

REFERÊNCIAS

ASSIS, J.; SOBRAL, M. d. C. M.; SOUZA, W. Análise de detecção de variabilidades climáticas com base na precipitação nas bacias hidrográficas do sertão de pernambuco. **Revista Brasileira de Geografia Física**, Recife, v. 3, p. 630–645, 2012.

ASSIS, J. M. O. **Análise de tendências de mudanças climáticas no semiárido de Pernambuco**. 2012. 166 p. Dissertação (Mestrado em Desenvolvimento e Meio Ambiente), Universidade Federal de Pernambuco, Recife, PE, 2012.

BAUTISTA, E. A. L. **A distribuição generalizada de valores extremos no estudo da velocidade máxima de vento em Piracicaba, SP**. 2002. 49 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica), Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, SP, 2002.

BAUTISTA, E. A. L.; 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. **Matemática e Estatística**, São Paulo, v. 22, n. 1, p. 95–111, 2004.

BEIJO, L. A. **Construção de uma priori para os parâmetros do modelo de valores extremos generalizado baseada em quantis com distribuição Gumbel**. 2006. 110 p. Tese (Doutorado em Estatística e Experimentação Agropecuária), Universidade Federal de Lavras, Lavras, MG, 2006.

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, v. 1, n. 1, p. 10–16, 2011.

BERGER, J. O. **Statistical decision theory and Bayesian analysis**. New York: Springer-Verlag, 1985.

BORTKIEWICZ, L. V. Variationsbreite and mittlerer fehler. **Sitzungsber. Berli. Math. Ges.**, S.1., v. 21, p. 3–11, 1922.

BOX, G. E.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: Wiley, 1992.

BRANCO, K. P.; OLIVEIRA, A. d. C. de; BEIJO, L. A. Predição da precipitação máxima de Manhuaçu-MG via abordagem bayesiana. **Pensar Acadêmico**, Manhuaçu, v. 20, n. 2, p. 452–469, 2022.

CARVALHO, D. F. *et al.* Espacialização do período de veranico para diferentes níveis de perda de produção na cultura do milho, na bacia do Rio Verde Grande, MG. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 4, n. 2, p. 172–176, 2000.

- COLES, S. **An introduction to statistical modeling of extreme values**. London: Springer, 2001.
- COLES, S.; PERICCHI, L. Anticipating catastrophes through extreme value modelling. **Journal of the Royal Statistical Society Series C: Applied Statistics**, Oxford University Press, v. 52, n. 4, p. 405–416, 2003.
- COLES, S. G.; DIXON, M. J. Likelihood-based inference for extreme value models. **Extremes**, Springer, v. 2, n. 1, p. 5–23, 1999.
- COLES, S. G.; POWELL, E. A. Bayesian methods in extreme value modelling: a review and new developments. **International Statistical Review**, JSTOR, v. 64, p. 119–136, 1996.
- COSTA, M. d. S.; BEIJO, L. A.; AVELAR, F. G. Bayesian modeling of the maximum streamflows from the furnas reservoir. **Engenharia Sanitaria e Ambiental**, SciELO Brasil, v. 27, n. 4, p. 693–699, 2022.
- COX, D. R.; STUART, A. Some quick sign tests for trend in location and dispersion. **Biometrika**, JSTOR, v. 42, n. 1/2, p. 80–95, 1955.
- CUPOLILLO, F. **Diagnóstico Hidroclimatológico da Bacia do Rio Doce**. 2008. 156 p. Tese (Doutorado em Geografia), Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2008.
- FARIAS, C. H. d. A. *et al.* Índices de crescimento da cana-de-açúcar irrigada e de sequeiro no estado da paraíba. **Revista Brasileira de Engenharia agrícola e ambiental**, SciELO Brasil, v. 12, p. 356–362, 2008.
- FERNANDES, F. B. P. **Disponibilidade hídrica para a cultura do feijão-de-corda em função do manejo de solo no semiárido cearense**. 2014. 112 p. Tese (Doutorado em Engenharia Agrícola), Universidade Federal do Ceará, Fortaleza, CE, 2014.
- FERREIRA, T. R. *et al.* Modelagem bayesiana da temperatura máxima do ar em divinópolis-mg. **Nativa**, v. 12, n. 3, p. 449–456, 2024.
- FERREIRA, V. d. P. M. *et al.* Predição da temperatura máxima de lavras-mg: comparação da normal climatológica com a teoria de valores extremos. **OBSERVATÓRIO DE LA ECONOMÍA LATINOAMERICANA**, v. 22, n. 5, p. e4896–e4896, 2024.
- FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: **Mathematical proceedings of the Cambridge philosophical society**. Cambridge: Cambridge University Press, 1928. v. 24, n. 2, p. 180–190.
- GAMERMAN, D.; LOPES, H. F. **Markov chain Monte Carlo: stochastic simulation for Bayesian inference**. Broken Sound Parkway: Chapman and Hall, 2006.
- GELFAND, A. E. *et al.* Illustration of bayesian inference in normal data models using gibbs sampling. **Journal of the American Statistical Association**, Taylor & Francis, v. 85, n. 412, p. 972–985, 1990.

