27th ABCM International Congress of Mechanical Engineering
December 04-08, 2023, Florianópolis-SC, Brazil

COB-2023-1823
COMPARATIVE ANALYSIS OF FIVE PROBABILITY DENSITY FUNCTIONS APPLIED TO WIND SPEED

Darío Gerardo Fantini
Mario Benjamim Baptista de Siqueira
Reginaldo Nunes da Silva
Antonio Brasil Junior
Rafael Castilho Faria Mendes
Energy and Environment Laboratory, Department of Mechanical Engineering, University of Brasilia, Brasilia, Federal District, Brazil.
fantinidario@gmail.com, mariosiqueira@unb.br, profreginaldo.n@gmail.com, brasiljr@unb.br, rafael.cfmendes@gmail.com

Abstract. *For feasibility analysis of a wind power plant, knowing the probability density function (PDF) of wind speed of the geographical location involved is relevant for an estimate of energy generation. Thus, this work compared the PDFs Weibull of two parameters, a two-component mixture Weibull, Gamma, Log-Normal, and Log-Logistic. - The wind speed data at 50 m height, acquired from NASA Power, with the hourly average for the sites of Aparecida de Goiânia and Itumbiara in Goiás state, and Fortim in Ceará state were used for analysis. To obtain more relevant indicators, which allow inferring which of the PDFs is the best fit for the distribution of the actual wind speed data, a new procedure was implemented to determine the PDF for different analyses (hourly, monthly, and annual). Finally, it is determined that the two-component mixture Weibull is the best fit for the wind speed distribution, giving a more reliable visibility of the real frequency of wind speed distribution, but this significant visual improvement is not reflected in the calculation of the average power density available, average energy and when compared to a real sample, compared to the Weibull of two parameters PDF.*

Keywords: *Two-component mixture Weibull, Mixture Weibull, Brazil's wind resources,* Wind Power Density, Wind Energy.

1. INTRODUCTION

Understanding the probability density function (PDF) of wind velocity with the highest possible accuracy for the location under consideration in the feasibility analysis of a wind farm is of great importance to obtain a reliable estimate of the energy the farm will produce. Typically, the PDF is determined for the specific location, and based on this, the choice of turbine type is made. The wind potential of the location is then determined using the turbine's power curve or power coefficient (CP). In most cases, the two-parameter Weibull distribution is the chosen function. However, according to Dhiman et al. (2020), the Weibull distribution does not accurately estimate all wind regimes, especially cases with bimodal behavior. For this reason, other PDFs have been proposed in the literature, such as the two-component mixture Weibull (or two-component Weibull), extended generalized Lindley, and combinations of density functions. For unimodal wind distributions, in addition to the Weibull PDF, the Rayleigh, Log-normal, and Gamma PDFs are considered, as highlighted by Ali et al. (2022), who analyzed nine types of two-parameter PDFs at three different locations in Libya, Africa, showing that the PDFs that best fit the wind velocity data distribution were different at each location (Gamma, Inverse Gaussian, and Log-Normal).

These are not the only comparative studies using different PDFs. Akdağ et al. (2010) compared the two-component Weibull PDF with the original two-parameter Weibull PDF and found a better fit to the data for the mixture PDF. Kollu et al. (2012) showed that the Weibull-Generalized Extreme Value (GEV), Weibull-Lognormal, and GEV-Log-Normal mixture probability functions are better alternatives than Weibull, two-component Weibull, Gamma, and Log-normal for describing wind speed characteristics. Nevertheless, new research continues to be conducted, adding types of proposed PDFs to be compared with the original Weibull PDF. Recently, Khaled Khamees et al. (2022) compared various types of PDFs fitted to wind speed data collected from a Texas site over 5 years. The results indicate that the two-component Weibull PDF provides the best fit among the Weibull of two parameters and other two-component mixture PDFs. However, the three-component mixture Weibull distribution provides the best fit among all the PDFs but requires a more complex parameter estimation method due to the higher number of parameters. Therefore, it necessitates using an artificial intelligence optimization method for estimation. However, as Reid (2022) points out, even though some data distributions may not be optimally fitted by a Weibull distribution when compared to others, it is unlikely that a mixture of data from two distributions (particularly if they are overlapping) would be noticeably better fitted by other types of combinations than by a combination of Weibull PDFs.

In the study by Arslan et al. (2017), the authors explore PDFs that have been relatively unexplored in wind resource modelings, such as the generalized Lindley PDF and power Lindley PDF. However, Han et al. (2018) propose the use of non-parametric PDFs, highlighting that they provide a better fit to the joint wind speed and direction data.

Mechanisms that contribute to increasing the accuracy of the calculation of the energy potential of the site or to improve the accuracy in the calculation of the PDF are of great relevance (Herrera, 2016), in this sense, works presenting new methods that improve the estimation of the parameters of the different PDFs, or comparing different estimation methods are proposed to improve the accuracy of the PDF adjustments (Abbas et al., 2023; Indhumathy et al., 2019; Indhumathy. et al., 2021; Khaled Khamees et al., 2022; Sumair et al., 2022).

While the two-parameter Weibull PDF is recommended by the international standard IEC 61400-12 for estimating annual energy production, numerous studies have emphasized the increased accuracy achieved by using other PDFs. This raises the question whether the use of alternative PDFs could be applied to wind resource analysis in Brazil, and whether improvements in data fitting are truly relevant for calculating the power density and energy production of a specific turbine specified for a wind project.

