JISSE

ISSN: 2636-4425

Journal of International Society for Science and Engineering

Vol. 6, No. 2, 47-51 (2024)

JISSE

E-ISSN:2682-3438

A Direct Iterative Technique for Weibull Parameters Estimation Optimized for lowest Root Mean Square Error

Mahmoud El-Bayoumi

Mechanical Engineering Department, Engineering and Renewable Energy Research Institute, National Research Centre, Egypt

A R T I C L E I N F O

Article history:
Received:03-10-2024
Accepted:08-10-2024
Online:09-10-2024

Keywords:
Weibull Distribution
Parameter estimation
Optimization
Iteration
RMSE

A B S T R A C T

Classical Weibull distribution (Weibull minima) is gaining momentum due to its employment in wind speed estimation for the growing renewable energy field. The main difficulty of Weibull distribution is that it is hard to accurately estimate its parameters for a set of wind speed data. Much research has been involved in this task with varying complexity, success, and accuracy. Statistical analysis is usually used to compare the accuracy of proposed Weibull parameters estimation methods. As the Weibull parameters of a set of data are not normally estimated, for simulation, more than once, this research considered the processing time to be of far less importance than accuracy. In this research, efforts were made to make the best use of the current PC processing power to achieve the best possible accuracy of the Weibull parameters. In the current research, an iterative technique is proposed to directly estimate the Weibull parameters that yield the best results with the most employed statistical analysis; least Root Mean Square Errors of estimation (RMSE). The results of the new technique were compared to that of the Maximum likelihood method (MLM), and a considerable improvement in accuracy, 9.3 % was achieved.

1. Introduction

For simulation purposes, there was a need for a technique that accepts a typical set of measurements as input and delivers a randomized set of data that exhibits the same probability density as the original measurements. Several distributions were developed for this purpose. Amongst these distributions is Weibull distribution which was developed by Swedish mathematician W. Weibull [1] in 1939.

The standard Weibull distribution employs a set of measured input data to produce a Probability Density Function (PDF) curve that was found to match to a high degree the probability distribution of many natural phenomena [1] such as wind speeds data [2] required for current work. This PDF is based on two parameters, a scale parameter and a shape parameter. Throughout research, scale and shape parameters were assigned different letters; a and b, b and k, c and k, λ and k, α and β, α and γ, η and β, and μ and σ respectively. In the current research, they will be called c and k, respectively due to their popularity. The scale and shape parameters are meant to configure Weibull's probability density function (PDF). For Weibull distribution, probability of wind speed equal to v is given in equation (1).

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

Due to different patterns of wind speed data, there were challenges in finding a consistent match of Weibull's scale and shape parameters [3]. This resulted in continuous research efforts to develop more consistent techniques for this purpose. At the moment the most famous techniques of Weibull parameters estimation are maximum likelihood method (MLM) [4],[5],[6], graphical method (GM) [7],[8], power density method (PDM) [7],[9], moment method (MM) [7], empirical method of Justus (EMJ) [10], modified maximum likelihood method (MMLM) [4],[5],[6], empirical method of Lysen (EML) [8],[10], alternative maximum likelihood method (AMLM) [8], equivalent energy method (EEM) [7],[8],[11], energy pattern factor method (EPFM) [8], modified energy pattern factor method (MEPFM) [7],

* Mahmoud El-Bayoumi, Mechanical Engineering Department, National Research Centre, Cairo, Egypt, +2 01151886211, el_bayoumi@hotmail.com

mabchour method (MMab) [8], WAsP method (WM) [4], l-moment estimation method (LMOM) [7], energy pattern factor method (EPF) [10], and percentile estimators (PE) [7].

