

Ajuste de distribuições de probabilidade à precipitação mensal no estado de Pernambuco – Brasil

Fit of probability distributions to monthly precipitation in the state of Pernambuco – Brazil

Ajuste de las distribuciones de probabilidad a la precipitación mensual en el estado de Pernambuco – Brasil

Recebido: 04/11/2020 | Revisado: 13/11/2020 | Aceito: 17/11/2020 | Publicado: 22/11/2020

Patricia de Souza Medeiros Pina Ximenes

ORCID: <https://orcid.org/0000-0003-2468-1683>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: patricia.medeiross10@gmail.com

Antonio Samuel Alves da Silva

ORCID: <http://orcid.org/0000-0002-8759-0036>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: samuelmathematical@gmail.com

Fahim Ashkar

ORCID: <https://orcid.org/0000-0002-1899-9049>

Université de Moncton, Canadá

E-mail: ashkarf@umoncton.ca

Tatijana Stosic

ORCID: <https://orcid.org/0000-0002-5691-945X>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: tastosic@gmail.com

Resumo

Este estudo teve como objetivo identificar modelos de distribuição de probabilidade que melhor se ajustam a dados de precipitação mensal para o estado de Pernambuco – Brasil. Foram analisados os ajustes de seis distribuições de probabilidade de 2 parâmetros: gama (GAM), log normal (LNORM), Weibull (WEI), Pareto Generalizado (PG), Gumbel (GUM) e normal (NORM) para dados de precipitação mensal de 40 estações pluviométricas distribuídas no estado de Pernambuco, no período de 1988 a 2017 (30 anos). O método de Máxima Verossimilhança (ML) foi utilizado para estimar os parâmetros dos modelos e a

seleção do modelo baseou-se em uma modificação da estatística de Shapiro-Wilk. Os resultados mostraram que as distribuições de 2 parâmetros são flexíveis o suficiente para descrever dados de precipitação mensal para o estado de Pernambuco e que os modelos log normal, gama, Weibull e PG se ajustaram melhor aos dados. Os modelos Gumbel e normal raramente se ajustaram aos dados independente do mês analisado.

Palavras-chave: Precipitação mensal; Pernambuco; Distribuições de probabilidade.

Abstract

This study aimed to identify probability distribution that best fit monthly rainfall data for the state of Pernambuco - Brazil. The fits of six 2-parameters probability distributions were analyzed: gamma (GAM), log normal (LNORM), Weibull (WEI), Generalized Pareto (GP), Gumbel (GUM) and normal (NORM) for monthly rainfall data of 40 rainfall stations across the state of Pernambuco, from 1988 to 2017 (30 years). The Maximum Likelihood (ML) method was used to estimate the model parameters and the model selection was based on a modification of the Shapiro-Wilk statistic. The results showed the 2-parameters distributions are flexible enough to describe monthly precipitation data for the state of Pernambuco and the log normal, gamma, Weibull and GP models fitted better to the data. The Gumbel and normal models rarely adjusted to the data regardless of the month analyzed.

Keywords: Monthly rainfall; Pernambuco; Probability distributions.

Resumen

Este estudio tuvo como objetivo identificar los modelos de distribución de probabilidad que mejor se ajustan a los datos de precipitación mensual para el estado de Pernambuco - Brasil. Se analizaron los ajustes de seis distribuciones de probabilidad de 2 parámetros: gamma (GAM), log normal (LNORM), Weibull (WEI), Pareto generalizado (PG), Gumbel (GUM) y normal (NORM) para los datos de precipitación mensual. 40 estaciones pluviométricas distribuidas en el estado de Pernambuco, en el período de 1988 - 2017 (30 años). Se utilizó el método de máxima verosimilitud (ML) para estimar los parámetros del modelo y la selección del modelo se basó en una modificación del estadístico de Shapiro-Wilk. Los resultados mostraron que las distribuciones de 2 parámetros son lo suficientemente flexibles para describir los datos de precipitación mensual para el estado de Pernambuco y que los modelos log normal, gamma, Weibull y PG se ajustan mejor a los datos. Los modelos Gumbel y normal rara vez se ajustan a los datos independientemente del mes analizado.

Palabras clave: Precipitación mensual, Pernambuco, Distribuciones de probabilidad.

1. Introdução

A análise de dados de precipitação é extremamente importante para o planejamento estratégico e de tomada de decisão em vários sistemas naturais e socioeconômicos. Inúmeras áreas como a agricultura (Stern & Coe, 1982; Hussain, Mahmood, & Hayat, 2010), engenharia civil (Bjureland, Johansson, Sjölander, Spross, & Larsson, 2019), gestão de recursos hídricos (Papalexiou & Koutsoyiannis, 2012), entre outros, analisam dados de precipitação com o intuito de prevenir riscos associados à água (secas, enchentes, etc), que podem comprometer o seu desempenho.

Um dos principais problemas encontrados na análise de dados hidrológicos é a identificação de uma distribuição de probabilidade estatística que se ajusta a um conjunto de dados observados (Ashkar & Aucoin, 2012). Alguns métodos de análise de precipitação mundialmente conhecidos e utilizados, como o *Standardized Precipitation Index* (SPI) (McKee, Doesken, & Kleist, 1993), e o *Standardized Precipitation Evapotranspiration Index* (SPEI) (Vicente-Serrano, Beguería, & López-Moreno, 2010), necessitam da identificação de um modelo estatístico que melhor se ajuste aos dados, em uma das etapas de sua aplicação.

A discriminação entre modelos estatísticos com base em dados observados tem recebido considerável atenção na literatura (Cheng, Chiang, & Hsu, 2007; Li, Brissette, & Chen, 2013; Haberlandt & Radtke, 2014; Bermudez, Abilgos, Cuaresma, & Rabajante, 2017; Zhu, Chen, & Chen, 2019). Muitos estudos relacionados à hidrologia estão focados em discriminar entre modelos de 2 parâmetros, um parâmetro de escala e um parâmetro de forma, pois estes tem apresentado excelentes resultados (Ashkar & Aucoin, 2012; Elsherpieny, Muhammed, & Radwan, 2017; Ashkar & Ba, 2017; Mazucheli & Emanuelli, 2019).

A precipitação é um dos principais temas encontrados na literatura por ser um dos principais componentes do ciclo da água e interferir diretamente na disponibilidade de recursos hídricos do planeta terra. O conhecimento da variabilidade da precipitação em diferentes escalas temporais e espaciais é essencial para o desenvolvimento de medidas de curto e longo prazo para suavizar riscos relacionados à água, como inundações e secas que podem comprometer diretamente a economia de uma região (Silva, Menezes, Telesca, Stosic, & Stosic, 2020; Santana, Silva, Menezes, & Stosic, 2020).