It is important to note that a bimodal distribution can be considered a mixture of data from two distributions. Therefore, in this study, the Weibull PDFs with two parameters, two-component Weibull, Gamma, Log-Normal, and Log-Logistic PDFs were chosen based on literature and ease of implementation, which is not a less relevant factor. The objective is to obtain more relevant indicators that allow determining which PDF is the most suitable for modeling the distribution of real wind speed data, and to verify if there is a significance that justifies its choice over the two-parameter Weibull PDF. A new procedure is proposed and implemented using the Reliability and SciPy APIs in the Python compiler to achieve this goal.

2. MATERIAL AND METHODS

To carry out this study, wind speed and direction data at a height of 50 m were obtained from NASA Power (The POWER Project supported by NASA Earth Science - https://power.larc.nasa.gov) for the following sites: (i) Aparecida de Goiânia (lat.: -16.7961, lon.: -49.2313), (ii) Itumbiara (lat.: -18.4084, lon.: -49.1010) in the state of Goiás, and (iii) Fortim (lat.: -4.4154, lon.: -37.8167) in the state of Ceará. The data consists of hourly averages from the years 2002 to 2020, as shown in Figure 1.

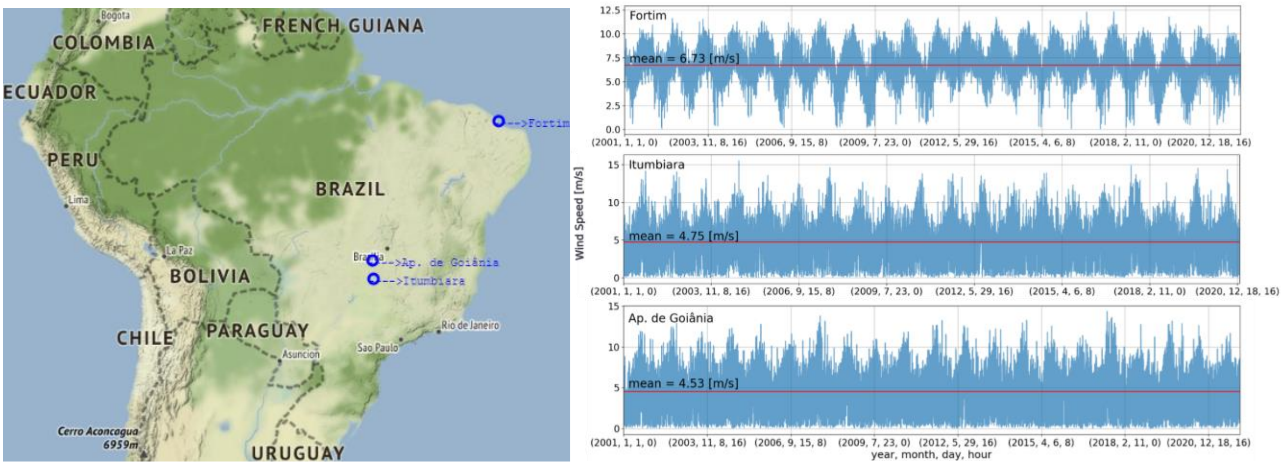

Figure 1. Temporal series of the three locations.

It is known that wind speed can be modeled as a non-stationary time series, and the additive model is commonly used for its analysis. The additive model is represented as $\boldsymbol{Y_t} = \{\dots, \boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}, \dots\} = \boldsymbol{\widehat{Y}_t} + \boldsymbol{T_t} + \boldsymbol{S_{nt}} + \boldsymbol{C_t} + \boldsymbol{\varepsilon_t}$, where $\boldsymbol{\widehat{Y}_t}$ is the mean of the series and represents the baseline level of the series. $\boldsymbol{T_t}$ is the trend component that describes the long-term behavior, which can be increasing, decreasing, or changing direction, but not frequently. $\boldsymbol{S_{nt}} = \{\boldsymbol{S_{1t}}, \boldsymbol{S_{2t}}, \boldsymbol{S_{3t}}, \dots\}$ is the seasonal component, which can have a unique or multiple character or may not exist at all. It represents different periodic variations of the series and is related to different frequencies (hourly, daily, weekly, monthly, etc.). $\mathbf{C_t}$ is the cyclic component, which can be described as a sequence of repeatable events in time. A cycle is defined by an initial and final point that belongs to consecutive local minima of the series. The final point of one cycle represents the initial point of the subsequent cycle. Unlike seasonality, cyclic variations are of low frequency and their amplitude can vary from one cycle to another. Lastly, $\boldsymbol{\varepsilon_t}$ is the stochastic component, also known as white noise, error, residual, innovations, or *shocks*.

Formally, white noise is defined as an independent and identically distributed (iid) stochastic process with mean μ (not necessarily zero, but ideally equal to zero) and variance σ^2, $\varepsilon_t \sim iid(\mu, \sigma^2)$. Although the aim of this study is not to model time series, it is important to understand the type of model that describes the behavior of the wind, as it allows for visual inference of some characteristics of the wind resource. Figure 1 shows the time series for the locations of Fortim, Itumbiara, and Aparecida de Goiânia, along with their means values. A pronounced annual seasonality is observed in Fortim, which is also present in Itumbiara and Aparecida de Goiânia, albeit with less intensity. In the case of the latter two, a higher standard deviation of data can be predicted compared to Fortim. Table 1 presents the characteristics of the locations and basic statistics related to the wind speed data, while Figure 1 indicates the geographical location of the analyzed sites.

Table 1. Site statistics.

