BENIN JOURNAL OF STATISTICS
ISSN 2682-5767
Vol. 7, pp. 113– 128 (2024)

Fitting the Distribution of Air Temperature for the State of Mato Grosso Do Sul, Brazil

E. Torsen[1, *] , Y. V. Mbaga [2], A. de Souza [3], and U. C. Dumka [4]

[1, 2]*Department of Statistics, Faculty of Physical Sciences, Modibbo Adama University, Yola - Nigeria*
[3]*Department of Physics, Federal University of Mato Grosso do Sul, Campo Grande, MS, Brazil*
[4]*Aryabhatta Research Institute of Observational Sciences, Manora Peak, Nainital, India*

(*Received: 3 April 2024; Accepted: 9 July 2024*)

Abstract. In this study, the objective was to verify the quality of the adjustment of the monthly average air temperature series in the State of Mato Grosso do Sul, using data retrieved from the National Institute of Meteorology networks in the state of Mato Grosso do Sul, Brazil. To achieve this objective, various distributions were fitted to the series. Goodness-of-fit test statistics (Kolmogorov-Smirnov, Anderson-Darling, Cramer-Von Mis), model selection criteria (AIC and BIC), and coefficient of Skewness and Kurtosis were employed, which showed that the air temperature is not uniformly distributed over the region. The result shows that the two-parameter Weibull distribution fits the data best. Furthermore, the air temperature of Mato Grosso do Sul, Brazil for the period of this study is slightly above average in most cases with few extremely low values. It is recommended that the two-parameter Weibull distribution be used when considering probability model for air temperature series in the state of Mato Grosso do Sul, Brazil.

Keywords: Air series, Clustering method, Temperature, Weibull distribution

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

Knowledge of the spatial variation of air temperature is essential to characterize and study the climate of a given region, perform agroclimatic zoning (Sediyama, 2002), assess climate risks for agricultural and forestry activities (Assad et al., 2003), characterize drought and desertification events (Gois et al., 2005), delimit ecological regions (Oliveira et al., 2002), analyze the distribution of native plant species (Buriol et al., 2007) and estimate global solar radiation (Meza

*Corresponding author. Email: torsen@mau.edu.ng

and Varas, 2000; Ihaddadene et al., 2019; de Souza et al., 2019; Oliveira et al., 2019). In addition to these, the knowledge of the current scenarios of the spatial distribution of air temperature are fundamental in the analysis of the impacts of climate change. This information mainly subsidizes socio-environmental policies, credit and rural and forestry security and energy generation.
Several studies that adjusted distributions of theoretical probability to climatic variables such as temperature data: Assis et al. (2004) fitted several probability distribution functions to the series of temperature and global radiation. In their study, the normal distribution was reported to have fitted the data better. Torsen et al. (2015) studied the reported maximum temperature in Adamawa state of Nigeria, by examining the PDF that fits the data best, Johnson S_B distribution was reported to outperform its counterparts. Assis et al. (2018) observed that out of the seven (7) PDFs that was fitted to the monthly average temperature data in Mossoro, Northern Brazil, the Normal PDF gave the best fit. Araujo et al. (2010) also reported the Normal distribution as the best PDF that adapts to the daily series of maximum temperature in Iguata city, Northeastern Brazil. de Souza et al. (2020) emphasized the benefits of modeling in planning activities to reduce climate risks, model adjustments for temperature and of interest to the public and should be studied.
The use of probability density functions is directly linked to the nature of the data to which these functions are related. Some have a good estimation capacity for a small number of data, others require a large series of observations. Due to the number of parameters in their equation, some can take different forms, fitting into a larger number of situations, that is, they are more flexible. As long as the representativeness of the data is respected, the estimates of its parameters for a given region can be established as general use, without prejudice to the precision in estimating the probability (Catalunha et al., 2002).
Continuous probability distributions are widely used in several probabilistic studies (Assis et al., 2004; Junqueira Junior et al., 2007; Lyra et al., 2006), due to the adjustment of their variables, which may not be perfect, but they describe a real situation well, providing answers to the hypotheses that may have been raised in the research.
The objective of this study is to verify the quality of the adjustment of monthly average air temperature series in the State of Mato Grosso do Sul, Brazil, with the following probability distributions: Normal, Log-Normal, 3-parameter Log-Normal, Gamma, 3-parameter Gamma, Weibull, 3-parameter Weibull, Gumbel and Rayleigh distributions. To do this, Goodness-of-fit statistics (Kolmogorov-Smirnov, Anderson-Darling test and Cramer-Von Mis) and model selection criteria (AIC and BIC) were used.

2. Materials and Methods

2.1 *The Study Area*

The state of Mato Grosso do Sul is located in the Midwest Region of Brazil, with approximately $358,159km^2$. The State is unique in the agricultural landscape since it emphasizes soy and cattle production and is the main source of revenue in the agricultural sector. The topography varies in elevation from 24 to 1,100 meters (Teodoro et al., 2016). The average yearly temperature and precipitation vary from 20 to 26°C and 1,000 to 1,900 mm, respectively. The Köppen classification represents numerous climatic types: Aw is found in the state's southeast and north, Am is in the middle region, Af is in the southwest region, and Cfa is in the southern state (de Souza et al., 2022; Dos Reis et al., 2022).

Tropical forest (Af) with year-round rainfall is the climate in the southwest of Mato Grosso do Sul, south of the Pantanal (between $-21°$ and $-22° latitudes$). With a brief winter dry season, the center part of the state has a monsoon climate (Am). The climate of Savannah (Aw), with dry winters and rainy summers, is found to the north, in addition to a small portion of the central region and the Southeastern state. The state's only region with a year-round humid climate is the south, where summers are hot and humid (Cfa) with highs of $< 22°$C (Abreu et al., 2021).

Mata Atlântica (14% of the state's total land), Cerrado (61% of the state's total area), and Pantanal (25% of the state's total area) are among the biomes that make up Mato Grosso do Sul (Abreu et al., 2021; Teodoro et al., 2016).

2.2 *Source of Data*

Data used in this study are monthly precipitation averages of 78 weather stations, retrieved from the National Institute of Meteorology (INMET) networks in the state of MS.

2.3 *Probability Distribution Functions (PDFs)*

In this study nine (9) probability distributions namely 2-parameter Weibull (W2P), 3-parameter Weibull (W3P), 2-parameter Rayleigh (RA2P), 2-parameter Gamma (G2P), 3-parameter Gamma (G3P), Normal (NORM), 2-parameter Lognormal (LN2P), 3-parameter Lognormal (LN3P) and Maximum Gumbel distributions (GUM) are used to model the temperature data in the of Mato Grosso do Sul. The probability distribution functions (PDF), corresponding cumulative distribution functions (CDF), domains and parameters for these distributions are discussed in Table 1 below.

