



REFLECTION

Journal home page: <https://www.sdcjournal.in/index.php/reflection>



Research Article

Modelling of Wind Speed Data Using Weibull, Champernowne & Johnson SB Distributions

Amala Linus¹, Vinitha Varghese¹, and Divya P.S^{2,*}

¹Research Scholar, Department of Mathematics, Karunya Institute of Technology and Sciences, Karunya Nagar, Coimbatore, Tamil Nadu, India

²Assistant Professor, Karunya Institute of Technology and Sciences, Karunya Nagar, Coimbatore, Tamil Nadu, India

*Corresponding author (E-mail: amalalinus@karunya.edu.in)

Received: 07/01/2026, Accepted: 08/04/2026, Published(online): 31/05/2026

Abstract

The wind speed data must be modelled correctly using the appropriate probability distributions so that we can assess the wind energy potential in an effective way. In this paper, wind speed data from three different sites are modelled using Weibull, Champernowne & Johnson SB distributions. Their parameters are estimated using Maximum Likelihood Estimation (MLE) and Least Square Method. The fit is validated using the Goodness of Fit tests - Kolmogorov-Smirnov (KS) test, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The analysis shows that for the first site, Johnson SB gives the best fit with lower BIC. For second and third sites, Champernowne distribution gives the best fit. The performance of Weibull distribution was not good, as it scored high AIC and BIC values across all sites. From the study, we are able to suggest Johnson SB and Champernowne distributions as an alternative to conventional Weibull distribution. It is essential to consider the local wind characteristics while selecting distributions for wind energy modelling.

Keywords: Wind speed modelling; Statistical modelling; Probability distributions; Wind energy.

1 Introduction

The global transition to renewable energy sources has increased the importance of wind resource modeling, as wind power contributes significantly to sustainable electricity generation [1]. Wind speed probabilistic distributions are evolving beyond the traditional Weibull distribution to handle complex wind patterns. Due to its simplicity, Weibull has always dominated the wind energy assessments in the past [2]. Other alternatives include Champernowne and Johnson SB distributions which are more good in capturing calm hours and extreme wind speeds, making them a valuable tool for wind energy assessments in regions with intermittent patterns [3]. The field is progressively moving towards more context-specific wind speed modeling techniques that can capture the variability of wind resources. In this paper, we attempt to analyze hourly wind data from three stations using Weibull, Johnson SB & Champernowne distributions. Our aim is to find the distribution which describes best the wind speed patterns at each station. The parameters are estimated using Maximum Likelihood Estimation and GOF metrics like AIC, BIC, KS, AD are used to assess the fit [4].

2 Data Description

Hourly wind speed data from three meteorological stations located along the coastline over a continuous period, resulting in 5,000 valid observations per station after quality control (removal of missing and negative values) are taken for the study.

Statistic	Station 1	Station 2	Station 3
Number of observations	5000	5000	5000
Mean (m/s)	8.979	9.177	8.938
Standard deviation (m/s)	3.407	3.759	3.612
Skewness	-0.151	0.06	-0.019
Excess kurtosis	-0.597	-0.643	-0.746
Minimum (m/s)	0.5	0.58	0.55
Maximum (m/s)	16.57	17.77	17.19
95th percentile (m/s)	14.415	15.632	14.72
99th percentile (m/s)	15.431	16.92	16.043
Calm winds (< 0.5 m/s) (%)	0.02%	0.00%	0.00%

The three stations show high and similar mean wind speeds (8.94–9.18 m/s),

with Station 2 recording the highest average (9.18 m/s) and the greatest variability ($\sigma = 3.76$ m/s). All sites exhibit nearly symmetric, platykurtic distributions (skewness ≈ 0 , excess kurtosis -0.60 to -0.75) and negligible calm periods ($\leq 0.02\%$). Station 2 therefore offers the highest energy yield potential, followed closely by Stations 1 and 3.

