Acta Scientiarum

http://periodicos.uem.br/ojs
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v45i1.60621

STATISTICS

Mixture models of probability distributions applied to rainfall in the state of Pernambuco, Brazil

Luciano Pereira da Silva[1], Edgo Jackson Pinto Santiago[1], Frank Gomes-Silva[1*], Antonio Samuel Alves da Silva[1] and Rômulo Simões Cezar Menezes[2]

[1]Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros, s/n, 52171-900, Dois Irmãos, Recife, Pernambuco, Brazil. [2]Universidade Federal de Pernambuco, Cidade Universitária, 1235, 50670-901, Recife, Pernambuco, Brazil. *Author for correspondence. E-mail: franksinatrags@gmail.com

ABSTRACT. The Brazilian semi-arid region is recurrently affected by the scarcity of water that marks the landscape as it prints periods of severe drought. Therefore, rainfall in this region greatly influences plant growth in regional hydrological processes that affect droughts or floods. It is of practical interest to assess how changes in rainfall patterns occur to anticipate hydrological dynamics. However, this is not easy as climate change reshapes global hydrology. Thus, assertive forecasting has become rare and imputed estimates of a reasonable degree of uncertainty. The objective of this work was to verify from the mixture of exponential, gamma, beta, log-normal, Weibull, normal, log-logistic, and exponentiated log-logistic distributions, which best fits the monthly rainfall of the state of Pernambuco, Brazil. The data used came from 133 monthly rainfall series (1950 to 2012) distributed over the state of Pernambuco. The Maximum Likelihood Method estimated all parameters. The Kolmogorov-Smirnov adherence test was applied at 5% probability to assess the adjustments. The least rejected distributions in the adherence test were Weibull, gamma, and beta; October presented the smallest number of distributions considered adequate to model monthly rainfall. More than 99% of the rain gauge stations had some adequate probabilistic distribution to model monthly rainfall in March. For most months, except for March, the Weibull distribution was the most suitable for modeling the monthly rainfall in the vast majority of rain gauge stations of Pernambuco.

Keywords: rain; semi-arid; distribution mixture; modeling.

Received on August 21, 2021.
Accepted on August 22, 2022.

Introduction

The Brazilian semi-arid region is recurrently affected by water scarcity, which in addition to being unevenly distributed, marks the landscape by printing periods of severe drought. In this region, rainwater represents the main form of water supply, and for this reason, it is an important meteorological parameter in hydrological processes (Ljungqvist et al., 2016).

It is of practical interest to assess how changes in the rainfall regime occur to improve the view of rainfall variability and anticipate hydrological dynamics. However, this is not trivial as climate change, including changing regional rainfall patterns, has modified hydrological processes across multiple scales and will continue to reshape global hydrology (Guo, Zhang, Meng, Xu, & Song, 2020). Therefore, assertive hydrological forecasts have become less frequent on a regional and global scale. In response to global climate change, rainfall patterns are becoming rare, and the imputed estimates of a reasonable degree of uncertainty.

Rainfall collected by pluviometers or recorded in pluviographs constitutes the main data source for measuring rainfall events. However, these data are not always available in loco at rainfall stations or meteorological stations due to failure to obtain the data or record it. The pluviometric network may be inefficient for demonstrating the spatial distribution of rainfall since rainfall measurements are scarce and measurement stations are unevenly distributed in space. However, this network can be useful for adjusting probabilistic distributions that adequately model such a phenomenon, even more so if it introduces interpolation to examine the spatial distribution of rainfall. According to Ávila, Mello, and Viola (2009), rainfall estimation with a certain level of probability is of great importance in agricultural planning. For this, Barreto, Pereira, Santos, Freire, and Maia (2015) state that the use of probability functions needs to be directly linked to the nature of the data to which it relates.

Rainfall does not have a linear or stationary pattern, and in semi-arid regions, its variability is more evident with months without any record of rainfall. This lack of rainfall has to be considered when adjusting the probability density functions by the maximum likelihood estimation method, under penalty of misrepresentation of the studied phenomenon. This fact is because some probability functions, such as the gamma, have supports that are not defined at zero (absence of rainfall), and these estimates by maximum likelihood are infeasible (Saft, Western, Lu, Peel, & Potter, 2015; Deb, Kiem, & Willgoose, 2019)

Several works, such as those by Ribeiro and Lunardi (1997), Lyra, Garcia, Piedade, Sediyama, and Sentelhas (2006), and Soccol, Cardoso, and Miquelluti (2010), present an alternative to this limitation, which is to replace null rainfall by 0.1 mm. Others, such as Silva, Heldwein, Martins, Trentin, and Grimm (2007), suggest using rainfall data greater than or equal to 1 mm, neglecting smaller or null rainfall. Other works, however, consider only days with some amount of rain, as suggested by Kist and Filho (2015).

In all situations described above, for the null rainfall event, there will be unreasonable modeling of the studied phenomenon. This can lead to fragile inferences and compromise the study's credibility, especially when developed in arid or semi-arid regions, where the absence of rainfall predominates throughout the year.

From a practical standpoint, great caution should be exercised when considering the 0 mm rainfall event equivalent to the 0.1 mm rainfall event. Modeling a phenomenon that attests rainfall of 0.1 mm when in fact, it is zero means that the volume precipitated in 10,000 m^2 is 1.0 m^3, which is equivalent to 500 m^3 in a small hydrographic basin of 5 km^2. Hence the importance of opting for methodologies that do not replace or even exclude null rainfall but take them into account at some point when adjusting the probabilistic models (Fernandes, Schmidt, & Migon 2009).

For data containing many zeros, the modeling technique of mixing probability distributions can be an excellent way to estimate, especially when the homogeneity in the data series cannot be categorically faced.

From the point of view of probability theory, the assumption of an independent and identically distributed random variable is critically discussed by several authors (e.g., Klemes, 2000; Avanzi, Boglioni, Lafaye de Micheaux, Ouimet, & Wong, 2021). In practice, an argument against this assumption may lie in the different origins of rainfalls. A given rainfall event can result from variations in hydrological processes and a combination of different meteorological conditions. Different atmospheric manifestations (e.g., heavy rain, moderate rain, light rain, no rain) are often mixed within a single series of monthly or even annual rainfall, which can alter the independent and identical distribution of the random variable. Furthermore, different meteorological systems of atmospheric circulation can act in the formation of rainfall, such as cold fronts, eastern waves, intertropical convergence zone, land and sea breezes, and cyclonic vortices of the upper atmosphere.