2.4 *Model Selection Criteria and Goodness of Fit Tests*

Assessing the performance of different probability distribution models is necessary to provide more accurate information about rainfall at a particular location. In this study, in order to assess the goodness of fit (GOF) of the selected pdfs for rainfall data the GOF tests such as the Cramer-von Mises (CvM), Kolmogorov-Smirnov test (KS) and Anderson-Darling test (AD) are first applied and next the information criteria such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used. The goodness of fit tests are briefly described below. The Cramer-von Mises, Kolmogorov-Smirnov and Anderson-Darling tests were used to decide if the air temperature series follow the specified distributions.

Table 1: List of the Distributions, PDFs, CDFs, domains and parameters

Distributions	Probability Density Function (PDF)	Cumulative Distribution Function (CDF)	Domains	Parameters
W2P	$f(x)=\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$	$F(x)=1-\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$	$x>0$	$\alpha>0$, $\beta>0$
W3P	$f(x)=\frac{\alpha}{\beta^{\alpha}}(x-\theta)^{\alpha-1}\exp\left(-\left(\frac{x-\theta}{\beta}\right)^{\alpha}\right)$	$F(x)=1-\exp\left(-\left(\frac{x-\theta}{\beta}\right)^{\alpha}\right)$	$x\geq\theta$	$\alpha>0$, $\beta>0$
RA2P	$f(x)=\frac{2}{\beta^{2}}(x-\theta)\exp\left(-\left(\frac{x-\theta}{\beta}\right)^{2}\right)$	$F(x)=1-\exp\left(-\left(\frac{x-\theta}{\beta}\right)^{2}\right)$	$x\geq\theta$	$\beta>0$
G2P	$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)$	$F(x)=\frac{\gamma\left(\alpha,\frac{x}{\beta}\right)}{\Gamma(\alpha)}$	$x>0$	$\alpha>0$, $\beta>0$
G3P	$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}(x-\theta)^{\alpha-1}\exp\left(-\frac{x-\theta}{\beta}\right)$	$F(x)=\frac{\gamma\left(\alpha,\frac{x-\theta}{\beta}\right)}{\Gamma(\alpha)}$	$x>\theta$	$\alpha>0$, $\beta>0$
NORM	$f(x)=\frac{1}{\sigma\sqrt{(2\pi)}}exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}\right\}$	$F(x)=\frac{1}{2}\left[1+erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$	$x\in\mathbb{R}$	$\mu\in\mathbb{R}$ $\sigma^{2}>0$
LN2P	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^{2}\right)$	$F(x)=\Phi\left(\frac{\ln x-\mu}{\sigma}\right)$	$x>0$	$\mu\in R$, $\sigma>0$
LN3P	$f(x)=\frac{1}{(x-\theta)\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln(x-\theta)-\mu}{\sigma}\right)^{2}\right)$	$F(x)=\Phi\left(\frac{\ln(x-\theta)-\mu}{\sigma}\right)$	$x>\theta$	$\mu\in R$, $\sigma>0$
GUM	$f(x)=\frac{1}{\beta}\exp\left(-\frac{x-\mu}{\beta}-\exp\left(-\frac{x-\mu}{\beta}\right)\right)$	$F(x)=\exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right)$	$x\in R$	$\mu\in R$, $\beta>0$

2.4.1 Cramer-von Mises (CvM) Test

Let $x_1, x_2, ..., x_n$ be the observed values in increasing order, then the CvM one sample test statistic as given by Pettitt (1976) is

$$\mathcal{T} = n\omega^2 = \sum_{i=1}^{n}\left[\frac{2i-1}{2n} - F(x_i)\right]^2 \tag{1}$$

where $\omega^2 = \int_{-\infty}^{\infty}\left[F_n(x) - F^*(x)\right]dF^*(x)$, F^* = theoretical distribution and F_n = empirical observed distribution.

If equation (1) is larger than critical value, the null hypothesis that the data came from the distribution F can be rejected.

2.4.2 Anderson-Darling (AD) Test

Given that $x_1, x_2, ..., x_n$ is the sample, then the AD test statistic (Anderson, 2011) is

$$\mathcal{A}_n^2 = -n - \sum_{i=1}^{n}\left[\frac{2i-1}{n}[log(F(x_i)) + log(1 - F(x_{n+1-i}))]\right] \tag{2}$$

where n = the sample size, $F(x)$ = the CDF of the specified distribution, and

i = the i^{th} sample, computed when the data is arranged in ascending order of magnitude.

2.4.3 Kolmogorov-Smirnov (KS) Test

The KS test statistic (Yazici and Yolacan, 2007) is defined as follows

$$\mathcal{D}_n = \max_{1\le i\le n}\left\{\left|\frac{i}{n} - \hat{F}(x_{(i)})\right|, \left|\hat{F}(x_{(i)}) - \frac{i-1}{n}\right|\right\} \tag{3}$$

Where, $x_{(1)}, x_{(2)}, ..., x_{(n)}$ are observations in ascending order, so that $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$. The hypothesis that, the data follow specified distribution is rejected at the significance level α if the test statistic D_n is greater than the critical value of the KS test $D_n(\alpha)$.

2.4.4 Akaike Information Criterion (AIC)

The AIC (Bozdogan, 1987) is known as commonly used model selection criterion that is calculated based on the maximized value of the log-likelihood function for the estimated model. The AIC can be calculated as follows

$$AIC = -2lnL + 2k \tag{4}$$

2.4.5 Bayesian Information Criterion (BIC)

The BIC (Neath and Cavanaugh, 2012) is another commonly used information criterion and it is closely related to the AIC. The BIC can be calculated as follows

$$BIC = -2lnL + kln(n) \tag{5}$$

2.5 *The W2P Distribution*

The W2P probability distribution function is the best PDF for modeling air temperature in Mato Grosso Do Sul, Brazil, hence, we briefly discussed W2P as obtainable in Nielsen (2011) is given below. If $X \sim Weibull(\alpha, \beta)$ then its PDF is defined as:

$$f(x|\alpha,\beta) = \frac{\alpha}{\beta^{\alpha}} x^{(\alpha-1)} exp\left\{-\left(\frac{x}{\beta}\right)^{\alpha}\right\}, x \ge 0, \alpha > 0 \quad and \quad \beta > 0 \tag{6}$$

Given this density function, α is the shape parameter and β represents the scale parameter.
The CDF is defined as:

$$F(x|\alpha,\beta) = 1 - exp\left\{-\left(\frac{x}{\beta}\right)^{\alpha}\right\}, x \ge 0, \alpha > 0 \quad and \quad \beta > 0 \tag{7}$$

The mean is

$$\mathrm{E}(X) = \beta\Gamma\left(1 + \frac{1}{\alpha}\right) \tag{8}$$

The median is

$$Med(X) = \beta\Gamma(log2)^{\frac{1}{\alpha}} \tag{9}$$

The mode is given as

$$Mod(X) = \begin{cases} \beta\Big(\dfrac{\alpha-1}{\alpha}\Big)^{\frac{1}{\alpha}}, & \alpha > 1 \\ 0, & \alpha \leq 1 \end{cases} \tag{10}$$

The variance is given by

$$Var(X) = \beta^2\Big\{\Gamma\big(1+\frac{2}{\alpha}\big) - \big[\big(1+\frac{1}{\alpha}\big)\big]^2\Big\} \tag{11}$$

The skewness is denoted by

$$Skew(X) = \gamma_1 = \frac{\Gamma\Big(1+\frac{3}{\alpha}\Big)\beta^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3} \tag{12}$$

Where μ is the mean and σ_2 is the variance.

