

Universidade Federal de Alfenas

Thales Rangel Ferreira

Estudo do ajuste de distribuições de probabilidade na
análise de precipitações máximas de Machado-MG

Alfenas - MG

2021

Thales Rangel Ferreira

Estudo do ajuste de distribuições de probabilidade na
análise de precipitações máximas de Machado-MG

Trabalho de Conclusão de Curso apresentado à Universidade Federal de Alfenas como parte dos requisitos para obtenção da Licenciatura em Matemática.

Área de concentração: Estatística Aplicada.

Orientador: Prof. Dr. Luiz Alberto Beijo

Coorientador: Prof. Dr. Fabricio Goecking Avelar

Alfenas - MG

2021

Thales Rangel Ferreira

Estudo do ajuste de distribuições de probabilidade na análise de precipitações máximas de Machado-MG

A banca examinadora abaixo-assinada, aprova o Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção do Certificado de Conclusão do Curso de Licenciatura em Matemática pela Universidade Federal de Alfenas.

Área de concentração: Estatística Aplicada.

Aprovado em: ___ de _____ de 2021.

Prof. Dr. Luiz Alberto Beijo

Orientador

Prof. Dr. Flávio Bittencourt

Universidade Federal de Alfenas

Dr. Carlos José Dos Reis

Universidade Federal de Lavras

Prof. Dr. Denismar Alves Nogueira (Suplente)

Universidade Federal de Alfenas

Agradecimentos

Gostaria de agradecer primeiramente a Deus, pela minha saúde de corpo e espírito. Agradeço ao meu Orientador Luiz Alberto Beijo por me dar suporte durante todos os 4 anos de minha graduação, pela paciência, dedicação, confiança e amizade. Ao meu Coorientador e amigo Fabrício Goecking Avelar, por fazer me apaixonar pela Estatística, por acreditar em mim, e me ajudar tanto em aspectos acadêmicos quanto pessoais.

À toda minha família, especialmente meus pais e meu irmão, que estiveram sempre ao meu lado, independente da distância e vida. Aos meus colegas de classe Amanda, Edson, Augusto, Mariana e Mesek, pela amizade, companheirismo, ajuda e a todos os muitos momentos de descontração. Agradeço em especial a Amanda, por estar sempre presente na minha vida, me dando suporte em momentos difíceis, e reciprocamente compartilhando nossos inúmeros momentos felizes. Você é a melhor amiga que eu poderia ter, obrigado.

Diversas pessoas foram importantes durante minha trajetória nesta graduação, mas agradeço em especial a Maria Eduarda, por estar do meu lado a vários anos sempre disposta a me ajudar em todas as situações. Ao Wallace, pela amizade e companheirismo desde a época de escola. Ao Matheus, que o considero como um orientador, por me auxiliar durante todas as minhas ICs. À Mari, pela amizade que se manteve mesmo após nos abandonar no curso. Á Cidinha, minha coordenadora de Ensino Médio, que se disponibilizou a me ajudar mesmo em horários extraescolares, sem você eu não teria entrado na faculdade. Á todas as pessoas que de alguma forma me ajudaram e desejaram o meu bem, fica meu agradecimento.

“Minha fé é no desconhecido, em tudo que não podemos compreender por meio da razão. Creio que o que está acima do nosso entendimento é apenas um fato em outras dimensões e que no reino do desconhecido há uma infinita reserva de poder”

(Charlie Chaplin)

Resumo

As precipitações pluviais, quando em excesso, podem causar danos como erosão de solos, inundações, prejuízos em obras hidráulicas e rompimentos de barragens e represas, entre outros. O conhecimento sobre a precipitação máxima esperada, numa determinada região, pode auxiliar no planejamento de atividades agrícolas e construções hidráulicas de forma a evitar danos e prejuízos. Para realizar previsão de chuvas intensas, encontram-se na literatura diversos modelos de probabilidade, dentre estes as distribuições Generalizada de Valores Extremos (GEV), Gumbel e Log-Normal. Neste estudo objetivou-se identificar, dentre as distribuições GEV, Gumbel e Log-Normal, qual apresenta as melhores estimativas de precipitação máxima mensal para a cidade de Machado, considerando os tempos de retorno de 3, 5, 10 e 15 anos, e aplica-la para calcular a precipitação máxima mensal esperada para os tempos de retorno de 5, 10, 25, 50 e 100 anos. A estimação dos parâmetros das distribuições se deu por meio do método da Máxima Verossimilhança. Foram utilizados dados de precipitação relativos aos anos de 1980 a 2016, obtidos no Banco de Dados Meteorológicos para Ensino e Pesquisa do Instituto Nacional de Meteorologia. A partir dos resultados, verificou-se pelo teste de aderência de Kolmogorov-Smirnov que somente a distribuição GEV, referente às séries de máximos de junho e julho, e a distribuição Log-Normal referente á série de máximo de agosto, não se ajustaram. A escolha da distribuição mais adequada para a estimativa da precipitação máxima dos períodos chuvoso e seco ocorreu pelo resultado do erro médio de previsão, que evidenciou que a distribuição GEV forneceu as melhores estimativas da precipitação máxima provável de ambos os períodos no município de Machado.

Palavras-chave: Distribuição GEV. Distribuição Gumbel. Distribuição Log-Normal. Precipitação extrema.

Abstract

Rainfall, when in excess, can cause damage such as soil erosion, flooding, damage to hydraulic works and rupture of dams and dams, among others. The knowledge about the maximum expected precipitation, in a given region, can help in the planning of agricultural activities and hydraulic constructions in order to avoid damages and losses. In order to forecast heavy rains, several probability models are found in the literature, among them the Generalized Extreme Value (GEV), Gumbel and Log-Normal distributions. This study aimed to identify, among the GEV, Gumbel and Log-Normal distributions, which presents the best estimates of maximum monthly precipitation for the city of Machado, considering the return times of 3, 5, 10 and 15 years, and applies it to calculate the maximum expected monthly precipitation for the 5, 10, 25, 50 and 100 year return times. The distribution parameters were estimated using the Maximum Likelihood method. Precipitation data for the years 1980 to 2016, obtained from the Meteorological Database for Teaching and Research of the National Institute of Meteorology, were used. From the results, it was verified by the Kolmogorov-Smirnov adherence test that only the GEV distribution, referring to the series of maxima of June and July, and the Log-Normal distribution referring to the series of maximum of August, not adjusted. The choice of the most suitable distribution for the estimation of the maximum precipitation of the rainy and dry periods was due to the result of the average forecast error, which showed that the GEV distribution obtained the best estimates of the maximum probable precipitation of both periods in the municipality of Machado.

Keywords: GEV distribution. Gumbel distribution. Log-Normal distribution. Extreme precipitation.

Sumário

1	Introdução	6
2	Revisão de Literatura	8
2.1	Eventos Climáticos Extremos	8
2.2	Distribuição Generalizada de Valores Extremos	8
2.3	Gumbel	9
2.4	Log-Normal	9
2.5	Estimação dos parâmetros	10
2.6	Ajuste das distribuições	11
3	Material e Métodos	13
3.1	Banco de dados	13
3.2	Composição das séries históricas de precipitações máximas mensais	13
3.3	Teste de pressuposições e aderência	14
3.4	Nível de retorno	14
3.5	Medidas utilizadas para selecionar as distribuições	15
3.6	Recursos Computacionais	15
4	Resultados e Discussões	16
5	Conclusões	25

1 Introdução

Dentre os diversos fenômenos naturais existentes há de se destacar a precipitação pluvial, um evento climático imprescindível e um tema de extrema relevância em qualquer sociedade, visto que a sua intensidade, frequência ou duração está diretamente ligada à agricultura, geração de energia, economia, entre outros fatores.

