WEIBULL PARAMETERS ESTIMATION FOR WIND SPEED PROBABILITY DISTRIBUTION IN KON DONG USING FIVE DIFFERENT NUMERICAL METHODS

ƯỚC TÍNH THAM SỐ WEIBULL CHO PHÂN PHỐI XÁC SUẤT TỐC ĐỘ GIÓ Ở KON DONG SỬ DỤNG NĂM PHƯƠNG PHÁP SỐ KHÁC NHAU

Nguyen Thi Hoai Thu[1,*], Pham Phong Ky[1]

DOI: http://doi.org/10.57001/huih5804.2025.118

ABSTRACT

Wind energy, a clean and sustainable renewable resource, plays a crucial role in advancing global sustainable development. Accurate wind speed estimation is essential for assessing the power output potential of wind turbines. This study models wind speed data from the Kon Dong wind farm using statistical methods based on the Weibull distribution. Five methods for estimating the Weibull shape and scale parameters were evaluated: the Energy Pattern Factor Method (EPFM), Empirical Methods by Lysen (EML) and Justus (EMJ), a hybrid EPFM-EMJ approach, and Method of Moments (MoM). Model performance was assessed using root mean square error (RMSE) under both non-seasonal and seasonal conditions. Results show that EPFM, EMJ, EPFM-EMJ, and MoM achieved good fits, with RMSE values ranging from 0.022 to 0.024 for non-seasonal data, 0.024 to 0.026 during the strong wind season, 0.005 in the rainy season, and 0.020 to 0.023 in the transitional season. In contrast, the EML method consistently produced the highest RMSE across all conditions, indicating the poorest fit. These findings highlight the effectiveness of EPFM, EMJ, EPFM-EMJ, and MoM in accurately estimating Weibull parameters for wind resource assessment.

Keywords: *Wind energy, Weibull parameters estimation, Numerical methods, Statistical analysis.*

TÓM TẮT

Năng lượng gió là một nguồn năng lượng tái tạo đóng vai trò quan trọng trong quá trình phát triển toàn cầu. Để đánh giá tiềm năng phát điện của các dự án năng lượng gió, cần ước lượng vận tốc gió 1 cách chính xác. Nghiên cứu này mô hình hóa dữ liệu tốc độ gió từ nhà máy điện gió Kon Dong bằng các phương pháp thống kê dựa trên phân bố Weibull. Năm phương pháp được đánh giá để ước tính các tham số hình dạng và tỉ lệ của phân bố Weibull bao gồm: hệ số mẫu năng lượng (EPFM), phương pháp thực nghiệm của Lysen (EML), phương pháp thực nghiệm của Justus (EMJ), phương pháp lai EPFM-EMJ và phương pháp Moment (MoM). Độ chính xác của các phương pháp được đánh giá thông qua sai số căn bậc hai trung bình bình phương (RMSE) trong trường hợp không phân mùa và có phân mùa. Kết quả cho thấy các phương pháp EPFM, EMJ, EPFM-EMJ và MoM đạt độ khớp tốt với RMSE dao động từ 0,022 đến 0,024 đối với dữ liệu không phân mùa; từ 0,024 đến 0,026 trong mùa gió mạnh; khoảng 0,005 trong mùa mưa; và từ 0,020 đến 0,023 trong mùa chuyển tiếp. Ngược lại, phương pháp EML cho kết quả kém nhất với RMSE cao trong cả hai trường hợp. Những kết quả này cho thấy các phương pháp EPFM, EMJ, EPFM-EMJ và MoM đạt độ chính xác cao trong việc ước lượng phân bố Weibull phục vụ đánh giá tiềm năng năng lượng gió.

Từ khóa: *Năng lượng gió, ước tính tham số Weibull, phương pháp số, phân tích xác suất.*

[1]PGRE. Lab., School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Vietnam

[*]Email: thu.nguyenthihoai@hust.edu.vn

Received: 25/2/2025

Revised: 15/5/2025

Accepted: 28/5/2025

1. INTRODUCTION

Climate change and the increasing frequency of extreme weather events are making the transition to renewable energy more important than ever. Among these renewable sources, wind power has consistently been recognized as one of the most significant, alongside solar power [1]. The growing importance of wind energy in the global energy mix necessitates accurate and reliable assessments of its potential, which are critical for the continued advancement of renewable energy technologies. One of the key aspects of wind energy assessment is the evaluation of wind speed variations, a task commonly performed using the Weibull distribution or, in some cases, its special variant, the Rayleigh distribution. These distributions are defined by two parameters: k, which shapes the distribution, and c, which scales the distribution.