Nos anos de 2012 e 2013, o Brasil registrou perdas de aproximadamente US\$ 1,6 bilhão no setor econômico e US\$ 1,5 bilhão com mortalidade de animais, causadas pela seca (Brito, Marengo, & Coutinho, 2017). A região Nordeste do Brasil (NB) é caracterizada por

apresentar um clima semi-árido e alta variabilidade sazonal e interanual de precipitação, com episódios extremos de umidade e seca.

Dentre os estados do NB, Pernambuco representa cerca de 6% do território nordestino e detém uma população de aproximadamente nove milhões de habitantes (IBGE, 2020). A economia do estado é movida pelos setores da agricultura, pecuária e prestação de serviços, além de o estado está classificado como o décimo mais rico do Brasil, apresentando um PIB de aproximadamente R\$181 bilhões (IBGE, 2020).

Diante da importância do estado de Pernambuco para a economia da região Nordeste do Brasil, o presente estudo tem como objetivo identificar distribuições de probabilidade que melhor se ajustam aos dados de precipitação mensal do estado, a fim de contribuir para um melhor entendimento dos padrões de chuva na região. Foram analisados dados de precipitação mensal de 40 estações pluviométricas espalhadas por Pernambuco em um período de 30 anos (1988 a 2017).

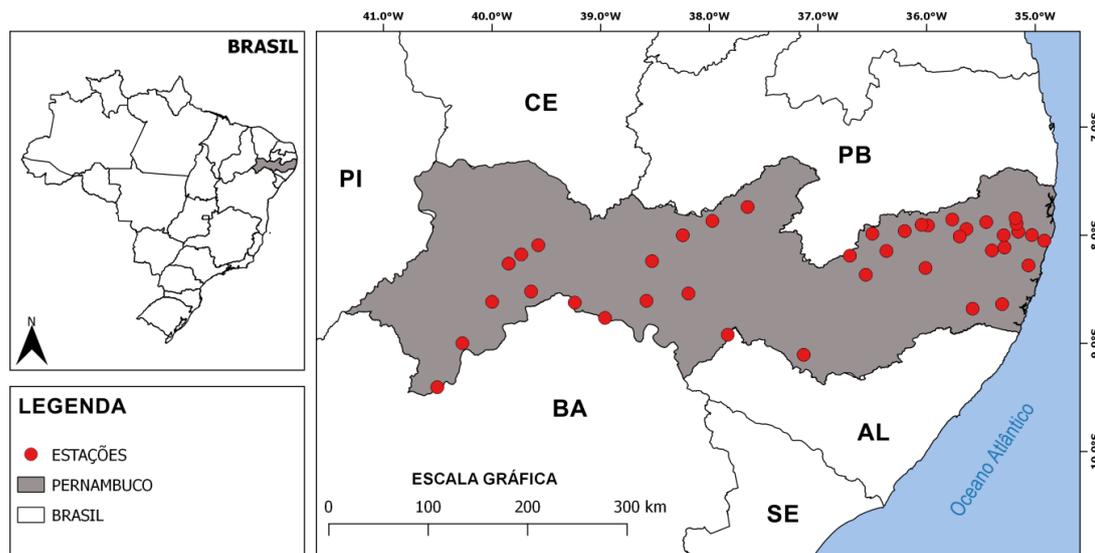
2. Metodologia

2.1 Área de estudo e dados

O estado de Pernambuco está dentre os nove estados que compõem a região Nordeste do Brasil. É banhado pelo oceano atlântico à sua direita e tem como limites os estados da Paraíba (N), Bahia (S), Alagoas (SE), Ceará (NO) e Piauí (O). Pernambuco possui um território de 98.149,119 km² e uma população de aproximadamente 9.616.621 habitantes.

Este estudo foi realizado com dados provenientes da plataforma online da Agência Nacional de Águas (ANA) (<http://hidroweb.ana.gov.br>, último acesso em novembro de 2020). Os dados foram coletados em 40 estações pluviométricas espalhados no estado e consiste em séries históricas de precipitação mensal no período de 1988 e 2017 (30 anos). A localização das estações pluviométricas pode ser visualizada na Figura 1.

Figura 1. Localização geográfica do estado de Pernambuco e distribuição espacial das estações pluviométricas utilizadas no estudo.



Fonte: Autores, (2020).

2.2 Estimação dos parâmetros via máxima verossimilhança

O método de estimação por Máxima Verossimilhança (MV) é um dos métodos de estimação de parâmetros mais utilizados na literatura por ser de fácil utilização, além de apresentar um excelente desempenho com diferentes distribuições de probabilidade (Sijbers, den Dekker, Scheunders, & Van Dyck, 1998; Ashkar & Tatsambon, 2007; Chang, 2011).

De forma resumida, e para melhor entendimento do método; seja X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n da variável aleatória X com função de densidade (ou função de probabilidade) $f(x|\theta)$ com $\theta \in \Theta$, em que Θ é o espaço paramétrico. A função de máxima verossimilhança de θ correspondente à amostra aleatória observada que é dada por:

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

O estimador de máxima verossimilhança de θ é o valor $\hat{\theta} \in \Theta$ que maximiza a função de verossimilhança $L(\theta; x)$ (Bolfarine & Sandoval, 2001). Na prática, o logaritmo de $L(\theta; x)$ é geralmente utilizado.

$$l(\theta; x) = \log L(\theta; x) = \sum_{i=1}^n \ln f(x_i | \theta). \quad (2)$$

As estimativas de MV $\hat{\theta}$ são aquelas que maximizam $l(\theta; x)$. $\hat{\theta}$ é chamado de estimador de máxima verossimilhança (EMV) de θ . O método de estimação MV possui algumas propriedades essenciais para sua aplicação. Essas propriedades não foram apresentadas porque estão fora do escopo deste estudo; mais detalhes sobre podem ser encontrados em Bolfarine & Sandoval (2001).

2.3 Estatística *TN.SW*

O critério de informação de Akaike (AIC) e a estatística de Anderson-Darling (AD) são métodos comumente utilizados na literatura para discriminação entre distribuições de probabilidade. No entanto, esses métodos são viesados, principalmente quando usados em amostras pequenas (Ashkar & Ba, 2017; Ashkar, Ba, & Dieng, 2019), que é a realidade da maioria dos estudos na área de hidrologia. Em 1997, Ashkar, Arsenault & Zoglat (1997) desenvolveram a estatística *TN.SW*, que é uma versão modificada da famosa estatística de teste de Shapiro-Wilk (Shapiro & Wilk, 1965).