O Sul de Minas Gerais apresenta uma forte economia agropecuária, possuindo grande influência econômica no estado. Entretanto, no período chuvoso, devido às chuvas intensas, a região sofre com deslizamentos de terra e alagamentos (OLIVEIRA et al., 2014), causador da desagregação do solo, que é o deslocamento de partículas ricas em nutrientes, que torna desse modo, o solo do local onde ocorreu alagamento, menos fértil, culminando em uma menor produtividade agrícola. Para compensar a perda de nutrientes, os produtores aumentam a quantidade de fertilizantes e agroquímicos aplicados no solo. Porém, mediante a recorrência de precipitações, os mesmos são deslocados para o leito dos rios, o que resulta na poluição das águas (SANTOS; NORI; OLIVEIRA, 2010).

A precipitação, quando extrema, além de ocasionar prejuízos nas atividades agrícolas, pode acarretar em catástrofes naturais, gerando danos sociais e econômicos, como ocorrido na Região Serrana do Rio de Janeiro em 2011, onde houve alagamentos e deslizamentos de terra, com um total de 916 mortos (PINHEIRO; ANDRADE; MOURA, 2011). Desse modo o conhecimento da distribuição de probabilidade que melhor se adéqua aos dados de precipitação máxima diária mensal se torna de suma importância (JUNQUEIRA JÚNIOR; MELLO; ALVES, 2015).

Diferentes distribuições de probabilidade são utilizadas na literatura para a modelagem de eventos extremos naturais como a precipitação máxima, permitindo desta forma, em uma localidade, conhecer a probabilidade de ocorrência de uma quantidade qualquer de precipitação, precedente ao acontecimento do evento (SANTOS et al., 2018).

As distribuições de probabilidade possuem parâmetros desconhecidos que, ao serem estimados, permitem a modelagem de eventos extremos. Para se obter as estimativas desses parâmetros, foi utilizado no presente trabalho o método da Máxima Verossimilhança (MV), que segundo Blain (2011) mostrou-se mais eficiente em comparação com método dos Momentos L.

Quando estimados os parâmetros das distribuições, é necessário verificar a qualidade do ajuste da mesma ao conjunto de dados trabalhado, antes de qualquer análise posterior.

Para verificar o ajuste dessa distribuição aos dados, pode-se utilizar testes de aderência, que verificam se uma amostra é proveniente de uma população com uma determinada distribuição teórica. Nesse trabalho, fora utilizado para tal, o teste de aderência de Kolmogorov–Smirnov (K-S).

Mediante os fatos mencionados, objetivou-se nesse trabalho verificar qual das distribuições, GEV, Gumbel ou Log-Normal, melhor se ajusta às séries de precipitações máximas da cidade de Machado. Pretende-se também, avaliar a qualidade de previsão das distribuições com base no erro médio de previsão (EMP) e viés relativo, considerando diferentes tempos de retorno.

2 Revisão de Literatura

2.1 Eventos Climáticos Extremos

De acordo com Albeverio, Jentsch e Kantz (2006), os extremos climáticos podem ser entendidos como valores de parâmetros meteorológicos, superiores ou inferiores a determinados limites, associados a probabilidade de ocorrência relativamente baixa e com impacto significativo na sociedade ou nos ecossistemas. Marengo et al. (2009) explicam que eventos climáticos extremos como chuvas intensas, vendavais, furacões, marés meteorológicas e grandes secas representam forças com maior poder de destruição.

Vários municípios de Santa Catarina, do Rio de Janeiro e de Minas Gerais vieram a decretar estado de calamidade devido a fortes chuvas ocorridas nos meses de novembro e dezembro de 2008. Essas chuvas fortes causaram destruição total de casas, deixando milhares de pessoas desabrigadas e centenas de mortos. De acordo com informações do Instituto Nacional de Pesquisas Espaciais (INPE, 2008), no Estado de Santa Catarina, os prejuízos causados por esse evento pluvial extremo, pelas subseqüentes enchentes e deslizamentos foram estimados em 350 milhões de dólares, principalmente pelo fechamento do Porto de Paranaguá.

O uso de funções de densidade de probabilidade está diretamente ligado à natureza dos dados a que elas se relacionam, em que algumas têm boa capacidade de estimação e dependendo do número de parâmetros, podem assumir diferentes formas, enquadrando-se em um número maior de situações, ou seja, são mais flexíveis. Desde que respeitado o aspecto da representatividade dos dados, as estimativas dos seus parâmetros, para uma determinada região, podem ser estabelecidas como de uso geral, sem prejuízo da precisão na estimação da probabilidade (CATALUNHA et al., 2002).

Diversas distribuições de probabilidade utilizadas para a análise de precipitações máximas, são encontradas na literatura, dentre elas, as distribuições GEV, Gumbel e Log-Normal são muito utilizadas, conforme Catalunha et al. (2002), Sansigolo (2008), Blain (2011), Franco et al. (2014), entre outros.

2.2 Distribuição Generalizada de Valores Extremos

Conforme Mendes (2004), a distribuição GEV foi introduzida por Jenkinson (1955) e incorpora as três formas assintóticas de valores extremos em uma única expressão. A

função de probabilidade acumulada da distribuição GEV é dada por:

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

em que ξ , σ e μ denotam, respectivamente, os parâmetros de forma, escala e posição.

O valor de ξ determina a forma assintótica de valores extremos, ou seja, se $\xi > 0$, a GEV representa a distribuição do Tipo II (Fréchet), definida apenas para $x > \mu - \sigma/\xi$, enquanto que, se $\xi < 0$, representa a distribuição do Tipo III (Weibull), definida apenas para $x < \mu - \sigma/\xi$ (COLES, 2001). Para o caso em que o limite de $F(x)$, com ξ tendendo à zero, a GEV corresponde à distribuição do Tipo I (Gumbel).

Derivando-se (1) em relação a x obtém-se a função densidade de probabilidade da GEV, dada por:

$$f(x|\xi, \mu, \sigma) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right)^{-\frac{1}{\xi}} \right] \right\}, \quad (2)$$

em que $-1 < x < \mu - \sigma/\xi$ para $\xi < 0$ e $\mu - \sigma/\xi < x < +\infty$ para $\xi > 0$.

2.3 Gumbel

A função de densidade de probabilidade Gumbel é dada por:

$$f(x|\mu, \sigma) = \frac{1}{\sigma} \exp \left[-\frac{x - \mu}{\sigma} - \exp \left(-\frac{x - \mu}{\sigma} \right) \right], \quad (3)$$

com $-\infty < x < \infty$.

A distribuição Gumbel, também conhecida como distribuição assintótica dos extremos do tipo I, distribuição tipo I de Fisher-Tippet ou dupla exponencial, desenvolvida por Gumbel (1958), é utilizada frequentemente em estudos de fluxo de água (rios, reservatórios), fenômenos meteorológicos, resistência de materiais, aeronáutica, geologia e engenharia naval, entre outros (BAUTISTA, 2002).

2.4 Log-Normal

De acordo com Mood, Graybill e Boes (1974) seja X uma variável aleatória positiva, e Y uma nova variável aleatória definida como $Y = \ln(X)$. Se Y tem distribuição Normal, então X tem uma distribuição Log-Normal com a seguinte função de densidade de

probabilidade:

$$f(x | \mu_y, \sigma_y) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp \left[-\frac{(\ln x - \mu_y)^2}{2\sigma_y^2} \right], \quad (4)$$

em que x é a variável aleatória associada a valores do período com $0 < x < 1$, $\sigma_y > 0$ é o desvio padrão dos logaritmos da variável x e μ_y a média dos logaritmos da variável x com $0 < \mu_y < 1$.

A média e a variância da distribuição Log-Normal são dadas, respectivamente, por $E(X) = \exp(\mu + \frac{1}{2}\sigma^2)$ e $V(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$. Também tem-se que, se X tem uma distribuição Log-Normal, então $E[\ln X] = \mu$ e $\text{var}[\ln X] = \sigma^2$.