Site	Region type	Mean [m/s]	Median [m/s]	Mode [m/s]	Variance [m/s]
Fortim	Coastal	6.73	6.75	6.19	2.47
Itumbiara	Savannah/Lakeside	4.75	4.68	4.93	4.59
Ap. de Goiânia	Savannah/Periurban	4.53	4.34	3.93	4.79

To perform the calculations, the Enercon E48/800 wind turbine model was chosen, which has the power curve Illustrated in Figure 2.

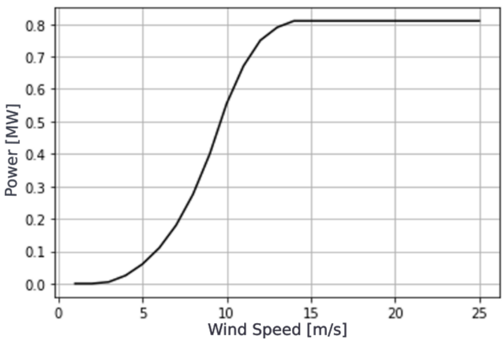

Figure 2. Turbine power curve - Enercon E48/800.

To obtain a more relevant indicator to determine which of the PDFs best represents the distribution of real wind speed data, the following procedures were implemented: 1) the complete time series (2002-2020) is grouped by month; 2) for each new grouping, different PDFs (for each month) are obtained, which are used to determine the energy generated by the chosen turbine (E_{turb}). 3) the original time series is then separated into samples (new time series) representative of each month of each year ($X_{a,m}$), which are used to determine the respective ($E_{turb\ a,m}$). Based on the values of $E_{turb\ a,m}$ and E_{turb} for each PDF, the root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are determined. The procedure described is illustrated in Figure 3.

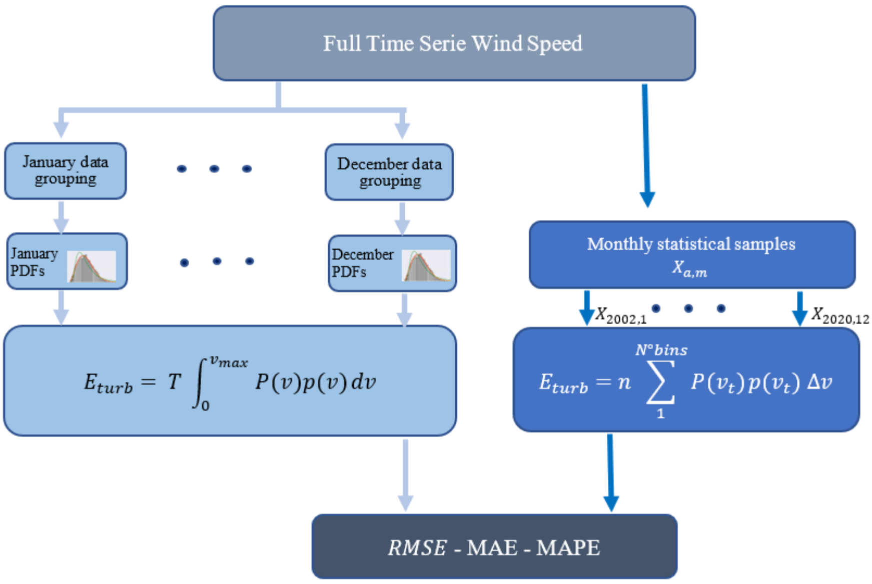

Figure 3. The new method for validation of the PDFs.

2.1 Theoretical framework

In this section, we present the equations of the probability density functions (PDFs) used to calculate the average available power density $[W/m^2]$ based on the time series of wind velocity at the selected locations.

- Two-parameter Weibull Distribution (W):

$$PDF_W = f(t) = \frac{\beta t^{\beta-1}}{\alpha^\beta} e^{-\left(\frac{t}{\alpha}\right)^\beta} = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{(\beta-1)} e^{-\left(\frac{t}{\alpha}\right)\beta}, \tag{1}$$

with the scale parameter $\alpha > 0$, shape parameter $\beta > 0$ and $t \geq 0$.

- Weibull distribution with two components (W-W) can be defined:

$$PDF_{W2C} = f(t) = \rho\, PDF_{W_1} + (1-\rho)\, PDF_{W_2}, \tag{2}$$

$$PDF_{W2C} = p \overbrace{\left[\frac{\beta_1}{\alpha_1}\left(\frac{t}{\alpha_1}\right)^{(\beta_1-1)} e^{-\left(\frac{t}{\alpha_1}\right)\beta_1}\right]}^{PDF_{W_1}} + (1-p) \overbrace{\left[\frac{\beta_2}{\alpha_2}\left(\frac{t}{\alpha_2}\right)^{(\beta_2-1)} e^{-\left(\frac{t}{\alpha_2}\right)\beta_2}\right]}^{PDF_{W_2}}, \tag{3}$$

the two-component Weibull model or Weibull mixture model consists of two Weibull distributions with scale parameters α_i and shape β_i with $i = 1{,}2$ and the proportionality parameter $0 \leq p \leq 1$.

- Gamma distribution (G):

$$PDF_G = f(t) = \frac{t^{\beta-1}}{\Gamma(\beta)\alpha^\beta} e^{-\frac{t}{\alpha}}, \tag{4}$$

where the scale parameter $\alpha > 0$, the shape parameter $\beta > 0$, $t \geq 0$ and $\Gamma(\beta) = \int_0^\infty t^{\beta-1}e^{-t}dt$ is the Gamma function.