In many cases, it is not an individual rainfall type that dominates rainfall records at regional levels but a mix of them. Another important question is whether rainfall events on a regional scale over time remain the same. All these questions point to the presence of heterogeneity in rainfall records due to the existence of natural and seasonal factors. This is because the different rainfall events occur to a greater or lesser degree only in a specific season (e.g., summer rains, winter rains, dry season), inevitably leading to non-rainfall events and zero-inflated data recording.

Suppose we accept that the rainfall series are heterogeneous and that several distributions must be combined to model the phenomenon correctly. In this sense, a mixed or mixed distribution is a distribution of a random variable obtained by a combination of other random variables. In this case, a convex model combining the distribution functions of these variables (Fischer, Schumann, & Schulte, 2016) can calculate the cumulative distribution function (cdf) of a finite mixture model.

In this context, this work aims to verify among mixtures of exponential, gamma, beta, log-normal, Weibull, normal, log-logistic, and exponentiated log-logistic distributions one best fit monthly rainfalls data from the state of Pernambuco, Brazil.

Material and methods

According to Andrade (2009), the Agreste of Pernambuco presents climatic types As' (hot and humid with rains from autumn to winter), BShs' (dry, low latitudes, with rains from autumn to winter), and Cs'a (mesothermal with hot summers and autumn-winter rains), according to the Köppen classification. In Sertão Pernambucano, according to the Köppen classification, the climate is semi-arid as Bswh type. In this region, the average annual rainfall in some municipalities further west of the state is 554.5 mm, distributed between

December and April. On the coast, the climate has characteristics of As' and BSh, according to Köppen, that is, precipitation in autumn and winter (Ramos, Silva, Sartori, Zimback, & Bassoi, 2011).

The data used to investigate the variability of rainfall in the state of Pernambuco came from 133 monthly rainfall series referring to rain gauge stations distributed throughout the state (Figure 1) collected during the period from 1950 to 2012. These data were provided by the Laboratory of Meteorology of Pernambuco (LAMEP), an agency belonging to the Pernambuco Institute of Technology (ITEP). The missing data were filled using the trend surface analysis interpolation method (Silva, Stosic, Menezes, & Singh, 2019).

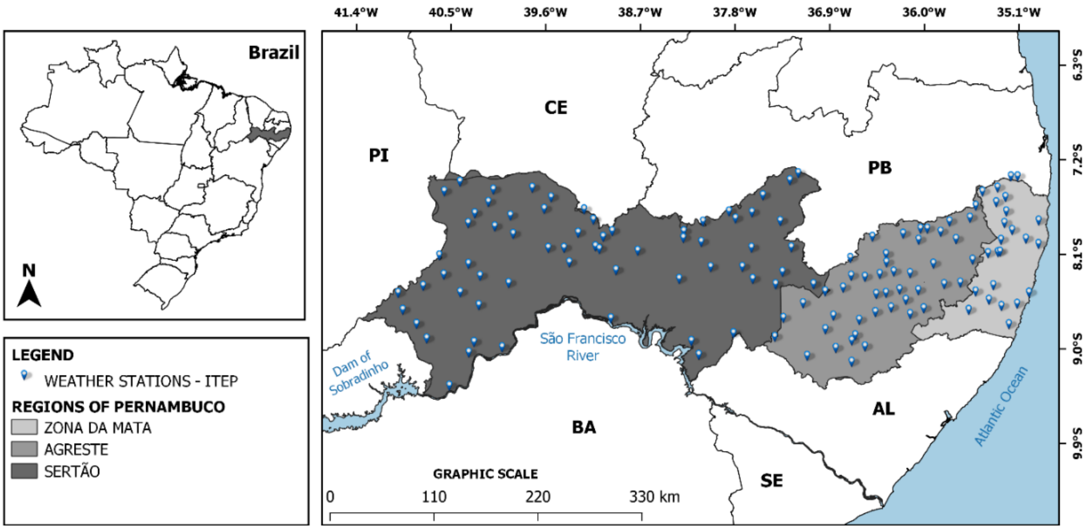

Figure 1. Spatial distribution of rainfall stations in Pernambuco State, Brazil.

With data from the series of each rainfall station and for each month of the year, eight probability distributions that are candidates for the best monthly rainfall modeling were adjusted. The distributions used in this study were exponential, gamma, beta, log-normal, Weibull, normal, log-logistic, and exponential log-logistic, whose probability density functions (pdf) are presented below.

i) Normal

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} I_{(-\infty,\infty)}(x) \tag{1}$$

where $\mu, \sigma \in R, \sigma > 0$, and I_A represents the indicator function of set A: $I_A(x) = \begin{cases} 1, \text{if} x \in A \\ 0, \text{if} x \notin A \end{cases}$;

ii) Log-Normal

$$f(x;\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{\frac{-(\log x-\mu)^2}{2\sigma^2}} I_{(0,\infty)}(x) \tag{2}$$

where μ is real, and σ is positive, are respectively the mean and standard deviation of the transformed variable Y = log X. In log-normal distribution, the random variable X is the antilog of a variable Y that follows a normal distribution;

iii) Gamma

$$f(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} I_{(0,\infty)}(x) \tag{3}$$

where α > 0 is the shape parameter, β > 0 is the scale parameter, and Γ is the gamma function. The gamma function is represented as follows:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx. \tag{4}$$

The pdf of the gamma distribution takes a wide variety of forms, depending on the value of the shape parameter;

iv) Exponential

$$f(x;\beta) = \frac{1}{\beta} e^{\frac{-x}{\beta}} I_{(0,\infty)}(x) \tag{5}$$

is a particular case of the gamma distribution when α = 1;

v) Beta

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)(b-a)^{\alpha+\beta+1}}(x-a)^{\alpha-1}(b-x)^{\beta-1}\, I_{(a,b)}(x) \tag{6}$$

where α > 0, and β > 0. In the model, the parameters α and β define the shape of the distribution, the values a (minimum) and b (maximum) represent the extremes of the distribution. In standard format a = 0 and b = 1. In general, if $\alpha \leq 1$, the probability is concentrated close to zero (for example, α = 0.25 and β=2 or α=1 and β=2), and for $\beta \leq 1$, the probability is concentrated close to 1. If both parameters are less than 1, the distribution is U-shaped. For α > 1 and β > 1 has a downward concavity between 0 and 1 (for example, α = 2 and β = 4 or α=10 and β = 2). If α=β, the distribution is symmetric; if α > β, the asymmetry is negative and, in the case of α < β, its asymmetry is positive;