- GELMAN, A. *et al.* **Bayesian data analysis**. Boca Raton: FL: CRC press, 2014.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, n. 6, p. 721–741, 1984.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. **Bayesian statistics**, Oxford University Press, v. 4, n. 1, p. 641–649, 1992.
- GNEDENKO, B. Sur la distribution limite du terme maximum d'une serie aleatoire. **Annals of mathematics**, JSTOR, p. 423–453, 1943.
- HAMMERSLEY, J. M.; HANDSCOMB, D. C. General principles of the monte carlo method. In: **Monte Carlo Methods**. [S.l.]: Springer, 1964. p. 50–75.
- HAMRA, G.; MACLEHOSE, R.; RICHARDSON, D. Markov chain monte carlo: an introduction for epidemiologists. **International journal of epidemiology**, Oxford University Press, v. 42, n. 2, p. 627–634, 2013.
- HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. **Operations Research**, INFORMS, v. 31, n. 6, p. 1109–1144, 1983.
- JEFFREYS, H. **Theory of probability**. Oxford: Clarendon, 1961.
- JEFFREYS, H. J. The times of P, S and SKS, and the velocities of P and S. **Geophysical Journal International**, Wiley Online Library, v. 4, p. 498–533, 1939.
- JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quarterly Journal of the Royal Meteorological Society**, Wiley Online Library, v. 81, n. 348, p. 158–171, 1955.
- KOLMOGOROV, A. Sulla determinazione empirica di una legge didistribuzione. **Giorn Dell'inst Ital Degli Att**, v. 4, p. 1–11, 1933.
- LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. **Biometrika**, Oxford University Press, v. 65, n. 2, p. 297–303, 1978.
- MARENGO, J. A. *et al.* Vulnerabilidade, impactos e adaptação à mudança do clima no semi-árido do brasil. **Parcerias estratégicas**, v. 13, n. 27, p. 149–176, 2008.
- MARTINS-FILHO, S. *et al.* Abordagem bayesiana das curvas de crescimento de duas cultivares de feijoeiro. **Ciência Rural**, SciELO Brasil, v. 38, n. 6, p. 1516–1521, 2008.
- MENDES, B. V. M. **Introdução à análise de eventos extremos**. Rio de Janeiro: Editora E-papers, 2004.
- MENEZES, H. E. A.; BRITO, J. I. B.; LIMA, R. A. F. A. Veranico e a produção agrícola no estado da paraíba, Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, SciELO

Brasil, v. 14, n. 2, p. 181–186, 2010.

MINUZZI, R. B. *et al.* El niño: ocorrência e duração dos veranicos do estado de Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, SciELO Brasil, v. 9, n. 3, p. 364–371, 2005.