- Log-Normal Distribution (LN):

$$PDF_{LN} = f(t) = \frac{1}{\sigma t\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2}, \tag{5}$$

where the scale parameter $-\infty < \mu < \infty$, the shape parameter $\sigma > 0$ and $t \geq 0$.

- Log-Logistic distribution (LL):

$$PDF_{LL} = f(t) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{t}{\alpha}\right)^\beta\right)^2},\text{em} \tag{6}$$

where the scale parameter $\alpha > 0$, the shape parameter $\beta > 0$ and $t \geq 0$.

The maximum likelihood estimation (MLE) method is used to fit the PDFs to the wind speed data, which works by calculating the probability of occurrence for each data point (we call this the likelihood) for a model with a given set of parameters (those we wish to estimate). These probabilities are summed for all data points. An optimizer is then used to change the parameters of the model to maximize the likelihood (sum of probabilities). Thus, MLE does not require the equation to be linearized (which is necessary for least squares estimation) for any equation to be modeled, however, least squares estimation is usually used to give an initial estimate for the parameter values.

The equation Eq. (7) was used to calculate the average available wind power density ($DPV - W/m^2$) with the time series data,

$$WPD_{data} = \frac{1}{2}\rho \frac{\sum_{t=0}^{n} v_t^3}{n}, \tag{7}$$

where n is the number of samples of the time series used, ρ the density of the air $[kg/m^2]$, and v_t is the wind speed $[m/s]$ of the sample t of the time series. To calculate the average power density based on the different PDFs, Eq. (8) was used,

$$WPD_{PDF} = \frac{1}{2}\rho \int_0^{v_{max}} p(v)v^3\, dv, \tag{8}$$

where v_{max} the maximum wind speed $[m/s\ ]$that introducing the time series and $p(v)$ is the chosen PDF.

The calculation of the energy generated by the turbine E_{turb} is based on Eq. (9) using directly the hourly values of the series and Eq. (10) for the different PDFs,

$$E_{turb} = n \sum_{1}^{N^\circ\, bins} P(v_t)\, p(v_t)\, \Delta v, \tag{9}$$

where $P(v_t)$ the power curve of the chosen turbine evaluated at v_t for a given ρ, and Δv is the width of the histogram bins.

$$E_{turb} = T \int_{0}^{v_{max}} P(v) p(v)\, dv, \tag{10}$$

where $P(v)$ the power curve function of the chosen turbine for a given ρ.

To choose the PDF that best fits the data, different indicators of the quality of the fit must be compared, the chosen ones being Log Likelihood (Loglik), Akaike Information Criterion (AICc), Bayesian Information Criterion (BIC), and Anderson Darling corrected statistic (AD). AICc and BIC are based on Loglik, and both penalize the score based on the number of parameters to be estimated in the PDF, and the closer to zero their values are the better the fit obtained. Loglik value can be between -∞ to +∞ so its value does not give any indication. However, Loglik is useful to compare several fitted models. It should be noted that models with more parameters to estimate or more complex tend to increase the value of Loglik, this means that only models that have the same number of parameters should be compared. The AD statistic is a modification of the Kolmogorov-Smirnov test and is shown to be an alternative to chi-squared and Kolmogorov-Smirnov that considers the specific PDF and emphasizes the tails, presenting the advantage of allowing a more sensitive test. The better the PDF fits the data, the lower the AD statistic. Eqs. (11) - (14) describe the different tests or indicators,

- Log Likelihood (Loglik):

$$\text{Loglik} = -2 \ln(Likelihood) = n \ln\left(\frac{2\pi R}{n}\right), \tag{11}$$

where n is the number sample of the time series, and R is the model-dependent estimate of the sum of squares error.

- Akaike Information Criterion (AICc)

$$AICc = \text{Loglik} + 2(p+1) + \frac{2(p+1)(p+2)}{n-p-2}, \tag{12}$$

where n is the number sample of the time series, and p is the number of coefficients in the PDF (the number of estimated parameters in the PDF), including the constant. $AICc$ is not calculated when $(n-p-2) \leq 0$.

- Bayesian Information Criterion (BIC)

$$BIC = \text{Loglik} + (p+1)\ln(n), \tag{13}$$

- Anderson Darling corrected statistic AD

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(x_i) + \ln\left(1 - F(x_{n-i+1})\right)\right], \tag{14}$$

where $\{x(1) < x(2) < \cdots < x(n)\}$ denotes an ordered sample size (arranged in increasing order), and F represents the cumulative distribution function (CDF) of the PDF in question.

3. RESULTS AND DISCUSSIONS

Figure 4 illustrates the histograms and PDFs fitted to the wind speed data for each month of the year for the Fortim region. The dataset for each month is constituted by the data of each month corresponding to each year of the total time series. It is observed that the PDF that best fits the data is Weibull two-component (W-W), presenting a closer fit for both the peak and both troughs of the histogram. However, there is a deficiency in achieving a good approximation of the peak in the months of March, April, and May. This is due to the increase in the standard deviation because of the increase in the right tail, indicating a greater record of low speeds. It can be concluded that none of the PDFs achieve a good fit when there is an increase in the histogram skewness with a bias to the left (negative skewness = negative asymmetry)

Table 2 presents the average values of the quality of fit indicators for the different PDFs and through these, the greater accuracy of the W-W PDF is validated for the adjustment of monthly data for Fortim.