The Excess Kurtosis is given by

$$Ex.Kurt(X) = \gamma_2 = \frac{-6\Gamma_1^4 + 12\Gamma_1^2\Gamma_2 - 3\Gamma_2^2 - 4\Gamma_1\Gamma_3 + \Gamma_4}{[\Gamma_2 - \Gamma_1^2]^2} \tag{13}$$

Where $\Gamma_i = \Gamma\Big(1+\frac{i}{\alpha}\Big)$.
Equation (13) can be written as as

$$Ex.Kurt(X) = \gamma_2 = \frac{\beta^4\Gamma(1+\frac{4}{\alpha}) - 4\gamma_1\sigma^3\mu - 6\mu^2\sigma_2 - \mu_4}{\sigma^4} - 3 \tag{14}$$

2.6 *Parameter Estimation*

It is known that several methods can be used to estimate the parameters of pdfs, the maximum likelihood estimation (MLE) was used for selecting the appropriate probability distribution. Furthermore, maximum likelihood (MLE), moment matching (MME), quantile matching (QME) and maximum goodness-of fit estimation (MGE) methods were compared for the selected probability distribution. Consequently, the MLE method presents better estimators in determining the estimates of parameters of each probability distribution. Hence, it is discussed below. For more information about the parameter estimation method see : Rao et al. (1973), Montgomery and Runger (2010), Coles et al. (2001), and Delignette-Muller and Dutang (2015).
The joint density of the maximum likelihood (Nielsen, 2011) is given as the product of the densities of each realization, thus from equation (6) we have

$$
\begin{aligned}
L(\alpha,\beta|x) &= \prod_{i=1}^{n} f(x_i|\alpha,\beta)\\
&= \prod_{i=1}^{n}\left\{\frac{\alpha}{\beta^{\alpha}}x^{(\alpha-1)}exp\big\{-\big(\frac{x}{\beta}\big)^{\alpha}\big\}\right\} \qquad (15)\\
&= \left(\frac{\alpha}{\beta^{\alpha}}\right)^{n}\prod_{i=1}^{n}x_i^{(\alpha-1)}\times exp\big\{-\big(\frac{\sum_{i=1}^{n}x_i}{\beta}\big)^{\alpha}\big\}
\end{aligned}
$$

Next, the log-likelihood transformation is given by

$$
\begin{aligned}
logL(\alpha,\beta|x) &= \left\{\big(\frac{\alpha}{\beta^{\alpha}}\big)^{n}\prod_{i=1}^{n}x_i^{(\alpha-1)}\times exp\big\{-\big(\frac{\sum_{i=1}^{n}x_i}{\beta}\big)^{\alpha}\big\}\right\}\\
&= log\big(\frac{\alpha}{\beta^{\alpha}}\big)^{n} + log\Big[\prod_{i=1}^{n}x_i^{(\alpha-1)}\Big] - \big(\frac{\sum_{i=1}^{n}x_i}{\beta}\big)^{\alpha}\\
&= nlog\big(\frac{\alpha}{\beta^{\alpha}}\big) + (\alpha-1)log\Big[\prod_{i=1}^{n}x_i\Big] - \sum_{i=1}^{n}\big(\frac{x_i}{\beta}\big)^{\alpha} \qquad (16)\\
&= nlog\alpha - nlog\beta^{\alpha} + (\alpha-1)\sum_{i=1}^{n}logx_i - \sum_{i=1}^{n}\big(\frac{x_i}{\beta}\big)^{\alpha}\\
&= nlog\alpha - n\alpha log\beta + (\alpha-1)\sum_{i=1}^{n}logx_i - \sum_{i=1}^{n}\big(\frac{x_i}{\beta}\big)^{\alpha}
\end{aligned}
$$

Differentiating equation (16) with respect to α and β and equating to zero, we have

$$
\frac{\partial logL(\alpha,\beta|x)}{\partial\alpha} = \big(\frac{n}{\alpha}\big) + \sum_{i=1}^{n}\big(\frac{x_i}{\beta}\big) - \sum_{i=1}^{n}\big(\frac{x_i}{\beta}\big)^{\alpha}log\big(\frac{x_i}{\beta}\big) = 0 \qquad (17)
$$

and

$$
\frac{\partial logL(\alpha,\beta|x)}{\partial\beta} = -n\big(\frac{\alpha}{\beta}\big) + \big(\frac{\alpha}{\beta}\big)\sum_{i=1}^{n}\big(\frac{x_i}{\beta}\big)^{\alpha} = 0 \qquad (18)
$$

From equation (18),

$$
\beta = \Big[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}\Big]^{\frac{1}{\alpha}} \qquad (19)
$$

$\hat{\alpha}$ can be determined using Newton-Raphson method discussed below. Once, $\hat{\alpha}$ is determined, $\hat{\beta}$ can be obtained from equation (19), see Guure et al. (2012) for details.

Let $h(\alpha)$ be as given in equation (17), differentiating $h(\alpha)$ gives

$$h'(\alpha) = \Big(\frac{n}{\alpha^2}\Big) - \sum_{i=1}^{n} \Big(\frac{x_i}{\beta}\Big)^{\alpha} log^2\Big(\frac{x_i}{\beta}\Big) \tag{20}$$

Putting equation (19) into equation (17) yields

$$h(\alpha) = \big(\frac{n}{\alpha}\big) - \sum_{i=1}^{n}\left[\frac{x_i}{[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}]^{\frac{1}{\alpha}}}\right] - \sum_{i=1}^{n}\left[\frac{(x_i)^{\alpha}}{\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}}\right] log\left[\frac{x_i}{[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}]^{\frac{1}{\alpha}}}\right] \tag{21}$$

Substitute equation (19) into equation (20), we have

$$h'(\alpha) = -\left\{\big(\frac{n}{\alpha}\big) + \sum_{i=1}^{n}\left[\frac{(x_i)^{\alpha}}{\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}} log\frac{x_i}{[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}]^{\frac{1}{\alpha}}}\right]\right\} \tag{22}$$

Hence, choosing an initial value for α_i carefully and repeating the process until it converges, $\hat{\alpha}$ can be obtained.

$$\begin{aligned} \alpha_{i+1} &= \alpha_i \\ &- \frac{\big(\frac{n}{\alpha}\big) + \sum_{i=1}^{n}\left[(x_i)/[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}]^{\frac{1}{\alpha}}\right]}{-\left\{\big(\frac{n}{\alpha}\big) + \sum_{i=1}^{n}\left[\Big((x_i)^{\alpha}/(\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha})\Big) log^2\big((x_i)/[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}]^{\frac{1}{\alpha}}\big)\right]\right\}} \\ &- \frac{\sum_{i=1}^{n}\left[(x_i)^{\alpha}/[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}]\right] log\left[(x_i)/[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}]^{\frac{1}{\alpha}}\right]}{-\left\{\big(\frac{n}{\alpha}\big) + \sum_{i=1}^{n}\left[\Big((x_i)^{\alpha}/(\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha})\Big) log^2\big((x_i)/[\frac{1}{n}\sum_{i=1}^{n}(x_i)^{\alpha}]^{\frac{1}{\alpha}}\big)\right]\right\}} \end{aligned} \tag{23}$$

