Rev. Roum. Sci. Techn.– Électrotechn. et Énerg.
Vol. **71**, *2*, pp. 223–228, Bucarest, 2026

SUSTAINABLE WIND RESOURCE ASSESSMENT USING ADAPTIVE WEIBULL ESTIMATION TECHNIQUES IN DATA-SCARCE REGIMES

CHATCHAWICH CHAIHONG[1], JUNTAKAN TAWEEKUN[2*]

Keywords: Wind resource exploitation; Missing wind speed data; Wind speed modeling; Performance analysis.

Accurate wind data is important for clean energy planning. However, many weather datasets have missing values due to sensor issues. This study explores how missing data affects Weibull parameter estimation and compares four methods: the empirical method of Justus (EMJ), maximum likelihood estimation (MLE), moment method (MM), and least squares method (LSM). Five years of data (2020–2024) from 14 stations in Southern Thailand were used. Wind speeds ranged from 0.35 to 2.18 m/s, with a peak of 32.98 m/s. Data completeness varied from 27.61% to 84.14%. Model accuracy was tested using MAPE, RMSE, Chi-square, and R². EMJ and MM gave strong results with complete data. MLE worked best with 50–80% completeness. LSM showed high R² but overfitted with missing data. MM was the most reliable overall. These findings support better wind modeling in areas with limited data, helping clean energy development in Southern Thailand.

1. INTRODUCTION

Reliable wind assessment is important for turbine design, cost analysis, and wind farm planning [1-4]. The Weibull distribution [1-6] is commonly used to model wind speed, based on two parameters, scale (c), which shows wind strength, and shape (k), which shows the spread and peak. Accurate estimation of these parameters is essential, as errors can lead to incorrect energy predictions [1–5].

Over the years, several numerical methods have been developed to estimate the Weibull parameters from observed wind speed data [1,2,4,5,7–10]. Common approaches include the graphical method (GM), the empirical methods of Justus (EMJ) and Lysen (EML), the moment method (MM), the energy pattern factor method (EPFM), and variations of the maximum likelihood method (MLM, MMLM, AMLM). Their performance is usually evaluated using validation metrics such as root mean square error (RMSE), the chi-square (X^2), and the coefficient of determination (R^2). However, results in the literature are often conflicting. Some studies suggest MLM and MMLM perform best, especially with large datasets [1,4], while others highlight MM as more reliable across diverse landscapes. By contrast, the GM method is often ranked lowest because it relies on cumulative frequencies and is sensitive to bin intervals [1, 4]. Sample size also plays a role; small samples ($n \leq 15$) tend to favor the least squares method (LSM), while larger samples ($n > 20$) typically improve MLM performance [10]. Similar research studies, studies by Yunn-Kuang Chu and Jau-Chuan Ke emphasize the importance of sample size. It was found that LSM significantly outperforms MLM when the sample size is small ($n \leq 15$), particularly for distributions with decreasing or constant failure rates [10]. This may be due to the lack of regularity conditions in the MLM estimating equations when applied to small samples [10]. Conversely, MLM tends to outperform LSM for distributions with increasing failure rates when sample sizes are large ($n > 20$) [10]. In contrast, A study on Jeju Island, South Korea, using five years of data from nine sites, found the MM to be the most accurate for estimating Weibull parameters, regardless of terrain [1]. Six methods were compared using RMSE, maximum error, WPD error, and R^2, with MM ranked first, followed by EPFM, EMJ, and GM. Another study using 10 years of data from Maldo Island tested 12 methods and found MM, EML, EMJ, and the Standard Deviation Method (STDM) to be the most stable, while others showed inconsistent results [11]. These findings confirm that terrain and wind patterns affect accuracy, and MM remains the most reliable method.

Despite advances in wind energy research, one important issue remains underexplored: the effect of data completeness on parameter estimation. In practice, meteorological datasets often have missing values due to sensor faults, transmission errors, or equipment maintenance [12]. These data gaps may reduce the accuracy of estimated parameters and increase errors in validation results [13–15]. Previous studies have focused mainly on data truncation or the removal of low wind speeds [1,2,10], but only a few have directly investigated how different levels of completeness influence the robustness of estimation methods. This study addresses this gap by comparing four commonly used methods for Weibull parameter estimation: the EMJ, MLE, MM, and LSM methods. A five-year dataset (2020–2024) from 14 automatic weather stations (AWS) operated by the Thai Meteorological Department (TMD) in Southern Thailand [16] was analyzed. The research evaluates how missing data affects the accuracy of the shape (k) and scale (c) parameters, as well as the performance of each method across key validation metrics, including MAPE, RMSE, X^2, and R^2. Southern Thailand is a suitable case study due to its diverse geography and wind patterns. This study tested real-world datasets with varying completeness to evaluate the strengths and weaknesses of different estimation methods. The results aim to improve wind energy modeling and support better planning in areas with limited or inconsistent wind data.