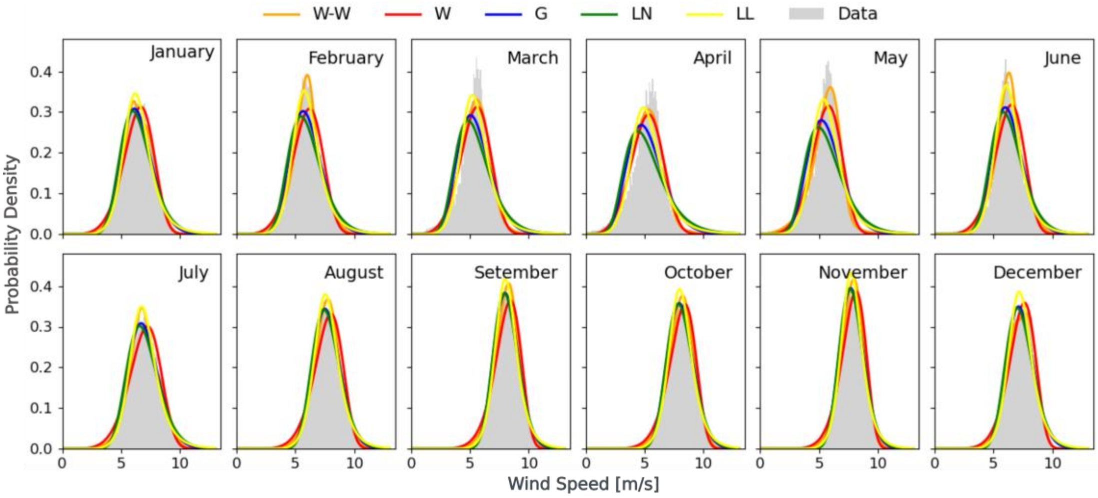

Figure 4. Monthly Fortim PDFs - Two-component Weibull (orange), Weibull (red), Gamma (blue), Log-Normal (green), and Log-Logistic(yellow).

Table 2. Average values of the Goodness of fit indicators of the monthly PDFs - Fortim.

	Goodness of fit			
PDF	***Loglik***	***AICc***	***BIC***	***AD***
W-W	*-24511.7*	*49033.45*	*49071.69*	*13.42929*
W	*-24737.9*	*49479.82*	*49495.12*	*51.83251*
G	*-25376.7*	*50757.43*	*50772.72*	*129.0855*
LN	*-26098.5*	*52201.06*	*52216.35*	*217.6773*
LL	*-25215.1*	*50434.19*	*50449.49*	*83.73141*

Table 3 shows the values of the average power density calculated based on Eqs. (7) and (8), WPD_{data} and WPD_{PDF} respectively (in this case the whole data was used) and the energy values per month for the dataset of the year 2020 ($X_{2020,m}$). ***Data*** in the Table 3 represents actual wind speed measurements.

As explained in the previous section, to obtain a more relevant indicator to verify which of the PDFs better represents the distribution of the real data, the procedure illustrated in Figure 3 was implemented, with the results obtained being the MAE, MAPE, and RMSE both for the comparison with the power density and for the energy generated by the turbine. It is again concluded that the two-component Weibull (W-W) PDF is the best fit for all samples of each month.

To exemplify how to obtain the samples $X_{a,m}$, and $E_{turb\ a,m}$ are presented in Table 3 on the right side, with the respective energy values ($E_{turb\ 2020,m}$) for the different monthly samples for the year 2020 ($X_{2020,m}$). It should be noted that these can be determined for any year of the series. Through this procedure, it is possible to verify that the two-component Weibull does not present a considerable improvement concerning the original Weibull (W) when we consider a single sample. However, the W-W better represented the power density data considering the total set of each month of the time series, which is equivalent to determining the average between the errors of each month.

Table 3. Wind Power Density and Energy per month of the year 2020 Fortim.

	Wind Power Density - kW/m^2						Energy per month of the year 2020 - MWh					
Month	***Data***	***W-W***	***W***	***G***	***LN***	***LL***	$X_{2020,m}$	***W-W***	***W***	***G***	***LN***	***LL***
January	153.14	153.06	153.06	155.84	158.95	162.97	115.80	113.88	113.88	115.95	118.26	121.25
February	122.78	122.38	122.15	126.07	131.36	132.01	81.97	82.24	82.09	84.72	88.27	88.71
March	92.38	91.39	91.45	96.62	103.17	104.97	64.65	68.00	68.04	71.89	76.76	78.10
April	83.47	81.85	81.93	89.14	99.31	99.11	66.85	58.93	58.99	64.18	71.50	71.36
May	101.80	101.31	100.42	107.63	116,32	116,78	60.83	75.37	74.71	80.08	86.54	86.89
June	141.60	141.25	141.19	144.98	149.36	150.21	62.25	101.70	101.66	104.39	107.54	108.15
July	196.00	195.97	196.62	197.87	200.56	201.73	111.16	145.80	146.28	147.22	149.22	150.09
August	260.90	260.98	262.09	261.04	261.66	265.68	178.11	194.17	195.00	194.22	194.67	197.67
September	304.22	304.52	305.22	303.73	303.77	308.36	175.10	219.25	219.76	218.68	218.71	222.02
October	304.28	304.39	305.14	303.43	303.63	309.73	182.28	226.46	227.03	225.75	225.90	230.44
November	270.80	270.93	271.46	270.63	270.95	276.13	150.19	195.07	195.45	194.85	195.09	198.81
December	222.19	222.17	222.33	223.21	224.50	230.62	156.40	165.29	165.42	166.07	167.03	171.58
MAE[kW/m²]		0.38	0.79	2.47	6.02	8.73		1.09	1.10	1.12	1.23	1.34
MAPE [%]		0.36	0.56	2.16	5.44	6.73		3.75	3.74	4.23	4.97	5.19
RMSE[kW/ m²]		0.59	0.90	3.15	8.03	9.51		6.23	6.28	6.34	6.64	7.09

