JTE HCMUTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Evaluation of Probability Distribution Models for Seasonal Wind Speed Considering Elevation Variability

Thai An Nguyen*, Thi Tham Khong, Thi Kim Thao Nguyen, Huy Anh Quyen, Thi Thu Hien Huynh

Ho Chi Minh City University of Technology and Education, Vietnam

*Corresponding author. *Email:* ntan@hcmute.edu.vn

ARTICLE INFO

Received:	16/06/2025
Revised:	27/06/2025
Accepted:	07/07/2025
Published:	

KEYWORDS

Probability distribution modeling;

Weibull distribution;

Rayleigh distribution;

Lognormal distribution;

Generalized Extreme Value (GEV) distribution;

Maximum Likelihood Estimation (MLE).

ABSTRACT

This study analyzes seasonal wind speed data at Kauai, Hawaii, at altitudes of 80m, 100m, and 120m to identify the most suitable probability distributions. Four commonly used distributions - Weibull, Rayleigh, Lognormal, and Generalized Extreme Value (GEV)—were examined. Distribution parameters were estimated using the Maximum Likelihood Estimation (MLE) method, and all modeling and analysis were performed in MATLAB. Statistical criteria employed to evaluate the performance of each distribution included the Kolmogorov-Smirnov (KS) test, which assesses the goodness-of-fit between the distribution and actual data; the Chi-square test, which determines the best frequency-based fit; the Akaike Information Criterion (AIC), which identifies the distribution with the optimal balance between model fit and complexity; and the Root Mean Square Error (RMSE), which indicates the lowest prediction error. Results indicate that the GEV distribution provided the best overall fit across all altitudes, with the lowest KS (0.0474), RMSE (0.0140), and AIC (42229.09) at 100 m. The Weibull distribution also demonstrated good performance, particularly at 80 m and 100 m, offering a balance between modeling accuracy and simplicity. Conversely, the Rayleigh distribution showed moderate performance, while the Lognormal model exhibited significantly inferior results. These findings underscore the importance of selecting appropriate probability distribution models for different altitudes in wind resource assessment and support effective wind energy system planning.

Doi: https://doi.org/10.54644/jte.2025.1952

1. Introduction

The increasing global energy demand is primarily met by fossil fuels [1]. However, diesel-based power generation in remote areas faces numerous challenges including high costs, volatile fuel prices, and environmental pollution, while these regions possess significant renewable energy potential, particularly wind resources [2]. Wind represents a clean, sustainable resource that is increasingly utilized to reduce dependence on fossil fuels. Accurate assessment of wind potential based on wind speed data is crucial [3], facilitating electricity generation forecasting and optimizing energy management [4].

In wind speed modeling, probability distributions serve as valuable tools [5]. Models commonly assume wind speed follows specific distributions such as Weibull or Gamma [6]–[10]. Patidar et al. [11] employed the Weibull function to assess offshore wind potential in Gujarat state, India, using six parameter estimation methods at two altitudes. Study [12] also utilized Weibull distribution, adjusting parameters to reduce errors at NEOM city, Saudi Arabia. Beyond Weibull, distributions such as Rayleigh, Gamma, and Log-normal are also employed depending on the region and research objectives. In [13], three distributions—Weibull, Rayleigh, and Champernowne—were compared, with results indicating that Weibull was most suitable for areas with complex wind characteristics. Aries et al. [14] evaluated eight distributions in Algeria and found that GEV and Gamma were most appropriate. In

JTE HCMUTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Malaysia, Masseran et al. [15] compared nine distributions and concluded that only five types were most suitable, depending on station location.

Research [16] emphasizes that each distribution represents different wind characteristics, and model selection should be based on criteria such as R^2, AIC, and K–S tests. Wind energy assessment also depends on local factors including climate, topography, and thermal differences between land and sea [17]. This study selected Kauai Island (Hawaii, USA) due to its high wind potential.

Previous research has focused solely on wind speed accuracy at a single point in time, without considering influential characteristics such as altitude. In this study, Weibull, Rayleigh, Lognormal, and Generalized Extreme Value (GEV) distributions are employed to simulate wind speeds at altitudes of 80m, 100m, and 120m. Data are analyzed monthly and seasonally to reflect temporal variations in wind speed, thereby improving upon previous studies. The Maximum Likelihood Estimation (MLE) method is utilized to determine input parameters for the distributions, ensuring models accurately reflect actual data characteristics. The objective of this research is to identify the most suitable probability distribution for describing wind speeds at Kauai, thereby supporting more effective planning of wind power systems for the region.

2. Methodology

In this study, the Weibull, Rayleigh, Lognormal, and Generalized Extreme Value (GEV) distributions were selected as they are widely used models with high effectiveness in describing wind speeds across numerous practical applications. The Weibull distribution is widely recognized for its flexibility and ability to represent various wind profiles, making it one of the most commonly used models in wind energy assessments. The Rayleigh distribution, a special case of the Weibull with a fixed shape parameter, provides a simplified modeling approach and is useful when statistical information is limited. The Lognormal distribution is suitable for positively skewed data with long right tails, and is often used in environmental and meteorological modeling. The GEV distribution is capable of modeling extreme values through its shape parameter, making it particularly useful in risk assessment and structural design. The inclusion of these four distributions allows for a comprehensive evaluation of wind speed behavior under different statistical assumptions. Collectively, they provide a balanced framework for assessing both the central tendencies and the extreme conditions present in wind speed data, thereby enhancing the robustness and applicability of the distribution selection process.