3. Results and Discussion

In this study, Nine (9) probability distributions listed in Table 1 were used to fit the average Air Temperature of each of the seven regions of the state of Brazil. Appendix A shows the density and cumulative density plot for the various regions while Appendix B contains goodness-of-fit statistic and model selection criteria. For all the regions the result is similar, 2-parameter Weibull distribution was found to perform better having lower goodness-of-fit criteria. The Average Temperature of Campo Grande region is hence adopted for further investigation. Appendix C is the density and cumulative density plots of the Average Temperature of Campo Grande based on four methods of estimation (Maximum Likelihood Estimation (MLE), Moment Matching Estimation (MME), Quantile Matching Estimation (QME) and Maximizing Goodness-of-fit Estimation (MGE)) and their parameter estimates. MLE seems to be preferred based on goodness-of-fit criteria (AIC and BIC) while MGE shows less distant in fitting 2-parameter Weibull distribution of average temperature of Campo Grande region. Eight (8) distance methods for MGE were further compared as shown in appendix D and ADL was selected for comparison with MLE. Finally, Appendix E suggests that, MLE is the preferred estimation method. Contrary to the view that empirical moments are influenced by large observed values (Delignette-Muller et al., 2020), trial with different sample sizes as low as 30 observation shows that MLE fits the data better.

Furthermore, simulated values based on shape and scale parameters obtained from Weibull distribution using three different samples of sizes ranging from small (10), medium (30) and large (50) shows negatively skewed Weibull distribution. Although Weibull distribution is said to be suited to accommodate positively skewed distributions; this study reveals that Weibull distribution is preferred even when the coefficient of skewness is negative (-1.19). The values of the coefficient of skewness and kurtosis for various sample size are presented in Table 2.

Table 2: Coefficient of Skewness and Kurtosis for Various Sample Sizes

	Simulation			**Observed Sample**
Statistics	n=10	n=30	n=50	n=4385
Min	18.31	15.45	13.45	7.18
Max	28.47	27.83	31.25	33.80
Mean	23.83	23.07	23.24	23.73
Standard Dev.	1.73	3.18	3.36	3.44
Skewness	-0.06	-0.056	-0.93	-1.19
Kurtosis	3.15	2.33	6.72	5.13

From Table 2, the simulated values are seen to be within the range of the values of the average temperature of Campo Grande with consistent mean across the four (4) samples. The standard deviations suggest that data generated with smaller sample size is less dispersed while the average temperature of Campo Grande is plagued with outliers compared to the simulated values as suggested by the coefficient of skewness. While the coefficients of skewness are negative (i.e longer left tail) and decreases with increase in sample observations, the kurtosis is closer to that of a symmetric distribution when the sample size is small. The kurtosis for the average temperature of Campo Grande is leptokurtic, indicating peaked curves compared to normal PDF. The same is applied to all the simulated samples except when the sample size is 30.

The scenario under study means the mode and median are higher than mean in both the observed and simulated samples. This is to say that the average air temperature for the period of this study is not uniformly distributed, rather slightly above average in most cases with few extremely low values.

4. Conclusion

In this paper, the air temperature series for the State of Mato Grosso Do Sul, Brazil was modeled using nine different probability distribution functions, out of the nine distributions, the two-parameter Weibull (W2P) probability distribution performed better. Hence, we recommend that the W2P probability distribution can be used for modeling air temperature series for the State of Mato Grosso Do Sul, Brazil.

References

Abreu, M. C., de Souza, A., Lyra, G. B., Pobocikova, I., and Cecilio, R. A. (2021). Analysis of monthly and annual rainfall variability using linear models in the state of mato grosso do sul, midwest of Brazil. International Journal of Climatology, 41:E2445–E2461.

Anderson, T.W. (2011). Anderson-darling tests of goodness-of-fit. International encyclopedia of statistical science, 1, 52-54.

Araujo, E. M., Silva, I. N., Oliveira, J. B. d., Cavalcante Junior, E. G., and Almeida, B. M. d. (2010). Aplicac¸ ˜ao de seis distribuic¸ ˜oes de probabilidade a s´eries de temperatura m´axima em iguatu-ce. Revista Ciˆencia Agronˆomica, 41, 36–45.

Assad, E. D., Macedo, M. A. d., Zullo J ´ unior, J., Pinto, H. S., and Brunini, O. (2003). Evaluation of geostatistics methods for interpolation of agrometeorological indices used to define climatic risks. Pesquisa Agropecu´aria Brasileira, 38, 161-171.

Assis, J. P. d., Neto, D. D., Manfron, P. A., Martin, T. N., Sparovek, G., and Timm, L. (2004). Ajuste de s´eries hist ´oricas de temperatura e radiac¸ ˜ao solar global di´aria ‘as func¸ ˜oes densidade de probabilidade normal e log-normal. Piracicaba, SP. Revista Brasileira de Agrometeorologia, Santa Maria, 12(1), 113-121.

Assis, J. P. d., Sousa, R. P. d., Batista, B. D. d. O., Linhares, P. C. F., Cardoso, E. d. A., Paula, J. d. A., and Neves, A. M. (2018). Probabilistic modeling of monthly temperature historical series in mossoro, northeastern brazil. Journal of Agricultural Science, 10(12).

Bozdogan, H. (1987). Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. Psychometrika, 52(3), 345-370.

Buriol, G. A., Estefanel, V., Chagas, A´ . C. d., and Eberhardt, D. (2007). Climae vegetac¸ ˜ao natural do estado do rio grande do sul segundo o diagrama clim´atico de walter e lieth. Ciˆencia Florestal, 17, 91–100.

Catalunha, M. J., Sediyama, G. C., Leal, B. G., Soares, C. P. B., and Ribeiro, A. (2002). Aplicac¸ ˜ao de cinco func¸ ˜oes densidade de probabilidade a series de precipitac¸ ˜ao pluvial no estado de minas gerais. Revista Brasileira de Agrometeorologia, 10(1), 153-162.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). An introduction to statistical modeling of extreme values, volume 208. Springer.

de Oliveira, G. V., Ferrari, F., and Emmendorfer, L. (2015). Investigac¸ ˜ao da aplicac¸ ˜ao de algoritmos de agrupamento para o problema astrofisico de classificac¸ ˜ao de galaxias. Revista Brasileira de Computa¸c˜ao Aplicada, 7(2), 52–61.

de Oliveira-Junior, J. F., Teodoro, P. E., da Silva Junior, C. A., Baio, F. H. R., Gava, R., Capristo-Silva, G. F., de Gois, G., Correia Filho, W. L. F., Lima, M., de Barros Santiago, D., et al. (2020). Fire foci related to rainfall and biomes of the state of mato grosso do sul, brazil. Agricultural and Forest Meteorology, 282, 107861.

de Souza, A., Abreu, M. C., Oliveira-Junior, J. F., Santos, C. M., Pobocikova, I., Fernandes,W., Torsen, E., Silva, E. B., and Mbaga, Y. V. (2020). Study of aerosol optical depth climatology using modis remote sensing data. Eur Chem Bull, 9(9), 291–299.

de Souza, A., de Oliveira-J ´ unior, J. F., Abreu, M. C., and Cavazzana, G. H. (2022). Spatial–temporal variability of the ozone column over the brazilian midwest from satellite data from 2005 to 2020. Water, Air, & Soil Pollution, 233(2), 59.

de Souza, A., Ihaddadene, R., Ihaddadene, N., and E. Oguntunde, P. (2019). Clarity index analysis and modeling using probability distribution functions in campo grande-ms, brazil. Journal of Solar Energy Engineering, 141(6), 061001.

