

# Analysis of Wind Speed Characteristics Using Probability Distribution in Johor

Nor Hafizah Hussin<sup>1,2</sup>, Fadhilah Yusof<sup>1,\*</sup>

<sup>1</sup>Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81300, Johor Bahru, Johor Darul Takzim, Malaysia

<sup>2</sup>Faculty of Electrical & Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100, Durian Tunggal, Melaka, Malaysia

Received August 24, 2021; Revised October 26, 2021; Accepted November 18, 2021

## Cite This Paper in the following Citation Styles

(a): [1] Nor Hafizah Hussin, Fadhilah Yusof, "Analysis of Wind Speed Characteristics Using Probability Distribution in Johor," *Environment and Ecology Research*, Vol. 10, No. 1, pp. 95 - 106, 2022. DOI: 10.13189/eer.2022.100109.

(b): Nor Hafizah Hussin, Fadhilah Yusof (2022). *Analysis of Wind Speed Characteristics Using Probability Distribution in Johor*. *Environment and Ecology Research*, 10(1), 95 - 106. DOI: 10.13189/eer.2022.100109.

Copyright©2022 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

**Abstract** Renewable energy and energy efficiency are the key factors to ensure a safe, reliable, affordable as well as sustainable energy system for a better future. One of the most congruous, environment-friendly, and renewable energy sources is wind energy. However, it is consequential to examine the suitable probability distribution function to study the wind speed characteristics before the element can be harnessed as a source of energy. In this study, five probability distributions, Gamma, Generalized Extreme Value (GEV), Lognormal, Rayleigh and Weibull distribution were selected to model the wind speed data from four wind stations in Johor in a ten-year period. In addition, the method of maximum likelihood estimation (MLE) was applied to obtain the parameter estimation for each selected distribution function, followed by the plotting the graphical representation of probability distribution function (PDF) and cumulative distribution function (CDF) for the theoretical distributions against the provided wind speed data. To determine the best-fitted model of the probability distribution, the Kolmogorov Smirnov (KS) test and Anderson Darling (AD) test were employed to assess the goodness-of-fit for each model distribution. Based on the plotted graph and calculated goodness-of-fit results, GEV distribution was found to be the best-fitted model for the wind speed dataset in Senai, Mersing, and Batu Pahat wind station, while Gamma distribution established the optimum model for the actual wind speed dataset in Kluang station.

**Keywords** Renewable Energy, Probability

Distributions, Parameter Estimation, Wind Speed

## 1. Introduction

Non-renewable energy sources such as fossil fuels and coal are reported to be rapidly depleting owing to the booming human population's increased need for energy [1]. Not only the decline of these natural resources would limit their availability, but it would also contribute to the greenhouse effect and global warming. An alternative source of energy consisted of wind, biomass, hydropower, and solar energy can help to sustain the availability of the non-renewable energy source and reduce the effect of global warming. Among the benefits of these renewable resources are their unlimited availability, lower technology costs, and generally do not release any harmful gas that may lead to pollution. As such, wind energy is one of the most often utilised renewable energy sources that have been proven to be effective in terms of energy generation. Therefore, an accurate evaluation on the wind characteristics at the targeted location is critical in order to convert wind energy into a usable resource. To that end, renewable energy, particularly wind energy, has been receiving a growing interest in Malaysia these past few years. It is also worth noting that the potential for wind energy utilisation in Malaysia is extremely reliant on the accessibility towards the natural resource, which varies according to the location since this country is

heavily influenced by the monsoon season [2].

Due to the highly unpredictable nature and the random variation of wind patterns, it is crucial to investigate the most appropriate distribution function that could represent the wind speed pattern in a certain area [3]. In this regard, various studies have reported towards the suitability of the wind speed distributions for modelling the actual data. A study conducted by Carrillo et al. [4] used Weibull distribution for wind energy analysis while Bidaoui et al. [5] applied Weibull and Rayleigh distribution expression to analyse the wind speed data series in Morocco. A comparative analysis study by Lawan et. al. [6] utilized five distribution functions to obtain the best model for the wind speed in Miri. The findings showed that two out of five distributions, namely Lognormal and Gamma distribution, were found to provide the best fitted model of the actual data.

Zhou et. al. [7] also conducted a study to evaluate the wind speed distribution model in North Dakota using six probability density functions: Weibull, Rayleigh, Gamma, Lognormal, inverse Gaussian, and maximum entropy principle (MEP). Sarkar et. al. [8] in their study adopted Weibull and GEV distribution to analyse the wind data in India, whereas Alayat et. al. [9] applied 10 different distribution functions to assess the wind energy potential in Northern Cyprus and discovered that the Generalized Extreme Value (GEV) distribution provided the best fit model of the actual data for most of the study location. Concerning the location in Malaysia, the study by Sanusi [10] operated on Weibull and Gamma distribution to model the wind speed in Mersing, while Saberi et. al. [11] utilized Weibull distribution to evaluate wind power potential in Kuala Terengganu.

Recent studies on the assessment of the suitable PDF for the wind speed data are noteworthy in determining the theoretical distribution that closely follows the observed distribution. Natarajan et. al. [12] studied the appropriateness of nine frequently used probability distributions to determine the best-fitted distribution of wind speed prediction model for 10 stations in Tamil Nadu, India. The results demonstrated that the GEV distribution provided the best-fitted model for the majority of the wind station, followed by the Kumaraswamy distribution. A research conducted by Chen et. al. [13] assessed the probabilistic modelling for wind speed data in the Norwegian Arctic region. Two distributions, namely Nakagami and GEV distributions were concluded to be the best distribution model for the numerical weather predicted and the actual wind speed model, respectively. They attributed these findings due to the superiority and stability of these models as compared to others. Meanwhile, the study by Suwarmo et. al. [14] utilised the Weibull, Gamma, and exponential distributions to analyse the wind speed characteristics in Medan city, Indonesia, whilst Khan et. al. [15] found that Weibull distribution was the most suitable model after they tested eight

probability distributions and performed a technical evaluation of wind characteristics to determine the optimal theoretical distribution in Jhimpir, Pakistan.

Based on the works of literature, it is derived that no one distribution model can fit all cases. Thus, an investigation to test the suitability of theoretical distribution is essential to be carried out for the purpose of obtaining the best fit model [16].