2.5 Estimação dos parâmetros

Ao estudar-se uma determinada variável aleatória, um dos interesses reside em determinar uma distribuição teórica que a descreva. Mas para isso torna-se necessário determinar os valores dos parâmetros que definem tal distribuição. Como não é possível contar com todas as informações da população, tem-se que os verdadeiros valores dos seus parâmetros não podem ser obtidos, ao menos no tocante a variáveis climáticas. O recurso a ser adotado é utilizar uma amostra que seja representativa em relação à população da variável em estudo (TUCCI, 2001).

Através de uma amostra representativa busca-se obter estimativas dos parâmetros, sendo esse processo conhecido por estimação. Existem vários métodos para se obter os estimadores dos parâmetros das distribuições de extremos. Segundo Coles (2001), dentre os métodos utilizados para a estimação dos parâmetros das distribuições de valores extremos, o mais utilizado é o da Máxima Verossimilhança. Uma das vantagens desse método é a fácil compreensão da teoria dos estimadores de Máxima Verossimilhança e suas boas propriedades assintóticas.

De acordo com Bolfarine e Sandoval (2001), sejam X_1, \dots, X_n uma amostra aleatória de tamanho n da variável aleatória X com função de densidade $f(x; \theta)$ com $\theta \in \Theta$, onde Θ é o espaço paramétrico. A função de verossimilhança de θ correspondente à amostra aleatória observada é dada por:

$$L(\theta; x) = \prod_{i=1}^n f(x_i | \theta). \quad (5)$$

Em posse do conhecimento da função de verossimilhança, o desejo reside em determinar o(s) valor(es) de $\hat{\theta}$ que maximizam a função de verossimilhança, ou seja, que maximizam a probabilidade da amostra ter ocorrido. O processo de maximização de $L(\theta; x)$ é feito através de derivadas parciais em relação ao(s) parâmetro(s) do vetor, igualando as derivadas parciais a zero.

Por simplicidade algébrica, muitas vezes é mais interessante trabalhar com somas ao invés de produtos. Segundo Stewart (2006), os cálculos de funções complicadas envolvendo produtos podem muitas vezes ser simplificadas tomando-se os logaritmos. Portanto, para obter-se o máximo de $L(\theta; x)$ é mais conveniente aplicar o logaritmo natural à função de verossimilhança, pela mudança de produto em somas. O logaritmo natural da função de verossimilhança, também conhecido como função suporte, é dado por:

$$l(\theta; x) = \ln L(\theta; x). \quad (6)$$

O sistema de equações obtidos de (6) não possui solução analítica, deste modo, faz-se necessário o uso de métodos numéricos para a obtenção dos estimadores de Máxima Verossimilhança. Para isso, utilizou-se métodos implementados no pacote *evd* do software R.

2.6 Ajuste das distribuições

Após obter-se as estimativas dos parâmetros que definem a distribuição teórica, sobre a qual reside a suspeita de ser a distribuição que represente os dados da variável aleatória, o próximo passo é verificar o seu ajuste a esse conjunto de informações. Um teste utilizado com essa finalidade é o teste não paramétrico de Kolmogorov-Smirnov.

Segundo Cotta, Correa e Albuquerque (2016), o teste K-S destaca-se na estatística, baseando-se na máxima diferença entre a função de distribuição acumulada assumida para os dados ($F(x)$) e a função de distribuição empírica dos dados ($F_n(x)$), dada por:

$$D = \sup |F(x) - F_n(x)|, \quad (7)$$

em que D representa a distância vertical máxima entre as imagens de $F(x)$ e $F_n(x)$.

A estatística de teste K-S (7) é utilizada para testar a hipótese de nulidade:

$$\begin{cases} H_0 : X \text{ tem função densidade de probabilidade } f_0 \\ H_1 : X \text{ não tem função densidade de probabilidade } f_0 \end{cases} \quad (8)$$

com f_0 sendo a função da distribuição desejada.

A hipótese de nulidade é rejeitada, dada uma distribuição específica com um nível de significância α , se D é maior que o valor crítico ($D_{(n,\alpha)}$). Pode-se testar a hipótese de nulidade também, comparando-se o p -valor com o nível de significância adotado no teste, rejeitando-se H_0 se $p\text{-valor} < \alpha$.

3 Material e Métodos

3.1 Banco de dados

Os dados de precipitação pluvial utilizados neste trabalho, dados pela altura de lâmina d'água (mm), foram obtidos a partir do Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) do Instituto Nacional de Meteorologia (INMET) referente a estação meteorológica de código A567, de coordenadas latitudinais de 21° 40' 50" S e longitudinais de 45° 56' 39" W, a 969 metros de altitude em relação ao nível do mar. Os dados obtidos são referentes aos anos de 1980 a 2016, totalizando um período de estudo máximo de 31 e mínimo de 30 anos, já que foram excluídos anos ou meses que apresentaram falhas no registro.

3.2 Composição das séries históricas de precipitações máximas mensais

Os dados de precipitação máxima de Machado foram agrupados em séries mensais, isto é, considerou-se como série, cada um dos meses de janeiro a dezembro. Para efeito de exemplificação da composição das séries de máximos de um dado mês, segue o procedimento realizado para o mês de janeiro. Dentre as 31 observações de precipitações diárias no mês de janeiro no ano de 1980, foi selecionada a ocorrência diária de maior valor para a composição da série. Esse processo foi repetido para os demais anos com observações no mês de janeiro, totalizando desta forma, a série de máximos mensal de janeiro, com 31 dados. Esse processo se repetiu de forma análoga para os demais meses, obtendo-se 31 dados para cada série dos meses restantes, com exceção à de junho, que apresentou 30 dados.

O conjunto de dados de máximos, de cada série mensal, foi dividido em dois grupos. Os dados do período de 1980 a 2000 foram reservados para estimação dos parâmetros dos modelos a serem testados. Os dados do período de 2001 a 2016, corresponde à fase de teste, foram reservadas para o cálculo do EMP e viés relativo das precipitações máximas prováveis, através do qual é possível comparar o desempenho dos modelos utilizados no trabalho.

No presente trabalho optou-se por agrupar os meses de acordo com os períodos de chuva, sendo esses o período seco, englobando os meses de maio, junho, julho agosto,

setembro e outubro, e o período chuvoso, composto por janeiro, fevereiro, março, abril, novembro e dezembro.

3.3 Teste de pressuposições e aderência

Para analisar independência das séries de máximos construídas foi utilizado o teste de Ljung-Box ao nível de significância de 1% assim como adotado por dos Reis, Beijo e Avelar (2017). De acordo com Ljung e Box (1978), o teste verifica se alguns grupos de autocorreção de uma série temporal são diferentes de zero.

O teste de K-S, apresentado na seção 2.6, foi utilizado para analisar a aderência das distribuições GEV, Gumbel e Log-Normal às série de máximos coletados. Para o teste de K-S adotou-se o nível de significância de 5%, assim como Martins et al. (2010).

3.4 Nível de retorno

Segundo Naghettini e Pinto (2007), o tempo de retorno T representa o inverso da probabilidade com que um dado evento E tenha ocorrido. Dada a ocorrência de um evento E , o tempo de retorno T é o tempo médio necessário (em anos) para que esse evento recorra, em um ano qualquer. Em termos práticos, seu significado é: se ocorrer um evento de intensidade x , qual é o tempo médio (T) esperado para que o evento de intensidade x ocorra novamente?

Por definição, segue que o tempo de retorno T associado ao evento E é dado por:

$$T = \frac{1}{P(E)}. \quad (9)$$

De acordo com a variável climatológica trabalhada neste estudo, o evento (E) é a precipitação máxima que excede um determinado valor x e a probabilidade de excedência de E é obtido por $1 - F(x)$. Assim:

$$T = \frac{1}{P(E)} = \frac{1}{1 - F(x)}, \quad (10)$$

em que $F(x)$ é a função cumulativa da distribuição GEV, Gumbel ou Log-Normal.