vi) Weibull

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}} I_{(0,\infty)}(x) \tag{7}$$

where α>0 is the shape parameter, β>0 is the scale parameter. The shape of the Weibull distribution is also controlled equal to the gamma distribution by the two parameters; if α < 1 produces inverted "J" shapes and strong positive asymmetry, for $\alpha = 1$, the Weibull distribution also reduces to the exponential distribution. For $\alpha \approx 3.6$, the Weibull is very similar to the normal distribution. However, for larger shape parameters beyond that, the Weibull density exhibits negative asymmetry;

vii) Log-Logistic

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{-2} I_{(0,\infty)}(x) \tag{8}$$

where α > 0 is the shape parameter and β > 0 is the scale parameter;

viii) Exponentiated Log-Logistic

$$f(x;a,\alpha,\beta) = \frac{a\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-2}\left[1+\left(\frac{x}{\beta}\right)^{-\alpha}\right]^{-(a-1)} I_{(0,\infty)}(x) \tag{9}$$

where α > 0 is the shape parameter, β > 0 is the scale parameter and a > 0 the second shape parameter. Exponentiated log-logistic was proposed by Rosaiah, Kantam, and Kumar (2006) and constitutes a generalization of the logistic model.

Most rainfall studies exclude zero values from the analysis, making the rainfall data likely to be adequately modeled by continuous distributions, especially by range, a distribution widely used for this purpose. However, by doing this, there is a mischaracterization of the rainfall phenomenon, which includes both continuous and discrete values (zeros), the latter being so frequent in semiarid regions such as the region where most of the state of Pernambuco is located.

Null values, which represent non-rainy days, can be modeled through discrete distributions, using appropriate methodologies that consider inflated zeros. In contrast, rainfall values (non-zero values) can be modeled by continuous distribution. Kedem, Chiu, and Karni (1990) and Yoo, Jung, and Kim (2005) recommend the use of the mixed distribution proposed by Thom (1951) for the analysis of rainfall data. Thus, the mixed distribution function with random variable X, must be zero in non-rainy periods, and its probability will be denoted by:

$$P(X = 0) = p \tag{10}$$

where p represents the probability of zeros, that is, of non-rainy days. However, on rainy days, X follows a continuous distribution, whose probability of being less than or equal to a rainfall measure x can be expressed by using a cdf F(x):

$$P(X \leq x) = F(x), x > 0 \tag{11}$$

Thus, the distribution function considering rainy and non-rainy days is expressed by the mixed distribution:

$$G(x) = p + (1-p)\,F(x) \tag{12}$$

where G(x) is the mixed cumulative distribution, p is the probability of zero in the series, and F(x) is any continuous cdf fitted with data after removing the zeros.

In the present study, as some series of observed monthly rainfall data contained many zeros, that is, no rainfall, the use of the mixed cdf (Equation 12) in all rain gauge stations was standardized.

All parameters of the distributions were estimated by the Maximum Likelihood Method. To assess the adjustment associated with the rainfall frequency values observed with the estimated ones the Kolmogorov-Smirnov adherence test was applied at the level of 5% probability of type I error. In this case, adherence was applied considering each data series, including the zeros that had eventually been removed during the adjustments to obtain each F(x). This is because an important feature of the KS test is that it should not be applied to the same data sample used to estimate the distribution parameters (Vlcek & Huth, 2009). The Stata MP software version 16.0 (StataCorp, College Station, TX, USA) was used to perform all analyzes related to the adjustment of probability distributions.

Results and discussion

Figures 2 and 3 show the box plot of p-values with significant adhesions for all distributions studied, in each month of the year, adjusted for the 133 rainfall stations in Pernambuco. It appears that there was not a single distribution that stood out hegemonically in all the months.

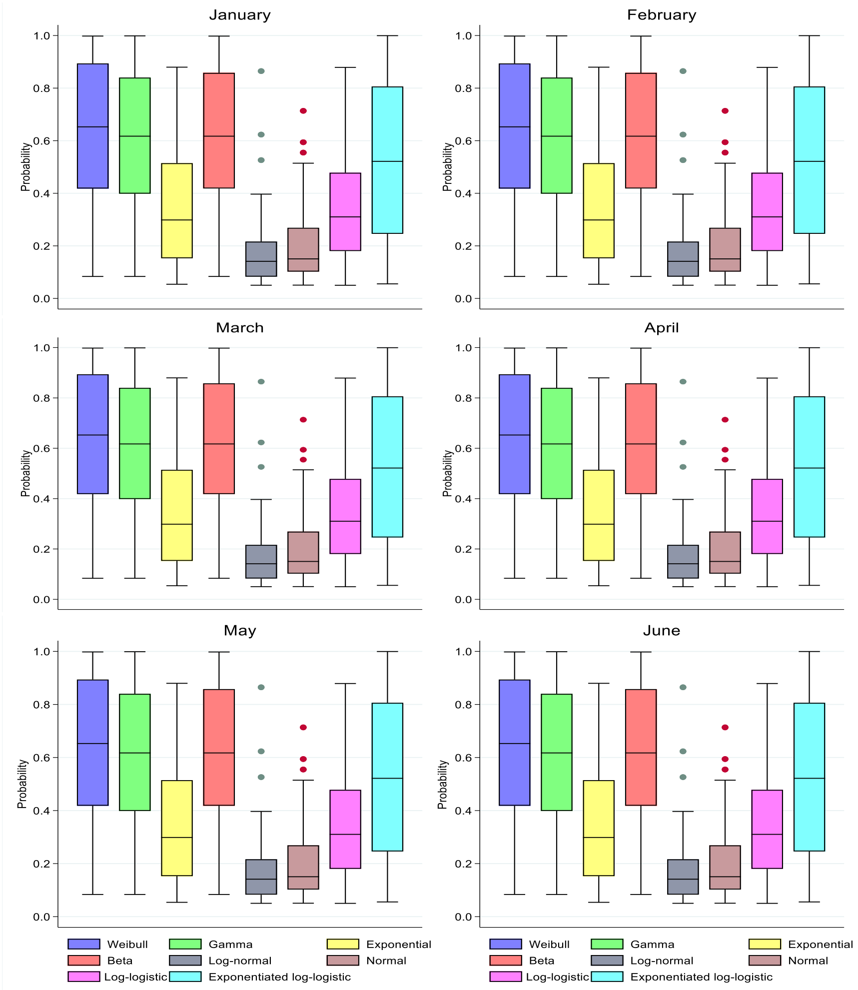

Figure 2. Box plots of the p-value K-S statistic from January to June.