The precise estimation of the k and c parameters in the Weibull distribution is crucial and has been extensively studied using a variety of numerical methods. Some popular numerical methods that can be mentioned include the energy pattern factor method (EPFM), empirical methods of Justus (EMJ), standard deviation method (STDM), graphical method (GM) or the empirical methods of Lysen (EML) [2-4]. For instance, Kapen et al. [5] conducted a comprehensive study employing ten distinct numerical methods to estimate these parameters, including EMJ, EML, the method of moments (MoM), GM, Mabchour's method (MMab), the energy pattern factor method (EPFM), the maximum likelihood method (MLM), the modified maximum likelihood method (MMLM), the equivalent energy method (EEM), and the alternative maximum likelihood method (AMLM). Their study focused on characterizing the wind energy potential of Hatiya Island in Bangladesh, demonstrating the versatility and effectiveness of these methods.

In this study, we selected five methods for estimating Weibull parameters, including EMJ, EML, EPFM, EPFM-EMJ, and MoM, to represent a diverse range of computational approaches while maintaining a balance between comprehensiveness and computational complexity. EMJ and EML are classical, straightforward methods widely applied in wind energy studies, suitable for datasets with stable distributions and minimal noise. EPFM is an improvement over empirical methods, aiming to enhance accuracy when data exhibit distributional deviations or contain outliers. EPFM-EMJ combines the approaches of EPFM and EMJ to leverage the strengths of both methods. MoM (Method of Moments) estimates Weibull parameters using statistical moments (mean and variance) of the data. It is known for its computational simplicity and ability to provide stable estimates, particularly effective for datasets of medium to large sizes. The selection of these five methods ensures diversity in estimation techniques, from empirical to statistical approaches, while avoiding excessive computational complexity that might arise from selecting too many methods and also ensuring that important trends are not overlooked by choosing only 2 - 3 methods. The effectiveness of these methods in fitting wind speed data will be rigorously evaluated using Root Mean Square Error (RMSE).

The main contributions of study can be summarized as follows:

1. Addressing the Gap in Wind Power Research in Vietnam: Despite Vietnam's significant wind power potential, research in this field remains limited. This study highlights the need for comprehensive wind analysis to support informed decision-making and attract investments in the sector.

2. Evaluating the Effectiveness of Numerical Methods for Weibull Estimation: One of the critical objectives of this research is to explore and compare various numerical methods for estimating Weibull distribution parameters, which are widely used in wind energy assessments. The methods examined in this study include EMJ, EML, EPFM, EPFM-EMJ, and MoM. The study provides a detailed evaluation of the effectiveness of these methods in accurately determine the Weibull parameters.

3. Data Preprocessing Through Seasonal Decomposition: To improve Weibull parameter estimation, this study introduces a novel data preprocessing approach that decomposes wind data into seasonal components. This enhances modeling accuracy; addresses data quality challenges and provides a practical framework for future wind energy assessments.

2. MATERIALS AND METHODS

2.1. Interquartile Range (IQR)

In this study, the Interquartile Range (IQR) method was employed to identify and manage outliers within the dataset. The IQR is a statistical measure that captures the dispersion of the middle 50% of the data, effectively reflecting the variability where the majority of the data points are concentrated. Specifically, it is defined as the range between the first quartile (Q1) and the third quartile (Q3), representing the spread of the central portion of the dataset.

The IQR method is particularly useful for assessing the variability of data because it provides insights into how tightly or broadly the central values are distributed. A larger IQR indicates that the middle 50% of the data is spread out over a wider range, suggesting greater variability. Conversely, a smaller IQR signifies that the central values are more closely clustered, indicating less variability.