Delignette-Muller, M., Dutang, C., Pouillot, R., Denis, J., and Siberchicot, A. (2020). Help to fit of a parametric distribution to non-censored or censored data. URL https://cran. r-project. org/web/packages/fitdistrplus/fitdistrplus. pdf.

Delignette-Muller, M. L. and Dutang, C. (2015). fitdistrplus: An r package for fitting distributions. Journal of statistical software, 64, 1–34.

Dos Reis, C. J., Souza, A., Graf, R., Kossowski, T. M., Abreu, M. C., de Oliveira-Junior, J. F., and Fernandes,W. A. (2022). Modeling of the air temperature using the extreme value theory for selected biomes in mato grosso do sul (brazil). Stochastic Environmental Research and Risk Assessment, 36(10), 3499–3516.

Ferreira-Filho, D. F. and Pessoa, F. C. L. (2022). Identification of homogeneous regions based on rainfall in the amazon river basin. International Journal of Climatology, 42(12), 6092-6108.

Field, C. B. (2012). Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change. Cambridge University Press.

Gois, P. M., Candeias, N. R., and Afonso, C. A. (2005). Preparation of enantioselective enriched -(dialkoxyphosphoryl) lactams via intramolecular chinsertion with chiral dirhodium (ii) catalysts. Journal of Molecular Catalysis A: Chemical, 227(1-2), 17-24.

Guure, C. B., Ibrahim, N. A., et al. (2012). Bayesian analysis of the survival function and failure rate of weibull distribution with censored data. Mathematical problems in engineering, 2012.

Hair, J. F. (2009). Multivariate data analysis.

Ihaddadene, R., Ihaddadene, N., Bemba, M. E. H. O. A., De Souza, A., et al.(2019). Daily global solar radiation estimation based on air temperature: case of study south of Algeria. In E3S Web of Conferences, 80, 01002.EDP Sciences.

Junqueira Junior, J. A., Gomes, N. M., Mello, C. R. d., and Silva, A. M. d. (2007). Probable rainfall for madre de deus county, alto rio grande region: distribution probability models and values characteristics. Ciencia e Agrotecnologia, 31, 842-850.

Lletı, R., Ortiz, M. C., Sarabia, L. A., and S ́anchez, M. S. (2004). Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. Analytica Chimica Acta, 515(1), 87-100.

Loikith, P. C.,Waliser, D. E., Lee, H., Kim, J., Neelin, J. D., Lintner, B. R., McGinnis, S., Mattmann, C. A., and Mearns, L. O. (2015). Surface temperature probability distributions in the narccap hindcast experiment: Evaluation methodology, metrics, and results. Journal of Climate, 28(3), 978-997.

Lyra, G. B., Garcia, B. I. L., Piedade, S. M. D. S., Sediyama, G. C., and Sentelhas, P. C. (2006). Regi ̃oes homog ̂eneas e func ̧ ̃oes de distribuic ̧ ̃ao de probabilidade da precipitac ̧ ̃ao pluvial no estado de t ́achira, venezuela. Pesquisa Agropecu ́aria Brasileira, 41:205–215.

Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. T., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., et al. (2007). Global climate projections. chapter 10.

Meehl, G. A. and Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 305(5686), 994–997.

Meehl, G. A., Tebaldi, C., Walton, G., Easterling, D., and McDaniel, L. (2009). Relative increase of record high maximum temperatures compared to record low minimum temperatures in the us. Geophysical Research Letters, 36(23).

Meza, F. and Varas, E. (2000). Estimation of mean monthly solar global radiation as a function of temperature. Agricultural and Forest Meteorology, 100(2-3), 231-241.

Montgomery, D. C. and Runger, G. C. (2010). Applied statistics and probability for engineers. John wiley & sons.

Neath, A. A. and Cavanaugh, J. E. (2012). The bayesian information criterion: background, derivation, and applications. Wiley Interdisciplinary Reviews: Computational Statistics, 4(2), 199–203.

Nielsen, M. A. (2011). Parameter estimation for the two-parameter Weibull distribution. Brigham Young University.

Oliveira, N. S. d., Reis, G. d., Reis, M. d. G., Leite, H., and Costa, J. d. (2002). Estimativa de temperaturas mınima, media e maxima do territorio brasileiro situado entre 16 e 24º latitude sul e 48 e 60º longitude oeste. Engenharia na Agricultura, 10(1-4), 8-17.

Oliveira, S. S. d., Cavazzana, G. H., and de Souza, A. (2019). Estimativa da radiac ̧ ̃ao solar global em func ̧ ̃ao da temperatura do ar e isolinhas para o estado de mato grosso do sul, brasil. Revista Brasileira de Gest ̃ao Ambiental e Sustentabilidade, 6(12), 93-108.

Pettitt, A. (1976). Cram ́er-von mises statistics for testing normality with censored samples. Biometrika, 63(3), 475-481.

Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R., and Rao, C. R. (1973). Linear statistical inference and its applications, volume 2. Wiley New York.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53–65.

Santos, E. B., Lucio, P. S., and Silva, C. M. S. e. (2015). Precipitation regionalization of the brazilian amazon. Atmospheric Science Letters, 16(3), 185–192.

Sediyama, G. C. (2002). Regionalizac¸ ˜ao do cafeeiro (coffea arabica l.) e mudanc¸as da composic¸ ˜ao agr´ıcola na regi˜ao zona da mata, mg. Revista Brasileira de Agrometeorologia, Santa Maria, 10(2), 353-362.

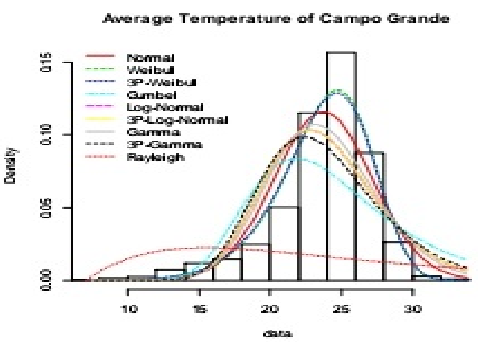
Teodoro, P. E., de Oliveira-J ´ unior, J. F., Da Cunha, E. R., Correa, C. C. G., Torres, F. E., Bacani, V. M., Gois, G., and Ribeiro, L. P. (2016). Cluster analysis applied to the spatial and temporal variability of monthly rainfall in mato grosso do sul state, brazil. Meteorology and Atmospheric Physics, 128, 197–209.