In general, an alternation over the months between the Weibull, gamma, and beta can be observed, which presented higher p-values, except for August, which stood out with normal distribution. August is a month in which significant precipitation events are still observed in the Zona da Mata region, and the rain data, due to symmetry and spatio-temporal regularity, could be adequately modeled by the Normal distribution. It is important to note that, even though the Normal distribution stood out in August for the highest p-value, this distribution had only two significant adhesions among the 35 observed for the month in question, representing less than 6% of adhesions with a p-value > 0.05 in every state.

Figures 4 and 5 show the spatial arrangement of the most appropriate probability distributions for each rainfall station in the state of Pernambuco according to the highest p-value.

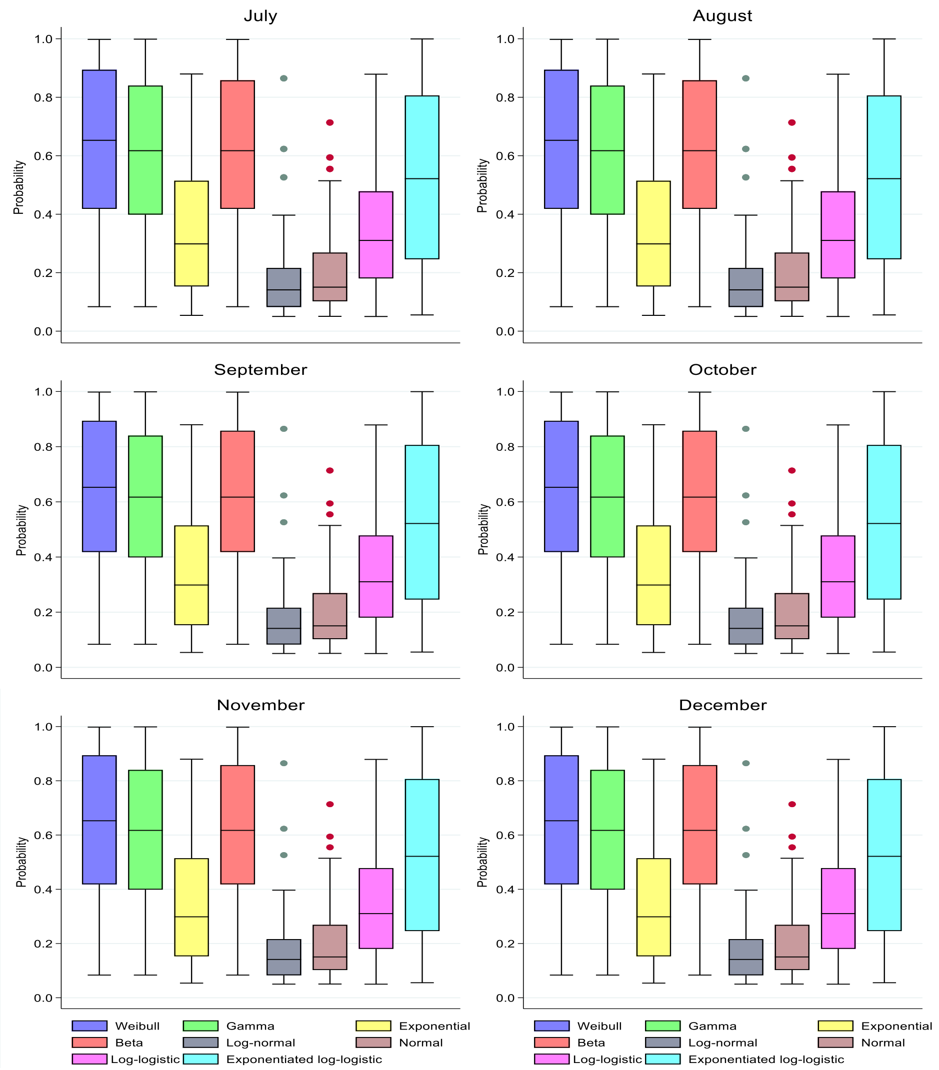

Figure 3. Box plots of the p-value K-S statistic from July to December.

Considering the results presented, March is the month with the highest number of adhesions in which the null hypothesis was accepted, considering a maximum probability of type I error of 5%. In this month, for 132 of the 133 pluviometric stations, some probability distribution, among the eight studied, was adequate to model the monthly rainfall. On the other hand, October was the month with the highest number of adhesions in which the null hypothesis was rejected; that is, the month with the lowest number of distributions considered adequate to model monthly rainfall. This month, across the state, there were 13 adhesions with p-values greater than 0.05 and a list of only three distributions, namely Weibull, beta, and log-logistic. These results may possibly be related to the fact that March is the month with the highest average rainfall distribution (335.9 mm) throughout the state and October the lowest (48.9 mm).

Historically, March is the wettest month in Sertão (Sertão of São Francisco and Sertão of Pernambuco) which represents more than 50% of the territory of Pernambuco and from which expressive rains have already been registered in Agreste region, despite the rainiest period in this region in months from April to August (Silva et al., 2022). The rainfall that occurs in Sertão and Agreste regions in March, together with that accumulated for that month on the coastal Zona da Mata region (more than 200 mm) and in Zona da Mata region (150 mm), according to the normal climatological average (Silva et al., 2022), contribute to a good rainfall coverage at the state level, favoring greater adequacy of the studied probability distribution models.