Figure 1 illustrates the overall methodology for wind speed distribution analysis. Wind speed data is first processed using Maximum Likelihood Estimation (MLE) to estimate parameters for four candidate distributions: Weibull, Rayleigh, Lognormal, and GEV. These distributions are then evaluated using RMSE, Kolmogorov-Smirnov test, Chi-square test, and AIC to identify the best-fitting model, which is subsequently used for wind speed simulation.

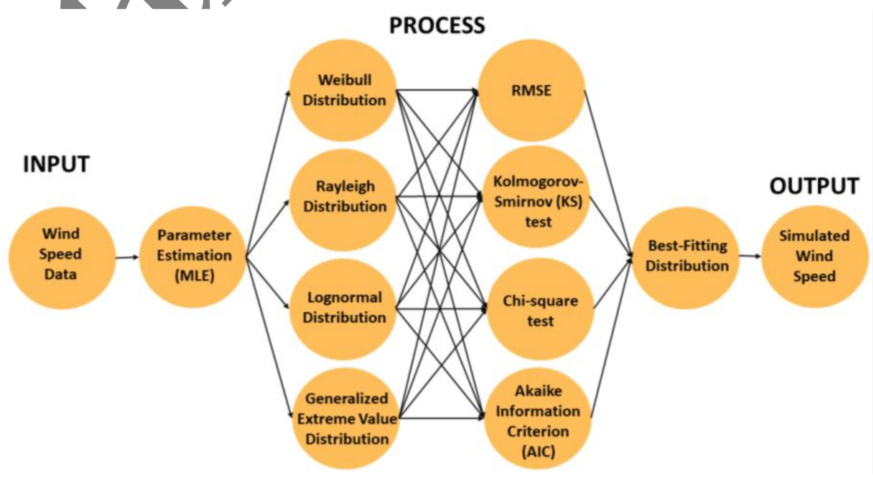

Figure 1. *Methodological framework for wind speed distribution fitting and evaluation.*

2.1. Maximum Likelihood Estimation (MLE) Method for Distribution Parameter Estimation.

To determine the parameters of a probability distribution that best fits experimental data, the Maximum Likelihood Estimation (MLE) method is employed. MLE identifies optimal parameter values by maximizing the probability of observing the given dataset [18]-[21].

The MLE procedure is implemented through the following steps: First, the probability density function (PDF) is identified. Subsequently, the likelihood function is constructed: The likelihood function is calculated based on $X = \{x_1, x_2, ..., x_n\}$ independent observed data points following a probability distribution with probability density function (PDF) $f(x;\theta)$, where θ represents the parameter set to be estimated. The likelihood function is then defined as: [21], [22]

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{1}$$

Since the logarithmic function has the property $\log(a\ .\ b\ .\ c) =$ log a +log b +log c for computational convenience, the likelihood function is expressed as the log-likelihood function: [21], [22]

$$\log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta) \tag{2}$$

The log-likelihood function is differentiated with respect to parameter θ, and the derivative equation is set to zero to find the maximum point, as the objective in MLE is to determine the parameter that maximizes the probability of the observed data. According to calculus, if a function $f(x)$ reaches an extremum at point θ', then: $f'(x) = 0$ [21], [22]

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0 \tag{3}$$

These steps ensure that parameters are accurately estimated from actual wind speed data.

2.2. Probability Distributions in Wind Modeling

2.2.1. Weibull Distribution

The Weibull distribution was introduced by Swedish physicist Weibull and has been utilized across multiple fields including physics, materials science, geography, medicine, and economics [17]. The Weibull distribution is commonly employed to model wind speeds as it accurately describes how wind speed varies in practice, with weak winds occurring more frequently than strong winds. The probability density function (PDF) f(x) of the Weibull distribution is [23]-[25]:

$$f(v) = \left(\frac{k}{c}\right)\left(\frac{v}{c}\right)^{(k-1)} \exp\left[-\left(\frac{v}{c}\right)^{k}\right] \quad 0 \le v \le \infty \tag{4}$$

Where n is the number of observations performed, c represents the scale parameter and k represents the shape parameter for the PDF of the Weibull function estimated using the Maximum Likelihood Estimation (MLE) method according to the following approximate formulas [5]:

$$k = \left[\frac{\sum_{i=1}^{n} v_i^k \ln(v_i)}{\sum_{i=1}^{n} v_i^k} - \frac{\sum_{i=1}^{n} \ln(v_i)}{n}\right] \tag{5}$$

JTE HCMUTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

$$c=\left(\frac{1}{n}\sum_{i=1}^{n}v_i^k\right)^{1/k} \tag{6}$$

2.2.2. Rayleigh Distribution

When the shape parameter of the two-parameter Weibull equals 2, the Rayleigh distribution is obtained. The probability density function (PDF) f(x) of the Rayleigh distribution is [26]-[29]:

$$f(v)=\left(\frac{v}{\sigma^2}\right)\exp\left[-\left(\frac{v}{2\sigma}\right)^2\right] \quad 0\le v\le\infty \tag{7}$$