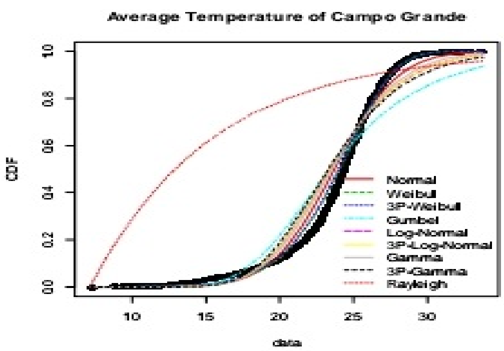
Torsen, E., Akinrefon, A., Rueben, B., and Mbaga, Y. (2015). Probability distribution of maximum temperature in adamawa state, nigeria. Journal of Mathematics (IOSR-JM), 11(4),1–6.

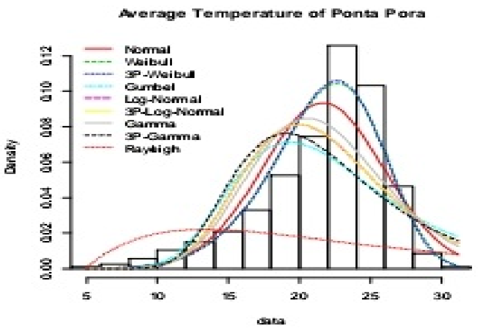
Wilks, D. S. (2011). Statistical methods in the atmospheric sciences, volume 100. Academic press.

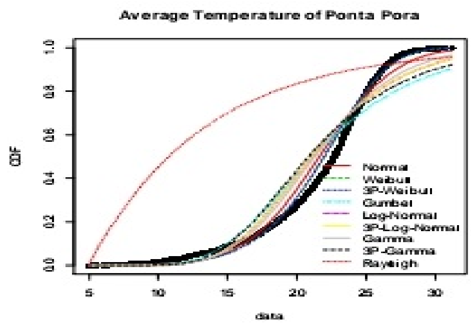
Yazici, B. and Yolacan, S. (2007). A comparison of various tests of normality. Journal of statistical computation and simulation, 77(2), 175-183.

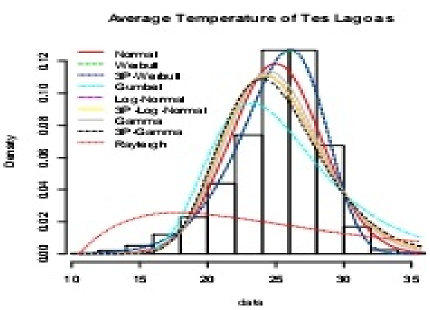
Appedix A: Density and Cumulative Density Plots for the various Regions

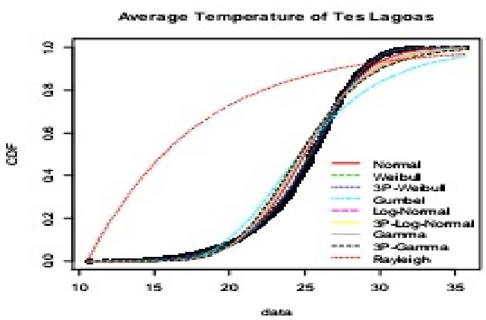


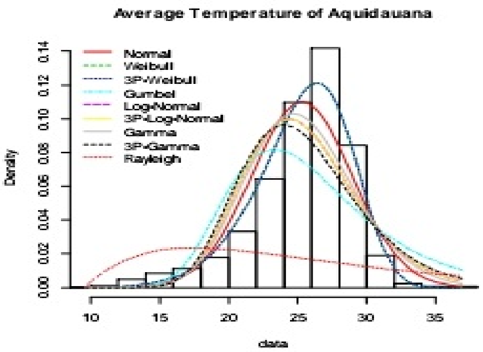


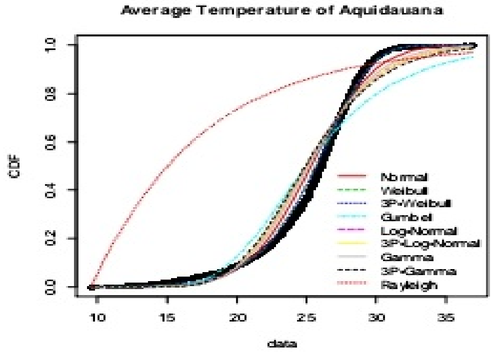


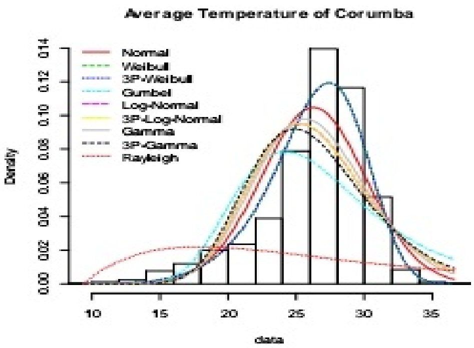


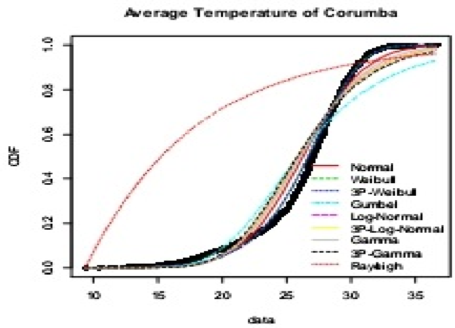


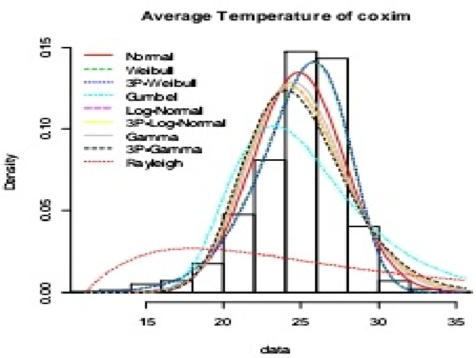


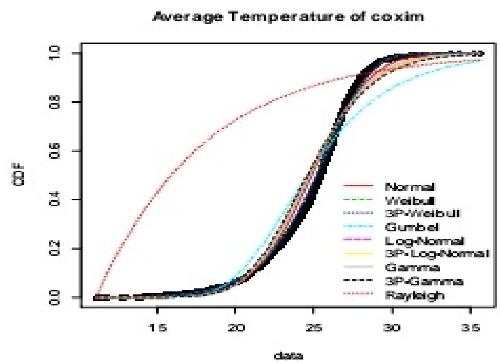

Average Temperature of Aquidauana
Normal
Weibull
3P-Weibull
Gumbel
Log-Normal
3P-Log-Normal
Gamma
3P-Gamma
Rayleigh
Density
data