NETO, P. C.; VILLELA, E. A. Veranico: um problema de seca no período chuvoso. **Informe Agropecuário**, Belo Horizonte, v. 12, n. 138, p. 59–62, 1986.

NOGUEIRA, D. A.; SAFADI, T.; FERREIRA, D. Avaliação de critérios de convergência para o método de monte carlo via cadeias de markov. **Revista Brasileira de Estatística**, IBGE, v. 65, n. 224, p. 59–88, 2004.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003.

PLUMMER, M. *et al.* Coda: Convergence diagnosis and output analysis for mcmc. **R News**, New Zealand, v. 6, n. 1, p. 7–11, 2006. Disponível em: <<https://journal.r-project.org/archive/>>. Acesso em: 10 nov. 2022.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <<http://www.R-project.org/>>. Acesso em: 15 set. 2022.

RAFTERY, A.; LEWIS, S. **How Many Iterations in the Gibbs Sampler Bayesian Statistics**. Oxford: Oxford University Press, 1992.

RAFTERY, A. E.; LEWIS, S. M. Implementing MCMC. in: Gilks, W. R.; Richardson, S.; Spiegelhalter, D. J. (Ed.). **Markov chain Monte Carlo in practice**, London, p. 115–130, 1995.

REIS, C. G. dos; BEIJO, L. A.; BUTTURI-GOMES, D. Espacialização da probabilidade de ocorrência de veranicos extremos em regiões de Minas Gerais: uma abordagem bayesiana. **Revista de Gestão e Secretariado**, v. 15, n. 12, p. e4440, 2024.

ROSSI, R. M. **Introdução aos métodos Bayesianos na análise de dados zootécnicos com uso do WinBUGS e R**. Maringá: Eduem, 2011. v. 1.

SILVA, M. M. **Variabilidade do início e fim da estação chuvosa e ocorrência de veranicos na bacia hidrográfica do Rio Paranaíba**. 2014. 99 p. Dissertação (Mestrado em Geografia), Universidade Federal de Uberlândia, Uberlândia, MG, 2014.

SILVA, S. V. C. d. *et al.* Modelagem bayesiana da precipitação máxima de Petrópolis (RJ) e Poços de Caldas (MG). **Engenharia Sanitária e Ambiental**, SciELO Brasil, v. 28, p. e20210342, 2023.

SLEIMAN, J. **Veranicos ocorridos na porção noroeste do estado do Rio Grande do Sul entre 1978 e 2005 e sua associação às condições climáticas na atmosfera**. 2008. 164 p. Dissertação (Mestrado em Desenvolvimento e Meio Ambiente), Universidade de São Paulo, São Paulo, SP, 2008.

SMIRNOV, H. Sur les fucarts de la courbe de distribution empirique. In **Recueil Mathematique: Matemnticeskii Sbornik**, v. 48, n. 1, p. 3–26, 1939.

SOARES, D. B.; NÓBREGA, R. S. Análise espacial e climatológica da ocorrência de veranicos no sertão de pernambuco. **Revista de Geografia**, v. 27, n. 1, p. 95–106, 2010.

SPIEGELHALTER, D. J. *et al.* Bayesian measures of model complexity and fit. **Journal of the royal statistical society: Series b (statistical methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002.

STEPHENSON, A.; RIBATET., M. **evdbayes: Bayesian Analysis in Extreme Value Theory**. [S.l.], 2023. Disponível em: <<https://CRAN.R-project.org/package=evdbayes>>. Acesso em: 20 abr. 2023.

STEPHENSON, A. G. evd: Extreme value distributions. **R News**, v. 2, n. 2, p. 0, June 2002. Disponível em: <<https://CRAN.R-project.org/doc/Rnews/>>. Acesso em: 12 abr. 2023.

TIERNEY, L. Markov chains for exploring posterior distributions. **Annals of Statistics**, Hayward, v. 22, n. 4, p. 1701–1728, 1994.