Hence, the primary target for this study is to determine the best-fitted distribution model that can represent the wind speed series in four wind stations in Johor: Senai station, Kluang station, Batu Pahat station and Mersing station. Based on previous studies, the selected probability distributions to perform this evaluation were Gamma, Generalized Extreme Value (GEV), Lognormal, Rayleigh and Weibull distribution. These distributions were proven to generate the most appropriate distribution model for the wind speed series and are widely adopted in wind power applications [17]. In determining the best-fitted probability distribution function to study the wind speed characteristics in Johor, the goodness-of-fit for each model was assessed using Kolmogorov Smirnov (KS) test and Anderson Darling (AD) test, while the MLE was applied to perform the parameter estimation for each distribution selected in this study.

## 2. Materials and Methods

### 2.1. Location Description

Johor is one of the states in Malaysia situated in the south of Peninsular Malaysia. There are four meteorological stations located in Johor, specifically in Senai, Kluang, Mersing, and Batu Pahat. The geographical coordinates of those stations are listed in Table 1.

**Table 1.** Coordinate of wind stations in Johor

Location	Latitude	Longitude
Senai	1°38'20.3"N	103°39'57.0"E
Kluang	2°01'41.6"N	103°19'14.0"E
Mersing	2°26'42.6"N	103°49'52.6"E
Batu Pahat	1°52'14.5"N	102°59'25.6"E

In Malaysia, the wind blow is considerably affected by the monsoon seasons, particularly the northeast and southwest monsoon. The northeast monsoon usually takes place between November to March with heavier rainfall recorded in Peninsular Malaysia, especially in the southern and eastern part, with Johor being one of the most affected areas. In this research, the daily wind speed series for four wind stations in Johor was provided by the Malaysia Meteorological Department for a period of ten years from January 2004 to December 2014.

**2.2. Modelling Methods**

Before conducting the modelling for the probability distribution, it is worth mentioning the descriptive analysis of the data series, comprised of the value of the mean, standard deviation, and the skewness of the data series, would help to provide the pattern of the overall data. As such, the value of the mean provided the average wind speed in each wind station. The standard deviation described the variation of the data and how much they differ from the mean, while the skewness value depicted the symmetry of the data distribution. All these values were calculated using the following equation, respectively.

$$\text{Mean, } m = \frac{1}{n} \sum_{i=1}^n v_i \tag{1}$$

Standard Deviation,

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (v_i - m)^2} \tag{2}$$

$$\text{Skewness, } g = \frac{1}{n \times s^3} \sum_{i=1}^n (v_i - m)^3 \tag{3}$$

The data analysis in this study was performed using the Minitab 18 for descriptive analysis and EasyFit 5.5 as the main software for modelling the probability distribution function of the daily wind speed series.

**2.3. Probability Distribution Function**

In this research, the Gamma, Generalized Extreme Value (GEV), Lognormal, Rayleigh and Weibull distribution were used to analyse the wind speed characteristics and develop a model for the daily wind speed data series as these distributions are commonly utilized by researchers in the wind energy applications field [17-20]. The formulation for probability density function (PDF) and cumulative density function (CDF) concerning the velocity variable  $v$  for each distribution are listed in Table 2 and Table 3, respectively.

**Table 2.** Probability Density Function (PDF) formulation for each distribution

Distribution	Parameter	PDF
Gamma	$\alpha$ : shape $\beta$ : scale	$f(v; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} v^{\alpha-1} \exp(-\frac{v}{\beta})$ $\Gamma(v)$ is the Gamma function
Generalized Extreme Value (GEV)	$\xi$ : shape $\beta$ : scale $\mu$ : location	$f(v; \xi, \beta, \mu) = \frac{1}{\xi} \left[ 1 - \frac{\beta(v-\mu)}{\xi} \right]^{\frac{1}{\beta}-1} \exp \left[ - \left( 1 - 1 - \frac{\beta(v-\mu)}{\xi} \right)^{\frac{1}{\beta}} \right]$
Lognormal	$\sigma$ : shape $m$ : scale	$f(v; \sigma, m) = \frac{1}{v\sigma\sqrt{2\pi}} \exp \left[ \left( - \frac{(\ln(v) - m)^2}{2\sigma^2} \right) \right]$
Rayleigh	$\kappa$ : scale Shape =2	$f(v; \kappa) = \frac{2v}{\kappa^2} \exp \left( - \left( \frac{v}{\kappa} \right)^2 \right)$
Weibull	$\sigma$ : shape $\lambda$ : scale	$f(v; \sigma, \lambda) = \frac{\sigma}{\lambda} \left( \frac{v}{\lambda} \right)^{\sigma-1} \exp \left[ \left( - \frac{v}{\lambda} \right)^\sigma \right]$

**Table 3.** Cumulative Distribution Function (CDF) formulation for each distribution

Distribution	Parameter	CDF
Gamma	$\alpha$ : shape $\beta$ : scale	$F(v; \alpha, \beta) = \int \frac{1}{\beta^\alpha \Gamma(\alpha)} v^{\alpha-1} \exp(-\frac{v}{\beta}) dv$
Generalized Extreme Value (GEV)	$\xi$ : shape $\beta$ : scale $\mu$ : location	$F(v; \xi, \beta, \mu) = \exp \left[ - \left( 1 - 1 - \frac{\beta(v) - \mu}{\xi} \right)^{\frac{1}{\beta}} \right]$
Lognormal	$\sigma$ : shape $m$ : scale	$F(v; \sigma, m) = \frac{1}{2} + \left[ \left( \frac{1}{\sqrt{\pi}} \int_0^v \exp(-t^2) dt \right) \left( \frac{\ln(v) - m}{\sigma\sqrt{2}} \right) \right]$
Rayleigh	$\kappa$ : scale Shape =2	$F(v; \kappa) = 1 - \exp \left[ - \left( \frac{v}{\kappa} \right)^2 \right]$
Weibull	$\sigma$ : shape $\lambda$ : scale	$F(v; \sigma, \lambda) = 1 - \exp \left[ \left( - \frac{v}{\lambda} \right)^\sigma \right]$