Conforme Bautista (2002), o nível de retorno (x_p) é obtido pela solução da equação:

$$\int_{-\infty}^{x_p} f(x)dx = 1 - p, \quad (11)$$

em que x_p está associado ao tempo de retorno T e $p = \frac{1}{T}$.

De (11) segue que $F(x_p) = 1 - p$. Dessa forma, considerando que as observações máximas seguem uma distribuição GEV, Gumbel ou Log-Normal, o estimador do p -quantil x_p é dado por:

$$x_p = F^{-1}(1 - p). \quad (12)$$

A estimativa \hat{x}_p do nível de retorno x_p para tempos de retorno $T = \frac{1}{p}$ é obtida pela substituição das estimativas de Máxima Verossimilhança dos parâmetros das distribuições GEV, Gumbel ou Log-Normal.

3.5 Medidas utilizadas para selecionar as distribuições

O desempenho das distribuições ajustadas aos dados de precipitação máxima de Machado, foi avaliado por meio do EMP para cada nível de retorno, dado por:

$$EMP = \frac{1}{n} \sum_{i=1}^n \left| \frac{o_i - \hat{o}_i}{o_i} \right| \times 100\%. \quad (13)$$

em que o_i é a precipitação máxima observada, \hat{o}_i é a precipitação máxima prevista para o tempo de retorno “ i ” e n é o número de previsões.

Outro critério utilizado para avaliar o desempenho dos modelos utilizados nesse trabalho é o viés relativo, dado pela expressão:

$$Viés = \frac{\hat{o}_i - o_i}{o_i}. \quad (14)$$

3.6 Recursos Computacionais

Todas as análises estatísticas foram realizadas utilizando-se o software R, conforme R Core Team (2018). Os pacotes *evd* (STEPHENSON, 2002), *fBasics* (WUERTZ, SETZ e CHALABI, 2020) e *tseries* (TRAPLETTI e HORNIK, 2019) foram utilizados para o estudo dos dados.

4 Resultados e Discussões

São apresentados nas Figura 1, 2 e 3 a representação gráfica das séries de precipitações máximas diárias mensais relativas ao período de estudo (1980 a 2016).

Pode-se observar nas Figuras 1, 2 e 3, que nas máximas de precipitação diária mensal, poucas foram as ocorrências que superaram 100 mm, o que fora observado também no trabalho de Beijo, Vivanco e Muniz (2009), os quais analisaram o tempo de retorno das precipitações máximas em Jaboticabal-SP via análise Bayesiana.

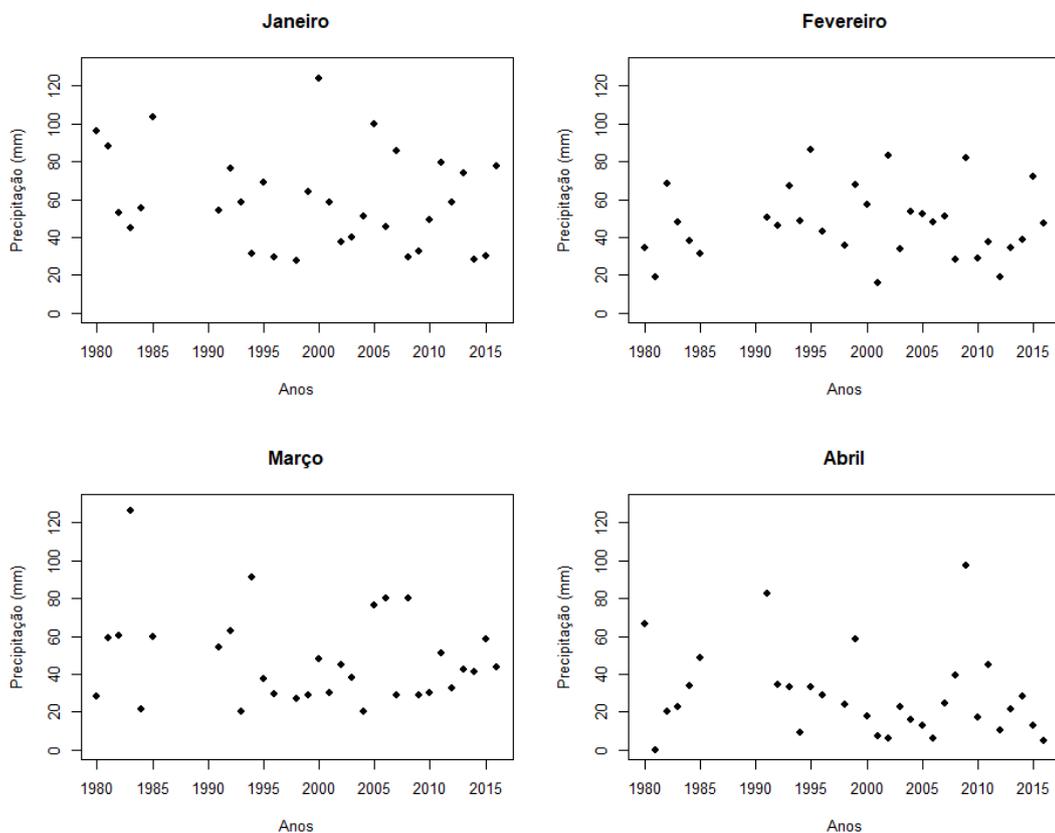


Figura 1: Representação gráfica das séries de precipitações diárias máximas (mm) da cidade de Machado-MG, dos meses de janeiro a abril no período de 1980 a 2016.

Analisando os resultados apresentados nas Figuras 1, 2 e 3, é possível observar que os meses do período seco (maio, junho, julho, agosto, setembro e outubro) apresentam observações de precipitação máxima diária notoriamente inferior aos meses do período chuvoso (janeiro, fevereiro, março, abril, novembro e dezembro). Ressaltando-se ainda a grande quantidade de ocorrências de precipitações máximas próximas de 0 mm nos meses de junho, julho e agosto.

Na Tabela 1 são apresentadas medidas descritivas das precipitações pluviais máximas