A estatística *TN.SW* é menos tendenciosa do que os tradicionais métodos AIC e AD e oferece uma vantagem sobre os métodos baseados em estimativa de máxima verossimilhança robusta, pois não favorece a seleção de um modelo sobre o outro, especialmente para tamanhos de amostra pequenos (Ashkar & Ba, 2017 ; Ashkar, Ba, & Dieng, 2019). Conforme descrito por Ashkar & BA (2017) o cálculo da estatística consiste em duas etapas:

1. Seja Φ^{-1} a função quantil da distribuição normal padrão $Z \sim N(0,1)$, que assume, o modelo A sendo a função de distribuição cumulativa F_A para ser a verdadeira distribuição para a amostra $X_n = \{X_1, X_2, \dots, X_n\}$. Se F_A fosse conhecido, a seguinte transformação modificaria exatamente X_n para uma amostra normal padrão $Z_n = \{Z_1, Z_2, \dots, Z_n\}$:

$$Z_i = \Phi^{-1}(F_A(X_i)) \quad (3)$$

No entanto, uma vez que os parâmetros do modelo A são realmente desconhecidos, usa-se $\hat{F}_A(X_i)$ para estimar $F_A(X_i)$, em que $\hat{F}_A(X_i)$ é obtido, por exemplo, por MV. Então a equação (3) torna-se:

$$Z_i = \Phi^{-1}(\hat{F}_A(X_i)) \quad (4)$$

em que Z_i são aproximadamente uma normal padrão, i.e. $Z_i \sim N(0,1)$.

- Utiliza-se a amostra transformada $Z_n = \{Z_1, Z_2, \dots, Z_n\}$ obtida na equação (4) para calcular a estatística *TN.SW*:

$$S_A = \left[\sum_{i=1}^n v_i Z_{(i)} \right]^2 / \sum_{i=1}^n [Z_{(i)} - \bar{Z}]^2 \quad (5)$$

em que Z_i é a i -ésima estatística de ordem de Z_n , \bar{Z} é o valor médio de Z , e v_i são coeficientes aproximados pelo método proposto por Royston (1982a, 1982b, 1995). Repetem-se as etapas (4) e (5) substituindo Modelo A pelo Modelo B, para obter S_B . A regra de decisão é escolher modelo A como o modelo correto se $S_A > S_B$, e escolher o modelo B caso contrário. Diante das vantagens oferecidas pela estatística *TN.SW*, esse método foi escolhido para ser utilizado neste estudo.

O presente estudo é uma pesquisa de natureza quantitativa (Pereira, Shitsuka, Parreira, & Shitsuka, 2018) e tem como objetivo identificar qual distribuição de probabilidade apresenta o ajuste mais adequado para dados de precipitação mensal do estado de Pernambuco. O cálculo da estatística *TN.SW*, bem como as estimativas dos parâmetros via MV para cada modelo, foram obtidos utilizando a linguagem de programação R (R Core Team, 2020). A linguagem de programação R é uma ferramenta que disponibiliza várias técnicas estatísticas e gráficas, já programadas e disponibilizadas a partir de pacotes. Vários pesquisadores utilizam esta ferramenta por ser de fácil manipulação, apresentar uma grande variedade de técnicas, pode ser executada em diferentes sistemas operacionais e é disponibilizada com zero custo para qualquer pessoa.

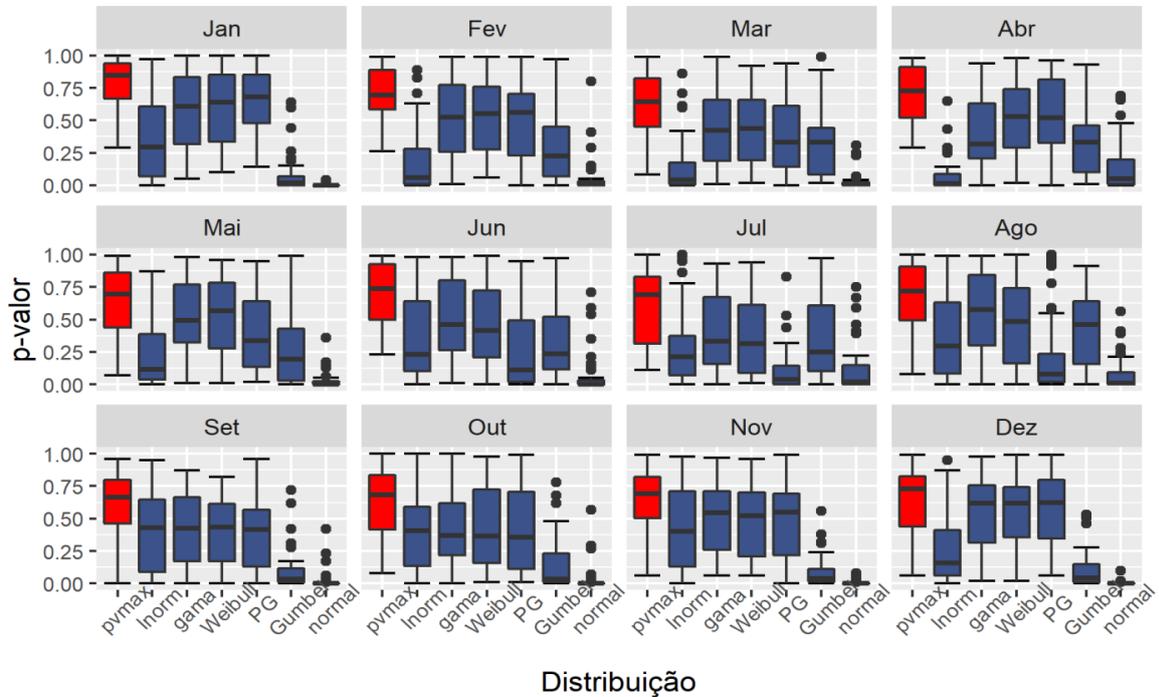
3. Resultados e Discussão

Nesta seção, serão apresentados os resultados obtidos a partir do ajuste das distribuições de probabilidade utilizadas neste estudo. O objetivo é identificar qual distribuição apresentou o ajuste mais adequado para cada mês do ano, utilizando dados de precipitação mensal de 40 estações pluviométricas espalhadas pelo estado de Pernambuco no período de 1988 - 2017.