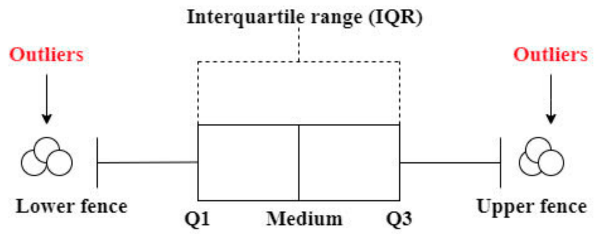

Figure 1. The Interquartile Range method

Outliers are identified using the IQR by calculating the upper and lower fences. Data points that exceed the upper fence or fall below the lower fence are considered outliers. This approach is advantageous because the IQR is relatively robust to the influence of outliers, helping to mitigate their impact on the analysis. Additionally, the IQR method does not rely on any assumptions regarding the underlying distribution of the data, such as normality, making it a flexible tool for outlier detection.

The upper and lower fences are determined using the following formulas:

$$\begin{aligned} \text{IQR} &= \text{Q3} - \text{Q1} \\ \text{Upperfence} &= \text{Q3} + 1.5 * \text{IQR} \\ \text{Lowerfence} &= \text{Q1} - 1.5 * \text{IQR} \end{aligned} \tag{1}$$

2.2. Weibull distribution

Weibull distribution, described in detail by Waloddi Weibull [6], is a popular statistical models used regularly for analyzing wind speed variations by using two functions, the probability density function (PDF) and the cumulative distribution function (CDF).

The Weibull distribution, characterized by two parameters k and c, can be illustrated by,

$$f(v) = \frac{k}{c}.\left(\frac{v}{c}\right)^{k-1}.e^{-\left(\frac{v}{c}\right)^k} \tag{2}$$

$$F(v) = 1 - e^{-\left(\frac{v}{c}\right)^k} \tag{3}$$

where f(v) is the Weibull distribution function, F(v) is the Weibull cumulative distribution function, v is the wind speed with the units in m/s, k is the shape parameter with dimensionless units, c is the scale parameter with the units in m/s.

2.3. Numerical Methods

2.3.1. Energy pattern factor method (EPFM)

EPFM is a technique grounded in the analysis of mean wind speeds [7], serving as a reliable approach for evaluating wind energy potential. This method facilitates the estimation of the key parameters, k and c, which characterize the wind speed distribution. These parameters are crucial for modeling the wind speed distribution accurately and are derived through the application of specific equations, as outlined below:

$$k = \frac{3.69}{(E_{pf})^2} \tag{4}$$

$$c = \frac{\bar{v}}{\Gamma\left(1 + \frac{1}{k}\right)} \tag{5}$$

$$E_{pf} = \frac{\overline{v^3}}{(\bar{v})^3} \tag{6}$$

$$\Gamma(t) = \int_0^\infty (e^{-x}x^{t-1})dx \tag{7}$$

where E_{pf} is the the energy pattern factor and Γ is the gamma function

2.3.2. Empirical Method of Justus (EMJ)

EMJ, introduced by Justus in 1977 [8], is a seminal approach in the field of wind energy analysis. His method leverages statistical measures, specifically the standard deviation (σ) and the mean wind speed ($\bar{v}$), to calculate the parameters k and c that define the Weibull distributions:

$$c = \frac{\bar{v}}{\Gamma\left(1 + \frac{1}{k}\right)} \tag{8}$$

$$k = \left(\frac{\sigma}{\bar{v}}\right)^{-1.086} \tag{9}$$

2.3.3. Method of Moments (MoM)

MoM is a method utilized by Justus et al. in 1977 [9]. Used by many researchers, this method utilized the mean wind speed $\bar{v}$ and the standard deviation σ to evaluate the Weibull parameters.