The scale coefficient σ is calculated as follows [27]:

$$\sigma=\sqrt{\frac{1}{2n}\sum_{i=1}^{n}v_i^2} \tag{8}$$

2.2.3. Lognormal Distribution

The lognormal distribution is utilized in situations modeling wind speeds with right-skewed distributions, particularly in areas with significant short-term wind speed variations. The probability density function (PDF) f(x) of the Lognormal distribution is [5], [26]:

$$f(v)=\frac{1}{v\sigma\sqrt{2\pi}}\exp\left[-\left(\frac{(\ln v-\mu)}{2\sigma}\right)^2\right] \quad 0\le v\le\infty \tag{9}$$

Where $v>0, \mu\in R, \sigma>0$. σ is the scale parameter and μ is the location parameter estimated using the MLE method [5]

$$\mu=\frac{1}{n}\sum_{i=1}^{n}\ln v_i \tag{10}$$

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(\ln v_i-\mu)^2 \tag{11}$$

2.2.4. Generalized Extreme Value (GEV) Distribution

The Generalized Extreme Value Distribution (GEV) is employed to model extreme events, such as maximum wind speeds, in wind studies [17]. The probability density function (PDF) f(x) of the GEV distribution is [30], [31]:

$$f(v)=\frac{1}{\sigma}\left[1+\xi\left(\frac{v-\mu}{\sigma}\right)\right]^{-1-1/\xi}\exp\left(-\left[1+\xi\left(\frac{v-\mu}{\sigma}\right)\right]^{-1/\xi}\right) 0\le v\le\infty \tag{12}$$

With the condition $1+\xi\left(\frac{v-\mu}{\sigma}\right)>0$. Where: μ: location parameter, σ: scale parameter, ξ: shape parameter.

2.2.5. Goodness-of-Fit Assessment

Root Mean Square Error (RMSE) [13], [31] is employed to evaluate model accuracy by measuring the mean square deviation between observed cumulative probabilities and estimated cumulative probabilities. Lower RMSE values indicate better model fit to actual data.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(F_i - F_i)^2} \tag{13}$$

Where F_i and F_i are the empirical cumulative probability of measured data and estimated cumulative probability in the i-th wind speed interval, respectively, and n is the number of samples.

The Kolmogorov-Smirnov (KS) test [6], [17] is used to determine whether a probability distribution model fits a wind speed observation dataset and to compare the goodness-of-fit of different models with the same dataset. Lower KS values indicate better model fit to actual data.

$$KS = \max\left\{F_i - F_i\right\} \tag{14}$$

Where F_i and F_i are the empirical cumulative probability of measured data and estimated cumulative probability in the i-th wind speed interval, respectively.

The Chi-square test [6], [17] is used to verify whether the frequency of measured wind speed data matches the frequency obtained from the assumed model. The chi-square test requires dividing observed data into multiple groups, with the frequency of each group represented as a frequency histogram. The test statistic is then calculated. Lower Chi² values indicate better model fit to actual data.

$$\chi^2 = \sum_{t=1}^{T}\frac{\left(O_i - E_i\right)^2}{E_i} \tag{15}$$

Where O_i : is the observed frequency of the i-th group

is the estimated frequency of the i-th group, calculated according to the following equation: $E_i = n\left[F\left(v_i\right) - F\left(v_{i-1}\right)\right]$, v_i and v_{i-1} are the upper and lower wind speed limits of the i-th group, respectively.

The Akaike Information Criterion (AIC) [32] is used to compare the goodness-of-fit of different distributions. Lower AIC values indicate better performance.

$$AIC = -2\log L + 2k \tag{16}$$

Where k is the number of model parameters and L is the maximum likelihood function value.

3. Proposed Methodology

This study employs statistical simulation to assess wind speed characteristics by altitude and season throughout the year. The procedure comprises:

Step 1: Collection of hourly wind speed data over one year at different altitudes.

Step 2: Data preprocessing, removal of missing or anomalous values, followed by division into four seasons (Spring, Summer, Fall, Winter).

Step 3: Application of four probability distributions (Weibull, Rayleigh, Lognormal, GEV) to simulate wind speeds at each altitude and season, based on the right-skewed characteristics and flexibility of the models.

Step 4: Parameter estimation for each distribution using the Maximum Likelihood Estimation (MLE) method to ensure the highest goodness-of-fit with the data.

Step 5: Assessment of model goodness-of-fit.

Step 6: Analysis and comparison of results.

Results from each distribution model are synthesized and compared by altitude and season to identify which model demonstrates the best simulation capability under specific conditions. This provides a

JTE
HCMUTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

scientific basis for selecting appropriate models in practical applications such as wind energy potential assessment and turbine design.

The implementation procedure for evaluating probability distribution function (PDF) models to simulate wind speed is illustrated in Figure 2:

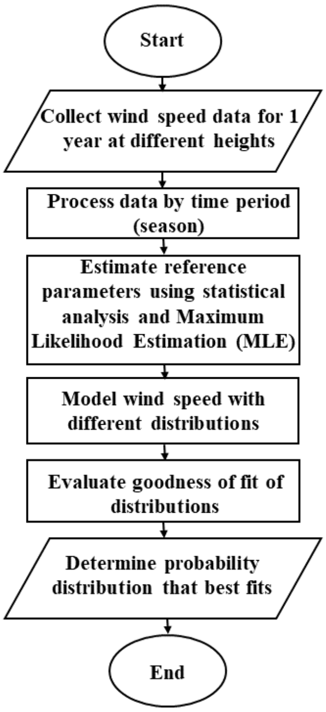

Figure 2. *Flowchart for evaluating probability distribution function (PDF) models.*

4. Results and Discussion

4.1. Statistical Descriptive Analysis of Wind Data

This section presents the wind speed simulation results at altitudes of 80m, 100m, and 120m, categorized by four seasons. The probability distribution models employed include Weibull, Rayleigh, Lognormal, and GEV.