A estatística *TN.SW* permite a comparação dos ajustes das distribuições, mas não indica se o ajuste é “bom” ou “ruim” (Ashkar & Aucoin, 2012). Então, para avaliar a qualidade do ajuste das distribuições mencionadas anteriormente, é necessário avaliar o *p-valor* da estatística *TN.SW*. Quanto maior o *p-valor* produzido pelo ajuste de um modelo, melhor é o seu ajuste aos dados. Baixo *p-valor* indica um ajuste de distribuição inadequado aos dados. As

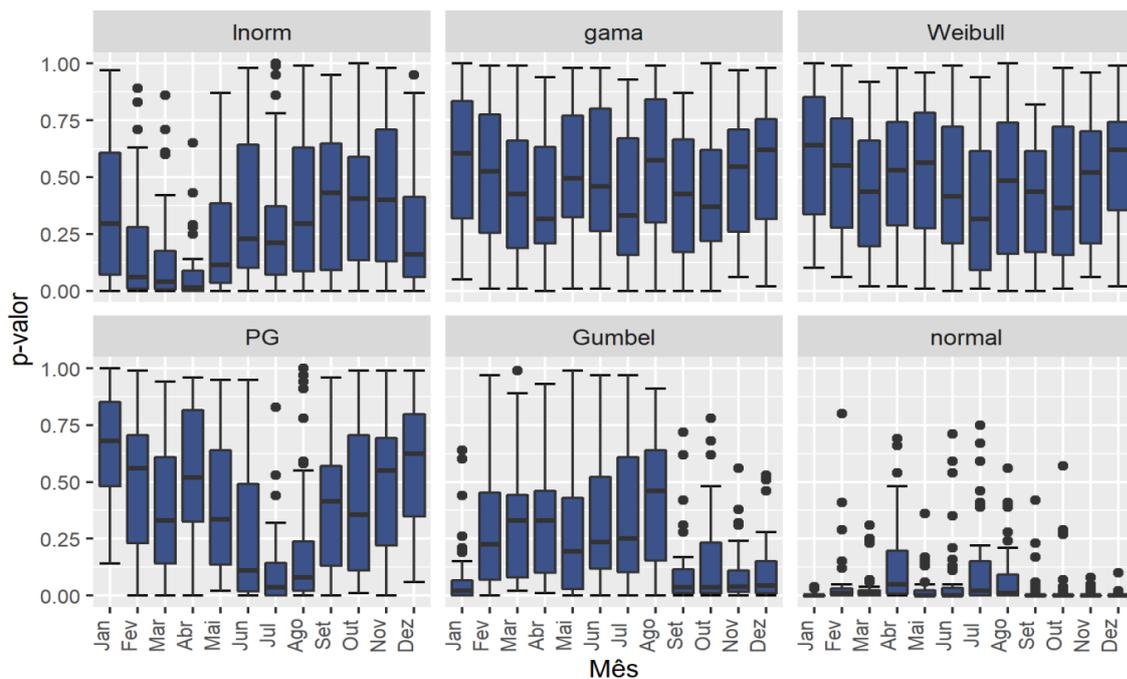
Figuras 2 e 3 e apresentam os boxplots do p -valor da estatística $TN.SW$ entre todas as distribuições de probabilidade consideradas.

Figura 2. Boxplots do p -valor da estatística $TN.SW$ para as distribuições analisadas. pvmax representa o maior p -valor obtido entre todas as distribuições consideradas.



Fonte: Autores, (2020).

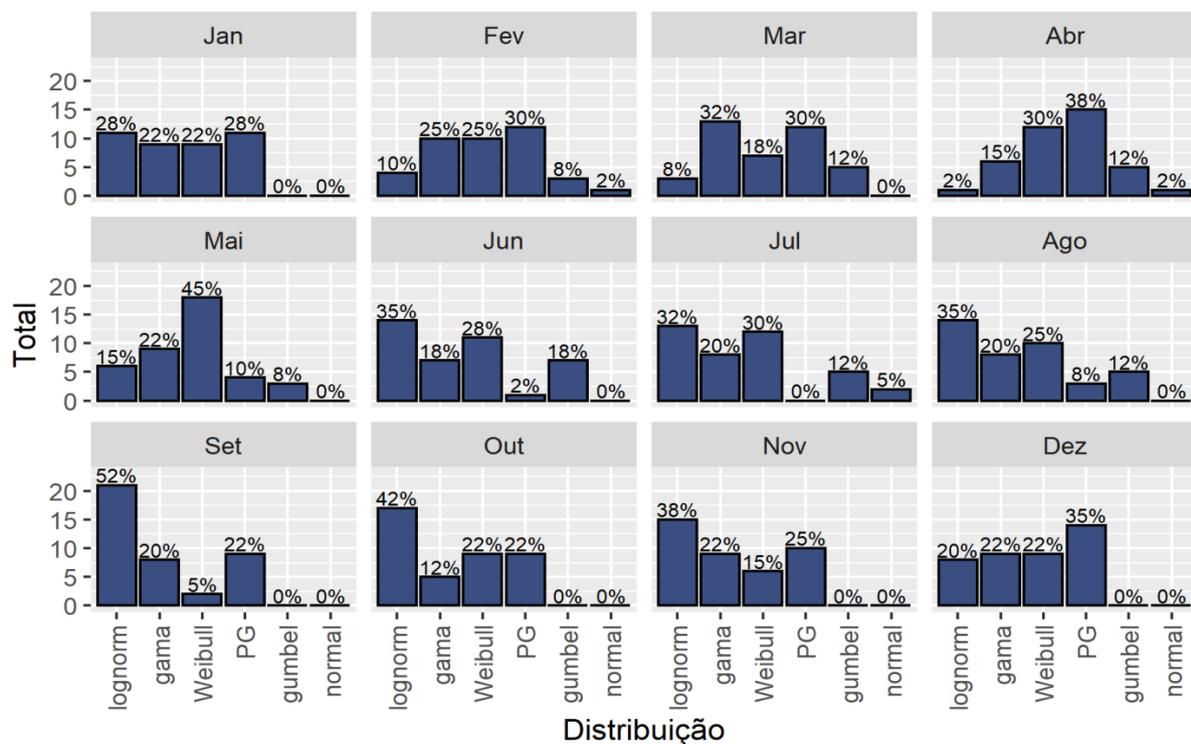
Figura 3. Boxplots do p -valor da estatística $TN.SW$, para as distribuições analisadas.



Fonte: Autores, (2020).

Na Figura 3, pode-se observar que, de forma geral, a maioria do p -valor esteve acima de 0,10. Isso significa que, a um nível de significância $\alpha = 0,05$, por exemplo, a maioria dos modelos não foram rejeitados. De forma geral, os modelos gama e Weibull apresentaram os melhores ajustes em comparação com os demais, mas o modelo log normal, durante um determinado período do ano, especialmente nos meses de janeiro e de junho até novembro, apresentou ajustes comparáveis aos apresentados pelos modelos gama e Weibull (Figuras 3 e 4). O modelo GP apresentou bons ajustes para os meses de janeiro, fevereiro e abril. Para quase todos os meses do ano, as distribuições Weibull e gama, apesar de grande variabilidade, apresentaram um primeiro quartil (Q1) acima de 0,125 (Figura 3), ou seja, menos de 25% de cada um desses modelos foram rejeitados à um nível de significância $\alpha = 0,125$.