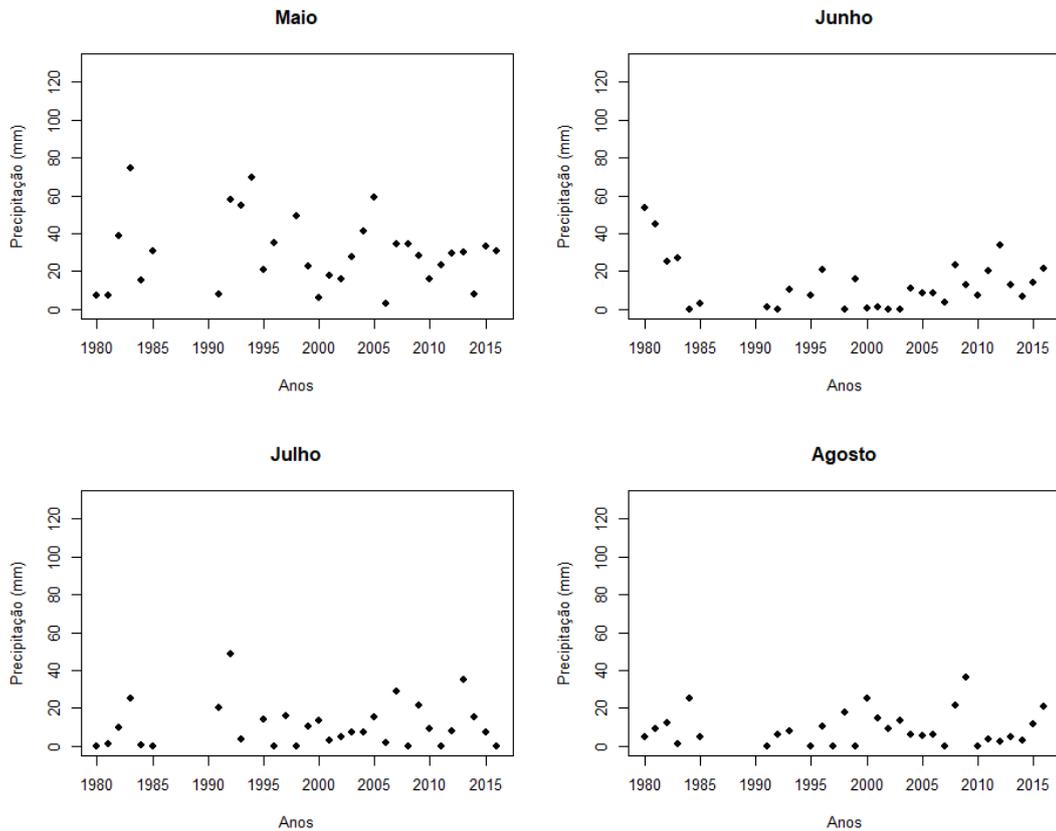


Figura 2: Representação gráfica das séries de precipitações diárias máximas (mm) da cidade de Machado-MG, dos meses de maio a agosto no período de 1980 a 2016.

ocorridas entre os anos de 1980 a 2016 da cidade de Machado. Pode-se observar que a menor precipitação máxima mensal registrada ocorreu em agosto, com valor de 36,2 mm, já a precipitação pluvial máxima mais elevada, ocorreu em março, atingindo 126,5 mm.

Observa-se na Tabela 1 que as médias de precipitação pluvial máxima mais elevadas ocorreram nos meses de janeiro e dezembro, com 59,8 e 50,4 mm, respectivamente. Resultado semelhante foi observado no estudo de Beijo, Muniz e Castro Neto (2005), que analisando o tempo de retorno de precipitações máximas na cidade de Lavras-MG, obtiveram resultados evidenciando médias de precipitações máximas mais elevadas nos meses de janeiro (57 mm) e dezembro (55 mm). Nota-se também, que as médias com os valores mais baixos, ocorreram nos meses de julho e agosto, com 10,6 e 9,1 mm, respectivamente.

Ainda sobre os resultados da Tabela 1, observa-se que o coeficiente de variação das precipitações máximas diárias dos meses de junho, julho e agosto foram relativamente altos em comparação com os demais meses, o que pode ser explicado devido a grande quantidade de observações próximas de zero, como pode-se observar na Figura 2.

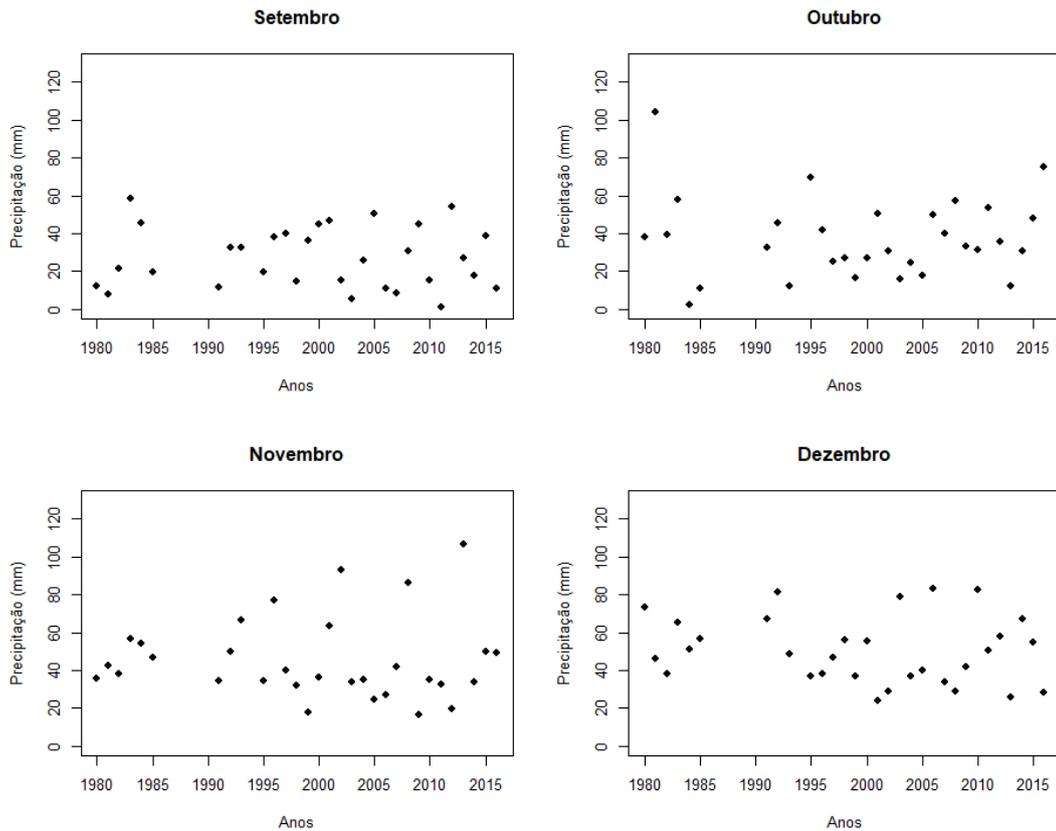


Figura 3: Representação gráfica das séries de precipitações diárias máximas (mm) da cidade de Machado-MG, dos meses de setembro a dezembro no período de 1980 a 2016.

Tabela 1: Estatística descritiva das precipitações máximas (mm) ocorridas no período chuvoso e seco em Machado-MG nos anos de 1980 a 2016.

Período	Meses	n	Máx	Med	\bar{X}	D.p	C.v	C.a
Chuvoso	Jan.	30	123,7	55,2	59,8	25,2	42,2	0,61
	Fev.	31	86,4	47,2	47,4	18,4	38,9	0,43
	Mar.	31	126,5	42,4	47,9	23,7	49,6	1,33
	Abr.	31	97,6	23,0	28,5	22,5	78,7	0,80
	Nov.	31	106,8	38,0	45,5	21,4	47,0	0,80
	Dez.	31	83,0	48,7	50,4	17,4	34,6	1,03
Seco	Mai.	31	74,8	29,6	30,0	18,6	62,0	0,74
	Jun.	30	53,4	9,7	13,2	13,4	101,0	1,37
	Jul.	31	48,5	7,4	10,6	11,6	109,2	0,81
	Ago.	31	36,2	6,2	9,1	8,9	97,2	0,72
	Set.	31	58,5	25,6	27,1	15,8	58,1	0,87
	Out.	31	104,0	33,2	37,3	21,0	56,3	0,95

Máx = Máximo; Med = Mediana; \bar{X} = Média; D.p = Desvio padrão ; C.v = Coeficiente de variação (%); n = Tamanho da amostra; C.a = Coeficiente de assimetria.

Analisando os resultados da média e mediana dos dados das séries, verifica-se que para todos os meses, os valores das médias foram superiores aos das medianas, o que sugere que

as distribuições empíricas dos dados de todas as séries mensais são assimétricas à direita (positiva), o que é reforçado pelos valores positivos do coeficiente de assimetria.

Na Tabela 2 são apresentados os resultados do teste de independência de Ljung-Box às observações das séries e do teste de aderência de Kolmogorov-Smirnov para verificar o ajuste das distribuições GEV, Gumbel e Log-Normal aos dados de precipitações máximas de Machado.

Por meio dos resultados obtidos, é possível concluir ao nível de significância de 1%, que as observações das séries de precipitações mensais máximas são independentes (p -valor $> 0,01$).