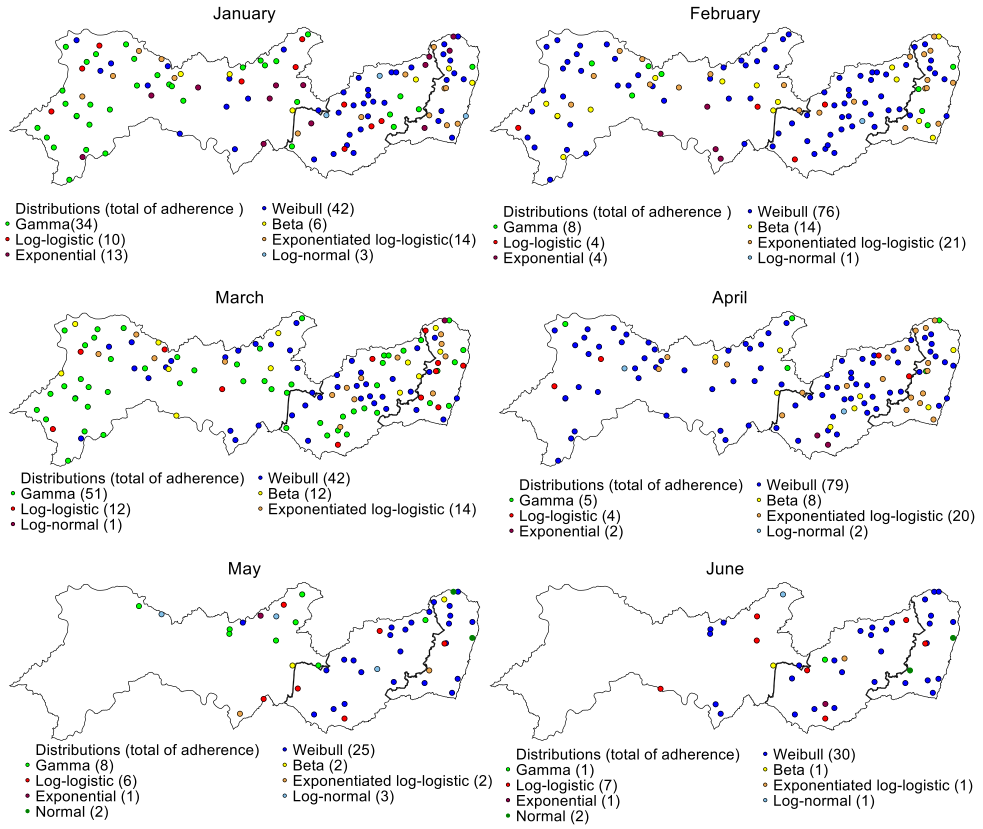

Figure 4. Spatial distribution of distribution adjustments in the regions of Pernambuco in from January to June.

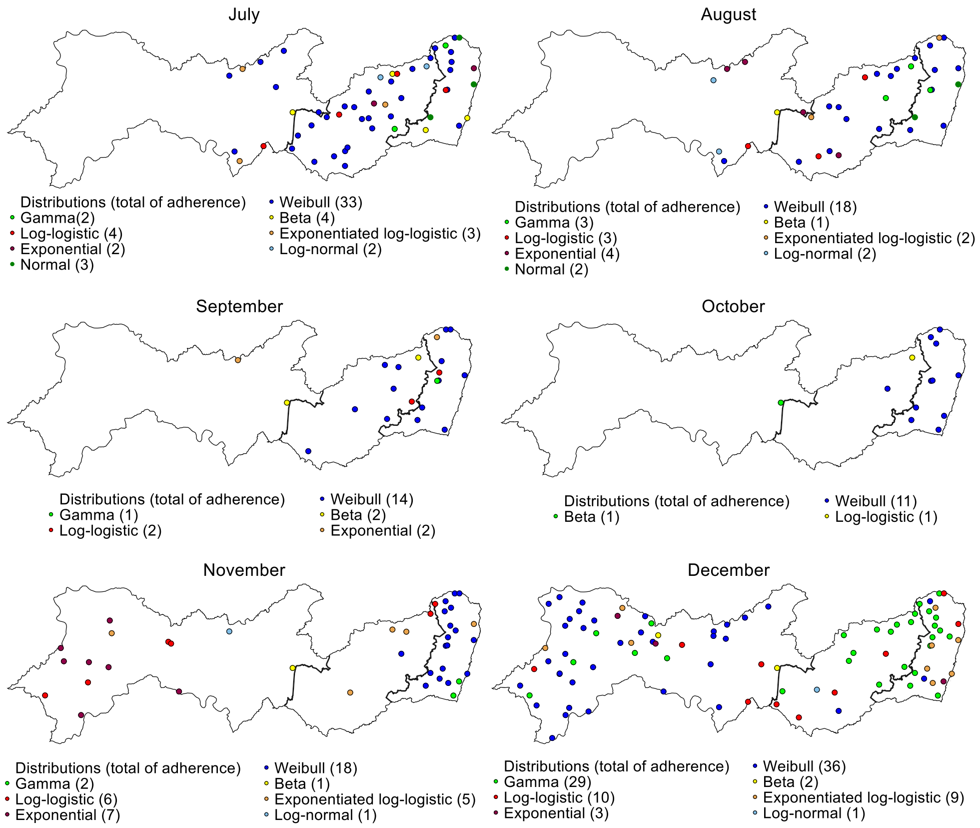

Figure 5. Spatial distribution of distribution adjustments in the regions of Pernambuco in from July to December.

Regarding the probability distributions (Figures 4 and 5), for most months, with the exception of March, the Weibull was used as the one that best suited the vast majority of rainfall stations. This distribution responded in April for 65.8% of all distributions with higher p-values among those that obtained p-values greater than 0.05 and is therefore classified as adequate to estimate monthly rainfall in 59.3% of all pluviometric stations studied in the state of Pernambuco. These results diverge from those obtained by Kist and Filho (2015) when analyzing probability distributions that best fit the historical series of rainfall in the state of Paraná. At the time, the authors reported that in all months, the gamma and Weibull distributions had higher p-values, so the gamma had a greater number of months (February, May, June, August, September, October, November, and December) and the other months assumed by Weibull. In Minas Gerais State, Catalunha, Sediyama, Leal, Soares, and Ribeiro (2002), when using rainfall data from 982 meteorological stations to fit six probability distributions, concluded that the Weibull best fit the rainfall data.

The Weibull distribution is a particular case of the generalized extreme value distribution, and as such, it models extreme phenomena well, being commonly used in hydrology for this purpose. Despite this, this work shows the great potential of this distribution in modeling monthly rainfall in the state of Pernambuco; that is, modeling a phenomenon, at first, inadequate for the purpose of its usual use. Ximenes, Silva, Ashkar, and Stosic (2021) by analyzing monthly precipitation data (in the period 1988–2017) in Northeast Brazil show both Gamma and Weibull models admit the best fits based on a modification of the Shapiro-Wilk test.