Table 1 presents the fundamental characteristics of wind speeds at three altitudes (80 m, 100 m, 120 m) by season. Results indicate that Winter exhibits the highest mean wind speeds (6.79 m/s at 80 m and 6.85 m/s at 100 m) along with the largest standard deviations (4.3–4.4 m/s), reflecting strong but highly variable winds over time. Conversely, Spring records the lowest mean wind speeds (4.93–4.96 m/s) and smallest standard deviations (~2.29 m/s), indicating weak but stable winds. Summer and Fall maintain mean wind speeds above 5 m/s with moderate standard deviations, suitable for electricity generation due to relatively stable wind conditions. Notably, maximum wind speeds in Winter at 120 m reach 26.83 m/s—exceeding the safe operating threshold of many turbines, while Summer, with its low variability, represents a suitable choice when prioritizing stability and operational safety.

Table 1. *Basic statistical characteristics of wind speed data at different heights*

Season	Height	Sample Size	Minimum Wind Speed (m/s)	Maximum Wind Speed (m/s)	Mean Wind Speed (m/s)	Standard Deviation (m/s)
Spring	80m	2208	0.29	15.42	4.95	2.29
	100m	2208	0.19	15.76	4.96	2.30
	120m	2208	0.12	15.99	4.93	2.30

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Summer	80m	2208	0.13	16.29	5.19	2.13
	100m	2208	0.05	16.47	5.21	2.14
	120m	2208	0.04	16.57	5.20	2.15
Fall	80m	2184	0.08	14.74	5.07	2.26
	100m	2184	0.09	15.12	5.09	2.28
	120m	2184	0.09	15.41	5.07	2.29
Winter	80m	2162	0.16	25.91	6.79	4.25
	100m	2162	0.14	26.44	6.85	4.35
	120m	2162	0.13	26.83	6.88	4.43

Figure 3 presents a bar chart illustrating mean wind speeds across four seasons at three altitudes: 80m, 100m, and 120m. The columns in the chart demonstrate distinct variations in mean wind speed by season. At all three altitudes, Winter exhibits the highest mean wind speeds, significantly surpassing other seasons, indicating the greatest wind energy exploitation potential during this period.

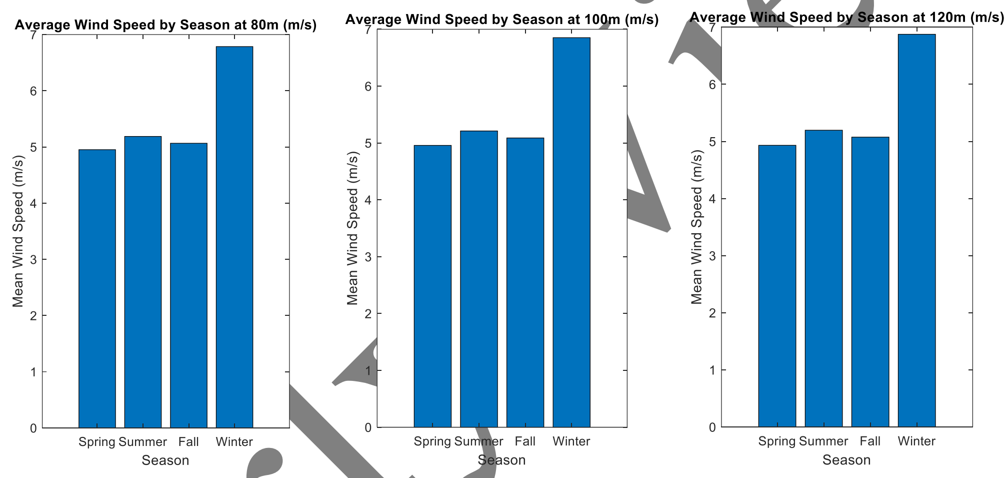

Figure 3. *Seasonal average wind speed at heights of 80m, 100m, 120m.*

Regarding altitude effects, wind speed increases moderately as altitude increases from 80m to 120m, which aligns with meteorological principles that wind speed increases with altitude. This demonstrates a consistent trend where mean wind speed increases proportionally with altitude. From the above analyses, it is evident that wind speed varies not only temporally but also depends on altitude. This constitutes an important foundation for assessing wind power potential and designing wind energy exploitation systems effectively and optimally. This phenomenon occurs because at higher altitudes, surface friction with terrain decreases, enabling wind to maintain higher velocities, demonstrating that installing wind turbines at greater heights provides more advantageous wind energy exploitation.