Tabela 2: p -valores do teste de independência Ljung-Box e do teste de aderência de Kolmogorov-Smirnov referente ao ajuste das distribuições GEV, Gumbel e Log-Normal aos dados de precipitações máximas mensais da cidade de Machado-MG no período de 1980 a 2016.

Mês	Ljung-Box	Kolmogorov-Smirnov		
		GEV	Gumbel	Log-Normal
Jan.	0,8491	0,9717	0,9647	0,9901
Fev.	0,1016	0,9707	0,8445	0,7285
Mar.	0,3073	0,7086	0,4468	0,5988
Abr.	0,3901	0,9905	0,9826	0,1398
Mai.	0,7523	0,8871	0,8773	0,6838
Jun.	0,0148	0,0057	0,3928	0,2491
Jul.	0,4267	0,0034	0,6824	0,2103
Ago.	0,0816	0,7479	0,8227	0,0077
Set.	0,4144	0,8385	0,8212	0,6181
Out.	0,8632	0,9929	0,9945	0,6035
Nov.	0,5995	0,9345	0,8525	0,8839
Dez.	0,6860	0,8729	0,8811	0,9101

De acordo com os resultados apresentados na Tabela 2, considerando um nível de significância de 5%, observa-se que apenas a distribuição Gumbel se ajustou a todas às séries de máximos mensais de precipitação (p -valor $> 0,05$). Os resultados do teste K-S evidenciaram que a distribuição Log-Normal não se ajustou apenas em relação a série de agosto (p -valor $< 0,05$). Verificou-se também, que houve ajuste da distribuição GEV aos dados de precipitação máxima de todas as séries, com exceção às de junho e julho.

Mediante os resultados do teste de independência e aderência, foram realizadas as estimativas da precipitação máxima mensal, com exceção às dos meses em que os testes anteriores não se adequaram.

Para uma melhor visualização da comparação das estimativas obtidas por meio das

distribuições e os valores reais observados, foi calculado o viés relativo da precipitação máxima mensal, apresentados nas Figuras 4 e 5.

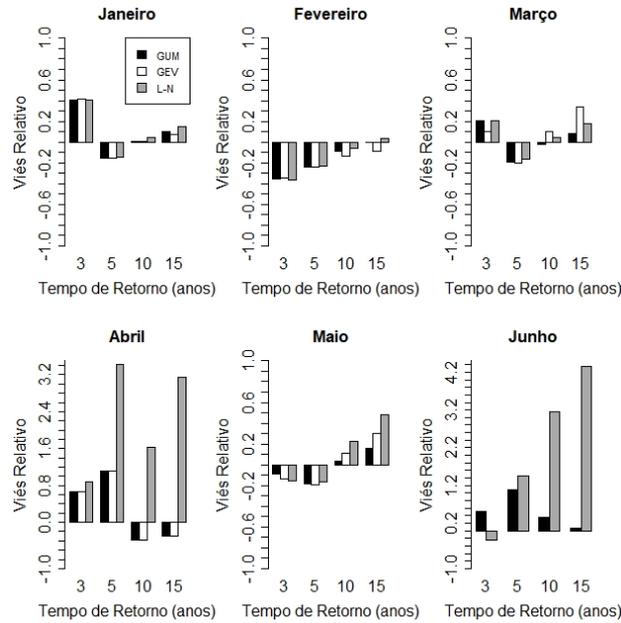


Figura 4: Viés relativo da precipitação máxima mensal (mm) dos meses de janeiro a junho da cidade de Machado-MG para diferentes tempos de retorno (anos), obtidos por meio das distribuições GEV, Gumbel e Log-Normal.

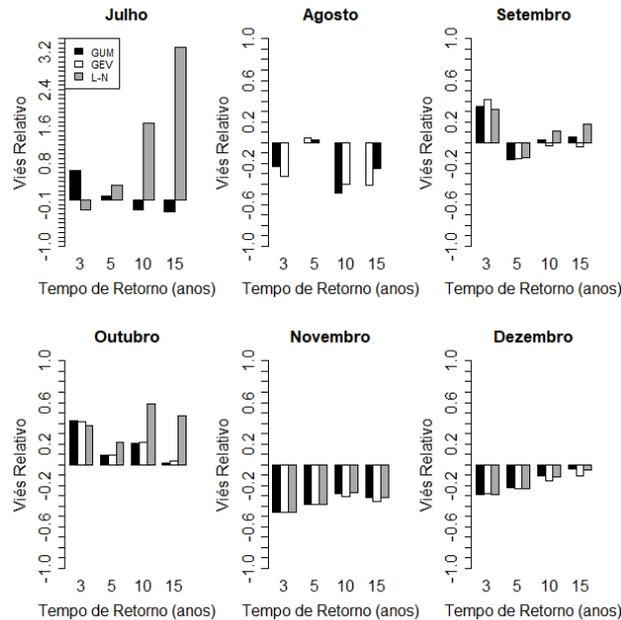


Figura 5: Viés relativo da precipitação máxima mensal (mm) dos meses de julho a dezembro da cidade de Machado-MG para diferentes tempos de retorno (anos), obtidos por meio das distribuições GEV, Gumbel e Log-Normal.

Nas Figuras 4 e 5, é possível observar que em determinados casos as distribuições

forneçam valores de viés relativo altos, excedendo o intervalo 1 e -1. Porém, é importante evidenciar que em todas as ocorrências o valor excedido foi positivo, isto é, a estimativa de precipitação máxima foi superior à observada.

A ocorrência de viés relativo superior a 1 no mês de abril foi verificada nos tempos de retorno de 5, 10 e 15 anos. Para o tempo de retorno de 5 anos, o viés relativo das estimativas de precipitação obtidas via distribuição Gumbel, GEV e Log-Normal foram respectivamente de 1,1, 1,1 e 3,42. Para os tempo de retorno de 10 e 15 anos, somente o viés relativo das estimativas obtidas via distribuição Log-Normal apresentaram valores superiores a 1, com 1,62 e 3,15, respectivamente.

Em junho, para o tempo de retorno de 5 anos os vieses relativos obtidos foram de 1,09 (Gumbel) e 1,46 (Log-Normal). Para 10 e 15 anos, somente o viés relativo das estimativas obtidas via distribuição Log-Normal excedeu o intervalo, com valores iguais a 3,16 e 4,34 respectivamente. Referente aos resultados da Figura 5, observa-se que no mês de julho, para os tempos de retorno de 10 e 15 anos o viés relativo das estimativas obtidas via distribuição Log-Normal apresentaram valores altos, iguais a 1,67 e 3,22, respectivamente.

De modo geral, a distribuição Gumbel obteve as estimativas de precipitações máximas que propiciaram os menores valores de viés relativo, o que pode ser confirmado pelos resultados apresentados na Tabela 3. Nessa, é apresentado em porcentagem, o EMP da precipitação máxima esperada para os períodos seco e chuvoso analisados no presente trabalho.

Tabela 3: Média dos erros médios de previsão (EMPs) (%) das precipitações máximas prováveis de cada tempo de retorno (3, 5, 10 e 15 anos) dos meses dos períodos seco e chuvoso, obtidas via distribuições GEV, Gumbel e Log-Normal ajustadas aos dados da cidade de Machado-MG.

Período	Distribuições	EMP (%)
Chuvoso	GEV	30,60
	Gumbel	28,43
	Log-Normal	51,19
Seco	GEV	24,13
	Gumbel	21,75
	Log-Normal	28,94

Os resultados da Tabela 3 evidenciam que a distribuição Gumbel, em comparação com as distribuições GEV e Log-Normal, apresentou os menores EMPs de precipitação máxima esperada nos períodos seco e chuvoso. Desta forma, pode-se concluir que a

distribuição Gumbel propiciou as melhores estimativas de precipitação máxima para a cidade de Machado nos dois períodos em estudo.