The PDF results for the Itumbiara and Aparecida de Goiânia sites are presented in Figure 5 and Figure 6. The behavior of the wind speed data distribution is similar in both sites. However, Aparecida de Goiânia presents a well-defined bimodal distribution in June, July, and August. The distribution of September presents a greater complexity with three peaks. For the site of Itumbiara, also in these months, there is a displacement of the peak without defining a bimodal distribution. This behavior signals an increase in the mean and standard deviation of wind speed. At both sites, the months of April and May mark a transition.

Table 4 presents the mean power density errors for the Itumbiara and Aparecida de Goiânia sites together with the mean values of the adjustment indicators for each PDF in each month. In these two cases, it is also determined that the W-W PDF performs better than the others. Comparing the errors related to the average power density (MAE, MAPE, and RMSE) of these two sites with the errors of the Fortim site, a significant increase in the errors related to the W, G, L, N, and LL PDFs is observed. On the other hand, the W-W PDF keeps the errors almost constant, and the metrics undergo a simultaneous increase and decrease that is not very significant, which allows us to infer a better performance of the W-W PDF in cases with a bimodal distribution and in cases in which there is a marked asymmetry with a shift of the peak to the right. When the distribution has a slight asymmetry (close to a normal distribution), as is the case of the Fortim data, the Weibull PDF can be considered without significant error in the calculation of energy or average power density. Another advantage of the W-W PDF is its flexibility to convert into a W PDF as observed in the months of February for the Itumbiara and Aparecida de Goiânia sites.

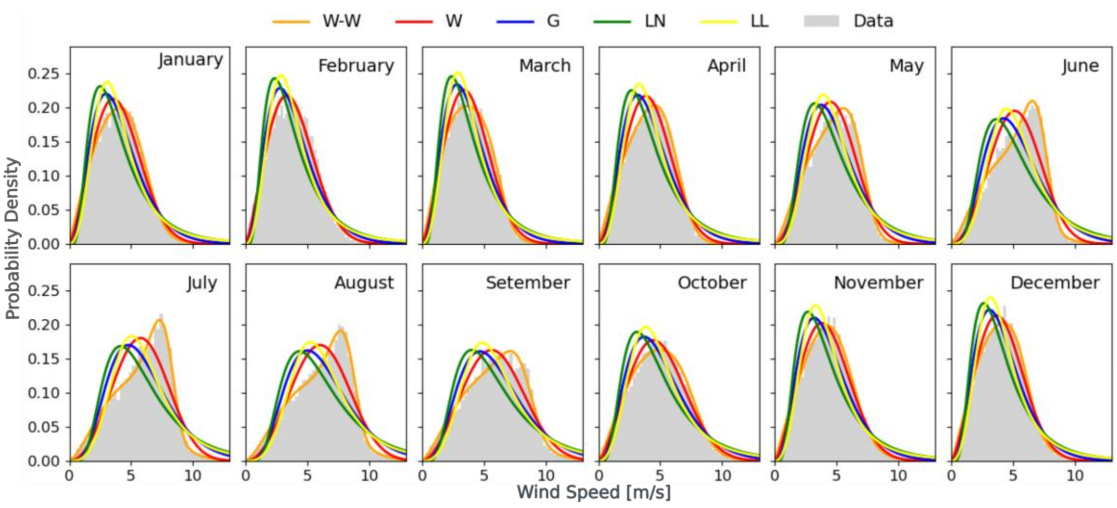

Figure 5. Monthly Itumbiara PDFs - Two-component Weibull (orange), Weibull (red), Gamma (blue), Log-Normal (green), and Log-Logistic(yellow).

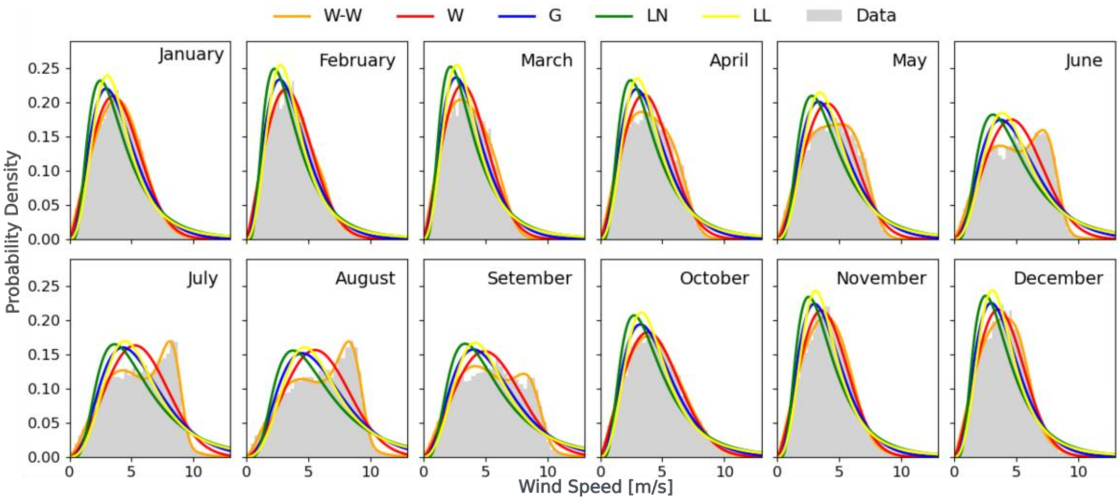

Figure 6. Monthly Aparecida de Goiânia PDFs - Two-component Weibull (orange), Weibull (red), Gamma (blue), Log-Normal (green), and Log-Logistic(yellow).