4.2. Parameter Estimation using MLE and Model Goodness-of-Fit Assessment

In wind speed modeling, the selection of appropriate probability distributions relies not only on the shape of observed data but also requires accurate estimation of characteristic parameters. The Maximum Likelihood Estimation (MLE) method is employed to estimate parameters for four distributions: Weibull, Rayleigh, Lognormal, and GEV. The likelihood function is constructed from the probability density function (PDF) according to equation (1) and converted to log-likelihood form to simplify calculations as shown in equation (2). The optimal parameters are determined through the system of partial derivative equations according to equation (3).

The values obtained from the MLE parameter estimation process are presented in Table 2. The Weibull distribution (equation (4)) with two parameters k and c, determined through equations (5) and

JTE HCMUTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

(6), shows that parameter c exhibits minimal variation, while k tends to decrease slightly with height, reflecting changes in distribution shape. The Rayleigh distribution (equation (7)) with the single parameter σ calculated using equation (8) demonstrates that σ increases slightly with height. The Lognormal distribution (equation (9)) has two parameters μ and σ, determined via equations (10) and (11), both of which increase slightly with height. For the GEV distribution (equation (12)), the three parameters μ, σ, ξ are estimated numerically due to the absence of closed-form solutions; these parameters fluctuate slightly and do not exhibit clear trends with height.

Following parameter estimation, the PDF functions are utilized to compare with experimental data. The goodness-of-fit is assessed through KS, RMSE, Chi², and AIC indices (equations (13) through (16)). At all three heights of 80 m, 100 m, and 120 m, the GEV distribution yields the best results with the lowest KS, Chi², RMSE, and AIC values. Particularly at 80 m, GEV achieves KS = 0.0428, Chi² = 417.61, RMSE = 0.0131, and AIC = 42075.00, reflecting high goodness-of-fit, low error, and a model that balances accuracy with complexity.

Although the Weibull distribution performs inferior to GEV, it still demonstrates considerable goodness-of-fit, especially in the peak and tail regions of the distribution. At 80 m, Weibull exhibits KS = 0.0698, RMSE = 0.0154, AIC = 42423.92, and maintains stable performance at greater heights. The Rayleigh distribution provides acceptable results at 80 m (RMSE = 0.0146, Chi² = 661.46) but shows poor effectiveness at 100 m and 120 m (KS > 0.08, high AIC). The Lognormal distribution performs worst with all indices being highest at all heights (e.g., at 80 m: KS = 0.0970, Chi² = 1270.77, RMSE = 0.0236, AIC = 43290.71).

A general trend is observed where KS, Chi², RMSE, and AIC values tend to increase with height for all distributions. This indicates a slight decrease in goodness-of-fit and accuracy of the models as height increases. However, the GEV distribution maintains the best performance at all three heights, demonstrating stability and superior modeling capability, particularly in the context of highly uncertain data with extreme value occurrences.

Table 2. *MLE estimation of parameters and goodness-of-fit at heights of 80m, 100m and 120m.*

Height	Distribution	MLM estimates of the parameters	Statistical test			
			KS	Chi²	RMSE	AIC
80m	Weibull	λ = 6.1971, k = 1.9467	0.0698	697.13	0.0154	42423.92
	Rayleigh	σ = 4.4104	0.0716	661.46	0.0146	42433.99
	Lognormal	μ = 1.5466, σ = 0.6094	0.0970	1270.77	0.0236	43290.71
	GEV	μ = 4.1867, σ = 2.2936, ξ = -0.0032	0.0428	417.61	0.0131	42075.00
100m	Weibull	λ = 6.2290, k = 1.9259	0.0755	779.81	0.0167	42631.35
	Rayleigh	σ = 4.4453	0.0781	758.89	0.0157	42653.23
	Lognormal	μ = 1.5498, σ = 0.6141	0.1008	1299.56	0.0241	43482.47
	GEV	μ = 4.1941, σ = 2.3053, ξ = 0.0035	0.0474	466.97	0.0140	42229.09
120m	Weibull	λ = 6.2193, k = 1.9025	0.0785	883.06	0.0177	42736.16
	Rayleigh	σ = 4.4529	0.0827	869.09	0.0165	42776.69
	Lognormal	μ = 1.5462, σ = 0.6173	0.1030	1345.17	0.0246	43509.32
	GEV	μ = 4.169, σ = 2.3, ξ = 0.01219	0.0490	528.09	0.0148	42276.83

4.3. Comparison of PDF Distributions

The probability density functions obtained from Weibull, Rayleigh, Lognormal, and GEV distributions at the study site are illustrated in Figure 4, reflecting distinct differences in wind speed modeling capabilities across different heights.

JTE HCMUTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

At heights of 80 m, 100 m, and 120 m, the wind speed histograms exhibit a right-skewed distribution with peaks ranging from 5 m/s to 6 m/s. The Weibull distribution demonstrates the highest goodness-of-fit at all three heights, effectively modeling both the peak and tail regions, particularly closely matching the actual data at 80 m and 100 m. Meanwhile, the GEV distribution shows clear superiority in the tail region, especially at 120 m where it accurately reproduces extreme values. Conversely, the Rayleigh distribution provides only moderate modeling in the central region but exhibits poor accuracy in the tail region, while the Lognormal distribution fails to reproduce both the peak and tail of the histogram. Overall, Weibull is the most suitable distribution for the entire data domain, while GEV is the preferred choice when analyzing extreme phenomena at greater heights.