Deste modo, mediante as conclusões obtidas a partir dos resultados apresentados na Tabela 3, foram calculadas as estimativas das precipitações máximas mensais via distribuição Gumbel para o tempo de retorno de 5, 10, 25, 50 e 100 anos com seus respectivos intervalos de confiança (95%).

As estimativas de precipitação máxima para cada tempo de retorno são apresentados na Figura 6. Ao analisá-las, nota-se que a medida que o tempo de retorno aumenta, a estimativa de precipitação referente ao mesmo, aumenta concomitantemente.

A partir dos resultados apresentados na Figura 6, observa-se que para todos os tempos de retorno, as estimativas de precipitações mais elevadas ocorreram no mês de janeiro, com os valores iguais a 55,3, 78, 93, 112,1, 126,1 e 140,1 mm referente aos tempos de retorno de 5, 10, 25, 50 e 100 anos, respectivamente.

Esses resultados foram semelhantes aos obtidos por Beijo, Muniz e Castro Neto (2005), que realizaram um estudo sobre tempo de retorno das precipitações máximas da cidade de Lavras-MG, que possui localização próxima ao município de Machado. Os autores obtiveram as estimativas de precipitação máxima mensal para diferentes tempo de retorno via distribuição de valores extremos tipo I e também verificaram que as mais altas estimativas ocorreram no mês de janeiro.

Pesquisas relacionados à precipitação pluvial máxima em regiões com climas semelhantes à aqui estudada, utilizam diferentes distribuições de densidade de probabilidade para estudos envolvendo essa variável climática. Beijo et al. (2003), ao estimarem a precipitação pluvial máxima esperada em diferentes níveis de probabilidade para a cidade de Jaboticabal-SP, por meio da distribuição Gumbel, com seus parâmetros estimados pelos Método dos Momentos e da Máxima Verossimilhança, chegaram a conclusão de que a distribuição Gumbel ajustou-se às series de precipitação máxima mensal.

Blain (2011) verificou na região de Campinas-SP, que a distribuição GEV com seus parâmetros estimados tanto pelo método da Máxima Verossimilhança quanto dos Momentos-L, apresentou um bom desempenho para a estimação da probabilidade de ocorrência associada aos totais máximos anuais de precipitação pluvial diária.

Barreto et al. (2015) ao analisarem o ajuste das distribuições Fréchet, Gumbel, Gama a 2 parâmetros, Log-Normal a 2 e 3 parâmetros e GEV, aos dados de precipitação máxima

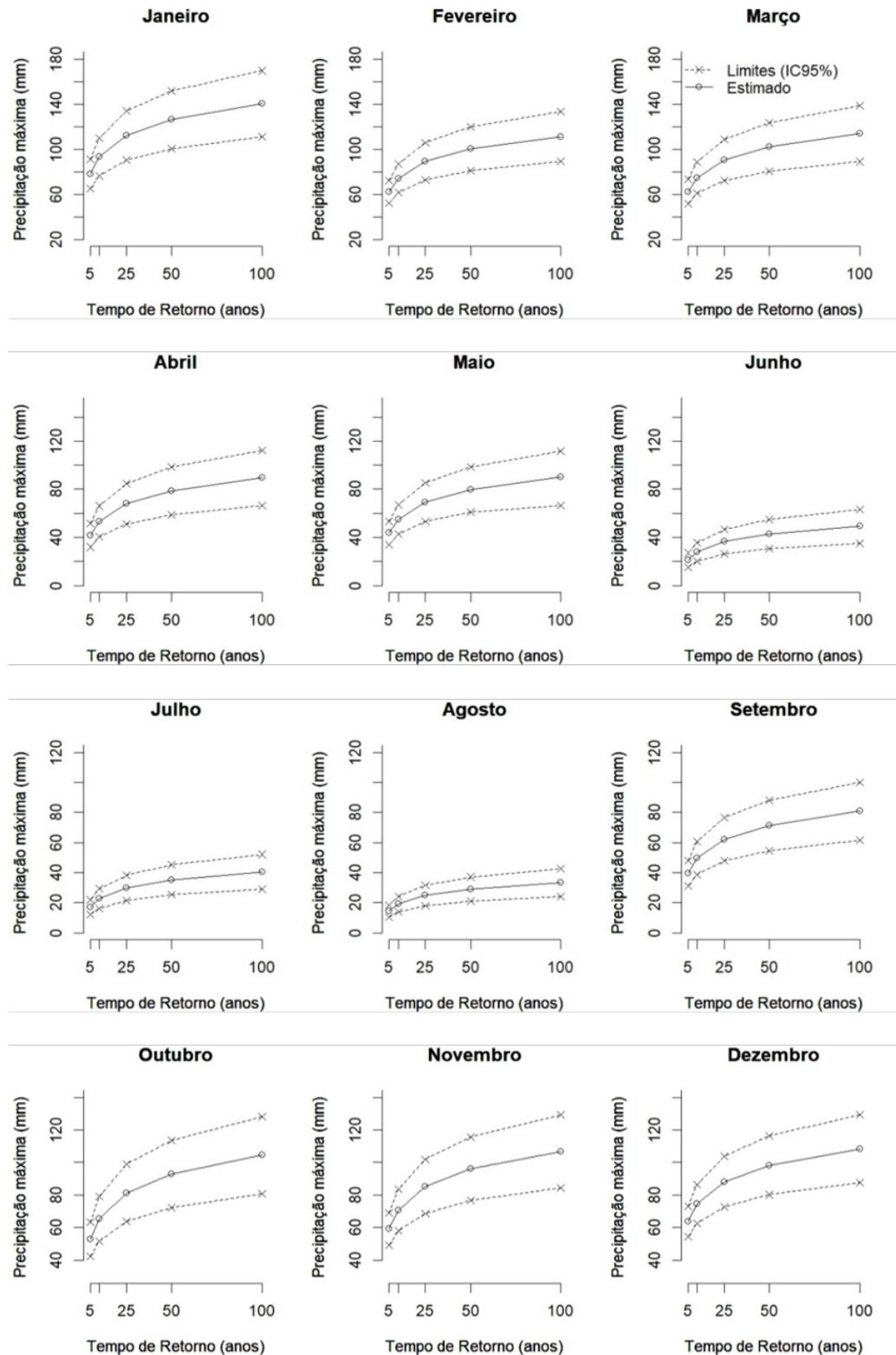


Figura 6: Valores da precipitação máxima mensal esperada (mm) para diferentes tempo de retorno (em anos) obtidos via distribuição Gumbel ajustada aos dados da cidade de Machado-MG e seus limitantes inferiores (LI) e superiores (LS) respectivos ao intervalo de confiança de 95%.

da cidade de Nepomuceno-MG, constataram que apenas a distribuição Gumbel obteve resultados satisfatórios para os testes de aderência Qui-Quadrado, Kolmogorov-Smirnov e Anderson Darling.

Ainda que a distribuição Gumbel tenha obtido as melhores estimativas de precipitação pluvial máxima para a região de Machado, esse resultado não pode ser generalizado, visto que uma região específica pode não ter somente uma distribuição que melhor se ajuste às séries de precipitação máxima do local ou que produza as melhores estimativas da precipitação máxima esperada. Pode-se obter resultados distintos para as diferentes distribuições dependendo do tamanho da amostra utilizada. Logo, aconselha-se testar outras distribuições para cada série de dados.

5 Conclusões

Em todas as séries de precipitação máxima da cidade de Machado verificou-se o ajuste da distribuição Gumbel. Já a distribuição GEV se ajustou a todas séries mensais, com exceção às de junho e julho, enquanto a distribuição Log-Normal não apresentou ajuste somente em relação a série de máximos de agosto.