It is likely that the good adequacy of the Weibull distribution to rainfall data is inherent in its mathematical model, which is capable of presenting a wide variety of shapes that can be quite asymmetric. This is because the asymmetry, which can be positive or negative, in the frequency distribution, may have made the model effectively monitor the sudden reduction in rainfall frequencies. That is why when the adherence tests were applied, in addition to the absence of sufficient evidence for its rejection, they provided higher p-values. Figures 6 and 7 show the percentage of adjustment of the distributions in each region over the months.

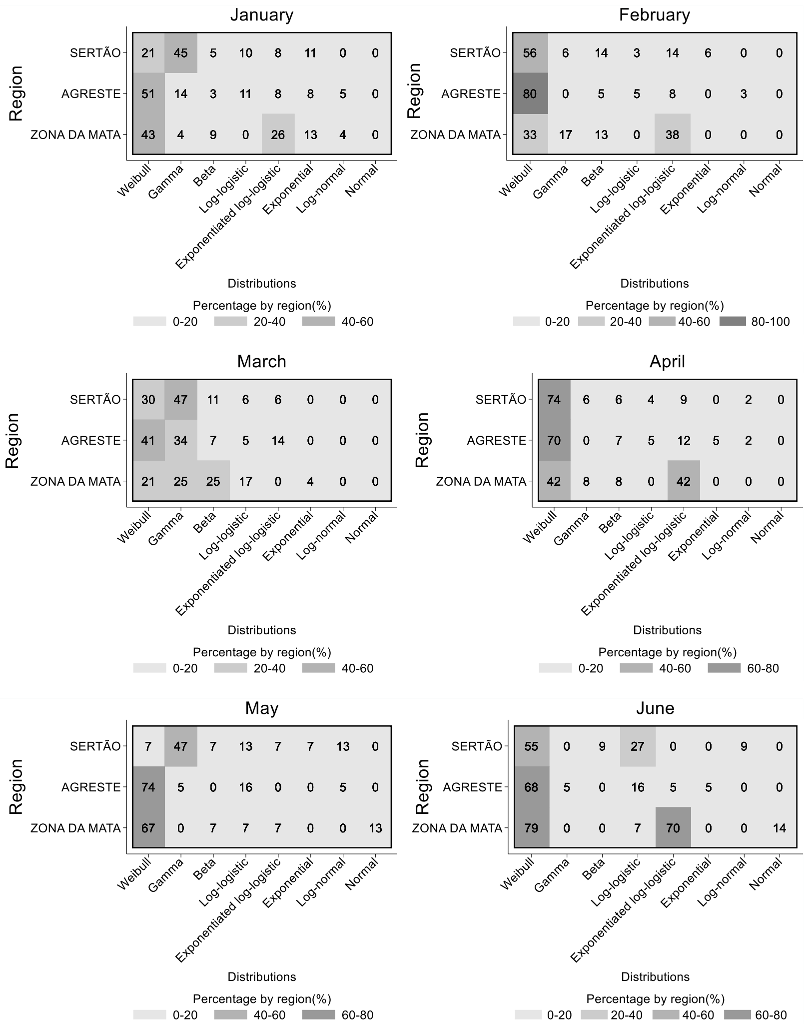

Figure 6. Percentage of distribution adjustments in the regions of Pernambuco from January to June.

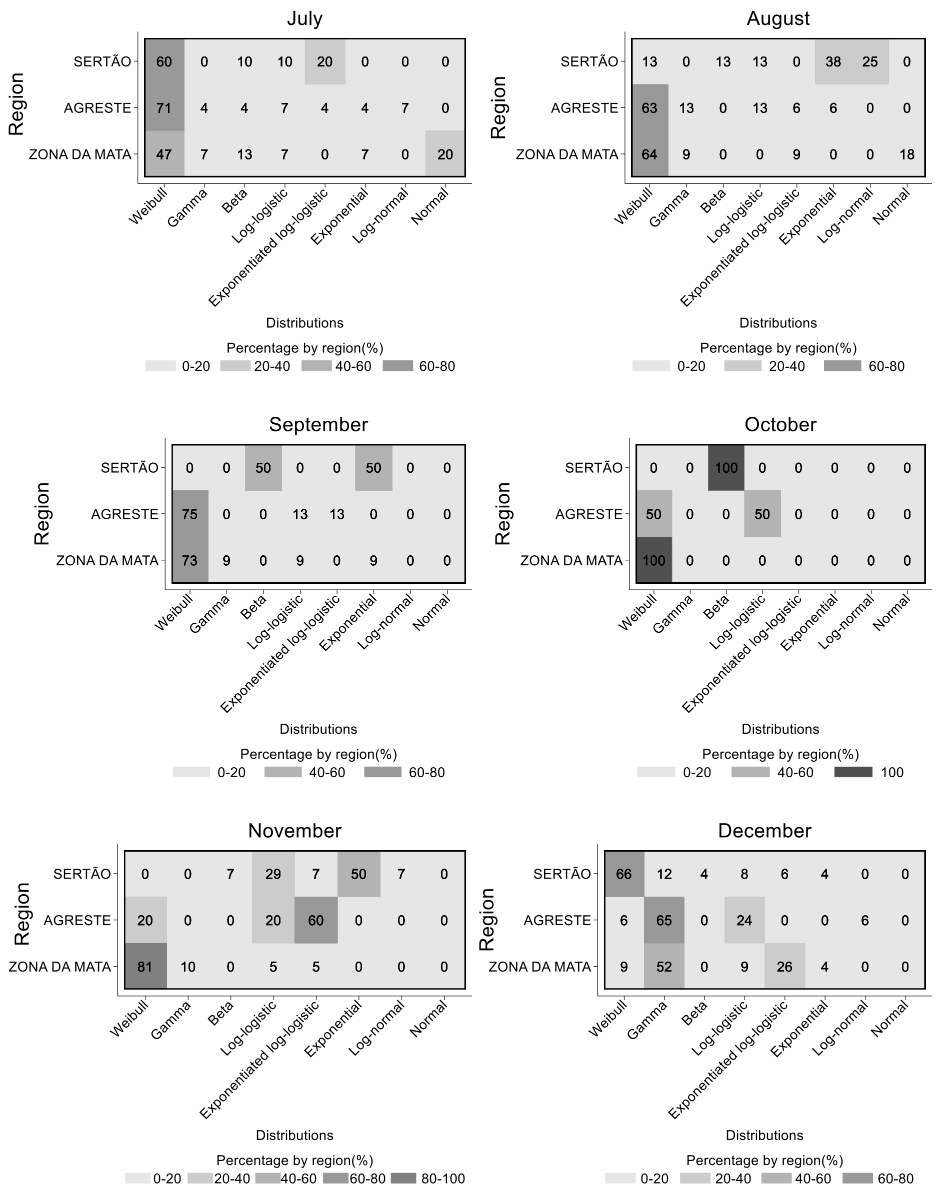

Figure 7. Percentage of distribution adjustments in the regions of Pernambuco from July to December.