To examine the accuracy of the match between estimated data based on a set of Weibull scale and shape parameters and original measured data, different statistical analysis techniques are normally employed; root mean square error (RMSE) [5],[6],[12], regression coefficient (R^2) [6],[12], chi-square (X2) [5],[12], mean absolute percentage error (MAPE) [5], Kolmogorov–Smirnov test (KS) [13], relative percentage error (RPE) [14], mean percentage error (MPE) [7],[15], correlation coefficient (R) [14], relative root mean square error (RRMSE) [10], index of agreement (IA) [7], mean absolute bias error (MABE) [10], relative bias (RB) [7],[16], and best bin size (B) [7].

Different Weibull parameters estimation techniques were found to achieve different ranking using different statistical analysis technique. They also ranked differently with different sets of data. The unsatisfactory results led to ongoing research efforts to establish more sound Weibull parameter estimation methodologies [17].

Maximum likelihood method could be considered the most important methodology for the estimation of Weibull parameters due to a number of reasons. The first is that while it is the second most researched methodology, it ranks the first in its results' accuracy with regard to lower RMSE, [7], which is the most employed method of evaluation. The second is its results' excellent coefficient of determination (R^2), the second most employed method of evaluation, [7].

MLM evaluation of Weibull parameters is rather straightforward methodology, however it requires iteration. In MLM the Weibull parameters are evaluated using equation (2) and equation (3). Where v_i is the wind speed and n is the number of data points.

$$k = \left[\frac{\sum_{i=1}^{n} v_i^k \ln v_i}{\frac{1}{n}\sum_{i=1}^{n} v_i^k} - \frac{\sum_{i=1}^{n} \ln v_i}{n}\right]^{-1} \tag{2}$$

$$c = \left[\frac{1}{n}\sum_{i=1}^{n} v_i^k\right]^{1/k} \tag{3}$$

The root mean square error (RMSE) is the most employed measure of the difference between real measurement data and its modeled predictions. The lower the value of RMSE the higher the conformity of the predicted data to measured one. RMSE = 0 means that both data sets are typical. RMSE is evaluated using equation (4), where y_i and $\hat{y}_i$ are values of measured and predicted data point number i, respectively.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

The coefficient of determination R^2 is usually employed to investigate the dependency of predicted data points' changes on measured data points' changes. The closer the values of R^2 to 1, the more dependant the data are, hence the better the prediction is. R^2 is evaluated using equation (5), where $\bar{y}$ is average of measured data.

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y_i})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y_i})^2} \tag{5}$$

The current work aims at developing an iterative approach to estimate parameters of the standard Weibull distribution that would directly yield Weibull parameters optimized for the lowest RMSE and comparing its results with the most researched Weibull parameters estimation techniques; MLM methodology with its established accuracy.

2. The new approach

2.1. Analysis of the problem

This work is based on hypotheses that the Weibull parameters estimation is challenged by two main problems. The first problem leads to the dependency of different techniques accuracy on the set of input data. It is caused by the techniques trying to convert the discrete randomness of the original input data into a sort of standard mathematical function. Different mathematical functions would represent different sets of input data with different accuracies, leading to accuracy uncertainties [3].

The second problem has to do with the interaction between the scale and the shape parameters. The scale parameter c modifies the PDF scale, Figure 1, but at the same time, it alters the shape of the PDF to a lesser degree. Moreover, the shape parameter k modifies the PDF shape, Figure 2, but alters the scale of the PDF to a lesser degree as well. Moreover, the effect of the parameters on the Weibull PDF is not linear as the effect of change of both c and k parameters diminishes as the parameters value increases as shown in Figure 1 and Figure 2. This would allow significantly different sets of scale and shape parameters to yield close Weibull PDF curves, which would complicate finding the optimal set.