2. DATA AND METHODOLOGY

The key steps taken to prepare and analyze wind speed data for parameter estimation. The process includes data processing, analytical techniques, estimation methods, and validation metrics.

Data Collection and Analysis

The data collection process was carried out in two main stages. The first stage selected the database of wind data sources through direct and indirect identification, guided by the key categories shown in Table 1. The second stage focused on analyzing and classifying the selected dataset [17].

---

[1] Energy Technology Program, Department of Interdisciplinary Engineering, Faculty of Engineering, Prince of Songkla University, Hat Yai campus, Songkhla, Thailand.
[2*] Department of Mechanical and Mechatronics Engineering, Faculty of Engineering, Prince of Songkla University, Hat Yai campus, Songkhla, Thailand.

Emails: 6610130021@email.psu.ac.th, juntakan.t@psu.ac.th (Corresponding Author)

DOI: 10.59277/RRST-EE.2026.2.9

Table 1

Knowledge database fields [17].

Field	Description
Wind data	Wind speed, Wind direction, and typical deviation
Methodology	Procedure is used to obtain wind variables and identify data truthfulness It is necessary to know the procedure used to obtain wind variables
Height	Height of measuring instrument at 10 meters above ground level (AGL)
Sampling campaign	A higher number of periods available, facilitating wind-resource analysis
Frequency	Reducing uncertainty, giving accurate estimations
Download format	Text, csv format, spreadsheet
Type of location	Onshore, offshore or both

2.1 STATION SELECTION FROM TMD

This study uses secondary wind speed data from AWS-TMD [16], collected from stations that meet international standards. The focus is on 14 southern provinces of Thailand, where diverse geography, including coasts, plains, and mountains, affects local wind patterns and supports wind energy potential.

Table 2

AWS Weather station from TMD.

Province	ID	Latitude/Longitude Decimal Degrees	UTM Zone	Altitude (m.)
Chumphon	55	10.332222/99.092499	47P	38.000
Ranong	54	9.954722/98.633614	47P	39.000
Surat Thani	88	9.144444/99.633057	47P	06.900
Phangnga	1083	8.683333/98.252220	47P	07.723
Krabi	1086	8.221389/99.196388	47P	89.000
Phuket	1085	7.891944/98.334724	47N	70.000
Nakhonsi-thammarat	89	8.177778/100.171669	47P	03.530
Phatthalung	50	7.613611/100.117500	47N	04.150
Songkhla	103	7.451111/100.452499	47N	03.600
Trang	1084	7.836111/99.691109	47N	118.000
Satun	63	6.653611/100.083054	47N	12.000
Pattani	85	6.789444/101.146942	47N	06.011
Yala	86	6.515278/101.273888	47N	36.040
Narathiwat	87	6.426667/101.824997	47N	05.130

Data from TMD

Data availability was verified for completeness. In this case, a 100% complete dataset must include continuous 10-minute intervals across 24 hours for 365 or 366 days, including leap years (2020 and 2024), depending on the year. All data were recorded at a height of 10 meters above ground level (AGL), following standard meteorological practices.

Table 3

Percent (%) of complete dataset.

Year	2020	2021	2022	2023	2024	2020-2024
Complete Dataset	52,704	52,560	52,560	52,560	52,704	263,088

For the five-year study period, a fully complete dataset from a single station would contain 263,088 data points. Based on this benchmark, the percentage of completeness for each station was calculated and compared, as shown in Table 3. The actual number of recorded data points for each station is detailed in Table 4, providing a clear view of how much data was available relative to the ideal total.

Table 4

Data from TMD.

Stations ID	Wind Speed in m/s at 10m. AGL			
	Range [Min (Max)]	Mean	Standard Deviation	Datasets
55	0.00 (6.17)	0.80	0.83	76.81%
54	0.00 (7.00)	1.30	0.99	83.17%
88	0.00 (7.77)	0.73	0.98	78.21%
1083	0.00 (9.12)	1.18	1.05	30.50%
1086	0.00 (12.80)	0.51	0.72	29.47%
1085	0.00 (6.96)	0.59	0.59	30.72%
89	0.00 (9.27)	0.82	1.11	84.14%
50	0.00 (8.64)	1.17	1.31	79.73%
103	0.00 (13.12)	2.18	1.71	73.93%
1084	0.00 (14.48)	0.35	0.52	27.61%
63	0.00 (8.59)	1.17	1.10	74.20%
85	0.00 (8.95)	1.16	1.13	76.49%
86	0.00 (10.49)	1.40	0.97	79.23%
87	0.00 (32.98)	1.93	1.25	84.10%

Tables 2 and 4 provide detailed information on each station's geographic coordinates, altitude, and UTM/USNG location, with altitudes ranging from near sea level in Nakhon Si Thammarat province (3.53 m) to higher terrain in Trang province (118 m). To complement the station-level data, Figure 1 presents the mean wind speed map of Thailand at 10 m height, sourced from the Global Wind Atlas [18]. This map offers a broader spatial context and supports the interpretation of localized wind patterns observed in the AWS data.