Figura 4. Percentual das distribuições selecionadas por apresentar melhor ajuste aos dados (maior p -valor), selecionadas por mês.



Fonte: Autores, (2020).

Ao analisar as Figuras 2 e 3, o modelo log normal não aparenta ser tão interessante quanto os modelos gama e Weibull, mas observando a Figura 4 é possível identificar sete meses em que este modelo apresentou o maior percentual de melhor ajuste, ou seja, para os meses de janeiro, e de junho a novembro, o modelo log normal se ajustou melhor, ou tão bem quanto os demais. Pela Figura 3 é possível observar que de setembro a novembro as medianas dos valores p foram maiores que 0,25, mostrando que em menos de 50% dos ajustes esse

modelo foi rejeitado com nível de significância $\alpha = 0,25$. Em geral, os modelos gama e Weibull apresentaram os melhores ajustes os dados, mas o modelo log normal se mostrou tão eficiente quanto. O bom desempenho destes três modelos também foi identificado por Sharma & Singh (2010) que identificaram a distribuição log normal como uma das melhores distribuições no ajuste de dados de precipitação anual, e os modelos gama e Weibull no ajuste de dados de precipitação mensal da cidade de Pantnagar, Índia.

A distribuição Pareto Generalizada apresentou bons ajustes em períodos distintos do ano. Nos meses de janeiro a abril, bem como nos meses de setembro a dezembro, o modelo teve um desempenho comparável ao dos modelos Weibull e gama apresentando as medianas do *p-valor* acima de 0,25 (Figura 3), o que significa que menos de 50% das vezes esse modelo foi rejeitado com nível de significância $\alpha = 0,25$. Já nos meses de junho, julho e agosto o modelo apresentou baixos *p-valor*, com mediana próxima a zero em todos os casos. Além disso, foi observada alta variabilidade entre o *p-valor*.

Os modelos Gumbel e normal, apresentaram os menores *p-valor* da estatística *TN.SW*. O modelo normal se mostrou ineficiente em todos os meses analisados. Na maioria dos meses, os valores medianos dessa distribuição foram bem próximos de zero, indicando rejeição ao nível de significância $\alpha = 0,05$. O modelo Gumbel, diferente do modelo normal, apresentou alguns meses (março, abril e agosto) em que o a mediana do *p-valor* esteve acima de 0,25 indicando que menos de 50% das vezes esse modelo foi rejeitado com nível de significância $\alpha = 0,25$ (Figura 3), mas ainda assim, não foi eficiente quanto os demais modelos em estudo.

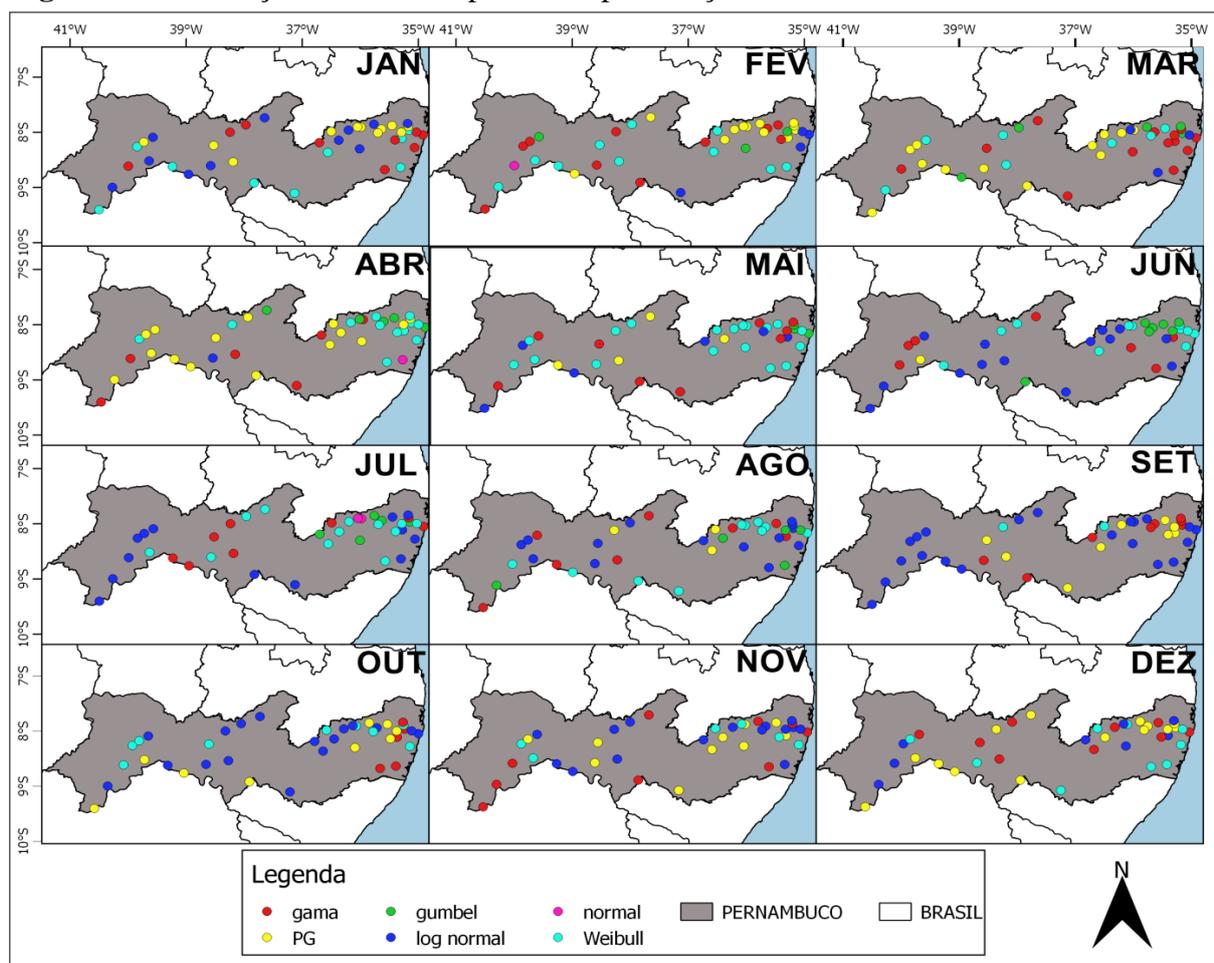
A Figura 4 apresenta o percentual das distribuições que melhor se ajustaram aos dados, selecionadas por mês. Assim como nas Figuras 2 e 3, as distribuições que mostraram bom potencial de ajuste foram gama, Weibull, log normal e PG. O modelo log normal apresentou melhor desempenho nos meses de janeiro e de junho a novembro, destacando-se entre os meses de agosto a setembro, em que apresentou um percentual de seleção bem acima das demais distribuições. A distribuição Pareto generalizada apresentou os maiores percentuais de seleção nos meses de janeiro a abril e de setembro a novembro. As distribuições Gumbel e normal não apresentaram desempenho superior em nenhum mês. Os modelos gama e Weibull estiveram entre as distribuições mais selecionadas para quase todos os meses do ano. O excelente desempenho dos modelos gama e Weibull não é uma surpresa neste estudo uma vez que estas distribuições já são reconhecidas na literatura por vários autores como excelentes candidatas a representar dados de hidrologia (Singh, 1987; Aksoy, 2000; Sharma & Singh, 2010; Elsherpieny, Muhammed, & Radwan, 2017; Ashkar, Ba, & Dieng, 2019).