WASHINGTON, B. T. **Up from slavery**. [S.l.]: Simon and Schuster, 2013.

APÊNDICE

APÊNDICE A - Rotina em R utilizada para realizar a mistura de *prioris*

```

require (evd)
library (mvtnorm)
require (mvtnorm)
rm(list=ls ())

setwd("C:\\Users\\Cliente\\Desktop\\Mestrado\\Dissertacao\\Dados_
    atualizados\\Veranicos_Maximos")

### Dados
dados=read.table ("Brejinho.txt", dec=".", h=T)
dados

## Armazenando cada ano do estudo em y

y=dados$Veranico[1:13];y
y<-as.numeric (y)
yop=dados$Veranico[14:21];yop

### Prioris
pv1 <- list (mean = c (19.754, 2.022, 0.233), cov=1*cbind (c (2.200, 0, 0),
    c (0, 0.027, 0), c (0, 0, 0.022)))
pv2 <- list (mean = c (22.612, 2.345, 0.187), cov=1*cbind (c (2.806, 0, 0), c
    (0, 0.017, 0), c (0, 0, 0.012)))
pv3 <- list (mean = c (21.795, 1.994, 0.198), cov=1*cbind (c (3.321, 0, 0), c
    (0, 0.045, 0), c (0, 0, 0.030)))
pv4 <- list (mean = c (19.891, 2.565, 0.212), cov=1*cbind (c (22.844, 0, 0),
    c (0, 0.125, 0), c (0, 0, 0.110)))
pv5 <- list (mean = c (24.723, 2.395, 0.248), cov=1*cbind (c (2.884, 0, 0),
    c (0, 0.017, 0), c (0, 0, 0.016)))
pv6 <- list (mean = c (18.863, 2.055, -0.014), cov=1*cbind (c (1.538, 0, 0),
    c (0, 0.014, 0), c (0, 0, 0.009)))
pv7 <- list (mean = c (21.692, 2.187, 0.051), cov=1*cbind (c (7.336, 0, 0), c

```

```

      (0,0.074,0),c(0,0,0.123)))
pv8 <- list(mean = c(24.563, 2.446,0.727), cov=1*cbind( c(9.249,0,0),
      c(0,0.089,0),c(0,0,0.085)))
pv9 <- list(mean = c(24.163, 2.295,0.233), cov=1*cbind( c(2.354,0,0),
      c(0,0.016,0),c(0,0,0.013)))
pv10 <- list(mean = c(21.404, 2.216,0.082), cov=1*cbind(c(2.737,0,0),
      c(0,0.019,0),c(0,0,0.016)))
pv11 <- list(mean = c(23.913, 2.100,0.017), cov=1*cbind(c(1.844,0,0),
      c(0,0.016,0),c(0,0,0.013)))
pv12 <- list(mean = c(34.973, 2.627,0.099), cov=1*cbind( c(8.052,0,0)
      ,c(0,0.027,0),c(0,0,0.015)))

minhas.prioris <- list(pv1, pv2, pv3, pv4, pv5, pv6, pv7, pv8, pv9,
      pv10, pv11, pv12)

### Pesos de cada priori
meus.pesos <- c(0.041900213, 0.033295998, 0.026719341, 0.058671262,
      0.267720724, 0.032846564, 0.023025538, 0.113123959, 0.228408873,
      0.021041819, 0.065973315,
      0.087272394)

### CONTROLES DA CADEIA
st      <- c(5,1,0.1)# valores de inicio
chain <- NULL      # objeto que armazena a cadeia
N.chain <- 100000  # tamanho da cadeia (numero de iteracoes), antes
      de burn e thin
burn   <- 2000     # burn-in
thin   <- 20       # thin
ac.nt  <- 0        # (apenas um contador)
it     <- 0        # (apenas um contador)

### CONTROLES DA PROPOSTA
Sig.pr   <- rbind( c(1e-3,0,0),c(0,.1,-.02),c(0,-.02,.08) )
upd.sig  <- T      # Se FALSE, desliga a adaptacao da matriz de