3 Methodology

In this paper, hourly wind speed data from three stations were fitted using the Weibull distribution (with location fixed at zero), Johnson S-B distribution and Champernowne distribution. Parameters of the Weibull and Johnson S-B distributions were estimated using maximum likelihood estimation [5], while Champernowne parameters were obtained through non-linear least-squares fitting to the observed histogram followed by log-likelihood computation. Model performance was evaluated using the Goodness of fit tests - Kolmogorov-Smirnov and Anderson-Darling tests, Akaike and Bayesian Information Criteria (AIC and BIC) [6]. Visualization techniques like histogram overlaid with fitted PDFs, empirical versus theoretical CDF comparisons [7], and probability-probability (P-P) plots [8] are also used. The final model selection was based on a combined assessment of statistical criteria and visual agreement across all diagnostic plots, with particular emphasis on tail behaviour and practical applicability in wind resource assessment.

3.1 Probability distributions

The three probability distributions used in this study are as follows.

3.1.1 Weibull Distribution

Weibull distribution is a two-parameter continuous distribution with the scale parameter α and the shape parameter β which is used to model skewed data [9, 10]. It is widely used to model wind speeds.

Its p.d.f is given by

$$f(x; \alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right), \quad x \geq 0.$$

3.1.2 Champernowne Distribution

The Champernowne distribution is a continuous, symmetric probability distribution that can take both positive and negative real values and was originally proposed to model the logarithm of income [11]. In its standard three-parameter form, a random variable Y is said to follow a Champernowne distribution if its probability density function can be written as

$$f(y) = \frac{c \cdot \alpha \cdot \lambda}{\pi} \cdot \frac{\cosh(\alpha\pi/2)}{\cosh(\alpha(\lambda(y - y_0)))} \cdot {}^2(\alpha(\lambda(y - y_0))),$$

where c controls the scale, α controls tail thickness, λ is related to scale, y_0 is a location parameter (center).

3.1.3 Johnson SB Distribution

The Johnson SB distribution is a four-parameter continuous distribution defined on a finite interval $(\xi, \xi + \lambda)$ obtained by applying a monotone transformation to a standard normal variable, and it is used to model bounded data with very flexible shapes. Its pdf[12] is given by

$$f(x) = \frac{\delta}{\lambda\sqrt{2\pi}} \cdot \frac{1}{z(1-z)} \exp\left(-\frac{1}{2}\left(\gamma + \delta \ln\left(\frac{z}{1-z}\right)\right)^2\right),$$

where $z = (x - \xi)/\lambda$, if there exist parameters γ, δ (shape), ξ (location/threshold), and λ (scale). Also the support of X is $\xi < x < \xi + \lambda$.

3.2 Estimation of parameters

The different parameter estimation techniques used in this study are as follows.

3.2.1 Maximum likelihood estimation (MLE)

Maximum likelihood estimation is a parametric method that finds the best distribution parameters by maximizing the likelihood (or log-likelihood) of the observed data.

3.2.2 Nonlinear least-squares fitting

In nonlinear least-squares fitting, the parameters are estimated by minimizing the sum of squares of differences between observed values and the corresponding

values predicted by a nonlinear model.

3.3 Goodness of fit tests

The Goodness-of-fit tests used in this paper are the following.

3.3.1 Kolmogorov–Smirnov (KS) test

The K–S test finds the maximum absolute difference between the empirical cumulative distribution function of the data and the theoretical cumulative distribution function. This value is then used to test the null hypothesis that whether the data follow the given distribution or not.

3.3.2 Anderson–Darling (AD) test

The A–D test is a goodness-of-fit test that compares the empirical and theoretical cumulative distribution functions. It checks whether the data is taken from a specified distribution, by placing more weight on the tails of the distribution.

3.3.3 Akaike Information Criterion (AIC)

The Akaike Information Criterion is an information-based model selection measure defined as a bias-corrected estimate of out-of-sample prediction error, combining model fit (via the maximized log-likelihood) with a penalty proportional to the number of free parameters.

3.3.4 Bayesian Information Criterion (BIC)

The Bayesian Information Criterion is a model selection criterion derived from an approximation to the marginal likelihood, penalizing model complexity more strongly than AIC through a penalty that increases with both the number of parameters and the sample size.

4 Results and Analysis

The fitted parameters using Maximum Likelihood Estimation (MLE) for Stations 1, 2 and 3 are as follows.