$$\sigma = c\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right]^{\frac{1}{2}} \tag{10}$$

$$\bar{v} = c\Gamma\left(1 + \frac{1}{k}\right) \tag{11}$$

2.3.4. Empirical method of Lysen (EML)

EML was introduced by Lysen in 1982 [10]. Although the method's approach to calculating the shape parameter k bears resemblance to the equation used in the EMJ, its calculation of the scale parameter c diverges, as it is determined through a distinct equation, as shown in Eq. 12:

$$c = \bar{v}\left(0.568 + \frac{0.433}{k}\right) - \frac{1}{k} \tag{12}$$

2.3.5. EPFM – EMJ

This method, proposed by Guenoukpati et al. [10], represents a hybrid approach that integrates elements from both the Energy Pattern Factor Method (EPFM) and the Empirical Method of Justus (EMJ). In this approach, the scale parameter c is calculated similarly to the procedure outlined in the EMJ method, as described in Eq. 8. However, the determination of the shape parameter k follows a distinct process, which is specified as follows:

$$k = \frac{1}{2}\left(1 + \frac{3.69}{E_{pf}^2} + \left(\frac{\sigma}{\bar{v}}\right)^{-1.086}\right) \quad (13)$$

2.4. Evaluation methods

To evaluate the efficiency of used methods for determining parameters of wind speed dataset distribution, Root Mean Square Error (RMSE) is used to validate the accuracy of the predicted wind speed distribution derived from the Weibull PDF function as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2} \quad (14)$$

Where:

N: The number of observations

x_i: The frequency of observations

y_i: The frequency of Weibull

3. DATASET

3.1. Dataset information

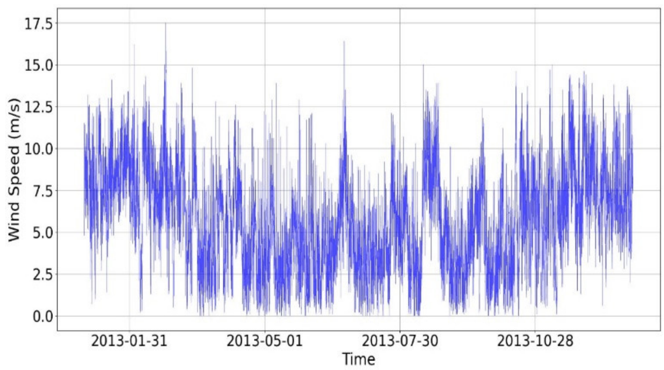

Figure 2. Original wind speed dataset from 2012-12-31 to 2013-12-31

In this study, the primary dataset utilized was derived from wind speed measurements collected at the Kon Dong station, located in Gia Lai province, Vietnam. The dataset encompasses a continuous time period starting from December 31, 2012, at 23:50:00, and extending through December 31, 2013, at 23:50:00. The data was recorded at ten-minute intervals, resulting in a comprehensive collection of 52572 individual data points.

3.2. Data preprocessing

The first step of the analysis involved thorough data cleaning to ensure the accuracy and reliability of the dataset [11]. Data points that showed little to no variation over time were removed, as they did not follow the overall trend of the data and could negatively impact the results. This initial process was essential for refining the dataset to focus on meaningful and relevant information.

Once the data was cleaned, the Interquartile Range (IQR) method was used to identify and remove outliers from the dataset. The IQR method, known for its effectiveness in detecting extreme values, was applied to ensure that the dataset was not influenced by irregular or unrepresentative data points. Figures 3 and 4 illustrate the overall process of the categorization and the usage of IQR, resulting in a cleaner and more reliable dataset.