**Table 4.** MLE formulation for parameter estimation of each distribution

Distribution	Parameter	MLE Formula
Gamma	$\alpha$ : shape $\beta$ : scale	$\ln(\alpha) - \psi(\alpha) = \ln\left(\frac{\sum_{j=1}^n v_j}{n}\right) - \frac{\sum_{j=1}^n \ln v_j}{n}, \quad \beta = \frac{m}{\alpha}$ <p>where <math>\psi</math> is the digamma function</p>
Generalized Extreme Value (GEV)	$\xi$ : shape $\beta$ : scale $\mu$ : location	$\ln[L(\phi v)] = -n \ln(\beta) + \sum_{j=1}^n \left[ \left(\frac{1}{\xi} - 1\right) \ln(y_j) - (y_j)^{\frac{1}{\xi}} \right]$ <p>where <math>\phi = (\xi, \beta, \mu)</math> and <math>y_j = [1 - (\frac{\xi}{\beta})(v - \mu)]</math></p>
Lognormal	$\sigma$ : shape $m$ : scale	$\sigma = \frac{1}{n} \sum_{j=1}^n \ln v_j, \quad m = \sqrt{\frac{1}{n} \sum_{j=1}^n (\ln v_j - \sigma)^2}$
Rayleigh	$\kappa$ : scale Shape =2	$\kappa^2 = \sqrt{\frac{1}{2n} \sum_{j=1}^n v_j^2}$
Weibull	$\sigma$ : shape $\lambda$ : scale	$\frac{1}{\sum_{j=1}^n v_j \sigma} \sum_{j=1}^n v_j \sigma \ln v_j - \frac{1}{\sigma} - \frac{1}{n} \sum_{j=1}^n v_j = 0, \quad \lambda = \left[ \frac{1}{n} \sum_{j=1}^n v_j \sigma \right]^{\frac{1}{\sigma}}$

**2.4. Parameter Estimation Method**

In determining the best-fitted distribution model, several methods for parameter estimation were proposed to allow the data to fit the distribution curve as close as possible. In this sense, the Maximum Likelihood Estimation (MLE) is the conventional and accurate method to be used [21]. In this study, it was observed that MLE operated by minimizing the mean square error associated with the estimated model parameter. For every function  $y$  with respect to  $v, f(v)$ , MLE of  $v$  represented the value of  $v$  that maximized the likelihood of  $f(v)$  or the logarithm of the likelihood of  $f(v)$ . This process would reduce the chances of  $v$  occurring in a significant number of improbable scenarios. Table 4 shows the formulation to estimate the model parameter using the MLE method for each distribution.

**2.5. Goodness-of-fit Test**

To determine the best distribution representing the actual data series, the evaluation for the goodness-of-fit of the developed model is important to be performed. A well-fitted model is a model that has the ability to explain the detailed information of the data. This includes the coefficient of the model, which would be able to be estimated with little uncertainty to describe the data variability, as well as predicting new observations with a higher degree of certainty.

This study applied two types of goodness-of-fit tests, namely the Kolmogorov-Smirnov (KS) test and the Anderson-Darling (AD) test. The AD test accommodates more weights towards the deviations on the tails of the distribution, while the KS test is more subtle towards the centre of the distribution curve [22]. Therefore, the

evaluation of goodness-of-fit using both tests would allow in obtaining the optimum theoretical distribution as they covered the centre and tail area of the distribution function [23]. A brief description of the methodology of the KS and AD tests is explained below.

(1) Kolmogorov-Smirnov (KS) test

The Kolmogorov Smirnov (KS) test was performed by applying the maximum value of the largest absolute deviation among the theoretical cumulative distribution and the empirical cumulative frequency distribution [24]. The KS test statistics is formulated as follows:

$$Q = \max|F(v) - O(v)| \tag{4}$$

where  $F(v)$ , which should be a continuous distribution, was the theoretical cumulative distribution for the tested distribution function, while  $O(v)$  represented the empirical cumulative frequency distribution calculated at  $v$ . Accordingly, a smaller value of KS test statistic would generate the best-fitted model of the theoretical distribution.

(2) Anderson-Darling (AD) test

The Anderson Darling (AD) test is a modification from the KS test and is one of the frequently used methods for the goodness-of-fit test. The AD test statistic is computed based on the following equation:

$$A^2 = -n - \frac{\sum_{j=1}^n (2j-1)}{n} [\ln F(v_j) + \ln(1 - F(v_{n-j+1}))] \tag{5}$$

where  $F(v_j)$  was the cumulative distribution function for the tested probability density function (PDF). Similar to the KS test, a smaller value of the AD test would indicate a better model of the theoretical distribution.

### 3. Results and Discussions

#### 3.1. Descriptive Statistics

The statistical description for all four wind stations selected in this study is listed in Table 5. Mersing station generated a higher average wind speed of 9.4238 m/s, given by (1). This could be due to the fact that the Mersing wind station is situated in the coastal area, hence, the difference between the land and sea breezes would cause higher wind blow intensity [9]. Meanwhile, Batu Pahat station recorded the lowest average wind speed of 6.8619 m/s. Additionally, the standard deviation, obtained from (2), for the Batu Pahat station also indicated a smaller variation of wind speed data from its average compared to other stations. It was also noticed that the distribution for all four wind stations was asymmetric based on the value of skewness, given by (3).

**Table 5.** Wind speed series description

Location	Mean (m)	Std. Deviation (s)	Skewness (g)
Senai	8.4262	2.0379	0.9700
Kluang	7.6291	2.5121	0.6641
Mersing	9.4238	2.4602	0.8958
Batu Pahat	6.8619	1.9314	1.1926

#### 3.2. Parameter Estimation

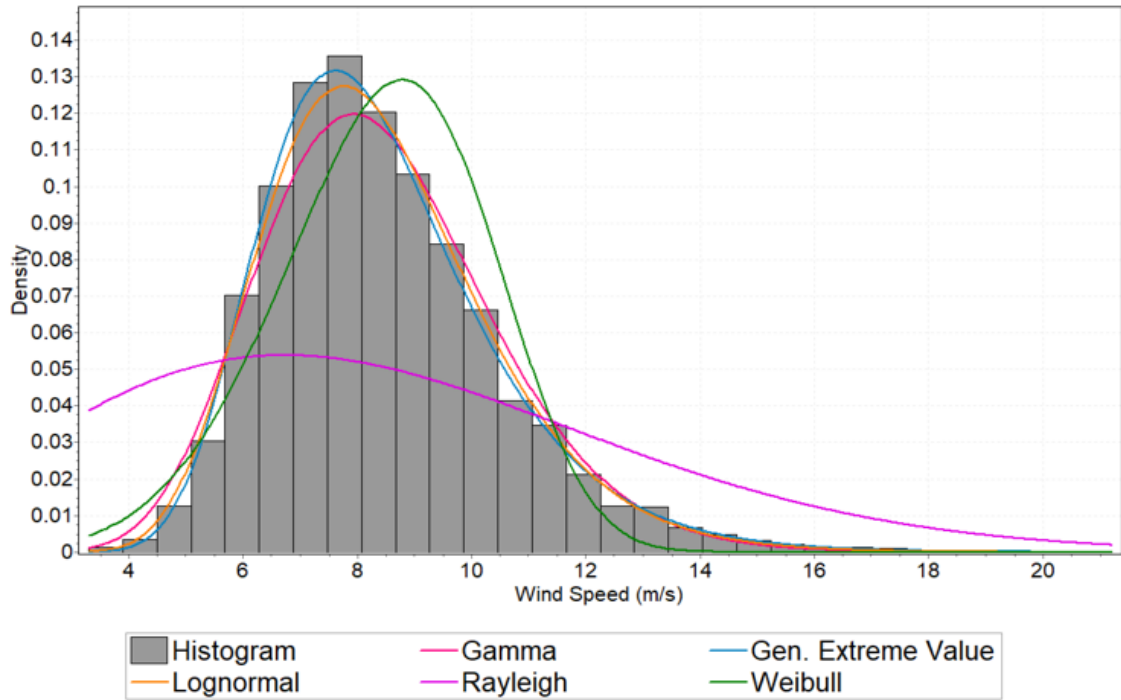
Five probability distributions were selected to perform the analysis of the wind speed characteristics in

determining the best-fitted model for the daily wind speed series obtained from four wind stations in Johor. Table 6 presented the estimated coefficient for each probability distribution parameter that was obtained based on the MLE method.