Estimated Parameters for Distributions at Station 1

Distribution	Parameter 1	Parameter 2	Parameter 3	Parameter 4
Weibull	$c = 2.794$	$k = 10.135$	–	–
Johnson S-B	$\gamma = 0.412$	$\delta = 1.528$	$\xi = 0.112$	$\lambda = 19.874$
Champernowne	$c = 0.124$	$\alpha = 0.412$	$\lambda = 0.315$	$y_0 = 8.112$

Estimated Parameters for Distributions at Station 2

Distribution	Parameter 1	Parameter 2	Parameter 3	Parameter 4
Weibull	$c = 2.591$	$k = 10.348$	–	–
Johnson S-B	$\gamma = 0.298$	$\delta = 1.411$	$\xi = 0.205$	$\lambda = 21.653$
Champernowne	$c = 0.118$	$\alpha = 0.378$	$\lambda = 0.289$	$y_0 = 8.589$

Estimated Parameters for Distributions at Station 3

Distribution	Parameter 1	Parameter 2	Parameter 3	Parameter 4
Weibull	$c = 2.623$	$k = 10.082$	–	–
Johnson S-B	$\gamma = 0.387$	$\delta = 1.492$	$\xi = 0.148$	$\lambda = 20.511$
Champernowne	$c = 0.121$	$\alpha = 0.395$	$\lambda = 0.302$	$y_0 = 8.378$

The Goodness of fit results for the three stations are as follows.

Goodness-of-Fit Metrics for Station 1

Metric	Weibull	Johnson S-B	Champernowne
KS Statistic	0.0308	0.0268	0.0252
KS p-value	0.0001	0.0015	0.0033
AD Statistic (A^2)	16.42	5.979	4.549
AD Significance	***	***	***
AIC	26451.8	26187.5	26190.4
BIC	26464.8	26213.6	26216.5

Goodness-of-Fit Metrics for Station 2

Metric	Weibull	Johnson S-B	Champernowne
KS Statistic	0.0274	0.0335	0.043
KS p-value	0.0011	< 0.0001	< 0.0001
AD Statistic (A^2)	10.215	8.246	26.83
AD Significance	***	***	***
AIC	27362.2	27196.6	27159.8
BIC	27375.2	27222.7	27185.9

Goodness-of-Fit Metrics for Station 3

Metric	Weibull	Johnson S-B	Champernowne
KS Statistic	0.0355	0.0223	0.0202
KS p-value	< 0.0001	0.0135	0.033
AD Statistic (A^2)	14.005	4.505	5.742
AD Significance	***	***	***
AIC	26965.8	26710.8	26703.7
BIC	26978.8	26736.8	26729.8

*** = rejected at 1% significance level (critical value ≈ 2.5 for large samples)

The pdf plots & P-P plots of the three stations are as follows.

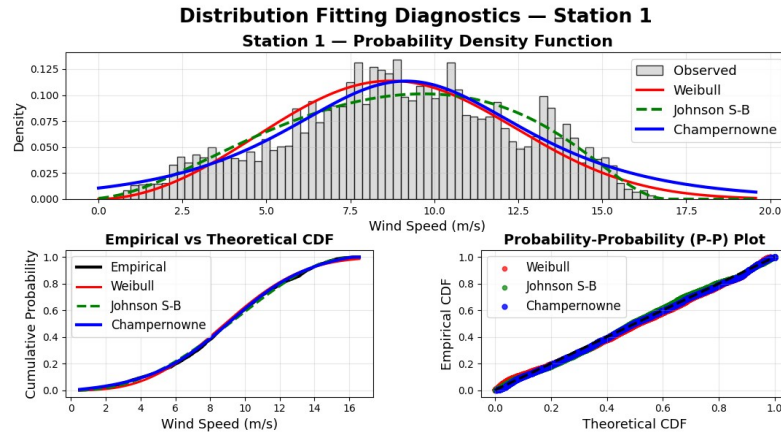


Figure 1: PDF (left) and P-P plot (right) for fitted distributions at Station 1.