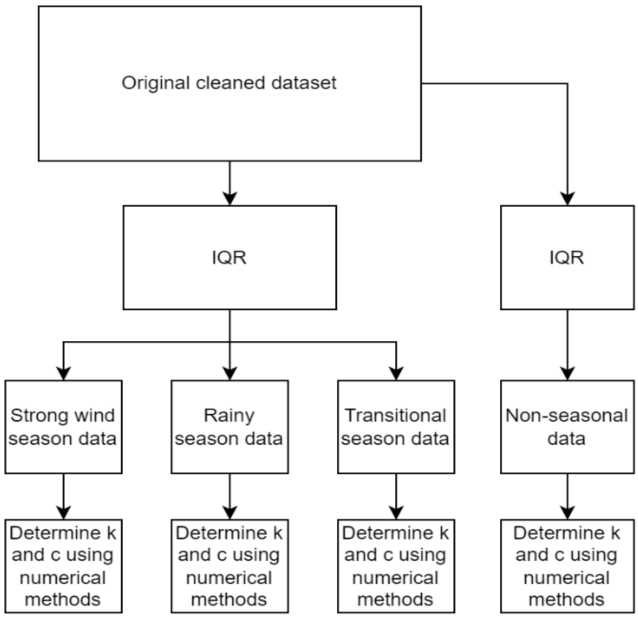

Figure 3. Parameters determination process

The Kon Dong area (Gia Lai) is located in the Central Highlands' tropical monsoon climate zone, characterized by two distinct seasons: the rainy season (May to October) and the dry season (November to April of the following year). After a preliminary analysis of statistical characteristics such as the mean, standard deviation, and probability distribution of wind speed by season, the results show a clear difference between the rainy and dry seasons. This indicates that separating the wind data by season is necessary for more accurate modeling of the Weibull distribution. Additionally, certain transitional periods between the two main seasons are categorized as transitional data to avoid introducing noise when analyzing the climatic characteristics of each season. Based on this, the dataset was divided into two main categories following IQR-based outlier removal: a seasonal dataset and a non-seasonal dataset. The seasonal dataset was further split into three distinct periods: the strong wind season, rainy season, and transitional season, each representing different climatic conditions. The non-seasonal dataset, meanwhile, retained the original cleaned data without accounting for seasonal differences. Figure 4 illustrates the dataset after being split into three seasons.

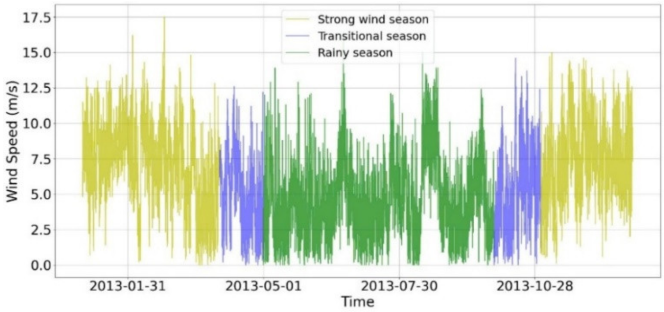

Figure 4. Processed dataset after being split into three seasons

4. RESULTS AND DISCUSSION

4.1. Comparison of Five Numerical Models Using Non-Seasonal Data

In this case study, the parameters k and c of the Weibull PDF were meticulously determined based on the wind speed data collected between December 2022 and December 2023. Table 1 presents the computed values of k and c using five distinct numerical methods applied to a non-seasonal dataset. Figure 5 provides a visual representation of these parameters to enhance clarity. Additionally, the RMSE for the CDF of each model using the corresponding k and c value are also illustrated in Table 1.

An examination of the data in Table 1 reveals that the k parameter remains relatively stable across all five numerical methods. The estimated shape parameter k across the five methods ranges from 2.051 to 2.109, with a mean value of approximately 2.074. Although slight variations exist, all methods produce k values within a narrow band of ±0.03 around the mean, indicating overall consistency in the estimation of wind speed distribution shape. Conversely, the c parameter exhibits noticeable variation. The EPFM, EMJ, EPFM-EMJ, and MoM methods yield closely aligned c values around 6.44, whereas the EML method produces a significantly lower c value of 3.95. This variation is also evident in Figure 5, which highlights the poor performance of the EML model in comparison to the other methods.

Figure 5 presents the PDF and the CDF for non-seasonal data. In the PDF, the EML method exhibits a sharper and higher peak compared to the other methods, deviating significantly from the overall trend of the actual data, particularly at the right tails. In contrast, the EPFM, EMJ, EPFM-EMJ, and MoM methods demonstrate better alignment with the actual data, showing lower peaks and more accurate distributions. A similar pattern is observed in the CDF figure, where the EML method again performs the worst, while the other four methods, EPFM, EMJ, EPFM-EMJ, and MoM, show comparable and more consistent performance.