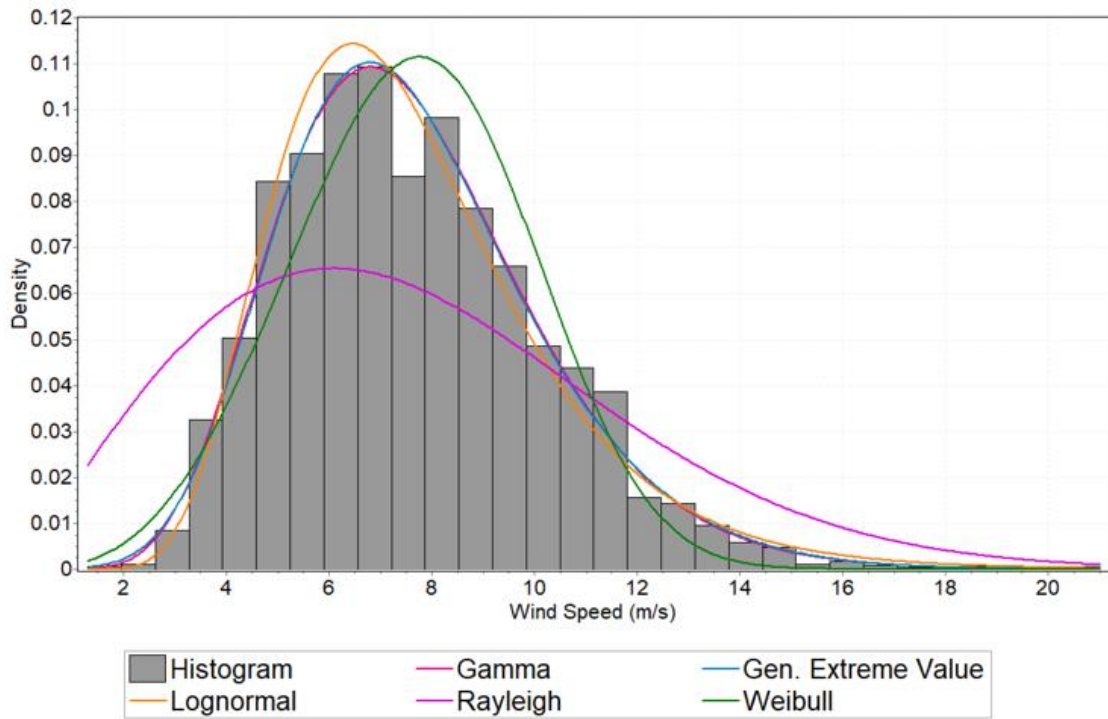
Figure 1 illustrates the graphical representation of PDF for theoretical distributions overlayed on the actual wind speed data in four wind stations. The x-axis in the PDF plot represented the wind speed (m/s) while the y-axis corresponded with the probability density. The PDF plot for all distributions is shown in different colour lines over the histogram of the observed distribution. Based on the PDF plot, GEV, Gamma, and Lognormal were visible to adhere to the pattern of the actual wind speed data for every station. For Senai station, GEV distribution was nearly in line with the peak density of the observed distribution, followed by Lognormal and Gamma distribution. For the Kluang station, both Gamma and GEV showed a close affinity with the actual distribution, marked by the Gamma distribution was virtually in line with the peak density of the observed distribution. For the Mersing station, GEV and Lognormal distribution roughly followed the peak density of the actual distribution. In the PDF plot for Batu Pahat station, it was noted that GEV and Lognormal distributions were able to represent the actual distribution very well although these distributions were not approximately in line with the peak density of actual wind speed distribution. On the other hand, the PDF of Weibull distribution appeared to be slightly overpredicting the data for all wind stations, whereas Rayleigh distribution seemed to underpredict the data as the peak density was at a distance from the actual distribution.