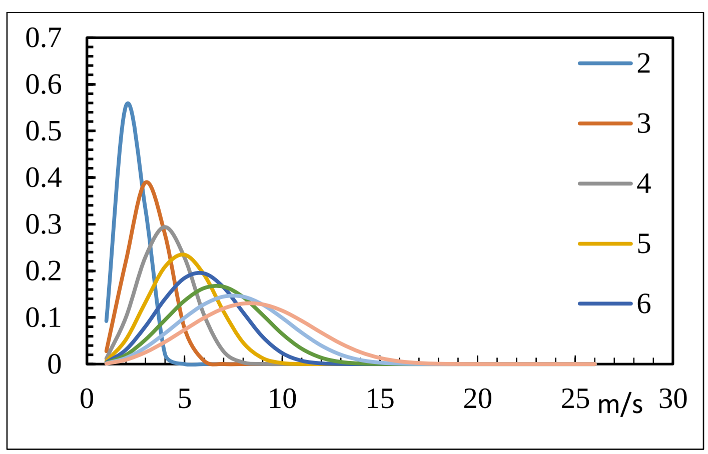

Figure 1: Effect of scale parameter c values (2 to 9) on the PDF of Weibull function at shape parameter k = 3

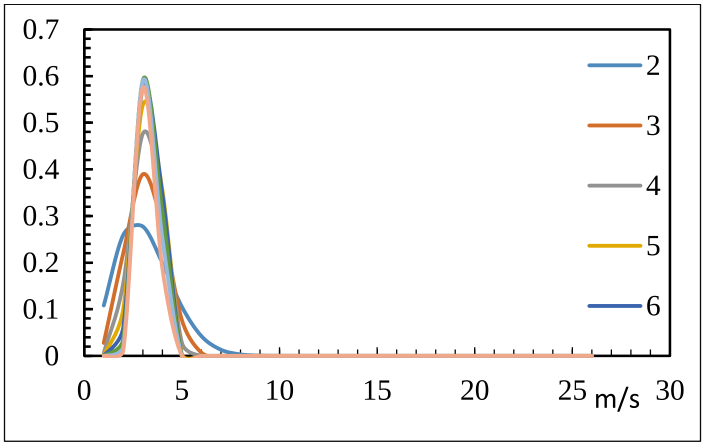

Figure 2: Effect of shape parameter k values (2 to 9) on the PDF of Weibull function at scale parameter c = 3

2.2. The iteration technique

Iteration techniques have the capacity to reach the best possible solution while making good use of the processing capacity of the modern computers. They also will achieve their objective of reaching the required optimal solution, even for non-conformal sets of data. On the other hand, iteration techniques would consume more time depending on the employed approach. However, Weibull parameters are found once for the set of data and then used forever. Hence accuracy of parameters estimation highly overweight longer time of parameters estimation. Moreover, the iteration times for the current case are usually within seconds.

To find the required optimal parameters the current study developed a direct approach for finding the required parameters. This approach starts by deciding the statistical analysis method for deciding the optimality of the required parameters. For the current research RMSE was selected due to its capacity to evaluate closeness of matching, hence being the most employed measure of Weibull parameters accuracy [7]. The second step makes use of the current immense processing power of modern computers to develop an iteration technique that would find the scale and shape parameter pair resulting in the absolute minimum RMSE.

The technique approaches the optimization of the two parameters in an alternating fashion. Starting with a set of initial values the iteration fixes the shape parameter and changes the scale parameter to find, to the nearest integer value, the scale parameter value that would achieve the lowest RMSE, and then do the same to the shape parameter. In the successive rounds the resolution is increased by a factor of 10 in each round. Accordingly, four rounds of iteration would yield the parameters resulting in the lowest RMSE with a resolution of ± 0.001.

The algorithm of the dedicated program required to apply the current technique is given in the following steps:-

1 – Input wind speed data

2 – Input initial c and k values that is larger than 1

3– Set initial search unit (1 is typical)

4 – Input required resolution of the answer (e.g 0.0001)

5 – Calculate RMSEs at current k and current c, c minus 1 unit and c plus 1 unit

6 – Compare RMSEs to find the direction (minus or plus) of minimal RMSE. If RMSE at c is the lowest move to step 8

7 – Advance in the minimum RMSE direction until the c value achieves the lowest RMSE (neither the minus nor the plus).