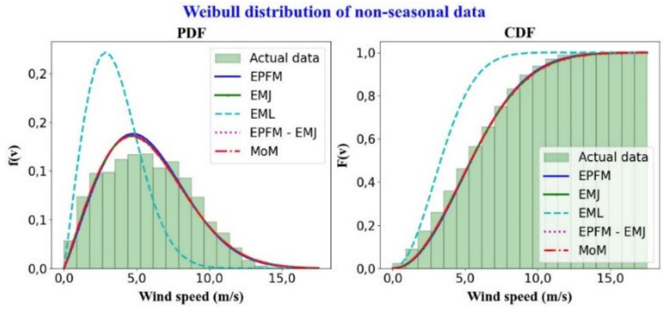

Figure 5. Weibull distribution for non-seasonal data of five numerical methods

Table 1. Weibull parameters and metric values for five numerical methods using non-seasonal data

Numerical method	Weibull parameters		Evaluation metrics
	k	c	RMSE
EPFM	2.109	6.440	0.024
EMJ	2.062	6.438	0.022
EML	2.062	3.952	0.254
MoM	2.051	6.438	0.022
EPFM-EMJ	2.086	6.439	0.023

Table 1 presents the Weibull parameters and RMSE values for five numerical methods applied to non-seasonal data. Among them, MoM and EMJ yields the lowest RMSE (0.022), indicating the best agreement with the observed data. EPFM and EPFM-EMJ also show relatively low RMSE values of 0.024 and 0.023, respectively. In contrast, EML shows a significantly higher RMSE of 0.2538, reflecting a poor fit. These findings suggest that the four models EPFM, EMJ, EPFM-EMJ and MoM have relatively similar performance with good fitting for the data while EML performed poorly with high RMSE value.

4.2. Comparison between Seasonal and Non-Seasonal Approaches

This study evaluates the accuracy of five numerical models and compares non-seasonal and seasonal analyses. In the non-seasonal approach, IQR is applied to the entire dataset, while in the seasonal approach, it is applied separately to each of the three seasons. Numerical methods in the seasonal analysis yield distinct k and c parameters for each season, whereas the non-seasonal analysis produces a single set for the overall wind distribution. Table 2, similar to Table 1, presents the k and c values for each season obtained from the dataset.

Table 2. Weibull parameters and metric values for five numerical methods using seasonal data

Seasons	Numerical method	Weibull parameters		Evaluation metrics
		k	c	RMSE
Strong wind season	EPFM	2.868	8.132	0.026
	EMJ	2.919	8.126	0.025
	EML	2.919	4.850	0.367
	MoM	2.914	8.127	0.024
	EPFM-EMJ	2.894	8.129	0.025
Rainy season	EPFM	1.864	4.987	0.005
	EMJ	1.869	4.988	0.005
	EML	1.869	3.006	0.242
	MoM	1.857	4.987	0.005
	EPFM-EMJ	1.867	4.988	0.005
Transitional season	EPFM	2.008	5.688	0.023
	EMJ	1.976	5.687	0.021
	EML	1.976	3.462	0.245
	MoM	1.964	5.686	0.020
	EPFM-EMJ	1.992	5.688	0.022

As can be seen from Table 2, the results obtained for the seasonal approach are relatively similar to Table 1. MoM remains the best-performing method, with RMSE values of 0.024, 0.005, and 0.020 for the strong wind season, rainy season, and transitional season, respectively. This is followed by EMJ, EPFM-EMJ, EPFM, and finally, EML. Additionally, the k and c values exhibit similar trends to those observed in Table 1. The k values for all methods remain relatively stable at approximately 2.90, 1.86, and 1.98 for the strong wind season, rainy season, and transitional season, respectively. Meanwhile, the c values for EML, similar to those in Table 1, are distinctly different from the other methods, with values of 4.85, 3.01, and 3.46, while the other methods cluster around 8.12, 4.98, and 5.68 for the respective seasons. Consequently, the EML method continues to exhibit the poorest performance among all models, as further demonstrated by the PDF and CDF visualizations in Figures 6, 7, and 8. This result can be explained by the fundamental difference in the EML approach, which maximizes the likelihood based on all data points, making it more sensitive to outliers and irregularities. In contrast, EPFM, EMJ, and MoM rely on summary statistics (e.g., moments or quantiles), which smooth the data and reduce sensitivity. Since these methods share similar assumptions and focus on aggregate features, their estimates are more consistent, whereas EML's dependence on raw data can lead to larger deviations, especially in the estimation of c.