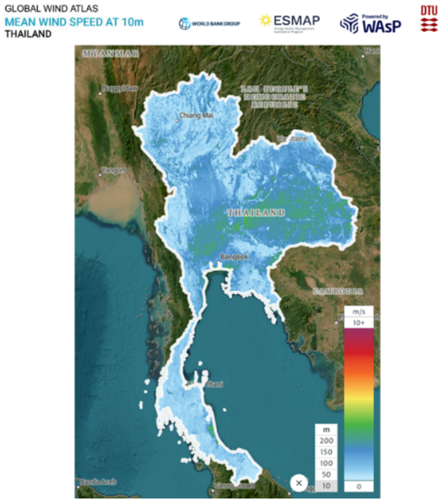


Fig. 1 – Mean Wind Speed Maps at Height 10 m. in Thailand (Source: https://globalwindatlas.info/en/area/Thailand) [18].

Wind characteristics in Southern Thailand vary by location. Mean wind speeds range from 0.35 m/s in Trang province to 2.18 m/s in Songkhla province, showing different energy potential. Songkhla province has the highest wind variation, while Trang and Phuket provinces are more stable. Narathiwat province recorded a peak of 32.98 m/s, suggesting the need for local risk checks. Most provinces have high data completeness (70–84%), but Trang, Krabi, and Phangnga provinces have lower completeness, which may affect accuracy.

2.2 DATA PROCESSING WORKFLOW

Data processing was carried out to transform raw meteorological records into structured datasets ready for analysis.

- Temporal: 10 min time interval
- Time extent: 1/1/2020 to 31/12/2024
- Parameters: timestamp, wind dir (deg), wind speed in knots at 10 m, max wind dir (deg), max wind speed in knots at 10m, temperature (°C), humidity %, heat index (°C), pressure (hPa), QFF. (hPa), precipitation (mm.), vis. m, weather. code
- Time standard: universal time coordinated (UTC) +7 UTC in Thailand.

Wind speed values were converted from knots to m/s using the formula Wind Speed (v) m/s = v knots × 0.51444. All data were saved in CSV format, compatible with Python tools. Pre-processing included timestamp standardization, error checking, and complete verification to ensure clean and consistent datasets.

2.3 WEIBULL DISTRIBUTIONS

Weibull distributions [19,20] are commonly used to model wind speed and estimate energy potential [21,22]. Special cases include the Rayleigh distribution (k = 2) for moderate wind and the exponential distribution (k = 1) for simpler patterns. These variations help assess suitability in different wind conditions [7,8].

The probability density function (PDF)

$$PDF\ or\ f(v) = \left(\frac{k}{c}\right)\left(\frac{v}{c}\right)^{k-1} exp\left[-\left(\frac{v}{c}\right)^{k}\right], \quad (1)$$
$$v \geq 0;\ k, c > 0$$

where $f(v)$ is the probability of witnessing wind speed (v), k denotes the dimensionless shape factor of the distribution, and c represents the scale parameter.

The cumulative distribution function (CDF)

PDF of wind speed v is the probability that the wind speed is less than or equal to v. Therefore, the CDF $F(v)$ is the integral of the PDF as follows

$$\text{CDF}\ or\ F(v) = \int_0^{\alpha} f(v)dv = 1 - \exp\left[-\left(\frac{v}{c}\right)^{k}\right], \quad (2)$$
$$v \geq 0;\ k, c > 0$$

where $F(v)$ is the cumulative distribution function of observing wind speed (v).

2.4 PARAMETER ESTIMATION TECHNIQUES

Statistical techniques are employed to estimate the Weibull parameters (k and c).

2.4.1. NORMAL PARAMETER ESTIMATION

This method is an empirical approximation for Weibull parameter estimation, often referred to as the "standard deviation method" or "Empirical method of Justus (EMJ)" when using a specific relationship between the mean, standard deviation, and Weibull parameters [4].