8 – Calculate RMSEs at the achieved c value and current k, k minus 1 unit, and k plus 1 unit.

9 – Compare RMSEs to find the direction (minus or plus) of minimal RMSE. If RMSE at k is the lowest move to step 11

10 – Advance in the minimum RMSE direction until the k value achieves the lowest RMSE (neither the minus nor the plus).

11 – If required resolution is achieved move to step 14.

12 – Change the value of the 1 unit (current resolution) to one-tenth of its value (increase iteration resolution)

13 – Go to step 5.

14 – The program prints c, k, and achieved RMSE.

15 – Program End

2.3. Testing of the developed approach

The current work is supposed to produce, by definition, the Weibull distribution parameters c and k pair that exhibits the lowest possible RMSE. However, for more rigor testing the results were compared with results obtained by MLM Weibull parameters estimation methodology, known for its excel with RMSE. For extra verification of the results, the coefficient of determination R^2 was evaluated for both the results of the work and the MLM-based results. R^2 is a statistical measure of how well the predicted regression results approximate the real data set. Accordingly, the planned test should be sufficient to demonstrate the capacity of the developed approach.

3. Results and discussion

For the validation of the current work, a dedicated program, based on the previously described algorithm, has been developed in the MATLAB ® environment version R2009a. A full year of wind speed measurements was employed. The results of the developed technique are finally compared with the results of the MATLAB function dedicated to the same purpose of evaluating Weibull's scale and shape parameters of a data set; wblfit(). The wblfit() function, employing the maximum likelihood method (MLM), is one of the most widely employed Weibull parameter evaluation methods. MLM is also known for producing Weibull parameters estimation with the lowest RMSE [7]. The function wblpdf() was also employed for recreation of data from evaluated Weibull parameters.

3.1. The employed data

The employed data is comprised of hourly measurements of a full year (2009) of wind speeds collected at Ras Sidr region, located in the Sinai Peninsula on the Gulf of Suez of the Red Sea in Egypt.

3.2. The Results

The developed program was used to evaluate Weibull's parameters. Through iteration, the c and k parameters pair with the lowest RMSE within resolution of 0.0001 was found in less than 0.25 seconds on a home laptop. The MATLAB function wblfit() was used to evaluate Weibull parameters using the MLM technique known for its high accuracy, especially when evaluated using RMSE and R^2.

The MATLAB's MLM results of the Ras Sidr data showed that a pair of Weibull scale and shape parameters of 9.2838 and 1.9051, respectively, was suitable to reproduce the PDF of Ras Sidr data with RMSE of 33.4506, Table 1. On the other hand, the developed iterative program evaluated the Weibull scale and shape parameters as 9.5177 and 1.8895, respectively, with an RMSE of 30.3474. Accordingly, the developed approach in the current test surpassed the MLM methodology, known for high accuracy, with 9.3%. The histogram of Ras Sidr wind speed was plotted against MLM predicted data and new approach data, derived for the Weibull parameters using MATLAB function wblpdf(). The results, as shown in Figure 3, show a better visual fit of the developed approach results.

The results were further test by evaluating the coefficient of determination R^2 using a specially developed MS Excel spreadsheet. The results, as shown in Table1, demonstrate the developed approach achieved R^2 results closer to 1 which is the ultimate outcome of the test. The developed approach scored R^2 = 0.985382 while the MLM achieved R^2 = 0.987801. The improvement of the new approach result's closeness to R^2 = 1 when compared to that of the MLM exhibiting an improvement of 16.5%. RMSE results of the new approach coupled with improvements in R^2 should be sufficient to prove the improvement of the new approach, against the more established MLM methodology.