```

```

    covariancias da proposta Sig.pr
upd.after <- 20      # comeca a adaptar a matriz de covariancias da
    proposta depois de 20 iteracoes
upd.each  <- 100    # adpata a matriz de covariancias a cada 100
    iteracoes
upd.p     <- 0.5    # usa 0.5 das 100 ultimas iteracoes para adaptacao
    da matriz de covariancias
upd.stop  <- 0.75  # depois de completar 75% do total de iteracoes (N
    .chain), para de adaptar
upd.n     <- 1      # (apenas um contador)

# Funcao para calcular log(verossimilhanca) + log( SOMA( pesos*prioris
    ) )
prior.llh <- function(pars,data,priors,weights){
  weights <- weights/sum(weights)
  prior.v <- 0
  if(length(weights)!=length(priors)) stop('Use_mesmo_numero_de_pesos
    _e_de_priors')
  for(i in 1:length(weights)){
    prior.v <- prior.v + weights[i]*dmvnorm(pars, mean = priors[[i]]$
      mean, sigma = priors[[i]]$cov, log = F)
  }
  llh.v  <- sum( dgev(data, loc=pars[1], scale=exp(pars[2]), shape=
    pars[3], log = T) )
  return(llh.v + log(prior.v) )
}

# Funcao para verificar se a matriz e positiva definida
is.positive.definite <- function(mat) {
  eigen(mat)$values > 0
}

### ALGORITMO
while( ac.nt < N.chain ){

```

```

prop.d <- rmvnorm(1,mean=st,sigma=Sig.pr)
R <- exp( prior.llh(prop.d,y,minhas.prioris,meus.pesos) - prior.llh
  (st,y,minhas.prioris,meus.pesos) )
while( is.nan(R) || !is.finite(R) ){
  prop.d <- rmvnorm(1,mean=st,sigma=Sig.pr)
  R <- exp( prior.llh(prop.d,y,minhas.prioris,meus.pesos) - prior.
    llh(st,y,minhas.prioris,meus.pesos) )
}
if( rbinom(1,size=1,prob=min(c(R,1)))==1 ){
  ac.nt <- ac.nt + 1
  st <- prop.d
  chain <- rbind(chain,st)
}
if(upd.sig &&
  ac.nt == (upd.after + upd.n*upd.each) &&
  ac.nt < upd.stop*N.chain
){
  Sig.pr <- var(chain[(ac.nt - upd.each*upd.p):(ac.nt),])
  upd.n <- upd.n + 1
}
it <- it + 1
ac.rt <- ac.nt/it
}

length(st) == ncol(Sig.pr) && nrow(Sig.pr) == ncol(Sig.pr)

chain <- chain[-c(1:burn),]
N.chain.eff <- nrow(chain)
chain <- chain[seq(1,N.chain.eff,thin), ]
N.chain.eff <- nrow(chain)
chain[,2] <- exp(chain[,2])
post <- chain
colnames(post) <- c("loc","sigma","xi")

```

```

### RESULTADOS
#=====
#   UTILIZANDO O PACOTE CODA para analisar convergencia das cadeias
#=====

library("coda") # Pacote para analisar as cadeias a posteriori

param.mcmcMIS <- as.mcmc(post)

summary(param.mcmcMIS)

HPDinterval(param.mcmcMIS)

raftery.diag(param.mcmcMIS)
geweke.diag(param.mcmcMIS)
heidel.diag(param.mcmcMIS)

# Analise grafica de convergencia
plot(param.mcmcMIS)

#=====
## Salvando as estimativas medias a posteriori
#=====

mu2<-mean(param.mcmcMIS[,1]); mu2    ## Estimativa do parametro
    posicao (Gev)
sig2<-mean(param.mcmcMIS[,2]); sig2  ## Estimativa do parametro
    escala (GEV)
xi2<-mean(param.mcmcMIS[,3]); xi2    ## Estimativa do parametro Forma(
    GEV)

#=====
#   Teste de Kolmogorov Smirnov
#=====