The parameters k and c are typically calculated using numerical methods

Shape parameter (k)

$$k = (\sigma_v/\bar{v})^{-1.086}, \quad (3)$$

where σ_v is the standard deviation of wind speed data (m/s), $\bar{v}$ is the mean of wind speed data (m/s)

Scale parameter (c)

After finding "k", this formula can find "c"

$$c = \bar{v}/\Gamma\left(1 + \frac{1}{k}\right), \quad (4)$$

where Γ is the gamma distribution in the form of the standard gamma function, defined by

$$\Gamma(x) = \int_0^{\infty} e^{-x} x^{n-1} dx. \quad (5)$$

2.4.2. MAXIMUM LIKELIHOOD ESTIMATION (MLE)

MLE is a widely used method for estimating Weibull parameters by maximizing the likelihood that the observed wind speed data fit the Weibull distribution [9]. It treats wind speed as a time series and applies a likelihood function to evaluate the probability of each data point, assuming unknown parameters [10].

$$k = \left[\left(\sum_{i=1}^{n} v_i^k \ln(v_i) \Big/ \sum_{i=1}^{n} v_i^k\right) - \left(\sum_{i=1}^{n} \ln(v_i)/n\right)\right]^{-1}, \quad (6)$$

$$c = \left(\sum_{i=1}^{n} v_i^k/n\right)^{\frac{1}{k}}, \quad (7)$$

where v_i is wind speed measured at the interval i (m/s), i is the measurement interval, and n is the number of non-zero values.

2.4.3. MOMENT METHOD (MM)

MM is a computational technique using moments of data distribution. Based on the mean and standard deviation of the Weibull distribution [7], there is another technique for estimating Weibull parameters [5]. It uses the mean wind speed (v_m) and the standard deviation of wind speed (σ_v) [4, 23], to determine accuracy, as expressed in

$$v_m = c\Gamma\left(1 + \frac{1}{k}\right), \quad (8)$$

$$\sigma_v = c\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{2}{k}\right)\right]^{1/2}, \quad (9)$$

where v_m is the mean of wind speeds and σ_v is the standard deviation of wind speeds

Alternatively, [23] σ_v stands for the determined wind speed standard deviation. Defined mathematically as

$$\sigma_v = \left[\frac{1}{n-1}\sum_{i=1}^{n}(v_i - v_m)^2\right]^{0.5}, \quad (10)$$

and mean wind speed v_m expressed as

$$v_m = \frac{1}{n}\sum_{i=1}^{n} v_i, \quad (11)$$

where v_i stands for the speed of the wind at time i, and n denotes the sum of the data set.

2.4.4. LEAST SQUARES METHOD (LSM)

LSM estimates Weibull parameters (k and c) [7] using a linear form of the CDF and minimizes squared differences between observed and modeled values [2]. It focuses on minimizing the sum of squared deviations between observed and modeled values, which is suitable for smaller datasets but depends on consistent data grouping.

$$\ln\left(-\ln\left(1 - F(v_i)\right)\right) = k\ln(v_i) - k\ln(c). \quad (12)$$

Here, the term $\ln\left(-\ln\left(1 - F(v_i)\right)\right)$ in eq. (12) is defined as y_i. Using this relationship, the least squares estimate for the shape and scale parameters of the Weibull distribution is calculated as follows

$$k = \frac{(n\sum_{i=1}^{n} y_i \ln(v_i) - \sum_{i=1}^{n}\ln(v_i)\sum_{i=1}^{n} y_i)}{(n\sum_{i=1}^{n}(\ln(v_i))^2 - (\sum_{i=1}^{n}\ln(v_i))^2)}, \quad (13)$$

$$c = exp\left(\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) - \left(\frac{\sum_{i=1}^{n} ln(v_i)}{kn}\right)\right). \quad (14)$$

The term y_i in eq. (13) and (14) represents the estimated value obtained by using the cumulative distribution function ($F(v_i)$) in the equation $y_i = \ln\left(-\ln\left(1 - F(v_i)\right)\right)$.

2.5. VALIDATION METRICS

The accuracy of Weibull parameter estimation was assessed using MAPE, RMSE, X^2, and R^2. These metrics show how well each method fits the wind speed data and how sensitive it is to be missing values. They help evaluate both model performance and data completeness, which are important for reliable wind resource analysis.

2.5.1. MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

MAPE [24] is a metric that evaluates the percentage of absolute errors by quantifying the average percentage deviation between observed and predicted values, calculated as

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \times 100, \tag{15}$$

where n is the number of data points., y_i is Observed (actual) value at data point for the $\text{i}-\text{th}$ data point, and $\hat{y}_i$ is the predicted value for the $i-th$ data point.

MAPE ranges from $[0,\infty]$, with lower values showing better accuracy. As a percentage, it is scale-independent but sensitive to small reference values, which can cause large errors. In wave-related data, MAPE may be undefined or too high, making it less suitable for such cases.

2.5.2. ROOT MEAN SQUARE ERROR (RMSE)

RMSE [24] evaluates errors between two datasets by calculating the square root of the mean squared error (MSE). MSE [24] is given by

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(F_i - \hat{F}_i\right)^2, \tag{16}$$

where F_i is observed value for data point i, $\hat{F}_i$ is predicted value for data point i., and n is Total number of data points.