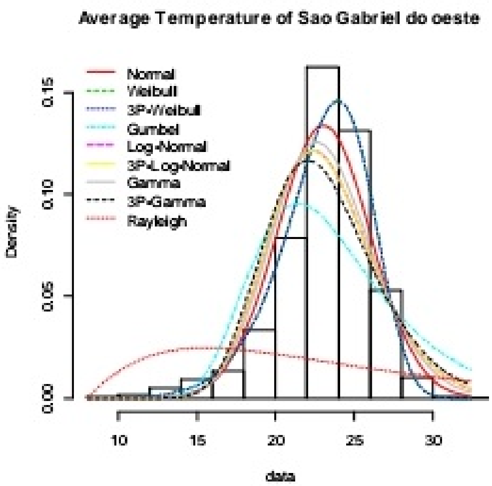

Average Temperature of Aquidauana
Normal
Weibull
3P-Weibull
Gumbel
Log-Normal
3P-Log-Normal
Gamma
3P-Gamma
Rayleigh
CDF
data

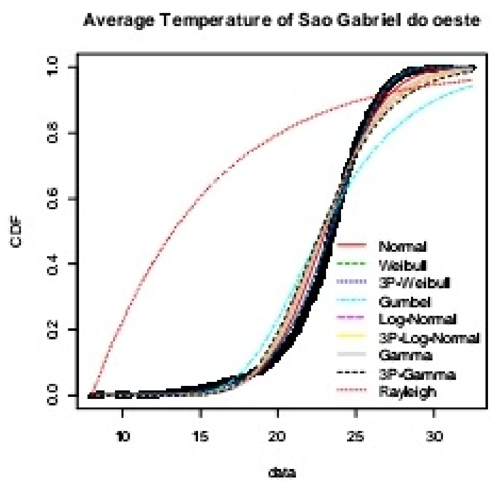

Average Temperature of Corumba
Normal
Weibull
3P-Weibull
Gumbel
Log-Normal
3P-Log-Normal
Gamma
3P-Gamma
Rayleigh
Density
data


Average Temperature of Corumba
Normal
Weibull
3P-Weibull
Gumbel
Log-Normal
3P-Log-Normal
Gamma
3P-Gamma
Rayleigh
CDF
data


Average Temperature of coxim
Normal
Weibull
3P-Weibull
Gumbel
Log-Normal
3P-Log-Normal
Gamma
3P-Gamma
Rayleigh
Density
data


Average Temperature of coxim
Normal
Weibull
3P-Weibull
Gumbel
Log-Normal
3P-Log-Normal
Gamma
3P-Gamma
Rayleigh
CDF
data

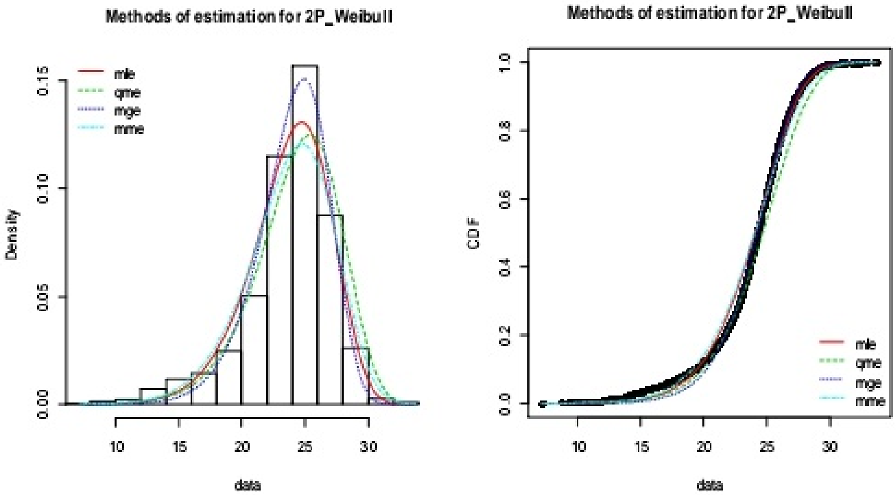

Average Temperature of Sao Gabriel do oeste
Normal
Weibull
3P-Weibull
Gumbel
Log-Normal
3P-Log-Normal
Gamma
3P-Gamma
Rayleigh
Density
data


Average Temperature of Sao Gabriel do oeste
Normal
Weibull
3P-Weibull
Gumbel
Log-Normal
3P-Log-Normal
Gamma
3P-Gamma
Rayleigh
CDF
data

Appedix B: Model Selection Criteria(The estimation method is MLE)