```

```
ks.test(y, "pgev", mu2, sig2, xi2, alternative="greater") # Dist GEV
```

```
#=====
#           Rotina R para PREDITIVA
#=====
```

```
k<-length(post[,1])
```

```
loc <- numeric(k)
```

```
escala <- numeric(k)
```

```
forma <- numeric(k)
```

```
TR=c(2,4,6,8) #INDICANDO tempos de retorno,
```

```
prob= 1/TR
```

```
np<-length(prob)
```

```
q1 <- numeric(k)
```

```
q2 <- numeric(k)
```

```
q3 <- numeric(k)
```

```
q4 <- numeric(k)
```

```
for (i in 1:k)
```

```
{
```

```
  loc[i] <- post[i,1]
```

```
  escala[i] <- post[i,2]
```

```
  forma[i] <- post[i,3]
```

```
  q1[i]<-(loc[i]+((escala[i]/forma[i])*((-log(1-prob[1]))^
                                                    (-forma[i]))-1)))
```

```
  q2[i]<-(loc[i]+((escala[i]/forma[i])*((-log(1-prob[2]))^
                                                    (-forma[i]))-1)))
```

```
  q3[i]<-(loc[i]+((escala[i]/forma[i])*((-log(1-prob[3]))^
                                                    (-forma[i]))-1)))
```

```
  q4[i]<-(loc[i]+((escala[i]/forma[i])*((-log(1-prob[4]))^
```

```

                                                                    (-forma[i]))-1)))
}
nit<-length(q1)
qoMIS<-matrix(0,nit,4)
qoMIS[,1]<-q1
qoMIS[,2]<-q2
qoMIS[,3]<-q3
qoMIS[,4]<-q4

#=====
#      Analise dos niveis de retorno
#=====

nivel_ret_mcmcMIS <- as.mcmc(qoMIS)

summary(nivel_ret_mcmcMIS)

HPDinterval(nivel_ret_mcmcMIS)

raftery.diag(nivel_ret_mcmcMIS)
geweke.diag(nivel_ret_mcmcMIS)
heideldiag(nivel_ret_mcmcMIS)

plot(nivel_ret_mcmcMIS)

#=====
#      CALCULANDO ERRO MEDIO DE PREDICAO
#=====

Vo = c(40,40,40,45) # Valores observados

q10i<-mean(nivel_ret_mcmcMIS[,1])
q20i<-mean(nivel_ret_mcmcMIS[,2])
q30i<-mean(nivel_ret_mcmcMIS[,3])
q40i<-mean(nivel_ret_mcmcMIS[,4])

```

```

pred_bayes_MIS=c(q10i,q20i,q30i,q40i);pred_bayes_MIS

EpB_MIS= abs ((Vo-pred_bayes_MIS)/Vo);EpB_MIS

round(mean(EpB_MIS)*100)

head(post)

#=====
#                               Calculo do DIC
#=====

# Funcao de log-verossimilhanca
log_verossim <- function(param, data) {
  versim <- dgev(data, loc = param[1], scale = param[2], shape =
    param[3], log = F)+1e-16
  return( log(versim) )
}

# Calculando a deviance para cada iteracao da cadeia
deviance <- apply(post, 1, function(param) {
  -2 * sum(log_verossim(param, y))
})

# Media da deviance
mean_deviance <- mean(deviance)

# Ponto de Maximo a Posteriori (MAP)
MAP_param <- post[which.max(apply(post, 1, function(param) {
  sum(log_verossim(param, y)) # Posterior maxima
})),]

# Deviance no MAP

```

```
MAP_deviance <- -2 * sum(log_verossim(MAP_param, y))

# Penalizacao de complexidade
p_D <- mean_deviance - MAP_deviance

# DIC
DIC <- mean_deviance + p_D

cat("O_DIC_do_modelo_e:", DIC, "\n")
```