The range of RMSE is $[0,\infty]$, where $MSE = 0$ indicates a perfect match between datasets. RMSE is highly sensitive to large errors, unlike MAE, and emphasizes outliers due to the squared term. RMSE is derived as

$$\text{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(F_i - \hat{F}_i\right)^2\right]^{\frac{1}{2}}. \tag{17}$$

Alternatively, RMSE can be expressed using the mean error $(\bar{e})$ and standard deviation (s) of errors

RMSE is a metric that shows both the average error and its variation. It gives more weight to large errors and is useful for checking model accuracy. When the mean error is zero, RMSE equals the standard deviation, which often happens in wave studies. This makes RMSE helpful for evaluating wind speed models.

2.5.3. CHI-SQUARE ERROR TEST (X^2)

X^2 [6] evaluates the goodness-of-fit between observed wind speed data and theoretical models.

$$x^2 = \sum_{i=1}^{T}\frac{(O_i-E_i)^2}{E_i}, \tag{18}$$

where O_i is Observed frequency for the $i-th$ group, E_i is expected frequency for the $i-th$ group, calculated as

$$E_i = n[F(v_i) - F(v_{i-1})]. \tag{19}$$

In this method, v_i and v_{i-1} are the wind speed bounds for each group. If the X^2 value is higher than the critical value, the model is rejected. The critical value depends on the chosen distribution. X^2 is less affected by outliers but can vary with class intervals, which must be selected carefully to avoid misleading results.

2.5.4. COEFFICIENT OF DETERMINATION (R^2)

R^2, which measures the goodness of fit of the fitted Weibull distribution to the observed data's PDF function [3]. It is not the adjusted R^2 or a standard correlation coefficient.

$$R^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 - \sum_{i=1}^{n}(y_i-x_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}, \tag{20}$$

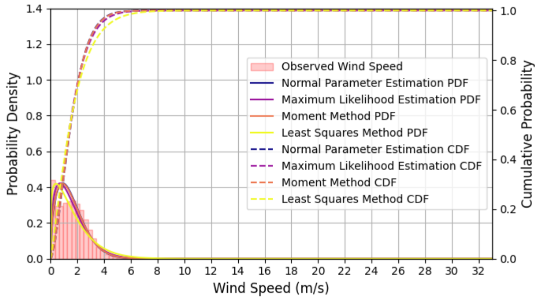
where y_i are the observed frequencies (the histogram values), z_i are the predicted frequencies from the fitted Weibull PDF and $\bar{y}$ are the mean of the observed frequencies.

This metric, often called "R^2 (distribution fit)" metric measures how well the Weibull curve matches the wind speed histogram. A value close to 1 indicates an excellent fit, while a value near 0 shows poor model accuracy.

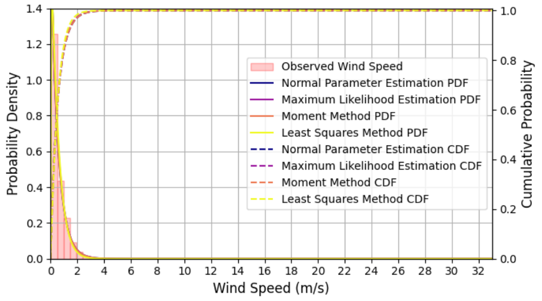
These methodologies (performance metrics with MAPE, RMSE, X^2, and R^2) ensure a comprehensive assessment of the statistical techniques, addressing challenges posed by missing or incomplete wind speed datasets.

3. RESULTS AND DISCUSSION

Observed Wind Speed PDF and CDF of the Weibull were plotted using a five-year dataset (2020–2024) from 14 AWS-TMD in provinces in Southern Thailand (Table 5). The histogram represents the observed wind speed data collected at 14 meteorological Stations. This histogram provides a visual representation of the frequency distribution of wind speeds at the station.



a. Nakhon Si Thammarat province (ID 89).



b. Trang province (ID 1084).

Fig. 2 – The PDF and CDF plotted from AWS-TMD data (fig. a.-b.)

Figure 2 shows PDF and CDF plots from AWS-TMD wind speed data across Nakhon Si Thammarat province and Trang province, highlighting wind variability and the fit of the Weibull model. Table 5 summarizes how data characteristics affect the performance of four estimation methods EMJ, MLE, MM, and LSM revealing each method's sensitivity to data quality.

Table 5

Report on data and impact on parameter estimation techniques.