It can be seen (Figures 6 and 7) that in Sertão, the gamma distribution stood out in the months of January, March, and May, whereas the Weibull in February, April, June, July, and December. In the other months, the beta and exponential distributions stood out. These results may be related to the occurrence of two distinct epochs in the region. The rainy season lasts from January to April, although there may already be occasional rains in November and the dry season in the other months of the year.

Nobrega, Farias, and Santos (2015), who studied the temporal and spatial variability of rainfall in Pernambuco from 1979 to 2010, found that in various parts of the Sertão, the annual rainy season comprises the months of December to April. In turn, Correia et al. (2011) and Reboita et al. (2016) highlighted in their studies on the causes of semi-aridity in the Northeast summer is the rainiest season of the year, comprising the months of January, February, and March, with some rainfall values already in December.

In general, it can be said that the gamma and Weibull distributions stood out in the rainy and transition seasons (wet to dry and dry to wet), while the beta and exponential in the period considered as dry. It is noteworthy that the dry season does not necessarily mean the absence of rainfall, but the occurrence of an average rainfall regime with low rates such as those that occur in August (6.8 mm month^{-1}), September (5.4 mm month^{-1}), and October (10.2 mm month^{-1}) in the Sertão.

Similar results were observed by Silva et al. (2007) and Silva, Oliveira, Fontes, and Arraes (2013), who found good behavior of the Weibull distribution in rainy periods and, in the case of the last author, good performance of the exponential distribution in the dry season. The exponential distribution models very well data with strong negative asymmetry, that is, data that, when distributed by their frequencies, present the highest values in the initial classes. The recurring situation with the rainfall of the dry season in the Sertão of Pernambuco.

Santiago, Silva, Gomes-Silva, and Silva (2019), when estimating the probable rainfall at different probabilistic levels in the Sertão of Pernambuco, they concluded that the gamma distribution also obtained a good fit for the data. At the time of the study, the highest historical monthly average of rainfall was 113.6 mm, very similar to the highest historical average (121.1 mm) for the month (March) in which the range stood out in the Sertão of the present study.

It is also observed that in the Agreste region (Figures 6 and 7), the Weibull distribution was hegemonically the one that stood out the most, but in the months of October, November, and December, the highlight was for the log-logistic, exponentiated log-logistic, and gamma distributions, respectively. In the coastal Zona da Mata region, the distributions with the highest percentages of adjustments followed a behavior similar to that observed in the Agreste region.

Conclusion

In general, the distributions with fewer rejections during the Kolmogorov-Smirnov adherence test were Weibull, gamma, and beta. October presented the smallest number of distributions considered adequate to model monthly rainfall in the state of Pernambuco. In March, more than 99% of the rainfall stations had some adequate probabilistic distribution to model monthly rainfall, being the month with the highest number of adhesions in which the null hypothesis was accepted. For most months, except for March, the Weibull distribution was the most suitable for modeling the monthly rainfall in the vast majority of rainfall stations in the state of Pernambuco.

Acknowledgments

This work is part of the National Observatory of Water and Carbon Dynamics in the Caatinga Biome - NOWCDCB, supported by FACEPE (grants: APQ-0498-3.07/17 INCT 2014, APQ-0500-5.01/22), CNPq (grants: INCT 465764/2014-2, 406202/2022-2, 440444/2022-5), and CAPES (grant: 88887.136369/2017-00).

References

Andrade, M. C. (2009). *Geografia de Pernambuco: ambiente e sociedade*. João Pessoa, PB: Grafset.

Avanzi, B., Boglioni, B. G., Lafaye de Micheaux, P., Ouimet, F., & Wong, B. (2021). A counterexample to the existence of a general central limit theorem for pairwise independent identically distributed random variables. *Journal of Mathematical Analysis and Applications*, *499*(1), 124982. DOI: https://doi.org/10.1016/j.jmaa.2021.124982

Ávila, L. F., Mello, C. R. D., & Viola, M. R. (2009). Mapeamento da precipitação mínima provável para o sul de Minas Gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *13*, 906-915. DOI: https://doi.org/10.1590/S1415-43662009000700013

Barreto, H. B. F., Pereira, G. M., Santos, W. O., Freire, F. G. C., & Maia, P. D. M. E. (2015). Avaliação de funções de probabilidade e relação-intensidade-duração-frequência para precipitação extrema em Nepomuceno-MG. *Global Science and Technology*, *8*(2), 118-127. DOI: https://doi.org/10.14688/1984-3801/gst.v8n2p118-1271

Catalunha, M. J., Sediyama, G. C., Leal, B. G., Soares, C. P. B., & Ribeiro, A. (2002). Aplicação de cinco funções densidade de probabilidade a séries de precipitação pluvial no Estado de Minas Gerais. *Revista Brasileira de Agrometeorologia*, *10*(1), 153-162.

Correia, R. C., Kiill, L. H. P., Moura, M. S. B., Cunha, T. J. F., Jesus Júnior, L. A., & Araújo, J. L. P. (2011). *A região semiárida brasileira*. Brasilia, DF: Embrapa Semiárido.