Finally, iteration methodology should not put off users as modern computers exhibit huge processing powers that are not usually recognized. Also, a carefully sculptured algorithm and code would further unleash the processing power of the computer. Moreover, having to process the data once and use the estimated Weibull parameters results thousands of times for years to come should emphasize the need to allow more time for a more accurate Weibull's parameter estimation process. This is further substantiated with the current approach and code processing time of less than a quarter of a second.

Table 1: Results of the new approach and MLM methodology

	MATLAB's MLM	New approach	Improvement
c	9.2838	9.5177	NA
k	1.9051	1.8895	NA
RMSE	33.4506	30.3474	9.3%
R^2	0.985382	0.987801	16.5%

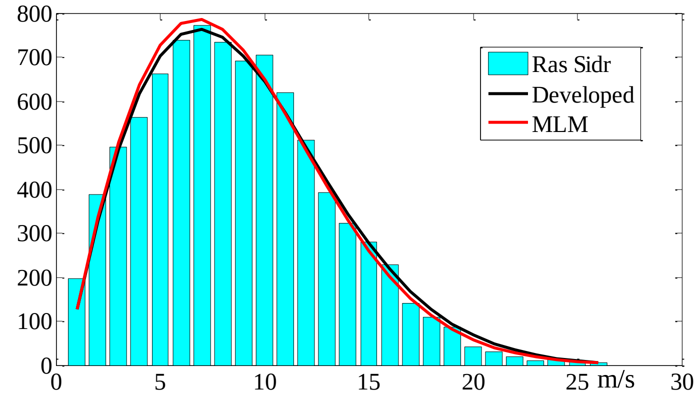

Figure 3: Histogram of Ras Sidr wind speed data compared with the Weibull's data generated by the developed approach and by the MATLAB's MLM

4. Conclusions

In the current work, assessment of problems facing Weibull parameters estimation techniques were discussed and an alternative approach was recommended. The new approach was developed from principles in the MATLAB environment. A detailed algorithm was introduced and the required code was developed.

The new approach takes an iterative direct approach to reach for the Weibull parameters that are optimized for the lowest possible RMSE of the given set of data.

The new approach was test against a full year of hourly measured wind speed data at Ras Sidr site in Egypt. The same data was evaluated using MLM methodology as carried out by the wblfit() function of MATLAB R2009a. The new results were also evaluated against the coefficient of determination R^2 and their results were compared. The new approach outperformed the MLM. I showed an improvement of 9.3% lower RMSE as well as 16.5% improvement in closeness to R^2 = 1. These results should demonstrate the soundness of the current approach to contribute to the Weibull parameters estimation field.

5. References

[1] I. Y. F. Lun and J. C. Lam, “A study of Weibull parameters using long-term wind observations,” *Renewable Energy*, vol. 20, no. 2, pp. 145–153, Jun. 2000, doi: 10.1016/S0960-1481(99)00103-2.

[2] I. Usta, “An innovative estimation method regarding Weibull parameters for wind energy applications,” *Energy*, vol. 106, pp. 301–314, Jul. 2016, doi: 10.1016/j.energy.2016.03.068.

[3] M. A. Saeed, Z. Ahmed, J. Yang, and W. Zhang, “An optimal approach of wind power assessment using Chebyshev metric for determining the Weibull distribution parameters,” *Sustainable Energy Technologies and Assessments*, vol. 37, p. 100612, Feb. 2020, doi: 10.1016/j.seta.2019.100612.

[4] A. Gautam, V. Warudkar, and J. L. Bhagoria, “A comparative analysis of wind characteristics for distinct terrains of India,” *Sustainable Energy Technologies and Assessments*, vol. 52, p. 102089, Aug. 2022, doi: 10.1016/j.seta.2022.102089.

[5] H. R. Alsamamra, S. Salah, J. A. H. Shoqeir, and A. J. Manasra, “A comparative study of five numerical methods for the estimation of Weibull parameters for wind energy evaluation at Eastern Jerusalem, Palestine,” *Energy Reports*, vol. 8, pp. 4801–4810, Nov. 2022, doi: 10.1016/j.egyr.2022.03.180.