Stations ID / Datasets	Method	Parameter		Validation Metrics			
		Shape (*k*)	Scale (*c*)	MAPE	RMSE	X^2	R^2
ID 55 76.81%	EMJ	1.096	0.934	0.000	0.088	0.000	0.890
	MLE	1.023	0.911	0.015	0.072	0.000	0.926
	MM	1.089	0.932	0.000	0.086	0.000	0.894
	LSM	1.032	0.900	1.501	0.073	0.000	0.924
ID 54 83.17%	EMJ	1.471	1.525	0.000	0.046	0.000	0.931
	MLE	1.453	1.525	0.131	0.045	0.000	0.933
	MM	1.449	1.522	0.000	0.045	0.000	0.933
	LSM	1.491	1.513	0.961	0.046	0.000	0.930
ID 88 78.21%	EMJ	1.189	1.251	0.000	0.069	0.000	0.875
	MLE	1.075	1.211	0.127	0.052	0.000	0.930
	MM	1.177	1.247	0.000	0.067	0.000	0.882
	LSM	1.038	1.209	0.995	0.046	0.000	0.943
ID 1083 30.50%	EMJ	1.420	1.522	0.000	0.038	0.000	0.943
	MLE	1.359	1.509	0.150	0.035	0.000	0.952
	MM	1.399	1.518	0.000	0.037	0.000	0.946
	LSM	1.318	1.513	0.674	0.033	0.000	0.955
ID 1086 29.47%	EMJ	1.138	0.894	0.000	0.008	0.000	0.998
	MLE	1.107	0.887	0.014	0.006	0.000	0.999
	MM	1.129	0.892	0.000	0.007	0.000	0.998
	LSM	1.116	0.880	1.001	0.007	0.000	0.999
ID 1085 30.72%	EMJ	1.249	0.759	0.000	0.018	0.000	0.996
	MLE	1.207	0.752	0.080	0.014	0.000	0.997

Stations ID / Datasets	Method	Parameter		Validation Metrics			
		Shape (k)	Scale (c)	MAPE	RMSE	X^2	R^2
	MM	1.234	0.757	0.000	0.016	0.000	0.996
	LSM	1.193	0.752	0.148	0.013	0.000	0.998
ID 89 84.14%	EMJ	1.550	1.782	0.000	0.054	0.000	0.850
	MLE	1.394	1.744	0.765	0.049	0.000	0.880
	MM	1.527	1.779	0.000	0.053	0.000	0.856
	LSM	1.199	1.796	5.416	0.045	0.004	0.896
ID 50 79.73%	EMJ	1.268	1.721	0.000	0.026	0.000	0.966
	MLE	1.188	1.691	0.189	0.019	0.000	0.982
	MM	1.253	1.716	0.000	0.024	0.000	0.970
	LSM	1.143	1.697	1.248	0.017	0.000	0.986
ID 103 73.93%	EMJ	1.419	2.541	0.000	0.025	0.000	0.942
	MLE	1.448	2.557	0.322	0.025	0.000	0.945
	MM	1.399	2.535	0.000	0.026	0.000	0.939
	LSM	1.539	2.537	1.186	0.024	0.000	0.950
ID 1084 27.61%	EMJ	0.963	0.535	0.000	0.023	0.000	0.991
	MLE	0.997	0.544	0.005	0.021	0.000	0.992
	MM	0.966	0.536	0.000	0.022	0.000	0.991
	LSM	1.066	0.533	4.583	0.016	0.001	0.995
ID 63 74.20%	EMJ	1.172	1.329	0.000	0.037	0.000	0.955
	MLE	1.112	1.308	0.028	0.030	0.000	0.971
	MM	1.161	1.325	0.000	0.036	0.000	0.959
	LSM	1.125	1.293	1.523	0.031	0.000	0.969
ID 85 76.49%	EMJ	1.546	1.738	0.000	0.034	0.000	0.943
	MLE	1.424	1.709	0.615	0.030	0.000	0.954
	MM	1.523	1.735	0.000	0.033	0.000	0.947
	LSM	1.258	1.752	4.218	0.032	0.003	0.948
ID 86 79.23%	EMJ	1.786	1.739	0.000	0.011	0.000	0.995
	MLE	1.734	1.732	0.212	0.010	0.000	0.995
	MM	1.761	1.738	0.000	0.010	0.000	0.995
	LSM	1.641	1.750	1.214	0.013	0.000	0.993
ID 87 84.10%	EMJ	1.594	2.148	0.000	0.010	0.000	0.985
	MLE	1.648	2.168	0.617	0.010	0.000	0.984
	MM	1.571	2.145	0.000	0.010	0.000	0.984
	LSM	1.895	2.128	2.003	0.017	0.001	0.957

The results in Table 5 show how four estimation methods, EMJ, MLE, MM, and LSM, perform across different provinces. Their impact on the Weibull parameters (*k* and *c*), as well as accuracy and fit metrics, is analyzed below.

Wind speed showed strong spatial variation, ranging from 0.35 m/s in Trang to 2.18 m/s in Songkhla, with a maximum of 32.98 m/s in Narathiwat. Data completeness also varied widely, from 27.61% in Trang to 84% in Nakhon Si Thammarat, influencing model reliability.