**Table 6.** The estimated parameters for each probability distribution

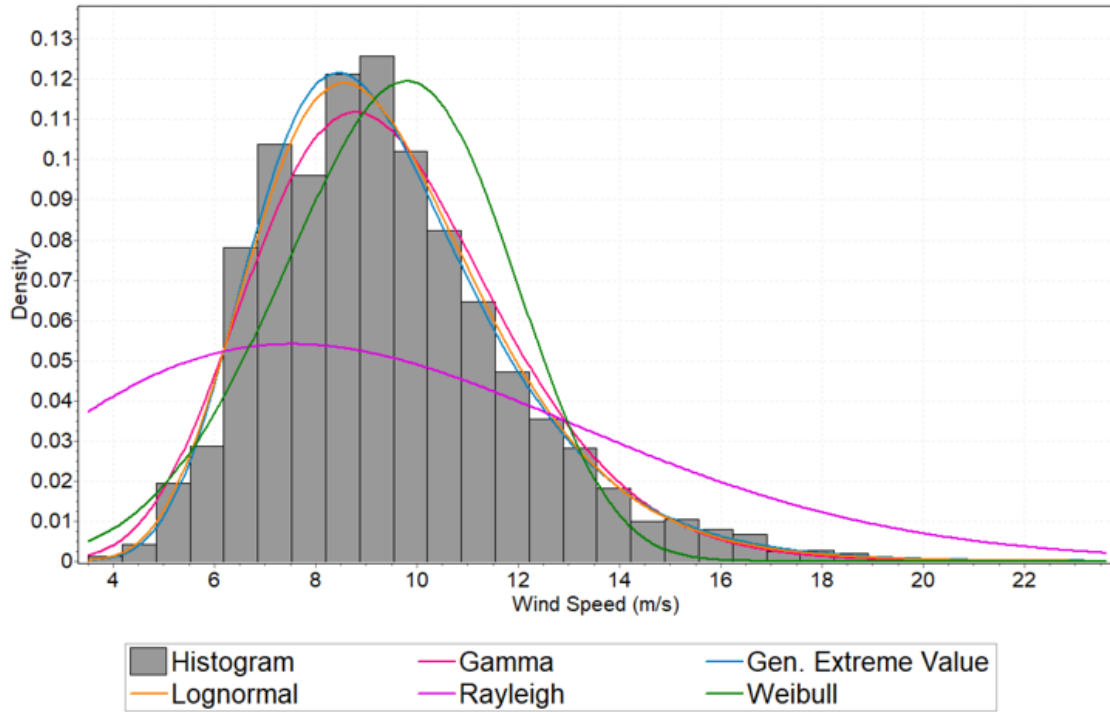
Distribution	Parameters			
	Senai	Kluang	Mersing	Batu Pahat
Gamma	$\alpha=17.096$ $\beta=0.4929$	$\alpha=9.2228$ $\beta=0.8272$	$\alpha=14.674$ $\beta=0.6422$	$\alpha=12.623$ $\beta=0.5431$
Generalized Extreme Value (GEV)	$\xi=-0.0459$ $\beta=1.6685$ $\mu=7.5359$	$\xi=-0.0991$ $\beta=2.2012$ $\mu=6.5565$	$\xi=-0.0472$ $\beta=2.0289$ $\mu=8.3435$	$\xi=-0.0536$ $\beta=1.5511$ $\mu=6.045$
Lognormal	$\sigma=0.23425$ $m=2.1036$	$\sigma=0.3355$ $m=1.9773$	$\sigma=0.25399$ $m=2.2108$	$\sigma=0.27724$ $m=1.8881$
Rayleigh	$\kappa=6.7232$	$\kappa=6.0872$	$\kappa=7.5191$	$\kappa=5.475$
Weibull	$\sigma=5.2768$ $\lambda=9.1394$	$\sigma=3.7384$ $\lambda=8.4245$	$\sigma=4.869$ $\lambda=10.266$	$\sigma=4.5197$ $\lambda=7.5026$



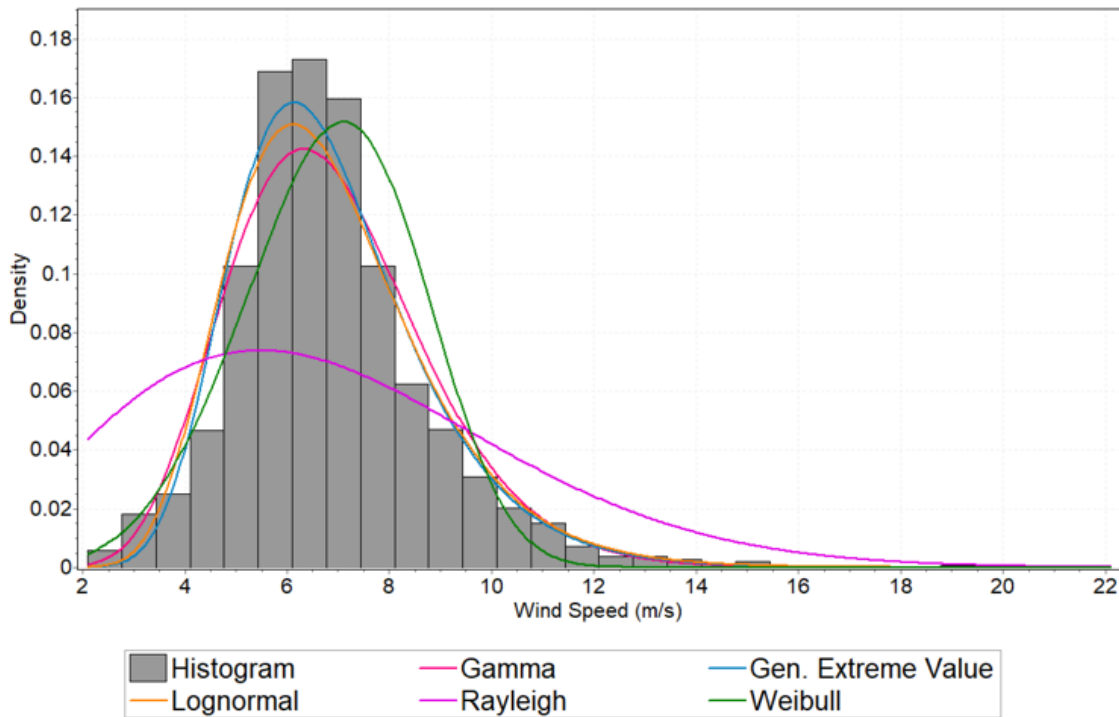
(a)



(b)



(c)

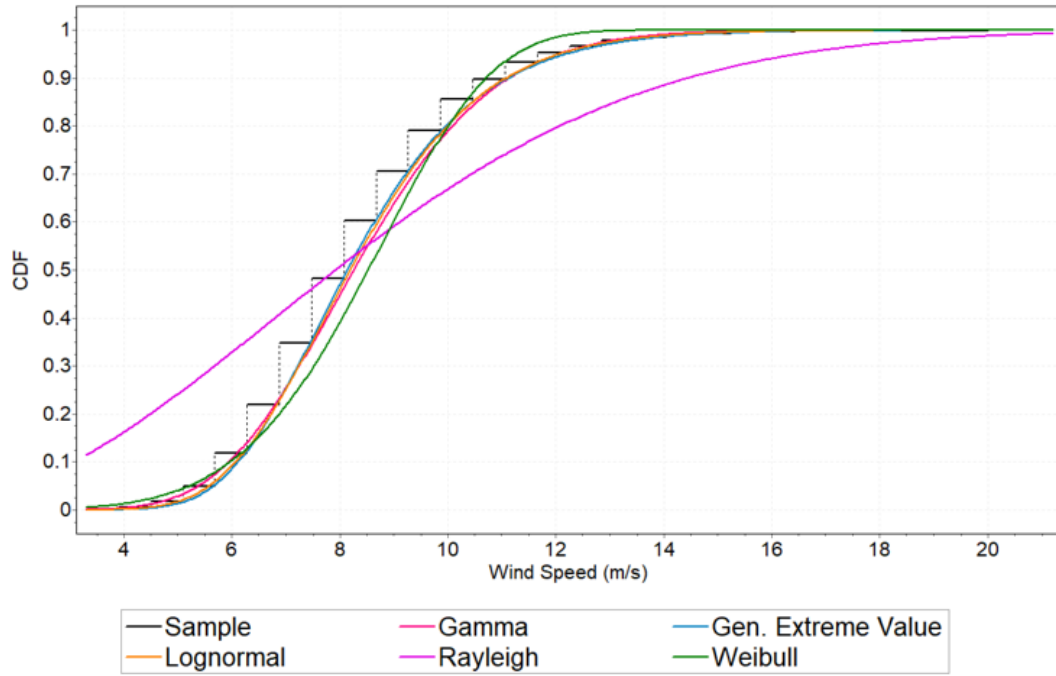


(d)