[6] F. G. Akgül, B. Şenoğlu, and T. Arslan, “An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution,” *Energy Conversion and Management*, vol. 114, pp. 234–240, Apr. 2016, doi: 10.1016/j.enconman.2016.02.026.

[7] S. Kang, A. Khanjari, S. You, and J.-H. Lee, “Comparison of different statistical methods used to estimate Weibull parameters for wind speed contribution in nearby an offshore site, Republic of Korea,” *Energy Reports*, vol. 7, pp. 7358–7373, Nov. 2021, doi: 10.1016/j.egyr.2021.10.078.

[8] P. Tiam Kapen, M. Jeutho Gouajio, and 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,” *Renewable Energy*, vol. 159, pp. 1188–1198, Oct. 2020, doi: 10.1016/j.renene.2020.05.185.

[9] M. Khalid Saeed, A. Salam, A. U. Rehman, and M. Abid Saeed, “Comparison of six different methods of Weibull distribution for wind power assessment: A case study for a site in the Northern region of Pakistan,” *Sustainable Energy Technologies and Assessments*, vol. 36, p. 100541, Dec. 2019, doi: 10.1016/j.seta.2019.100541.

[10] K. Mohammadi, O. Alavi, A. Mostafaeipour, N. Goudarzi, and M. Jalilvand, “Assessing different parameters estimation methods of Weibull distribution to compute wind power density,” *Energy Conversion and Management*, vol. 108, pp. 322–335, Jan. 2016, doi: 10.1016/j.enconman.2015.11.015.

[11] A. H. Shaban, A. K. Resen, and N. Bassil, “Weibull parameters evaluation by different methods for windmills farms,” *Energy Reports*, vol. 6, pp. 188–199, Feb. 2020, doi: 10.1016/j.egyr.2019.10.037.

[12] C. F. de Andrade, H. F. Maia Neto, P. A. Costa Rocha, and M. E. Vieira da Silva, “An efficiency comparison of numerical methods for determining Weibull parameters for wind energy applications: A new approach applied to the northeast region of Brazil,” *Energy Conversion and Management*, vol. 86, pp. 801–808, Oct. 2014, doi: 10.1016/j.enconman.2014.06.046.

[13] A. Saboor, H. S. Bakouch, and M. Nauman Khan, “Beta Sarhan–Zaindin modified Weibull distribution,” *Applied Mathematical Modelling*, vol. 40, no. 13, pp. 6604–6621, Jul. 2016, doi: 10.1016/j.apm.2016.01.033.

[14] W. S. A. Hasan, A. S. M. Hassan, and M. A. Shukri, “Assessing the performance of several numerical methods for estimating Weibull parameters for Wind Energy Applications: A case study of Al-Hodeidah in Yemen,” *Energy Reports*, vol. 10, pp. 2725–2739, Nov. 2023, doi: 10.1016/j.egyr.2023.09.081.

[15] Y. A. Kaplan, “Performance assessment of Power Density Method for determining the Weibull Distribution Coefficients at three different locations,” *Flow Measurement and Instrumentation*, vol. 63, pp. 8–13, Oct. 2018, doi: 10.1016/j.flowmeasinst.2018.07.004.

[16] T. C. Carneiro, S. P. Melo, P. C. M. Carvalho, and A. P. de S. Braga, “Particle Swarm Optimization method for estimation of Weibull parameters: A case study for the Brazilian northeast region,” *Renewable Energy*, vol. 86, pp. 751–759, Feb. 2016, doi: 10.1016/j.renene.2015.08.060.

[17] M. Sumair, T. Aized, S. A. R. Gardezi, S. U. ur Rehman, and S. M. S. Rehman, “A novel method developed to estimate Weibull parameters,” *Energy Reports*, vol. 6, pp. 1715–1733, Nov. 2020, doi: 10.1016/j.egyr.2020.06.017.