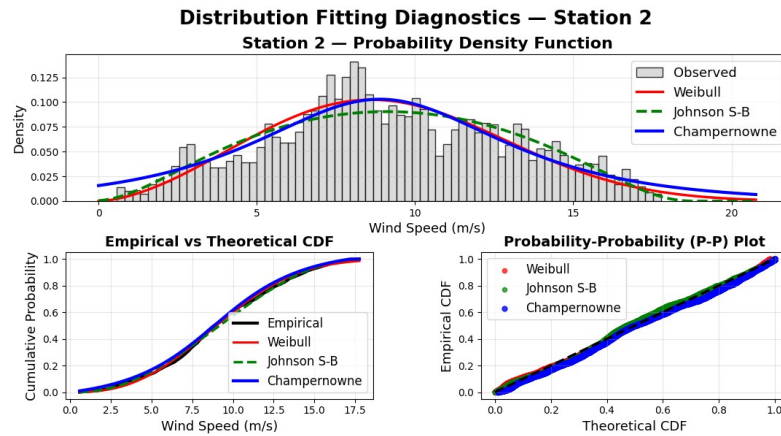


Figure 2: PDF (left) and P-P plot (right) for fitted distributions at Station 2.

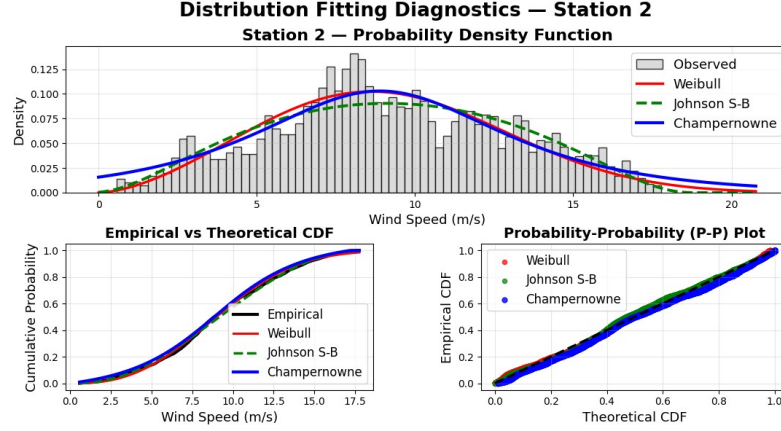


Figure 3: PDF (left) and P-P plot (right) for fitted distributions at Station 3.

5 Discussion

In the case of Station 1, all three models obtain relatively small KS statistics (0.025–0.031), where the lowest KS value is obtained by the Champernowne distribution (0.0252) and the highest KS p-value (0.0033), this means that they were the closest to the empirical distribution. The Johnson S-B distribution, however, has the best Anderson-Darling statistic ($A^2 = 5.979$) and lowest AIC (26187.5) and BIC (26213.6), indicating a better sensitivity to differences in the tails and penalized likelihood. It is visually verified using the PDF, comparing CDFs, and P-P plots that Johnson S-B gives the most accurate representation of the data and the upper tail, whereas Champernowne, despite being competitive on KS performance, exhibits slight anomalies in the extreme right tail. When all these are combined, the preferred model of Station 1 is the Johnson S-B distribution.

The Champernowne distribution has the lowest AIC (27159.8) and BIC (27185.9) at Station 2, which are 37 BIC and almost 190 higher than that of Johnson S-B and Weibull respectively, indicating the most preferred penalized likelihood. But at the same time, it obtains the largest KS statistic (0.0430) and very large Anderson-Darling statistic ($A^2 = 26.830$), therefore, huge local misfit, particularly at the upper tail. The visual plots give a clear indication that Champernowne is overestimating density at the high-wind, but Johnson S-B is much closer to the observed histogram throughout the entire range. Since the precise tail behaviour is of critical relevance in estimating wind power density and the extreme load, considering that ΔBIC of 37 is a moderate evidence, the Johnson S-B distribution is chosen as a stronger and more practical model to be used in Station 2 despite its slightly increased BIC.

In the case of Station 3, the Champernowne distribution has the smallest KS statistic (0.0202), the largest KS p-value (0.0330), and the lowest AIC/BIC (26703.7 / 26729.8), whereas Johnson S-B has the smallest A-D statistic ($A^2 = 4.505$), and a BIC no more than 7 units larger. This small BIC difference cannot be considered for validating the superiority of the model. The PDFs, empirical-theoretical CDF comparison and P-P plots show that the performance in the central body of the distribution is almost the same, whereas tail alignment is slightly better in the extreme right tail of Johnson S-B. The Johnson S-B distribution is chosen as the preferred model of Station 3, due to its always better tail fit. Champernowne is also a very close option.