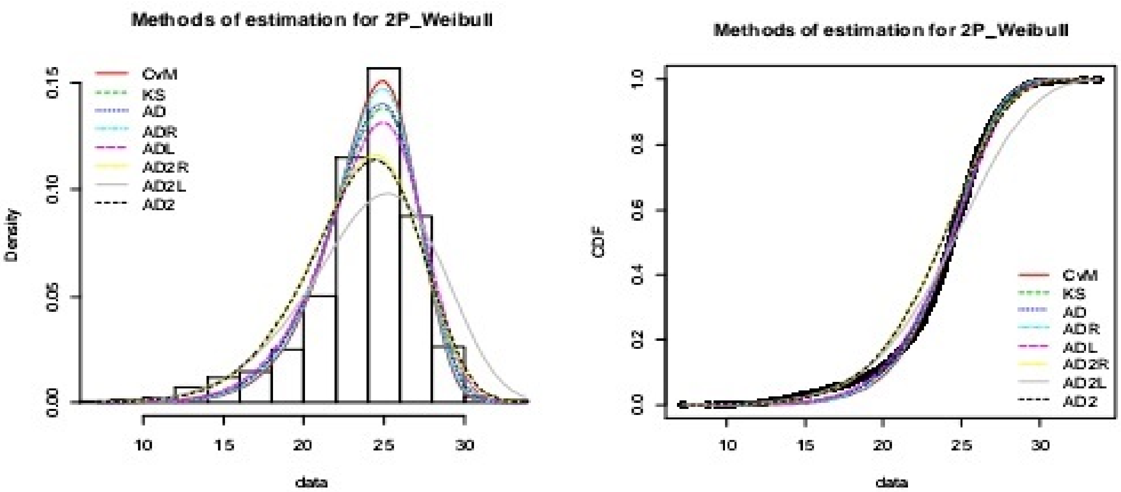
		Campo Grande	Ponta Pora	Tes Lagoas	Aquidauna	Corumba	Coxim	Sao Gabriel do oeste
Normal	KS	0.11	0.11	0.07	0.10	0.11	0.09	0.09
	CvM	15.39	14.85	6.31	12.83	16.68	9.19	9.68
	AD	91.09	86.23	37.05	76.41	96.39	51.40	57.45
	AIC	23262.82	25152.81	23072.29	23722.91	24171.74	21951.03	21996.96
	BIC	23275.59	25165.58	23085.06	23735.68	24184.51	21963.80	22009.73
2P-w	KS	0.06	0.07	0.03	0.05	0.06	0.05	0.05
	CvM	4.17	6.11	0.79	3.31	4.46	2.72	2.73
	AD	26.13	38.15	4.33	20.42	28.53	15.70	15.64
	AIC	22587.97	24607.55	22710.20	23159.64	23470.54	21628.79	21482.01
	BIC	22600.74	24620.32	22722.97	23172.41	23483.31	21641.56	21494.78
3P-W	KS	0.07	0.08	0.03	0.05	0.06	0.04	0.05
	CvM	5.37	6.90	0.65	3.02	4.05	2.48	2.82
	AD	30.54	41.30	3.84	19.64	27.36	15.63	16.04
	AIC	22590.29	24609.65	22710.73	23159.91	23470.99	21630.39	21482.07
	BIC	22603.06	24622.42	22723.50	23172.68	23483.76	21643.16	21494.84
Gumbel	KS	0.16	0.16	0.13	0.16	0.16	0.15	0.15
	CvM	51.93	41.86	29.61	44.12	47.65	36.19	45.64
	AD	297.18	243.96	179.23	258.07	269.72	215.47	268.46
	AIC	25519.06	27094.77	24756.64	25782.58	26202.78	23857.30	24263.56
	BIC	25531.83	27107.54	24769.41	25795.35	26215.55	23870.07	24276.33
Log-Normal	KS	0.15	0.14	0.10	0.13	0.15	0.11	0.12
	CvM	31.92	34.80	14.69	26.21	31.36	17.41	21.72
	AD	182.85	197.60	85.77	152.51	176.30	98.53	126.96
	AIC	24403.13	26643.37	23714.68	24690.71	25151.80	22585.24	22935.76
	BIC	24415.90	26656.14	23727.45	24703.48	25164.57	22598.01	22948.53
3P-Log-Normal	KS	0.15	0.14	0.10	0.13	0.15	0.11	0.12
	CvM	31.95	34.77	14.67	26.21	31.36	17.39	21.77
	AD	182.98	197.50	85.69	152.53	176.31	98.47	127.15
	AIC	24403.13	26643.37	23714.69	24690.71	25151.80	22585.24	22935.76
	BIC	24415.90	26656.14	23727.46	24703.48	25164.57	22598.01	22948.53
Gamma	KS	0.13	0.13	0.09	0.12	0.14	0.10	0.11
	CvM	25.75	27.18	11.50	21.23	26.02	14.36	17.06
	AD	148.74	155.24	67.21	124.24	147.29	80.92	100.16
	AIC	23964.83	26045.31	23463.04	24318.08	24779.18	22339.75	22570.73
	BIC	23977.60	26058.08	23475.81	24330.85	24791.95	22352.52	22583.50
3P-Gamma	KS	0.16	0.16	0.11	0.14	0.16	0.12	0.13
	CvM	39.86	46.84	18.40	32.21	37.43	20.91	27.88
	AD	225.05	260.66	106.81	185.33	208.14	118.40	161.32
	AIC	24954.41	27560.19	24003.84	25126.93	25570.08	22867.30	23406.34
	BIC	24967.18	27572.96	24016.61	25139.70	25582.85	22880.07	23419.11
Rayleigh	KS	0.67	0.63	0.65	0.65	0.64	0.67	0.69
	CvM	770.47	715.26	743.99	745.37	748.34	763.70	781.85
	AD	3942.46	3676.88	3776.17	3813.76	3835.81	3880.24	3980.36
	AIC	36329.02	36575.82	35190.12	35933.93	36522.83	34694.89	35427.00
	BIC	36341.79	36588.59	35202.89	35946.70	36535.60	34707.66	35439.77

Appedix C: Density and Cumulative Density Plots of four methods of estimation of 2 parameter Weibull



	Estimates shape	scale	KS	CvM	AD	AIC	BIC
MLE	8.869	25.088	0.061	4.166	26.125	22587.970	22600.740
MQE	8.710	25.766	0.108	16.097	98.151	22821.970	22834.740
MGE	10.275	25.160	0.042	1.017	24.802	22762.560	22775.330
MME	8.200	25.167	0.071	6.756	40.107	22642.100	22654.870

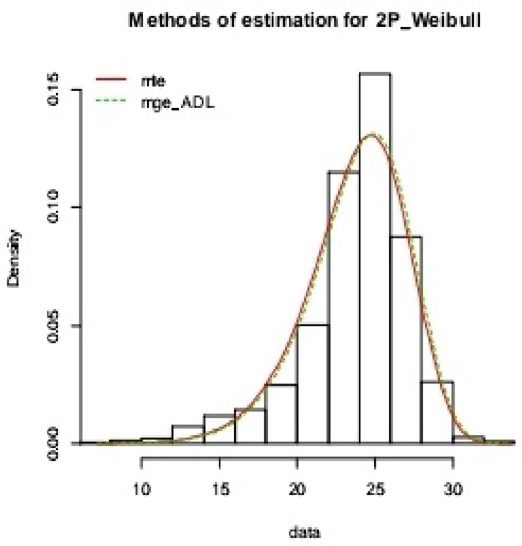
Appedix D: Density and Cumulative Density Plots of eight distances methods of MGE

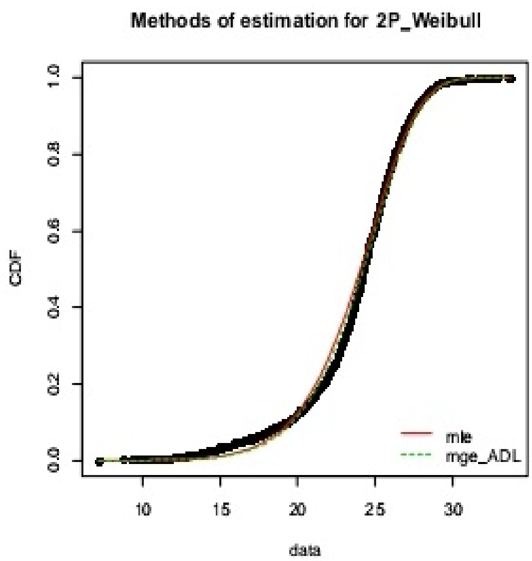


Distance Method	Parameter Estimates		AIC	BIC
CvM	10.275(1)	25.160(2)	21780.050	21792.820
KS	9.419(1)	25.223(2)	21739.040	21751.810
AD	9.556(1)	25.157(2)	21705.920	21718.690
ADR	10.039(1)	25.161(2)	21766.420	21779.190
ADL	8.973(1)	25.265(2)	21655.250	21668.020
AD2R	7.798(1)	24.864(2)	22190.730	22203.500
AD2L	6.821(1)	25.830(2)	21705.790	21718.560
AD2	7.642(1)	24.891(2)	21959.030	21971.800

(1)shape parameter, (2)scale parameter

Appedix E: Now compare MLE with MGE (ADL)





Estimation Method	Parameter Estimates		AIC	BIC
MLE	8.869(1)	25.088(2)	22587.970	22600.740
MGE (ADL)	8.973(1)	25.265(2)	22603.520	22616.290

(1)shape parameter, (2)scale parameter