A distribuição Gumbel apresentou os menores valores do viés relativo e EMP. Logo, forneceu as melhores estimativas de precipitação máxima mensal de Machado, para o tempo de retorno de 3, 5, 10 e 15 anos.

Portanto, recomenda-se neste trabalho, o uso da distribuição Gumbel para a análise de precipitações máximas mensais da cidade de Machado-MG.

Deste modo, os resultados obtidos no presente trabalho podem ser utilizados para o planejamento de atividades sujeitas aos efeitos adversos da precipitação extrema, como a implantação de projetos de engenharia hidráulica e agrícola.

Referências

- ALBEVERIO, S.; JENTSCH, V.; KANTZ, H. *Extreme events in nature and society*. New York: Springer Science & Business Media, 2006.
- BARRETO, H. B. et al. Avaliação de funções de probabilidades e relação-intensidade-duração-frequência para precipitação extrema em Nepomuceno-MG. *Global Science and Technology*, Rio Verde, v. 8, n. 2, p. 118-127, 2015.
- BAUTISTA, E. A. L. *A distribuição generalizada de valores extremos no estudo da velocidade máxima de vento em Piracicaba, SP*. Dissertação (Mestrado) - Departamento de Estatística e Experimentação Agronômica, Universidade de São Paulo, Piracicaba, 2002.
- BEIJO, L. A.; MUNIZ, J. A.; CASTRO NETO, P. Tempo de retorno das precipitações máximas em Lavras (MG) pela distribuição de valores extremos do tipo I. *Ciência e agrotecnologia*, Lavras, v. 29, n. 3, p. 657-667, 2005.
- BEIJO, L. A.; MUNIZ, J. A.; VOLPE, C. A.; PEREIRA, G. T. Estudo da precipitação máxima em Jaboticabal (SP) pela distribuição de Gumbel utilizando dois métodos de estimação dos parâmetros. *Revista Brasileira de Agrometeorologia*, Santa Maria, v. 11, n. 1, p. 141-147, 2003.
- BEIJO, L. A.; VIVANCO, M. J. F.; MUNIZ, J. A. Análise Bayesiana no estudo do tempo de retorno das precipitações pluviárias máximas em Jaboticabal (SP). *Ciência e Agrotecnologia*, Lavras, v. 33, n. 1, p. 261-270, 2009.
- BLAIN, C. G. Cento e vinte anos de totais extremos de precipitação pluvial máxima diária em Campinas, Estado de São Paulo: Análises estatísticas. *Bragantia*, Campinas, v. 70, n. 3, p. 722-728, 2011.
- BOLFARINE, H.; SANDOVAL, M. C. *Introdução à inferência estatística*. Rio de Janeiro: Sociedade Brasileira de Matemática-SBM, 2001.
- CATALUNHA, M. J. et al. Aplicação de cinco funções densidade de probabilidade a séries de precipitação pluvial no Estado de Minas Gerais. *Revista Brasileira de Agrometeorologia*, Santa Maria, v. 10, n. 1, p. 153-162, 2002.
- COLES, S. G. *An introduction to statistical modeling of extreme values*. London: Springer, 2001.
- COTTA, H. H.; CORREA, W. S. C.; ALBUQUERQUE, T. T. A. Aplicação da distribuição de Gumbel para valores extremos de precipitação no município de Vitória-ES. *Revista Brasileira de Climatologia*, Curitiba, v. 19, n. 12, p. 202-217, 2016.
- DOS REIS, C. J.; Beijo, L. A.; AVELAR, F. G. Temperatura mínima esperada para Piracicaba-SP via distribuições de valores extremos. *Revista Brasileira de Agricultura Irrigada*, Fortaleza, v. 11, n. 4, p. 1639-1650, 2017.

FRANCO, C. S. et al. Distribuição de probabilidades para precipitação máxima diária na Bacia Hidrográfica do Rio Verde, Minas Gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, Campina Grande, v. 18, n. 7, p. 735-741, 2014.

GUMBEL, E. J. *Statistical of Extremes*. New York: Columbia University Press, 1958. 375p.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). 2008. *As chuvas de novembro de 2008 em Santa Catarina: um estudo de caso visando à melhoria do monitoramento e da previsão de eventos extremos*. Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 71p.

JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quartely Journal of the Royal Meteorological Society*, London, v. 81, n. 348, p. 158-171, 1955.

JUNQUEIRA JÚNIOR, J. A.; MELLO, C. R.; ALVES, G. J. Eventos extremos de precipitação no Alto Rio Grande, MG: Análise probabilística. *Revista Brasileira de Engenharia Agrícola e Ambiental*, Campina Grande, v. 19, n. 4, p. 301-308, 2015.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, London, v. 65, n. 2, p. 297-303, 1978.

MARENGO, J. A. et al. Mudanças climáticas e eventos extremos no Brasil. *Rio de Janeiro: FBDS*, 2009.

MARTINS, J. A. et al. Probabilidade de precipitação para a microregião de Tangará da Serra, estado do Mato Grosso. *Pesquisa Agropecuária Tropical*, Goiânia, v. 40, n. 3, p. 291-296, 2010.

MENDES, B. V. M. *Introdução à análise de eventos extremos*. Rio de Janeiro: E-papers Serviços Editoriais Ltda, 2004. 232p.

MOOD, A. L.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the Theory of Statistics*. Tokio: McGraw-Hill Kogakusha 3. ed., 1974. 564p.

NAGHETTINI, M.; PINTO, E. J. A. *Hidrologia Estatística*. Belo horizonte: CRPM, 2007. 552p.

OLIVEIRA, A. S. et al. Aplicabilidade da distribuição GEV ao estudo da precipitação máxima diária anual na região sul de Minas Gerais. *Revista Agrogeoambiental*, Pouso Alegre, v. 6, n. 1, p. 31-44, 2014.

PINHEIRO, H. R.; ANDRADE, K. M.; MOURA, C. W. A maior catástrofe climática do Brasil sob a visão operacional do CPTEC/INPE. In: *Simpósio Internacional de Climatologia*, 4., 2011, João Pessoa-PB. Mudanças Climáticas e seus Impactos nas Áreas Urbanas: anais. João Pessoa-PB: SBMET, 2011.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org/>>. Acesso em: 02 nov. 2020.

SANTOS, G. S.; NORI, P. G.; OLIVEIRA, L. F. C. Chuvas intensas relacionadas à erosão hídrica. *Revista Brasileira de Engenharia Agrícola e Ambiental*, Campina Grande, v. 14, n. 2, p. 115-123, 2010.

SANSIGOLO, C. A. Distribuições de Extremos de precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, SP (1917-2006). *Revista Brasileira de Meteorologia*, v. 23, n. 3, p. 341-346, 2008.

SANTOS, R. S. et al. Caracterização de extremos mensais de precipitação em Cacoal (RO). *Revista Brasileira de Climatologia*, Curitiba, v. 22, n. 14, p. 267-280, 2018.

STEPHENSON, A. G. evd: Extreme Value Distributions. *R News*, v. 2, n. 2, p. 31-32, 2002. Disponível em: <<https://CRAN.R-project.org/doc/Rnews/>>. Acesso em: 08 mar. 2020.

STEWART, J. *Cálculo*, Volume I. 5 ed. São Paulo: Thomson Learning, 2006. 196p.

TRAPLETTI, A.; HORNIK, K. *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-47, 2019. Disponível em: <<https://CRAN.R-project.org/package=tseries>>. Acesso em: 08 mar. 2020.

TUCCI, C. M. *Hidrologia: Ciência e aplicação*. Porto Alegre: UFRGS, 2001. 943p.

WUERTZ, D.; SETZ, T.; CHALABI, Y. *fBasics: Rmetrics - Markets and Basic Statistics*. R package version 3042.89.1, 2020. Disponível em: <<https://CRAN.R-project.org/package=fBasics>>. Acesso em: 08 mar. 2020.