A seleção da distribuição PG mostrou-se bem espacializada. É possível observar que este modelo foi mais selecionado nos meses de janeiro a abril e de outubro a dezembro. Este comportamento pode ser justificado, uma vez que esta distribuição é particularmente adequada para ajuste de dados com cauda direita pesada, e nos meses de janeiro a abril, algumas regiões do estado apresentam baixa média de chuvas.

Os modelos log normal e gama se ajustaram bem aos dados, principalmente para os meses de junho a novembro. Esses meses são caracterizados por uma precipitação média mais elevada no estado. O modelo log normal se destacou nos meses de setembro e outubro apresentando o melhor ajuste para a maioria das estações, independente da região em que a estação se localizava.

A distribuição Weibull apresentou maior concentração de seleção nos meses de abril e maio, concentrando-se principalmente próximo a costa litorânea do estado. Os modelos Gumbel e normal raramente foram selecionados em quase todos os meses analisados.

Figura 5. Distribuição selecionada por mês e por estação.



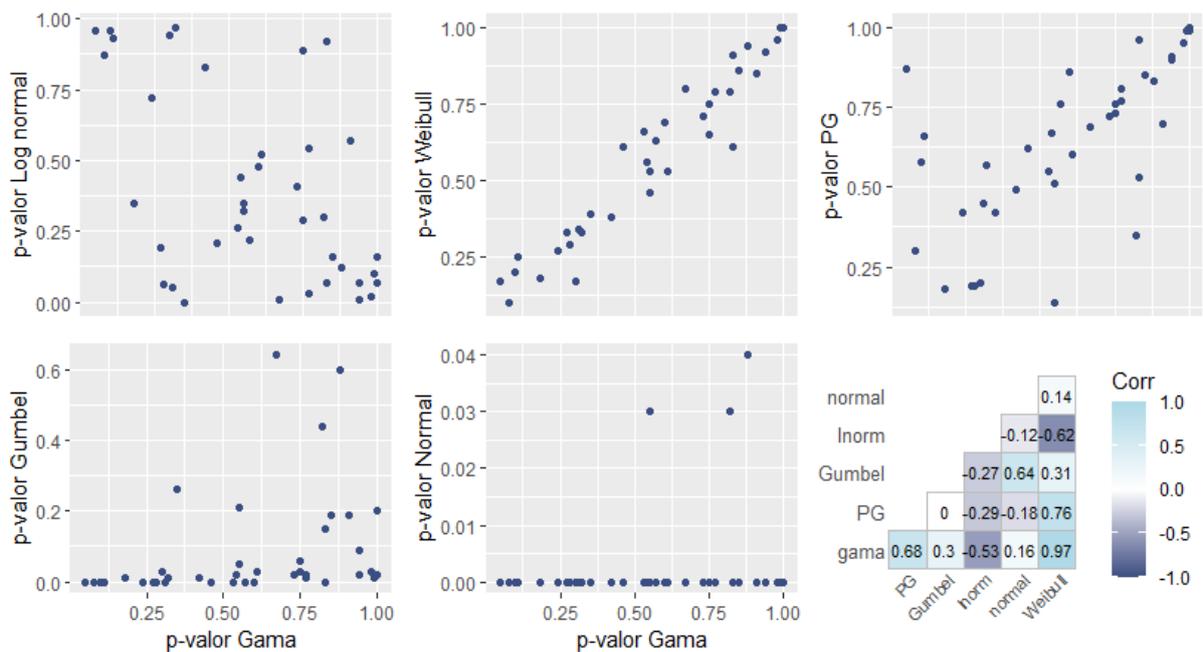
Fonte: Autores, (2020).

Algumas distribuições de probabilidade apresentaram comportamento semelhante na análise do *p-valor*. Para observar se existe alguma correlação entre os modelos, as Figuras 6 e 7 foram criadas. A partir dessas figuras, comparamos o *p-valor* entre os modelos analisados, e apresentamos um gráfico de dispersão mostrando a correlação entre essas variáveis, com base no coeficiente de correlação de Pearson (Pearson, 1896). Todos os meses foram analisados, porém os resultados foram semelhantes, dessa forma apresentaremos os resultados apenas para o mês de janeiro.

A Figura 6 fornece uma comparação entre o *p-valor* obtidos nos ajustes da distribuição gama e os demais modelos. É fácil ver uma forte relação linear positiva entre a distribuição gama e Weibull bem como o modelo gama e PG. A análise de correlação confirma que existe uma correlação forte e positiva (coeficiente de correlação $\rho = 0,97$) entre gama e Weibull e também entre gama e PG ($\rho = 0,68$).

Ao comparar o *p-valor* do modelo gama com os demais modelos, percebe-se que os pontos estão mais dispersos no gráfico, o que indica que não existe uma relação linear entre as distribuições ou que essa correlação é baixa. Os coeficientes de correlação calculados confirmam esta suspeita (gama x normal: $\rho = 0,16$, gama x Gumbel: $\rho = 0,3$).