The EMJ method provided stable estimates where completeness exceeded 80%. For example, in Chumphon (k = 1.096, c = 0.934) and Ranong (k = 1.471, c = 1.525), EMJ achieved low RMSE and R^2 above 0.89. However, it was sensitive to missing data, leading to reduced accuracy below 70% completeness. The MM method produced similar *k* and *c* values but showed greater robustness with incomplete datasets.

The MLE method is well adapted under moderate completeness. In Phatthalung, MLE improved the fit compared to EMJ (R^2 = 0.982 vs. 0.966; RMSE = 0.019 vs. 0.026) while slightly lowering *k*. Minor MAPE errors (<1%) indicated good predictive consistency.

In contrast, the LSM method displayed high R^2 but poor parameter stability. For instance, in Nakhon Si Thammarat (k = 1.199, MAPE = 5.42%) and Trang (MAPE = 4.58%), LSM results suggested overfitting despite R^2 values exceeding 0.99. Overall, MLE and MM performed most reliably, while EMJ and LSM were more affected by data gaps and uneven completeness.

Table 6

Comparative Table of Estimation Methods

Method	Strengths	Weaknesses	Best Applied When	Suitability by Data Size
Normal or Empirical Method (EMJ)	Simple, stable across most provinces, good for complete datasets.	Sensitive to missing data may underestimate variability.	Baseline comparison of wind speed distributions.	Best for ≥70% completeness.
Maximum Likelihood (MLE)	Handles incomplete data well, provides low RMSE and strong R^2.	Slightly biased in low-completeness cases.	Estimation under moderate missing data conditions.	Reliable for 50–80% completeness.
Moment Method (MM)	Very robust, consistent across stations, low MAPE, stable parameters.	May smooth out extremes (underestimates gusts).	General-purpose method for uneven datasets.	Suitable for all dataset sizes (>30%).
Least Squares (LSM)	Achieves very high R^2 values, good for trend fitting.	Overfits incomplete data, produces high MAPE, unstable under low completeness.	Testing theoretical fits or small pilot datasets.	Best for >80% completeness only.

The EMJ and MM methods give stable and low-error results, making them suitable for general use. MLE offers better precision but may slightly increase MAPE, while LSM is highly variable and sensitive to missing data. In Southern Thailand, accurate wind assessment depends on both method and data completeness. MM works well even with incomplete data, EMJ is best with ≥70% completeness, and MLE suits 50–80%. LSM should only be used with >80% completeness. Provinces like Nakhon Si Thammarat, Narathiwat, and Ranong are ready for feasibility studies, while Trang, Krabi, and Phangnga provinces need more data. Overall, MM is the most consistent method, MLE balances accuracy and fit, and LSM should be used with caution.

4. **CONCLUSIONS**

The results show that both wind characteristics and dataset completeness had a major influence on parameter stability and model accuracy. Mean wind speeds varied widely. Dataset completeness also differed significantly. This variability directly affected the accuracy of the estimated Weibull parameters (k and c) and validation metrics. The EMJ and MM methods consistently produced stable estimates with 0% MAPE and high R^2 values above 0.94 when data completeness was sufficient, confirming their strong reliability. The MLE method demonstrated adaptability to moderate completeness levels (50–80%), reducing RMSE and often improving R^2, although it introduced small MAPE errors such as 0.32% in Songkhla provinces and 0.62% in Narathiwat provinces. In contrast, the LSM method displayed the highest sensitivity to missing data, producing the largest MAPE values, peaking at 5.42% in Nakhon Si Thammarat provinces and 4.58% in Trang provinces, despite achieving extremely high R^2 values (up to 0.999). These findings show that such high distribution fits may reflect overfitting rather than true accuracy.

In conclusion, the MM method proved most robust across all completeness levels, while MLE provided balanced accuracy and fit under moderate data conditions. EMJ remained effective for highly complete datasets, whereas LSM should be used cautiously and only when completeness exceeds 80%. Reliable wind assessment in southern Thailand, therefore, depends on both method selection and effective handling of missing data, providing essential guidance for future wind farm planning and renewable energy development.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

CHATCHAWICH CHAIHONG: Conceptualization, data collection, experimental work, initial draft preparation, literature review, and visualization.

JUNTAKAN TAWEEKUN: Supervision, project administration, conceptual guidance, review & editing.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support from the Energy Technology Program, Department of Interdisciplinary Engineering, Faculty of Engineering, Prince of Songkla University (PSU). This research was funded by the Engineering Graduate Scholarship (2023) and the Research Grant for Thesis (Fiscal Year 2025) from the Graduate School, PSU. The authors also appreciate the valuable wind data provided by the Thai Meteorological Department (TMD).