Deb, P., Kiem, A. S., & Willgoose, G. (2019). Mechanisms influencing non-stationarity in rainfall-runoff relationships in southeast Australia. *Journal of Hydrology*, *571*, 749-764. DOI: https://doi.org/10.1016/j.jhydrol.2019.02.025

Fernandes, M. V. M., Schmidt, A. M., & Migon, H. S. (2009). Modelling zero-inflated spatio-temporal processes. *Statistical Modelling*, *9*(1), 3-25. DOI: https://doi.org/10.1177/1471082X0800900102

Fischer, S., Schumann, A., & Schulte, M. (2016). Characterisation of seasonal flood types according to timescales in mixed probability distributions. *Journal of Hydrology*, *539*, 38-56. DOI: https://doi.org/10.1016/j.jhydrol.2016.05.005

Guo, B., Zhang, J., Meng, X., Xu, T., & Song, Y. (2020). Long-term spatio-temporal precipitation variations in China with precipitation surface interpolated by ANUSPLIN. *Scientific Reports*, *10*(1), 1-17. DOI: https://doi.org/10.1038/s41598-019-57078-3

Kedem, B., Chiu, L. S., & Karni, Z. (1990). An analysis of the threshold method for measuring area-average rainfall. *Journal of Applied Meteorology and Climatology*, *29*(1), 3-20. DOI: https://doi.org/10.1175/1520-0450(1990)029<0003:AAOTTM>2.0.CO;2

Kist, A., & Filho, J. S. V. (2015). Análise probabilística da distribuição de dados diários de chuva no estado do Paraná. *Revista Ambiente & Água*, *10*(1), 172-181. DOI: https://doi.org/10.4136/ambi-agua.1489

Klemeš, V. (2000). Tall tales about tails of hydrological distributions. I. *Journal of Hydrologic Engineering*, *5*(3), 227-231. Doi: https://doi.org/10.1061/(ASCE)1084-0699(2000)5:3(227)

Ljungqvist, F. C., Krusic, P. J., Sundqvist, H. S., Zorita, E., Brattström, G., & Frank, D. (2016). Northern Hemisphere hydroclimate variability over the past twelve centuries. *Nature*, *532*(7597), 94-98. DOI: https://doi.org/10.1038/nature17418

Lyra, G. B., Garcia, B. I. L., Piedade, S. M. D. S., Sediyama, G. C., & Sentelhas, P. C. (2006). Regiões homogêneas e funções de distribuição de probabilidade da precipitação pluvial no Estado de Táchira, Venezuela. *Pesquisa Agropecuária Brasileira*, *41*(2), 205-215.

Nóbrega, R. S., Farias, R. F. D. L., & Santos, C. A. C. D. (2015). Variabilidade temporal e espacial da precipitação pluviométrica em Pernambuco através de índices de extremos climáticos. *Revista Brasileira de Meteorologia*, *30*(2), 171-180. DOI: https://doi.org/10.1590/0102-778620130624

Ramos, C. M. C., Silva, A. F., Sartori, A. A. C., Zimback, C. R. L., & Bassoi, L. H. (2011). Modelagem da variação horária da temperatura do ar em Petrolina, PE, e Botucatu, SP. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *15*(9), 959-965. DOI: https://doi.org/10.1590/S1415-43662011000900012

Reboita, M. S., Rodrigues, M., Armando, R., Freitas, C., Martins, D., & Miller, G. (2016). Causas da semi-aridez do sertão nordestino (the causes of semi-aridity in the northeast "sertão"). *Revista Brasileira de Climatologia*, *19,* 254-277. DOI: https://doi.org/10.5380/abclima.v19i0.42091.

Ribeiro, A. M. A., & Lunardi, D. M. C. (1997). A precipitação mensal provável para Londrina-PR, através da função gama. *Energia na Agricultura*, *12*(4), 37-44.

Rosaiah, K., Kantam, R., & Kumar, S. (2006). Reliability test plans for exponentiated loglogistic distribution. *Economic Quality Control*, *21*(2), 279-289. DOI: https://doi.org/10.1515/EQC.2006.279

Saft M., Western, A. W., Lu, Z., Peel, M. C., & Potter, N. J. (2015). The influence of multiyear drought on the annual rainfall-runoff relationship: an Australian perspective. *Water Resources Research*, *51*(4), 2444-2463. DOI: https://doi.org/10.1002/2014WR015348

Santiago, E. J. P., Silva, L. P., Gomes-Silva, F., & Silva, A. S. A. (2019). Precipitação provável para o município de Afrânio-PE em diferentes níveis probabilístico. *Sigmae*, *8*(2), 437-448.

Silva, A. S. A., Stosic, B., Menezes, R. S. C., & Singh, V. P. (2019). Comparison of interpolation methods for spatial distribution of monthly precipitation in the state of Pernambuco, Brazil. *Journal of Hydrologic Engineering*, *24*(3), 04018068. DOI: https://doi.org/10.1061/(ASCE)HE.1943-5584.0001743

Silva, Í. N., Oliveira, J. B. D., Fontes, L. D. O., & Arraes, F. D. D. (2013). Distribuição de frequência da chuva para região Centro-Sul do Ceará, Brasil. *Revista Ciência Agronômica*, *44*(3), 481-487. DOI: https://doi.org/10.1590/S1806-66902013000300009

Silva, J. C. D., Heldwein, A. B., Martins, F. B., Trentin, G., & Grimm, E. L. (2007). Análise de distribuição de chuva para Santa Maria, RS. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *11*(1), 67-72. DOI: https://doi.org/10.1590/S1415-43662007000100009

Silva, T. R. B. F., Santos, C. A. C. dos, Silva, D. J. F., Santos, C. A. G., da Silva, R. M., & de Brito, J. I. B. (2022). Climate Indices-Based Analysis of Rainfall Spatiotemporal Variability in Pernambuco State, Brazil. *Water*, *14*(14), 2190. https://doi.org/10.3390/w14142190

Soccol, O. J., Cardoso, C. O., & Miquelluti, D. J. (2010). Análise da precipitação mensal provável para o município de Lages, Santa Catarina. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *14*(2), 569-574. DOI: https://doi.org/10.1590/S1415-43662010000600001

Thom, H. C. (1951). A frequency distribution for precipitation. *Bulletin of the American Meteorological Society*, *32*(10), 397.

Vlček, O., & Huth, R. (2009). Is daily precipitation Gamma-distributed? Adverse effects of an incorrect use of the Kolmogorov–Smirnov test. *Atmospheric Research*, *93*(4), 759-766. DOI: https://doi.org/10.1016/j.atmosres.2009.03.005

Ximenes, P. S. M. P., Silva, A. S. A., Ashkar, F., & Stosic, T., (2021). Best-fit probability distribution models for monthly rainfall of Northeastern Brazil. *Water Science and Technology*, *84*(6), 1541-1556. DOI: https://doi.org/10.2166/wst.2021.304

Yoo, C., Jung, K. S., & Kim, T. W. (2005). Rainfall frequency analysis using a mixed Gamma distribution: evaluation of the global warming effect on daily rainfall. *Hydrological Processes: An International Journal*, *19*(19), 3851-3861. DOI: https://doi.org/10.1002/hyp.5985