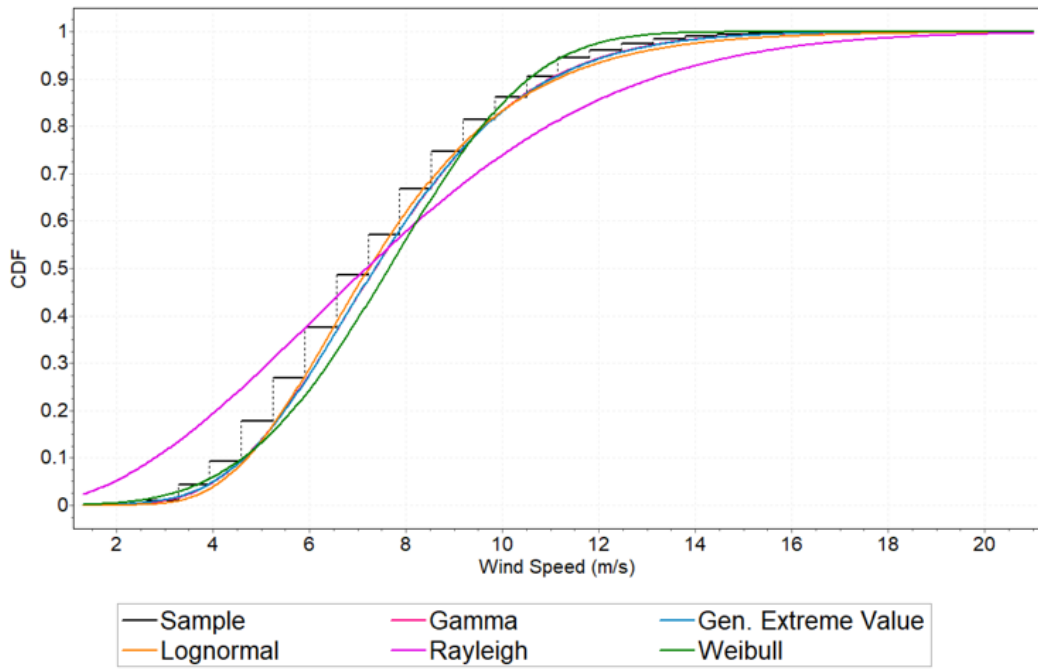
**Figure 1.** The probability density function (PDF) plot of wind speed series in (a) Senai, (b) Kluang, (c) Mersing, and (d) Batu Pahat

To support the findings, the CDF plots for each distribution function were compared with the actual wind speed data as illustrated in Figure 2. The CDF plot showed that Gamma, GEV, and Lognormal distribution were practically following the actual wind speed series in

all stations. Therefore, based on the PDF and CDF plots, it was concluded that these three distributions; Gamma, GEV, and Lognormal, were proven to provide a good data representation as compared to Weibull and Rayleigh distributions.

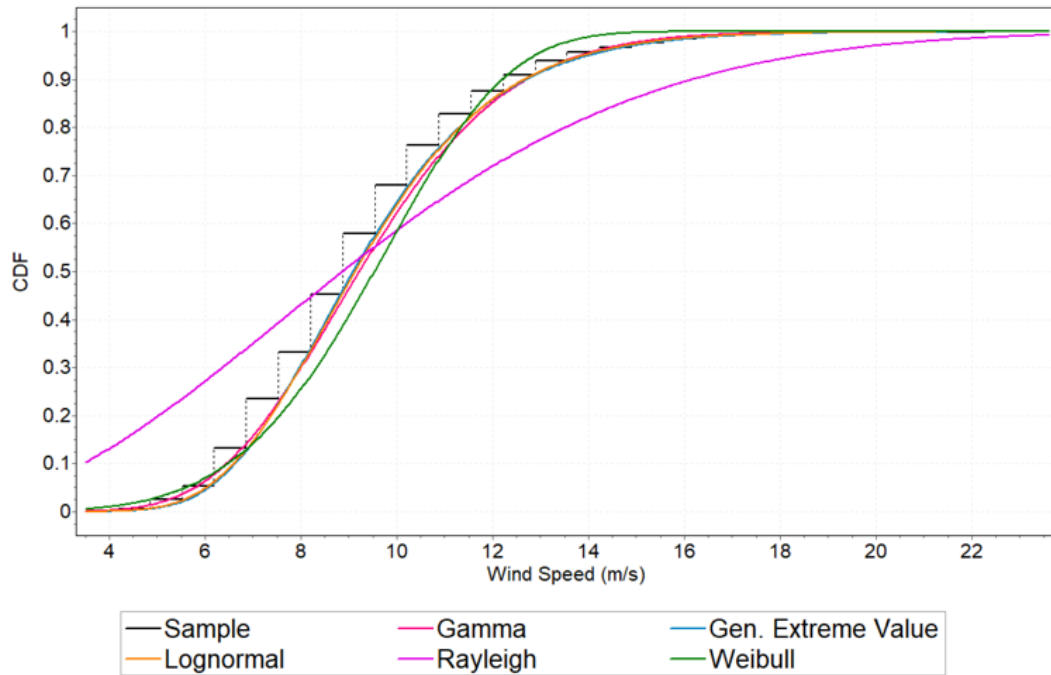


(a)