To facilitate a more detailed comparison, the results from the non-seasonal analysis are also disaggregated by season, mirroring the seasonal model. From there, the RMSE of the best-performing model, MoM, is used for comparison, as illustrated in Table 3. The results from Table 3 indicate a significant disparity between the RMSE values of the seasonal and non-seasonal datasets using MoM, with the seasonal data showing markedly lower error rates. Specifically, for the strong wind season, the RMSE of the seasonal dataset is 0.024, whereas the non-seasonal dataset yields much higher values of 0.189. Similarly, for the transitional season, the RMSE and MSE for the seasonal data is 0.020, compared to 0.064 for the non-seasonal data. During the rainy season, the seasonal dataset has an RMSE of 0.005, while the non-seasonal dataset exhibits substantially higher values of 0.140.

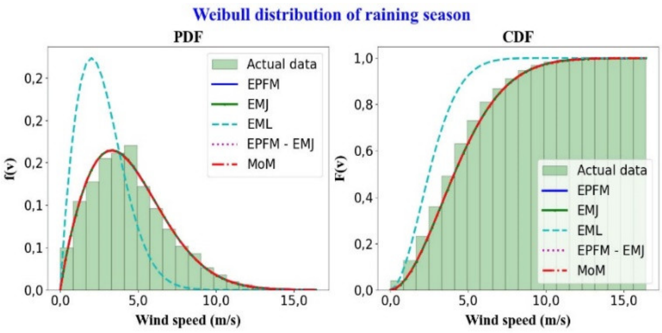

Figure 6. Weibull distribution for raining season of five numerical methods

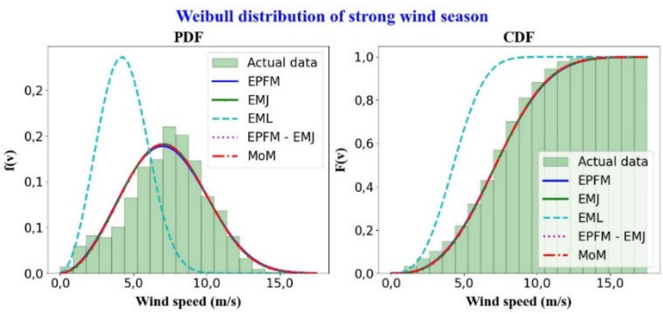

Figure 7. Weibull distribution for strong wind season of five numerical methods

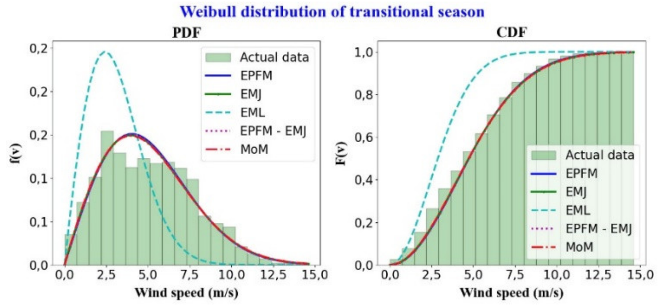

Figure 8. Weibull distribution for transitional season of five numerical methods

Table 3. Comparison between a seasonal and non-seasonal approach using the MoM method

Evaluation metrics	Data properties	Seasons		
		Strong wind season	Transitional season	Rainy season
RMSE	With season	0.024	0.020	0.005
	Without season	0.189	0.064	0.140

These findings clearly demonstrate that segmenting the dataset into seasonal components, doing preprocessing and applying numerical models tailored to each season significantly enhances model performance. The reduction in error rates underscores the value of this approach, validating the positive impact of season-specific modelling on the accuracy and reliability of wind speed predictions.