In terms of parameter interpretation, the Weibull distribution's shape parameter (k) controls the peakedness of the curve, and its scale parameter (c) is closely related to the average wind speed — both showing consistent adaptability across different heights. For the Rayleigh distribution, which is a special case of Weibull with fixed k = 2, the scale parameter tends to overestimate the probability in the lower wind speed ranges. The Lognormal distribution is governed by its mean (μ) and standard deviation (σ) of the underlying normal distribution in logarithmic scale, which leads to significant skewness and misalignment at both the head and tail of the data. The GEV distribution, characterized by its location (μ), scale (σ), and shape (ξ) parameters, shows better flexibility in modeling extreme wind speeds due to its capacity to account for heavier tails, especially evident in the 120m dataset.

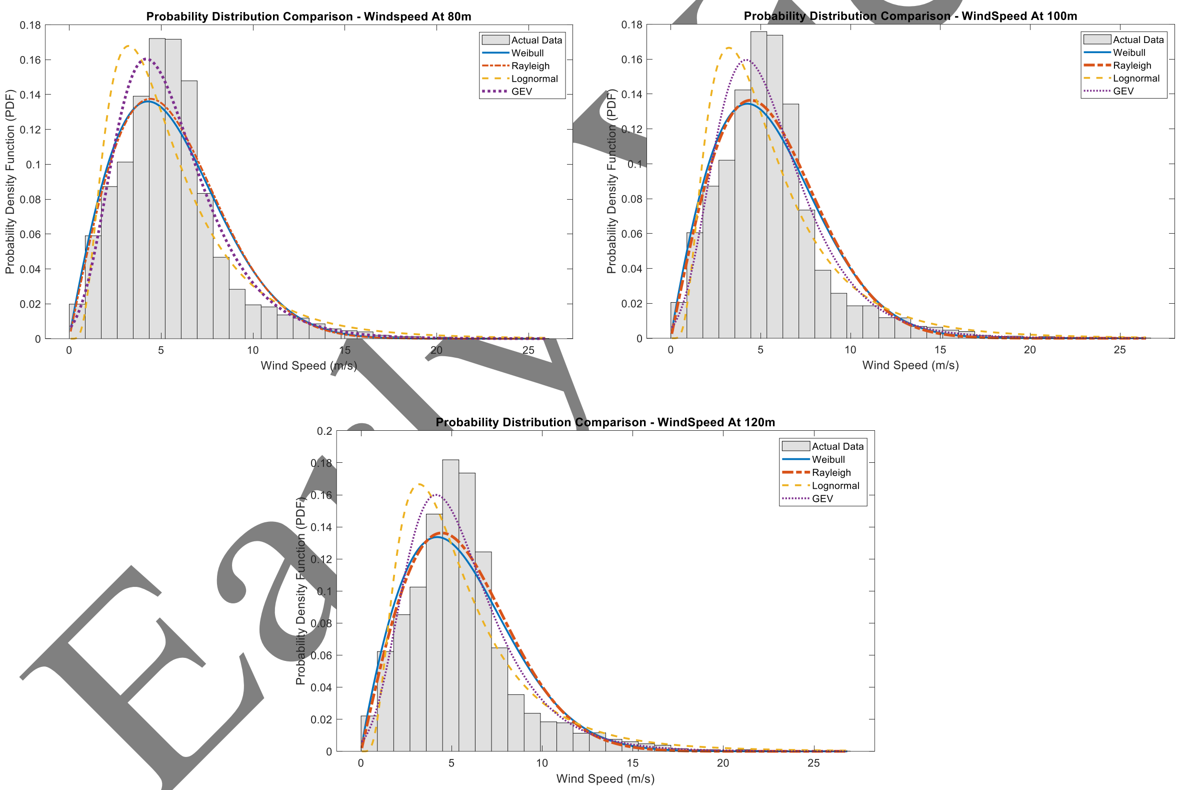

Figure 4. *Probability density functions of distribution models.*

In general, both Weibull and GEV distributions effectively model the actual data. GEV excels in the right tail region, particularly at 120 m, where the distribution curve nearly coincides with the histogram. Conversely, Weibull maintains high goodness-of-fit in the central region and peak, especially at 80 m and 100 m. Meanwhile, Rayleigh and Lognormal distributions exhibit significant deviations, reflecting poor modeling capabilities.

5. Conclusions

This study analyzes wind speed data from Kauai, Hawaii, to identify the most suitable probability distribution for modeling. Four distributions—Weibull, Rayleigh, Lognormal, and Generalized Extreme

Value (GEV)—were evaluated, with parameters estimated using Maximum Likelihood Estimation (MLE) in MATLAB. Model performance was assessed using the Kolmogorov-Smirnov and Chi-square tests, Akaike Information Criterion (AIC), and Root Mean Square Error (RMSE).

Among the candidates, the GEV distribution exhibited the best overall fit across all evaluated heights (80 m, 100 m, and 120 m), accurately capturing both peak and tail behaviors. Weibull also performed reasonably well, balancing model accuracy and computational simplicity. Rayleigh's single-parameter model showed limited accuracy, while Lognormal yielded the poorest fit with significant deviations from observed data.

The significance of this study lies in its practical implications for renewable energy system design, particularly in wind energy forecasting, turbine placement, and system reliability analysis. By identifying GEV as the most reliable distribution and Weibull as a suitable alternative, the study provides a robust statistical foundation for modeling wind behavior in island environments. This contributes to improving the efficiency, safety, and feasibility of wind energy integration in power systems, especially under conditions of high variability and extreme events.