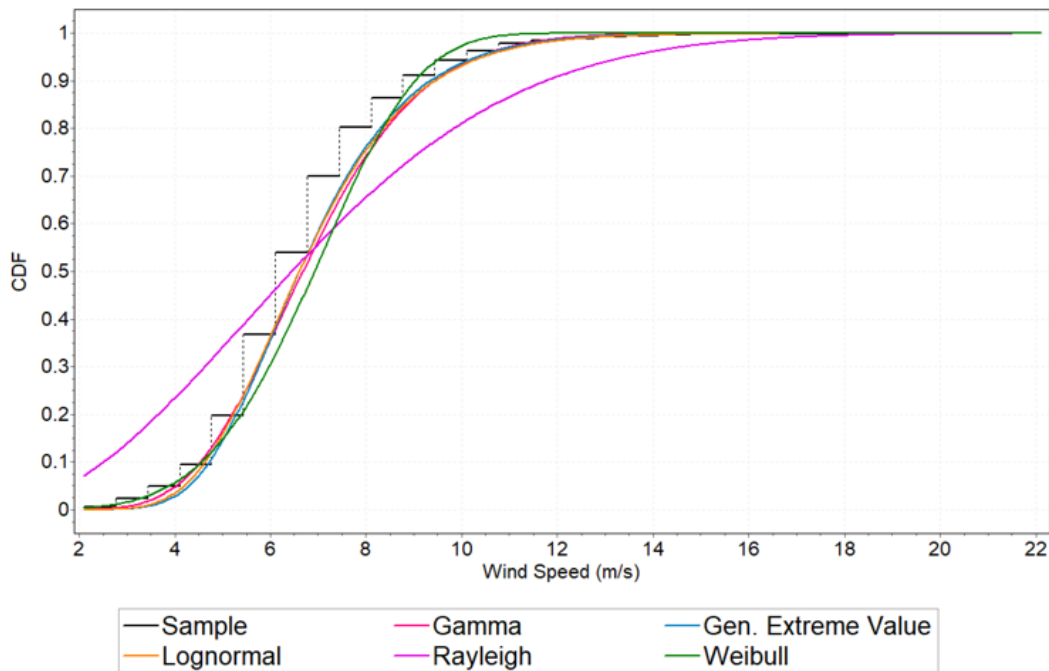


(b)





(c)



(d)

**Figure 2.** The cumulative distribution function (CDF) plot of wind speed series in (a) Senai, (b) Kluang, (c) Mersing, and (d) Batu Pahat

### 3.3. Goodness-of-Fit Test

In this study, the KS test and AD test were used to measure the goodness-of-fit for five probability distributions in determining the best-fitted distribution model to represent the actual wind speed data in Johor. The statistical values of the KS and AD tests for each distribution model are presented in Table 7 and Table 8,

respectively. These models were further ranked in an increasing order based on the values ascertained for each distribution model and station. For both tests, a smaller value denoted the best-fitted model of the theoretical distribution. In this respect, the highest rank would have the lowest statistical value, followed by an increasing order of rank associated with the higher value.

**Table 7.** Statistical values for goodness-of-fit test using KS test

Distribution	Senai		Kluang		Mersing		Batu Pahat	
	Statistical Value	Rank	Statistical Value	Rank	Statistical Value	Rank	Statistical Value	Rank
Gamma	0.0391	3	0.0216	1	0.0333	3	0.0583	3
GEV	0.0174	1	0.0224	2	0.0188	2	0.0445	1
Lognormal	0.0239	2	0.0377	3	0.0176	1	0.0561	2
Rayleigh	0.2540	5	0.1558	5	0.2365	5	0.2278	5
Weibull	0.0931	4	0.0641	4	0.0825	4	0.0995	4

**Table 8.** Statistical values for goodness-of-fit test using AD test

Distribution	Senai		Kluang		Mersing		Batu Pahat	
	Statistical Value	Rank	Statistical Value	Rank	Statistical Value	Rank	Statistical Value	Rank
Gamma	8.8286	3	1.6219	1	6.5726	3	20.364	3
GEV	0.83709	1	1.954	2	0.7330	1	11.103	1
Lognormal	1.8988	2	5.109	3	1.072	2	15.921	2
Rayleigh	432.49	5	194.37	5	374.08	5	357.29	5
Weibull	74.013	4	30.422	4	69.955	4	68.127	4

Based on the outcomes of the Kolmogorov-Smirnov (KS) test presented in Table 7, it was found that the GEV distribution provided the best-fitted model (first ranking) for two stations, namely Senai and Batu Pahat. Meanwhile, the Gamma and Lognormal distribution were observed to be the best fit representation (first ranking) for Kluang and Mersing stations, respectively. This was followed by the Lognormal distribution, ranked second for the Senai and Batu Pahat station, while GEV distribution was positioned at the second rank for Kluang and Mersing station. As for the results for the Anderson-Darling (AD) test presented in Table 8, it was observed that the GEV distribution yielded the best fit model (first ranking) for three stations: Senai, Mersing, and Batu Pahat, while the Gamma distribution produced the optimum theoretical distribution (first ranking) for Kluang station. This was followed by the Lognormal distribution which ranked second for three stations, specifically Senai, Mersing, and Batu Pahat, while GEV was in the second rank for only the Kluang station. For the Weibull and Rayleigh distributions, it was remarked that both distributions did not generate a good fit model (lower rank) for any of the stations based on the two tests.

For Senai, Kluang, and Batu Pahat stations, the distribution models with the highest ranking (ranked first) from the KS and AD statistical test also showed a best fit PDF simulation against the histogram of actual distribution, as shown in Figures 1(a), (b), and (d). Therefore, the best-fitted model for Senai and Batu Pahat stations was concluded to be the GEV distribution, followed by Lognormal distribution. For Kluang station, Gamma distribution was concluded as the best-fitted distribution model, while the GEV came in second. As for the Mersing station, the PDF illustrated in Figure 1(c)

showed that the GEV distribution was virtually following the peak density of the histogram of the actual distribution. Based on the findings in Tables 7 and 8, GEV distribution was identified to provide the best-fitted distribution model for the Mersing station, followed by Lognormal distribution. Thus, it was established, based on these results, that the GEV, Lognormal, and Gamma distributions yielded a good-fitted model and considered as the best model to study the wind speed characteristics for the selected wind stations in Johor.

## 4. Conclusions

This study was conducted to perform an analysis of the wind speed characteristics and determine the most suitable distribution model for the wind speed data set in four wind stations in Johor. Five different theoretical distributions were selected for this study by applying the MLE method to estimate the model parameter for each distribution. The graphical representation was plotted for the PDF and CDF for each of the theoretical distributions against the observed data of wind speed series in four wind stations. The GEV, Gamma, and Lognormal distributions were at a borderline with the peak density of the empirical data series for each station. With this respect, GEV and Lognormal distributions were nearly in line with the peak density of the observed distribution for Senai, Mersing, and Batu Pahat stations, while Gamma and GEV distribution closely followed the peak density of actual distribution in Kluang station. This was supported by the goodness-of-fit results by employing KS and AD test whereby the GEV was found to be in the highest ranking in investigating the wind speed criteria and was concluded

to be the best-fitted model for the wind speed data set in Senai, Mersing, and Batu Pahat station, while Lognormal distribution was ranked at second. The Gamma distribution was in the highest rank in studying the wind speed characteristics for the Kluang station and yielded the most appropriate fit for actual wind speed series, followed by the GEV distribution. Hence, for future work, these distributions can be used to obtain the quantitative measure for the wind power density in each studied location to assess the potential of wind power harvesting in Johor.