Received on 26 September 2025

REFERENCES

1. D. Kang, K. Ko, J. Huh, *Comparative study of different methods for estimating Weibull parameters: a case study on Jeju Island, South Korea*, Energies, **11**, *2*, pp. 356–356 (2018).
2. N. Aras, U. Erisoglu, H.D. Yıldızay, *Optimum method for determining Weibull distribution parameters used in wind energy estimation*, Pakistan Journal of Statistics and Operations Research, **16**, *4*, pp. 635–648 (2020).
3. S. Ali, S.-M. Lee, C.-M. Jang, *Statistical analysis of wind characteristics using Weibull and Rayleigh distributions in Deokjeok-do Island – Incheon, South Korea*, Renewable Energy, **123**, pp. 652–663 (2018).
4. T.P. Chang, *Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application*, Applied Energy, **88**, *1*, pp. 272–282 (2011).
5. F. Aristone et al., *Evaluation of Weibull parameters by different methods for farms*, Acta Scientiarum - Technology, **46**, *1*, pp. 1–10 (2024).
6. M.F. Zambak, C.I. Cahyadi, J. Helmi, T.M. Sofie, Suwarno, *Evaluation and analysis of wind speed with the Weibull and Rayleigh distribution models for energy potential using three models*, International Journal of Energy Economics and Policy, **13**, *2*, pp. 427–432 (2023).
7. A. Aziz, D. Tsuanyo, J. Nsouandele, I. Mamate, R. Mouangue, P. Elé Abiama, *Influence of Weibull parameters on the estimation of wind energy potential*, Sustainable Energy Research, **10**, *1*, pp. 1–18 (2023).
8. B.A. Çakmakçı, E. Hüner, *Evaluation of wind energy potential: a case study*, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, **44**, *1*, pp. 834–852 (2020).
9. A. Chumapan, P. Neranon, J. Taweekun, *Statistical analysis of wind resource assessment for different locations in South-Western Thailand*, Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, **118**, *2*, pp. 87–100 (2024).
10. Y.K. Chu, J.C. Ke, *Computation approaches for parameter estimation of Weibull distribution*, Mathematical and Computational Applications, **17**, *1*, pp. 39–47 (2012).
11. S. Kang, A. Khanjari, S. You, J.H. Lee, *Comparison of different statistical methods used to estimate Weibull parameters for wind speed contribution in a nearby offshore site, Republic of Korea*, Energy Reports, **7**, pp. 7358–7373 (2021).
12. A. Amin, M. Mourshed, *Weather and climate data for energy applications*, Renewable and Sustainable Energy Reviews, **192**, pp. 1–10 (2024).
13. H. Erol, A. Arslan, *Wind turbine pitch angle control with artificial neural networks*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **70**, *2*, pp. 235–240 (2025).
14. C. Popa, N.-S. Popa, F. Deliu, O. Cristea, I. Ciocioi, M.-O. Popescu, *Analysis of wind turbine power output via modeling, simulation, and validation*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **70**, *2*, pp. 175–180 (2025).
15. S. Bellarbi, A. Boufertella, *Maximum power wind extraction with feedback linearization control approach*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **67**, *3*, pp. 237–240 (2022).
16. ***Thai Meteorological Department, *Automatic weather system (AWS)*, Thai Meteorological Department, pp. 1–10 (2025).
17. A.S.D. Rey, I.C. Gil-García, M.S. García-Cascales, Á. Molina-García, *Online wind-atlas databases and GIS tool integration for wind resource assessment: a Spanish case study*, Energies, **15**, *3*, pp. 852–852 (2022).
18. ***DTU-Wind-Energy, *Global wind atlas*, pp. 1–10 (2025).
19. I. Kamdar, S. Ali, J. Taweekun, H.M. Ali, *Wind farm site selection using WAsP tool for application in the tropical region*, Sustainability, **13**, *24*, pp. 13718–13718 (2021).
20. C. Chaihong, J. Taweekun, *Techno-economic assessment of wind energy in urban environments: a case study in Pattaya, Thailand*, Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, **104**, *1*, pp. 169–184 (2023).
21. L. Wang, X. Zhang, *A novel method for studying the wind speed probability distribution and estimating the average wind energy density*, Engineering Research Express, **6**, *2*, pp. 1–10 (2024).
22. G.K. Gugliani, C. Ley, N. Nakhaei Rad, A. Bekker, *Comparison of probability distributions used for harnessing the wind energy potential: a case study from India*, Stochastic Environmental Research and Risk Assessment, **38**, *6*, pp. 2213–2230 (2024).
23. E.Y. Kombe, J. Muguthu, *Wind energy potential assessment of Great Cumbrae Island using Weibull distribution function*, Original Research Article Kombe and Muguthu, **2**, *2*, pp. 1–8 (2019).
24. T. Iida, *Identifying causes of errors between two wave-related data using performance metrics*, Applied Ocean Research, **148**, pp. 1–10 (2024).