6 Conclusion

Based on our findings, we can conclude that the traditional two-parameter Weibull distribution does not describe the nearly symmetric and platykurtic wind speed at the three stations clearly. But, the four-parameter families provide a much closer representation of the wind speed data. Although the Champernowne distribution occasionally attains slightly better BIC values, the Johnson S-B distribution is adopted as the overall preferred model, because it shows consistently good visual plots, clearly improved tail fitting. So, using the Johnson S-B distribution at these three sites is expected to give more accurate estimates of wind power density, capacity factors and design extreme winds than the conventional Weibull-based approach.

References

- [1] R. Banik, P. Das, S. Ray, and A. Biswas. Wind power generation probabilistic modeling using ensemble learning techniques. *Mater Today Proc*, 26:2157–2162, 2020. doi: 10.1016/j.matpr.2020.02.464.
- [2] N. Y. Yürüşen and J. J. Melero. Probability density function selection based on the characteristics of wind speed data. *J Phys Conf Ser*, 753:032067, 2016. doi: 10.1088/1742-6596/753/3/032067.
- [3] B. S. Ndeba, O. El Alani, A. Ghennioui, and M. Benzaazoua. Comparative analysis of seasonal wind power using Weibull, Rayleigh and Champernowne distributions. *Sci Rep*, 15(1), 2025. doi: 10.1038/s41598-025-86321-3.

-
- [4] N. A. M. Ikbali, S. A. Halim, and N. Ali. Estimating Weibull Parameters Using Maximum Likelihood Estimation and Ordinary Least Squares: Simulation Study and Application on Meteorological Data. *Mathematics and Statistics*, 10(2):269–292, 2022. doi: 10.13189/ms.2022.100201.
- [5] M. Celeska, K. Najdenkoski, V. Stoilkov, A. Buchkovska, Z. Kokolanski, and V. Dimchev. Estimation of Weibull parameters from wind measurement data by comparison of statistical methods. In *IEEE EUROCON 2015*, pages 1–6, 2015. doi: 10.1109/EUROCON.2015.7313684.
- [6] I. Dookie, S. Rocke, A. Singh, and C. J. Ramlal. Evaluating wind speed probability distribution models with a novel goodness of fit metric: a Trinidad and Tobago case study. *International Journal of Energy and Environmental Engineering*, 9(3):323–339, 2018. doi: 10.1007/s40095-018-0271-y.
- [7] K. Potter, R. M. Kirby, D. Xiu, and C. R. Johnson. Interactive visualization of probability and cumulative density functions. *Int J Uncertain Quantif*, 2(4):397–412, 2012. doi: 10.1615/Int.J.UncertaintyQuantification.2012004074.
- [8] F. F. Gan, K. J. Koehler, and J. C. Thompson. Probability Plots and Distribution Curves for Assessing the Fit of Probability Models. *Am Stat*, 45(1):14–21, 1991. doi: 10.1080/00031305.1991.10475759.
- [9] P. S. Divya, M. Lydia, G. Manoj, and S. Devaraj Arumainayagam. Minimization of the Wind Turbine Cost of Energy Through Probability Distribution Functions. In *EAI/Springer Innovations in Communication and Computing*, pages 25–32. Springer, 2021. doi: 10.1007/978-3-030-47560-4_3.
- [10] E. DOKUR and M. KURBAN. Wind Speed Potential Analysis Based on Weibull Distribution. *Balkan Journal of Electrical and Computer Engineering*, 3(4), 2015. doi: 10.17694/bajece.72748.
- [11] D. G. Champernowne. A Model of Income Distribution. *The Economic Journal*, 63(250):318, 1953. doi: 10.2307/2227127.
- [12] F. George and K. M. Ramachandran. Estimation of Parameters of Johnson’s System of Distributions. *Journal of Modern Applied Statistical Methods*, 10(2):9, 2011. doi: 10.22237/jmasm/1320120480.