## Acknowledgments

The authors would like to thank the Ministry of Higher Education (MOHE) for financial funding under the FRGS, Registration Proposal No.:FRGS/1/2018/STG06/UTM/02 /2, Universiti Teknologi Malaysia (UTM) and Universiti Teknikal Malaysia Melaka (UTeM).

## REFERENCES

- [1] J. I. Kindangen, V. A. Kumurur, and O. H. Rogi. (2021). Integration of Solar Panels as the Shading Devices to Lower the Indoor Air Temperatures. *Civil Engineering and Architecture*, Vol. 9, No.1, 115-123, 2021
- [2] A. A. Rahman, N. A. Bakar, F. Hanaffi, and A. Khamis. Study of renewable energy potential in Malaysia. *IEEE Conference on Clean Energy and Technology (CET)*, 170–176, 2011.
- [3] V. Sohoni S. Gupta, and R. Nema. A comparative analysis of wind speed probability distributions for wind power assessment of four sites, *Turkish Journal of Electrical Engineering & Computer Sciences*, Vol. 24, No.6, 4724-4735, 2016.
- [4] M. A. Obando. An Approach to determine the Weibull parameters for wind energy analysis: the case of Galicia (Spain), *Energies*, Vol. 7, No. 4, 2676-2700, 2014.
- [5] H. Bidaoui E. A. Ikram, A. Bouardi, and D. Moumen. Wind speed data analysis using Weibull and Rayleigh distribution functions, Case study: five cities northern Morocco. *Procedia Manufacturing*, Vol. 32, 786-793, 2015.
- [6] S. M. Lawan, W. A. W. Z. Abidin, W. Y. Chai, A. Baharun, and T. Masri. Statistical modelling of long-term wind speed data, *American Journal of Computer and Information Technology*, Vol. 3, No.1, 1-6, 2015.
- [7] J. Zhou, E. Erdem, G. Li, and J. Shi, Comprehensive evaluation of wind speed distribution models: A case study for North Dakota sites, *Energy Conversion and Management*, Vol.51, No.7, 1449-1458, 2010.
- [8] A. Sarkar, S. Deep, D. Datta, A. Vijaywargiya, R. Roy, and V. S. Phanikanth. Weibull and Generalized Extreme Value distributions for wind speed data analysis of some locations in India, *KSCE Journal of Civil Engineering*. Vol.23, No.8, 3476-3492, 2019.
- [9] M. M. Alayat, Y. Kassem, and H. Çamur. Assessment of wind energy potential as a power generation source: A case study of eight selected locations in Northern Cyprus, *Energies*, Vol.11, No.10, 2697, 2018.
- [10] N. Sanusi, A. Zaharim, and S. Mat. Wind energy potential: A case study of Mersing, Malaysia. *ARPN Journal of Engineering and Applied Sciences*, Vol.11, No.12, 7712-7716, 2013.
- [11] Z. Saberi, A. Fudholi, and K. Sopian, K. Potential evaluation of wind energy in Kuala Terengganu, Malaysia through Weibull distribution method, *IOP Conference Series: Earth and Environmental Science*, Vol. 268, No. 1, 012074, 2019.
- [12] N. Natarajan, M. Vasudevan, and S. Rehman. Evaluation of suitability of wind speed probability distribution models: a case study from Tamil Nadu, India. *Environmental Science and Pollution Research*, 1-14, 2021.
- [13] H. Chen, Y. Birkelund, S. N. Anfinsen, R. Staupe-Delgado, and F. Yuan. Assessing probabilistic modelling for wind speed from numerical weather prediction model and observation in the Arctic, *Scientific Reports*, Vol.11, No.1, 1-11, 2021.
- [14] I. Y. Suwarno, M. Irwanto, and A. Hiendro. Analysis of wind speed characteristics using different distribution models in Medan City, Indonesia, *International Journal of Power Electronics and Drive System*, Vol.12, No.2, 1103-1113, 2021.
- [15] M. A. Khan, Y. Zhang, J. Wang, J. Wei, M. A. Raza, A. Ahmad, and Y. Yuan. Determination of Optimal Parametric Distribution and Technical Evaluation of Wind Resource Characteristics for Wind Power Potential at Jhimpir, Pakistan, *IEEE Access*, Vol.9, 70118-70141, 2021.
- [16] M. Sumair, T. Aized, S. A. R. Gardezi, M. M. A. Bhutta, S. Ubaid ur Rehman, and S. M. Sohail Rehman. Weibull parameters estimation using combined energy pattern and power density method for wind resource assessment, *Energy Exploration and Exploitation*, 1-18, 2020.
- [17] I. Tizgui, F. El Guezar, H. Bouzahir, and B. Benaïd. Wind speed distribution modeling for wind power estimation: Case of Agadir in Morocco, *Wind Engineering*. Vol.43, No.2, 190-200, 2019.
- [18] A. Zaharim, S. K. Najid, A. M. Razali, K. Sopian. Analyzing Malaysian wind speed data using statistical distribution. *Proceedings of the 4th IASME/WSEAS International conference on energy and environment*, Vol. 2426, 360370, 2009.
- [19] K. W. Rajapaksha and K. Perera. Wind speed analysis and energy calculation based on mixture distributions in Narakkalliya, Sri Lanka, *Journal of the National Science Foundation of Sri Lanka*. Vol.44, No.4, 409-416, 2016.
- [20] A. S. Badawi, N. F. Hasbullah, S. H. Yusoff, S. Khan, A. Hashim, A. Zyoud, M. Elamassie. Weibull probability distribution of wind speed for Gaza strip for 10 years. *Applied Mechanics and Materials*, Vol.892, 284-291, 2019.
- [21] E. S. Martins and J. R. Stedinger. Generalized Maximum - likelihood Generalized Extreme - Value quantile

- estimators for hydrologic data, *Water Resources Research*, Vol.36, No.3, 737-744, 2000.
- [22] F. W. Piegorsch. Maximum Likelihood Estimation. *Encyclopedia of Statistical Sciences*. 2006.
- [23] T. H. Soukissian and F. E. Karathanasi. On the selection of bivariate parametric models for wind data, *Applied Energy*, Vol. 188, 280–304, 2017.
- [24] C. Jung, D. Schindler, J. Laible, and A. Buchholz. Introducing a system of wind speed distributions for modeling properties of wind speed regimes around the world, *Energy Conversion Management*, Vol.144, 181–192, 2017.
- [25] J. M. Frank. The Kolmogorov-Smirnov Test for Goodness of Fit, *Journal of the American Statistical Association*, Vol.46, No.253, 1951.