5. CONCLUSION

This paper presents a statistical analysis of one year of wind speed data collected in Kon Dong, Gia Lai, using the Weibull distribution. To estimate the distribution parameters k and c, five numerical methods were applied: EMJ, EML, EPFM, EPFM-EMJ, and MoM. The performance of each method was evaluated based on RMSE under two conditions: non-seasonal and seasonal. In the non-seasonal case, the EMJ, EPFM, EPFM-EMJ, and MoM methods all achieved RMSE values below 0.025, indicating good fit, while EML produced a significantly higher RMSE of 0.254. Under the seasonal condition, all methods except EML again yielded similar and low RMSE values around 0.024, with EML lagging behind due to a higher error. These results suggest that, with the exception of EML, the tested methods performed consistently well across both conditions. Therefore, EMJ, EPFM, EPFM-EMJ, and MoM are recommended for future wind speed analyses in similar contexts.

REFERENCES

[1]. World Bank, Going Global*: Expanding Offshore Wind to Emerging Markets*. [Online], 2019. Available: https://documents1.worldbank.org/curated/en/716891572457609829/pdf/Going-Global-Expanding-Offshore-Wind-To-Emerging-Markets.pdf

[2]. Hussain, A. Haider, Z. Ullah, M. Russo, G. M. Casolino, B. Azeem, "Comparative Analysis of Eight Numerical Methods Using Weibull Distribution to Estimate Wind Power Density for Coastal Areas in Pakistan," *Energies*, 16, 3, 1515, 2023. doi: 10.3390/en16031515.

[3]. S. Vega-Zuñiga, J. G. Rueda-Bayona, A. Ospino-Castro, "Evaluation of Eleven Numerical Methods for Determining Weibull Parameters for Wind Energy Generation in the Caribbean Region of Colombia," *MMEP*, 09, 01, 194-199, 2022. doi: 10.18280/mmep.090124.

[4]. Y. W. Koholé, R. H. T. Djiela, F. C. V. Fohagui, T. Ghislain, "Comparative study of thirteen numerical methods for evaluating Weibull parameters for solar energy generation at ten selected locations in Cameroon," *Cleaner Energy Systems*, 4, 100047, 2023. doi: 10.1016/j.cles.2022.100047.

[5]. P. Tiam Kapen, M. Jeutho Gouajio, D. Yemélé, "Analysis and efficient comparison of ten numerical methods in estimating Weibull parameters for wind energy potential: Application to the city of Bafoussam, Cameroon," *Renew. Energy*, 159, 1188-1198, 2020. doi: 10.1016/j.renene.2020.05.185.

[6]. W. Weibull, *A Statistical Theory of the Strength of Materials*. Generalstabens litografiska anstalts förlag, 1939.

[7]. T. C. Carneiro, S. P. Melo, P. C. M. Carvalho, A. P. D. S. Braga, "Particle Swarm Optimization method for estimation of Weibull parameters: A case study for the Brazilian northeast region," *Renew. Energy*, 86, 751-759, 2016. doi: 10.1016/j.renene.2015.08.060.

[8]. C. G. Justus, W. R. Hargraves, A. Mikhail, D. Graber, "Methods for Estimating Wind Speed Frequency Distributions," *J. Appl. Meteorol.*, 17, 3, 350-353, 1978. http://www.jstor.org/stable/26178009

[9]. E. H. Lysen, *Introduction to Wind Energy: Basic and Advanced Introduction to Wind Energy with Emphasis on Water Pumping Windmills*. in Publication CWD. CWD- Consultancy Services Wind Energy Developing Countries, 1983. [Online]. Available: https://books.google.com.vn/books?id=8fqbswEACAAJ

[10]. A. Guenoukpati, A. A. Salami, M. K. Kodjo, K. Napo, "Estimating Weibull Parameters for Wind Energy Applications using Seven Numerical Methods: Case studies of three costal sites in West Africa," *Int. J. Renew. Energy Dev.*, 9, 2, 217-226, 2020. doi: 10.14710/ijred.9.2.217-226.

[11]. C. Fan, M. Chen, X. Wang, J. Wang, B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery from Building Operational Data," *Front. Energy Res.*, 9, 2021. doi: 10.3389/fenrg.2021.652801.

THÔNG TIN TÁC GIẢ

Nguyễn Thị Hoài Thu, Phạm Phong Kỳ

Phòng thí nghiệm Hệ thống điện và Năng lượng tái tạo, Khoa Điện, Trường Điện - Điện tử, Đại học Bách khoa Hà Nội