Table 4. Metrics (errors and goodness of fit) of the PDFs – Itumbiara and Aparecida e Goiânia.

		Itumbiara					***Aparecida de Goiânia***				
	Metrics	***W-W***	***W***	***G***	***LN***	***LL***	***W-W***	***W***	***G***	***LN***	***LL***
Wind Power Density kW/m²	***MAE***	0.38	1.38	5.52	13.58	16.26	0.42	1.78	4.39	12.61	12.82
	MAPE[%]	0.48	1.72	7.53	21.05	23.59	0.51	1.80	6.62	20.89	21.94
	RMSE	0.44	1.46	6.11	15.04	16.97	0.51	2.52	4.76	13.51	14.26
Goodness of fit	***Loglik***	-30469	-30774	-31673	-33167	-32357	-34911	-35339	-36134	-37041	-37699
	AICc	61552	61552	63351	66339	64719	69834	70682	72273	74086	75402
	BIC	60987	61568	63367	66354	64734	69875	70699	72290	74103	75419
	AD	2.56	51.96	193.63	389.85	232.04	1.59	61.84	168.35	233.72	363.13

The above analysis is applied to the annual data set for the three sites as shown in Figure 7 and for all data, the W-W PDF is the best fit visually. Table 5 shows the average power density and annual energy produced with the selected turbine for the three sites, with the best results achieved by the W-W PDF. The column ***Data*** represents the average annual value obtained with the real data. Table 6 shows the errors in the annual energy produced by the turbine at each of the sites. An inexpressive difference is observed between the W-W PDF and the W PDF, determining that there is an advantage in using the original Weibull PDF due to its simplicity of implementation and the quantities of parameters to be determined.

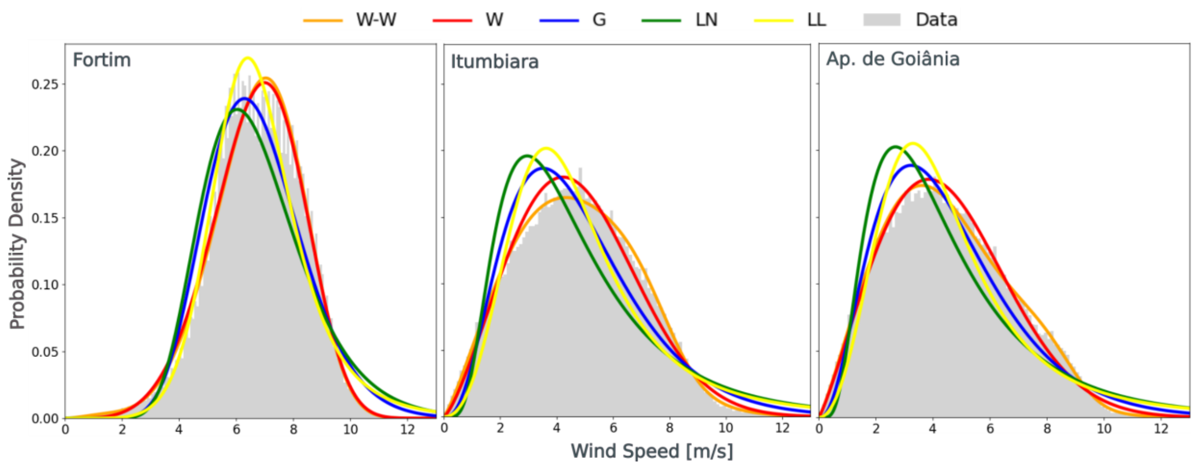

Figure 7. Yearly PDFs of Fortim, Itumbiara and Aparecida de Goiânia - Two-component Weibull (orange), Weibull (red), Gamma (blue), Log-Normal (green), and Log-Logistic(yellow).

Table 5. Energy density and annual energy per PDF.

	Wind Power Density – kW/m²						Annual Energy - MWh					
Site	***Data***	***W-W***	***W***	***G***	***LN***	***LL***	***Data***	***W-W***	***W***	***G***	***LN***	***LL***
Fortim	187.42	187.31	186.77	191.22	196.47	199.05	1649.15	1640.81	1636.13	1675.05	1721.09	1743.66
Itumbiara	90.03	90.06	89.44	95.61	102.60	103.51	800.62	791.09	785.61	839.83	901.23	909.27
Ap. de Goiânia	84.06	83.63	83.33	87.63	93.85	94.15	743.69	732.63	729.95	767.68	822.12	824.72

Table 6. Annual energy errors.