Acknowledgments

This work belongs to the project grant No: T2024-02NCS funded by Ho Chi Minh City University of Technology and Education, Vietnam.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

[1] R. Sharma, P. Dutta, and S. S. Murthy, "Application of hydrogen storage in polygeneration microgrids: case study of wind microgrid in India," *Energy*, vol. 311, p. 133331, 2024, doi: 10.1016/j.energy.2024.133331.

[2] Z. Ren, Q. Yu, D. Lin, and Y. Dong, "Optimal capacity configuration of a wind-solar-battery-diesel microgrid system using continuous grey wolf optimization," *Journal of Energy Storage*, vol. 113, pp. 115630, 2025, doi: 10.1016/j.est.2025.115630.

[3] X. Li, Y. Chen, K. Li, S. Gao, and Y. Cui, "The optimal wind speed product selection for wind energy assessment under multi-factor constraints," *Cleaner Engineering and Technology*, 2025, doi: 10.1016/j.clet.2025.100883.

[4] G. R. Sheu *et al.*, "Distribution of atmospheric mercury in northern Southeast Asia and South China Sea during Dongsha Experiment," *Atmos. Environ.*, vol. 78, pp. 174–183, Oct. 2013, doi: 10.1016/J.ATMOSENV.2012.07.002.

[5] I. Pobočíková, Z. Sedliačková, and M. Michalková, "Application of four probability distributions for wind speed modeling," *Procedia Engineering*, vol. 192, pp. 713–718, 2017, doi: 10.1016/j.proeng.2017.06.123.

[6] Y. Wang, Y. Li, R. Zou, and D. Song, "Bayesian infinite mixture models for wind speed distribution estimation," *Energy Conversion and Management*, vol. 236, p. 113946, 2021, doi: 10.1016/j.enconman.2021.113946.

[7] F. Chellali, A. Khellaf, A. Belouchrani, and R. Khanniche, "A comparison between wind speed distributions derived from the maximum entropy principle and Weibull distribution: Case study of six regions of Algeria," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 1, pp. 379–385, 2011, doi: 10.1016/j.rser.2011.08.002.

[8] W. Werapun, Y. Tirawanichakul, and J. Waewsak, "Comparative study of five methods to estimate Weibull parameters for wind speed on Phangan Island, Thailand," *Energy Procedia*, vol. 79, pp. 976–981, 2015, doi: 10.1016/j.egypro.2015.11.596.

[9] S. M. Aljeddani and M. Mohammed, "A novel approach to Weibull distribution for the assessment of wind energy speed," *Alexandria Engineering Journal*, vol. 78, pp. 56–64, 2023, doi: 10.1016/j.aej.2023.07.027.

[10] D. R. Lins, K. S. Guedes, A. R. P. Neto, P. A. C. Rocha, and C. F. De Andrade, "Comparison of the performance of different wind speed distribution models applied to onshore and offshore wind speed data in Northeast Brazil," *Energy*, vol. 278, p. 127787, 2023, doi: 10.1016/j.energy.2023.127787.

[11] H. Patidar, V. Shende, and P. Baredar, "Comparative study of offshore wind energy potential assessment using different Weibull parameters estimation methods," *Environ. Sci. Pollut. Res.*, vol. 29, pp. 46341–46356, 2022, doi: 10.1007/s11356-022-19109-x.

[12] F. Alfawzan, J. E. Alleman, and C. R. Rehmann, "Wind energy assessment for NEOM city, Saudi Arabia," *Energy Sci. Eng.*, vol. 8, pp. 755–767, 2020, doi: 10.1002/ese3.548.

[13] B. Ndeba, O. el Alani, A. Ghennioui, and M. Benzaazoua, "Comparative analysis of seasonal wind power using Weibull, Rayleigh and Champernowne distributions," *Scientific Reports*, vol. 15, 2025, doi: 10.1038/s41598-025-86321-3.

[14] N. Aries, S. M. Boudia, and H. Ounis, "Deep assessment of wind speed distribution models: A case study of four sites in Algeria," *Energy Conversion and Management*, vol. 155, pp. 78–90, 2017, doi: 10.1016/j.enconman.2017.10.082.

[15] N. Masseran, A. Razali, and K. Ibrahim, "An analysis of wind power density derived from several wind speed density functions: The regional assessment on wind power in Malaysia," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 8, pp. 6476–6487, 2012, doi: 10.1016/j.rser.2012.03.073.

[16] N. Masseran, "Integrated approach for the determination of an accurate wind-speed distribution model," *Energy Conversion and Management*, vol. 173, pp. 56–64, 2018, doi: 10.1016/j.enconman.2018.07.066.

[17] H. Shi, Z. Dong, N. Xiao, and Q. Huang, "Wind Speed Distributions Used in Wind Energy Assessment: A Review," *Frontiers in Energy Research*, vol. 9, 2021, doi: 10.3389/fenrg.2021.769920.

[18] W. W. Gould, J. Pitblado, and W. Sribney, "Maximum likelihood estimation with STATa," *ResearchGate*, Jan. 2006

[19] C. K. Enders, "Maximum likelihood estimation," *Encyclopedia of Statistics in Behavioral Science*, Apr. 2005, doi: 10.1002/0470013192.bsa200.