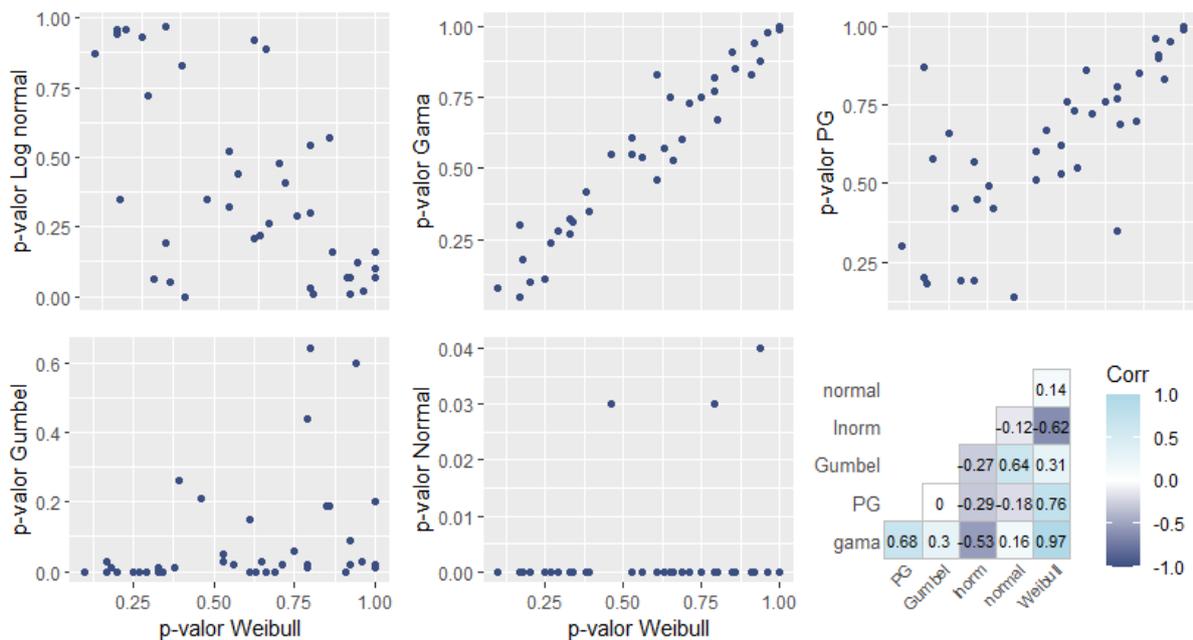
Figura 6. Comparação entre o *p-valor* das estatísticas *TN.SW* e correlações, janeiro (gama versus outros).



Fonte: Autores, (2020).

A Figura 7 fornece uma comparação entre o *p*-valor da distribuição Weibull e as demais. Assim como na Figura 6, o modelo Weibull também apresentou correlação linear com os modelos log normal e PG. É possível observar uma relação linear positiva entre Weibull e gama (já observado na Figura 6), bem como Weibull e PG ($\rho = 0,76$). No caso da distribuição log normal, a correlação é forte e negativa ($\rho = -0,62$). Comparando o modelo Weibull com os demais, é possível perceber pontos aleatórios no gráfico, significando que não há relação linear entre as variáveis.

Figura 7. Comparação entre o *p*-valor das estatísticas *TN.SW* e correlações, janeiro (Weibull versus outros).



Fonte: Autores, (2020).

Estes resultados estão de acordo com estudos anteriores que avaliaram ajustes de distribuições de probabilidade a dados de precipitação. Netto, Souza, & Lundgren, 2010 avaliaram o ajuste de seis distribuições de probabilidade para dados de precipitação mensal, em 28 municípios da região semi-árida do estado de Pernambuco e concluíram que as distribuições log normal, gama e Weibull se ajustam bem aos dados em alguns meses do ano, além disso, os pesquisadores também identificaram a ineficiência do modelo normal para o ajuste de dados de precipitação mensal. Lundgren, Souza & Netto (2015), avaliaram o ajuste de seis distribuições de probabilidade a dados de precipitação mensal em 72 estações pluviométricas distribuídas pelo estado de Sergipe, Brasil. Os resultados mostraram que os

modelos gama e Weibull, são ótimos candidatos para o ajuste de dados de precipitação mensal da região, em alguns meses do ano.

4. Considerações Finais

De forma geral, os modelos gama, Weibull, PG e log normal apresentaram os melhores ajustes em comparação com os demais modelos analisados. As distribuições gama e Weibull se ajustaram bem aos dados em todas as regiões do estado e apresentaram comportamento semelhante para quase todos os meses. O modelo gama se destacou principalmente nos meses de junho a novembro.

A distribuição Weibull apresentou maior percentual de ajuste nos meses de abril e maio, concentrando-se principalmente próximo a costa litorânea do estado. Embora os modelos gama e Weibull, de forma geral, tenham apresentado os melhores ajustes, o modelo log normal apresentou um ajuste comparável durante quase todo o ano, especialmente nos meses de setembro e outubro.

O modelo Pareto generalizado se ajustou melhor aos dados em regiões com baixa precipitação média. É possível observar que este modelo apresentou os melhores ajustes nos meses de janeiro a abril e de outubro a dezembro. Este comportamento pode ser justificado, uma vez que esta distribuição é particularmente adequada para ajuste de dados com cauda direita pesada, e nesse período, algumas regiões do estado de Pernambuco apresentam baixa média de chuvas.

Os resultados apresentados neste trabalho provaram que as distribuições de 2 parâmetros são flexíveis o suficiente para descrever dados de precipitação mensal para o estado de Pernambuco. De forma geral, este estudo identificou quais distribuições de probabilidade melhor se ajustam aos dados de precipitação mensal, relacionando os modelos mais adequados para cada mês do ano. Estes resultados serão úteis para estudos relacionados à seca na região, bem como na análise de precipitação.

Referências

Aksoy, H. (2000). Use of gamma distribution in hydrological analysis. *Turkish Journal of Engineering and Environmental Sciences*, 24(6), 419-428.

Ashkar, F., & Aucoin, F. (2012). Choice between competitive pairs of frequency models for use in hydrology: a review and some new results. *Hydrological sciences journal*, 57(6), 1092-106. <https://doi.org/10.1080/02626667.2012.701746>

Ashkar, F., & Ba, I. (2017). Selection between the generalized Pareto and kappa distributions in peaks-over-threshold hydrological frequency modelling. *Hydrological Sciences Journal*, 62(7), 1167-1180. <https://doi.org/10.1080/02626667.2017.1302089>

Ashkar, F., & Tatsambon, C. N. (2007). Revisiting some estimation methods for the generalized Pareto distribution. *Journal of Hydrology*, 346(3-4), 136-143. <https://doi.org/10.1016/j.jhydrol.2007.09.007>

Ashkar, F., Arsenault, M., & Zoglat, A. (1997). On the discrimination between statistical distributions for hydrological frequency analysis. In *The 1997 Annual Conference of the Canadian Society for Civil Engineering. Part 3(of 7), Sherbrooke, Can, 05/27-30/97* (pp. 169-178).

Ashkar, F., Ba, I., & Dieng, B. B. (2019, May). Hydrological Frequency Analysis: Some Results on Discriminating between the Gumbel or Weibull Probability Distributions and Other Competing Models. In *World Environmental and Water Resources Congress 2019: Watershed Management, Irrigation and Drainage, and Water Resources Planning and Management* (pp. 374-387). Reston, VA: American Society of Civil Engineers.