	Fortim			Itumbiara			Ap. de Goiânia		
PDF	***RMSE [MWh]***	***MAE [MWh]***	***MAPE [%]***	***RMSE [MWh]***	***MAE [MWh]***	***MAPE [%]***	***RMSE [MWh]***	***MAE [MWh]***	***MAPE [%]***
W-W	171.12	146.67	9.09	90.17	75.29	9.47	87.20	68.46	9.23
W	171.41	146.67	9.07	90.91	75.58	9.44	87.58	68.46	9.20
G	172.87	148.04	9.36	97.86	79.31	10.54	89.76	73.21	10.37
LN	185.44	157.25	10.17	134.77	111.34	15.08	118.52	98.49	14.34
LL	195.31	161.77	10.57	140.87	118.10	15.96	116.76	96.71	14.08

4. CONCLUSION

The two-component Weibull distribution provides the best fit to the wind speed distribution in different analyses (monthly and yearly), offering a more accurate representation of the frequency distribution of wind speed. This improved visibility is reflected in the calculation of the average available power density and the calculation of the average energy. When compared to a random sample from a specific month of a particular year, there may not be a substantial difference in the use of the two-parameter Weibull distribution function (PDF), but on average, it is superior.

The comparative analysis of PDFs using annual data sets showed that the Weibull PDF may be a more interesting alternative to the two-component Weibull PDF, as the latter is more complex and has more parameters, albeit with slightly superior performance.

Both Weibull PDFs, when compared to other PDFs, are significantly more suitable for describing the behavior of wind speed in cases where data exhibit a nearly symmetrical distribution in the positive or negative direction (Fortim) or a positive symmetry. However, for cases with a bimodal distribution or negative symmetry (Itumbiara and Aparecida de Goiânia), the two-component or mixture Weibull PDF is superior and provides a more realistic view of the data, allowing for the inference of wind behavior without the need for a deep analysis of the time series and its components.

Finally, greater flexibility is determined for the two-component Weibull PDF, as it can assume the behavior of a two-parameter Weibull PDF.

5. ACKNOWLEDGEMENTS

This work is part of the R&D project "Optimization of the BRAMS meteorological model, with experimental validation, to support improvements in wind power modeling systems" - PD-0394-1709/2017, regulated by the Brazilian National Electric Energy Agency (ANEEL), developed by Eletrobras FURNAS and the University of Brasília.

6. REFERENCES

Abbas, G., Gu, J., Asad, M. U., Balas, V. E., Farooq, U., & Khan, I. A. (2023). Estimation of Weibull Distribution Parameters by Analytical Methods for the Wind Speed of Jhimpir, Pakistan - A Comparative Assessment. 1–5. https://doi.org/10.1109/etecte55893.2022.10007311

Akdağ, S. A., Bagiorgas, H. S., & Mihalakakou, G. (2010). Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean. *Applied Energy, 87*(8), 2566–2573. https://doi.org/10.1016/j.apenergy.2010.02.033

Ali, H. M., Ahmed, M. Z., Khamiss, R. E., & Karim, H. A. (2022). Evaluation of Nine Two-Parameter Probability Distributions for Modeling Wind Speed Data at Three Sits in Western Mountain-Libya. *2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering, MI-STA 2022 - Proceeding*, 767–771. https://doi.org/10.1109/MI-STA54861.2022.9837731

Arslan, T., Acitas, S., & Senoglu, B. (2017). Generalized Lindley and Power Lindley distributions for modeling the wind speed data. *Energy Conversion and Management*, *152*, 300–311. https://doi.org/10.1016/j.enconman.2017.08.017

Dhiman, H. S., Deb, D., & Balas, V. E. (2020). Supervised Machine Learning in Wind Forecasting and Ramp Event Prediction. Elsevier - Academic Press.

Han, Q., Hao, Z., Hu, T., & Chu, F. (2018). Non-parametric models for joint probabilistic distributions of wind speed and direction data. *Renewable Energy*, *126*, 1032–1042. https://doi.org/10.1016/j.renene.2018.04.026

Herrera, J. C. (2016). Análisis del Recurso Eólico mediante la Transformada Wavelet con Aplicación a la Estimación de la Producción Eléctrica en Aerogeneradores. *Revista Técnica "Energía". No*, *12*, 56–63.

Indhumathy., D., Narmatha, D., Indirani, C., & Meenambika, K. (2021). Wind Resource Assessment Using Improved Mixture Weibull Distribution. *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, ICAECA 2021*. https://doi.org/10.1109/ICAECA52838.2021.9675701

Indhumathy, D., Seshaiah, C. V., & Vennila, B. (2019). Wind speed using improved mixture Weibull distribution. *International Journal of Engineering and Advanced Technology*, *8*(6), 1871–1874. https://doi.org/10.35940/ijeat.F7896.088619

Khaled Khamees, A., Abdelaziz, A. Y., Ali, Z. M., Alharthi, M. M., Ghoneim, S. S. M., Eskaros, M. R., & Attia, M. A. (2022a). Mixture probability distribution functions using novel metaheuristic method in wind speed modeling. *Ain Shams Engineering Journal*, *13*(3). https://doi.org/10.1016/j.asej.2021.10.008

Kollu, R., Rayapudi, S. R., Narasimham, S. V. L., & Pakkurthi, K. M. (2012). Mixture probability distribution functions to model wind speed distributions. *International Journal of Energy and Environmental Engineering*, *3*(1), 1–10. https://doi.org/10.1186/2251-6832-3-27

Reid, M. (2022). Reliability – a Python library for reliability engineering (Version 0.8.2) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.3938000

Sumair, M., Aized, T., Aslam Bhutta, M. M., Siddiqui, F. A., Tehreem, L., & Chaudhry, A. (2022). Method of Four Moments Mixture-A new approach for parametric estimation of Weibull Probability Distribution for wind potential estimation applications. *Renewable Energy*, *191*, 291–304. https://doi.org/10.1016/j.renene.2022.04.054

7. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.