[20] R. B. Millar, "Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB", Chichester, U.K.: Wiley, 2011.

[21] R. H. Norden, "A survey of maximum likelihood estimation," *International Statistical Review / Revue Internationale de Statistique*, vol. 40, no. 3, pp. 329–354, Dec. 1972.
[22] S. S. Kutty, M. G. M. Khan, and M. R. Ahmed, "Analysis of wind characteristics and wind energy resource assessment for Tonga using eleven methods of estimating Weibull parameters," *Heliyon*, vol. 10, no. 9, p. e30047, Apr. 2024, doi: 10.1016/j.heliyon.2024.e30047.
[23] S. Jahan, N. Masseran, and W. Zin, "Wind speed analysis using Weibull and lower upper truncated Weibull distribution in Bangladesh," *Energy Reports*, vol. 11, pp. 5456–5465, 2024, doi: 10.1016/j.egyr.2024.05.029.
[24] A. F. Zobaa and S. H. E. A. Aleem, "Uncertainties in Modern Power Systems," *Academic Press*, Chapter 1, pp. 3–4.
[25] I. Pobocikova, Z. Sedliackova, M. Michalkova, and F. George, "Monte Carlo Comparison of the methods for estimating the WEIBULL distribution parameters - Wind Speed application," *Communications - Scientific Letters of the University of Zilina*, vol. 19, no. 2A, pp. 79–86, 2017, doi: 10.26552/com.c.2017.2a.79-86.
[26] Y. Bai and W. Jin, "Random variables and uncertainty analysis," in *Elsevier eBooks*, 2015, pp. 615–625.
[27] C. Kongnam, S. Nuchprayoon, S. Premrudeepreechacharn, and S. Uatrongjit, "Decision analysis on generation capacity of a wind park," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 2, pp. 2126–2133, Jan. 2009, doi: 10.1016/j.rser.2009.01.023.
[28] G. Carpinelli, G. Celli, F. Pilo, and A. Russo, "Distributed generation siting and sizing under uncertainty," in *IEEE Conference Publication*, vol. 4, 2001, doi: 10.1109/PTC.2001.964856.
[29] S. Zhang, P. Huang, and Y. Sun, "A multi-criterion renewable energy system design optimization for net zero energy buildings under uncertainties," *Energy*, vol. 94, pp. 654–665, 2016, doi: 10.1016/j.energy.2015.11.044.
[30] Z. Zhou, J. Zhang, P. Liu, Z. Li, M. C. Georgiadis, and E. N. Pistikopoulos, "A two-stage stochastic programming model for the optimal design of distributed energy systems," *Applied Energy*, vol. 103, pp. 135–144, Jan. 2013, doi: 10.1016/j.apenergy.2012.09.019.
[31] S. S. Kutty, M. G. M. Khan, and M. R. Ahmed, "Analysis of wind characteristics and wind energy resource assessment for Tonga using eleven methods of estimating Weibull parameters," *Heliyon*, 2024
[32] K. P. Burnham and D. R. Anderson, "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements," *Wiley Interdisciplinary Reviews: Computational Statistics*, 2020, Art. no. e1460.

Nguyen Thai An received B.Sc. and M.Sc. degree in electrical engineering from Ho Chi Minh City University of Technology and Education (HCMUTE), Vietnam, in 2018 and 2020, respectively. Currently, he is a lecturer in the Faculty Electrical and Electronics Engineering, HCMUTE. His main areas of research interests are load shedding in power systems and Microgrid, power systems stability, and load forecasting, distribution network. He can be contacted at ntan@hcmute.edu.vn. ORCID: https://orcid.org/0000-0002-4435-5327

Khong Thi Tham is a student in the Faculty Electrical and Electronics Engineering, HCMUTE. Her main areas of research interests are load shedding in power systems and Microgrid, power systems stability, and load forecasting, distribution network. She can be contacted at 21142380@student.hcmute.edu.vn. ORCID: https://orcid.org/0009-0000-3590-6417

Nguyen Thi Kim Thao is a student in the Faculty Electrical and Electronics Engineering, HCMUTE. Her main areas of research interests are load shedding in power systems and Microgrid, power systems stability, and load forecasting, distribution network. She can be contacted at 21142376@student.hcmute.edu.vn. ORCID: https://orcid.org/0009-0009-2266-9470

Quyen Huy Anh received his PhD degree in power system from MPIE, Russia in 1993. Currently, he is a professor and lecturer in the Faculty Electrical and Electronics Engineering, HCMUTE. His main areas of research interests are modeling power systems, pattern recognition in dynamic stability of power systems, artificial intelligence. Email: anhqh@hcmute.edu.vn. ORCID: https://orcid.org/0000-0001-8946-1302

Huynh Thi Thu Hien received B.Sc. and M.Sc. degree in electrical engineering from Ho Chi Minh City University of Technology and Education (HCMUTE), Vietnam.. Currently, she is a lecturer in the Faculty Electrical and Electronics Engineering, HCMUTE. Her main areas of research interests are power systems and Microgrid, Renewable Energy. She can be contacted at hienhtthu@hcmute.edu.vn. ORCID: https://orcid.org/0009-0004-3158-9177