Bermudez, V.A.B.; Abilgos, A.B.B.; Cuaresma, D.C.N. & Rabajante, J.F (2017). Probability Distribution of Philippine Daily Rainfall Data. *Preprints*, 2017120150. <https://doi.org/10.20944/preprints201712.0150.v1>

Bjureland, W., Johansson, F., Sjölander, A., Spross, J., & Larsson, S. (2019). Probability distributions of shotcrete parameters for reliability-based analyses of rock tunnel support. *Tunnelling and Underground Space Technology*, 87, 15-26. <https://doi.org/10.1016/j.tust.2019.02.002>

Bolfarine, H., & Sandoval, M. C. (2001). *Introdução à inferência estatística* (Vol. 2). São Paulo: SBM.

Brito, S., Marengo, J., & Coutinho, M. (2017). Reduction of vulnerability to disasters: from knowledge to action. *Climate change and drought in Brazil* (pp. 361-376). São Paulo: Editora RIMA.

Chang, T. P. (2011). Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. *Applied Energy*, 88(1), 272-282. <https://doi.org/10.1016/j.apenergy.2010.06.018>

Cheng, K. S., Chiang, J. L., & Hsu, C. W. (2007). Simulation of probability distributions commonly used in hydrological frequency analysis. *Hydrological Processes: An International Journal*, 21(1), 51-60. <https://doi.org/10.1002/hyp.6176>

Elsherpieny, E. A., Muhammed, H. Z., & Radwan, N. U. M. M. (2017). On discriminating between gamma and log-logistic distributions in case of progressive type II censoring. *Pakistan Journal of Statistics and Operation Research*, 157-183. <https://doi.org/10.18187/pjsor.v13i1.1524>

Haberlandt, U., & Radtke, I. (2014). Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows. *Hydrology and Earth System Sciences* 18 (2014), Nr. 1, 18(1), 353-365. <http://dx.doi.org/10.5194/hess-18-353-2014>

Hussain, Z., Mahmood, Z., & Hayat, Y. (2010). Modeling the daily rainfall amounts of north-west Pakistan for agricultural planning. *Sarhad J. Agric*, 27(2), 313-321.

IBGE (2020). Instituto Brasileiro de Geografia e Estatística. Recuperado de <https://www.ibge.gov.br/>

Li, Z., Brissette, F., & Chen, J. (2013). Finding the most appropriate precipitation probability distribution for stochastic weather generation and hydrological modelling in Nordic watersheds. *Hydrological Processes*, 27(25), 3718-3729. <https://doi.org/10.1002/hyp.9499>

Lundgren, W. J. C., SOUZA, I. D., & NETTO, A. (2015). Uso de distribuições de probabilidades para ajuste aos dados de precipitação mensal do estado de Sergipe. *Revista Brasileira de Geografia Física*, 8(01), 071-080.

Mazucheli, J., & Emanuelli, I. P. (2019). The Nakagami Distribution Applied in Precipitation Data Analysis. *Revista Brasileira de Meteorologia*, 34(1), 1-7. <https://doi.org/10.1590/0102-77863340011>

McKee, T. B., Doesken, N. J., & Kleist, J. (1993, January). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* (Vol. 17, No. 22, pp. 179-183).

Netto, A. D. O. A., Souza, I. F. D., & Lundgren, W. J. C. (2010). Comparação entre distribuições de probabilidades da precipitação mensal no estado de Pernambuco. *Scientia Plena*, 6(6).

Papalexiou, S. M., & Koutsoyiannis, D. (2012). Entropy based derivation of probability distributions: A case study to daily rainfall. *Advances in Water Resources*, 45, 51-57. <https://doi.org/10.1016/j.advwatres.2011.11.007>

Pearson, K. (1896). VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, (187), 253-318. <https://doi.org/10.1098/rsta.1896.0007>

Pereira, A. S., Shitsuka, D. M., Parreira, F. J., & Shitsuka, R. (2018). Metodologia da pesquisa científica.[e-book]. Santa Maria. Ed. UAB/NTE/UFSM. Recuperado de https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-pesquisa-Cientifica.pdf.

R Core Team (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. Recuperado de <https://www.R-project.org/>.

Royston, J. P. (1982). Algorithm AS 181: the W test for normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 31(2), 176-180. <https://doi.org/10.2307/2347986>

Royston, J. P. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(2), 115-124. <https://doi.org/10.2307/2347973>

Royston, P. (1995). Remark AS R94: A remark on algorithm AS 181: The W-test for normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(4), 547-551. <https://doi.org/10.2307/2986146>

Santana, L. I. T., Silva, A. S. A., Menezes, R. S. C., & Stosic, T. (2020). Recurrence quantification analysis of monthly rainfall time series in Pernambuco, Brazil. *Research, Society and Development*, 9(9). e637997737. <https://doi.org/10.33448/rsd-v9i9.7737>

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591-611. <https://doi.org/10.2307/2333709>

Sharma, M. A., & Singh, J. B. (2010). Use of probability distribution in rainfall analysis. *New York Science Journal*, 3(9), 40-49.

Sijbers, J., den Dekker, A. J., Scheunders, P., & Van Dyck, D. (1998). Maximum-likelihood estimation of Rician distribution parameters. *IEEE Transactions on Medical Imaging*, 17(3), 357-361. <https://doi.org/10.1109/42.712125>

Silva, A. S. A. da, Menezes, R. S. C., Telesca, L., Stosic, B., & Stosic, T (2020). Fisher Shannon analysis of drought/wetness episodes along a rainfall gradient in Northeast Brazil. *International Journal of Climatology*, 1-14. <https://doi.org/10.1002/joc.6834>

Singh, V. P. (1987). On application of the Weibull distribution in hydrology. *Water Resources Management*, 1(1), 33-43. <https://doi.org/10.1007/BF00421796>

Stern, R. D., & Coe, R. (1982). The use of rainfall models in agricultural planning. *Agricultural Meteorology*, 26(1), 35-50. [https://doi.org/10.1016/0002-1571\(82\)90056-5](https://doi.org/10.1016/0002-1571(82)90056-5)

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7), 1696-1718. <https://doi.org/10.1175/2009JCLI2909.1>

Zhu, B., Chen, J., & Chen, H. (2019). Performance of multiple probability distributions in generating daily precipitation for the simulation of hydrological extremes. *Stochastic Environmental Research and Risk Assessment*, 33(8-9), 1581-1592. <https://doi.org/10.1007/s00477-019-01720-z>

Porcentagem de contribuição de cada autor no manuscrito

Patricia de Souza Medeiros Pina Ximenes – 25%

Antonio Samuel Alves da Silva – 25%

Fahim Ashkar – 25%

Tatijana Stosic – 25%