Journal of Wind Engineering
and Industrial Aerodynamics 85 (2000) 75–84

www.elsevier.com/locate/jweia

Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis

J.V. Seguro[a,*,1], T.W. Lambert[b]

[a]Department of Mechanical Engineering, Colorado State University, Fort Collins, CO, USA
[b]T.W. Lambert Consulting Inc., Calgary, AB, Canada

Received 17 July 1998; received in revised form 13 July 1999

Abstract

Three methods for calculating the parameters of the Weibull wind speed distribution for wind energy analysis are presented: the maximum likelihood method, the proposed modified maximum likelihood method, and the commonly used graphical method. The application of each method is demonstrated using a sample wind speed data set, and a comparison of the accuracy of each method is also performed. The maximum likelihood method is recommended for use with time series wind data, and the modified maximum likelihood method is recommended for use with wind data in frequency distribution format. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Wind energy; Weibull distribution; Maximum likelihood method

1. Introduction

The Weibull distribution is a two-parameter function commonly used to fit the wind speed frequency distribution. This family of curves has been shown to give a good fit to measured wind speed data [1]. The Weibull function provides a convenient representation of the wind speed data for wind energy calculation purposes. It is

* Correspondence address: 500 Remington St. # 7, Fort Collins, CO 80524 - USA. Tel.: + 1-970-484-9563; fax: + 1-970-491-3827.

E-mail address: js441789@engr.colostate.edu (J.V. Seguro)

[1] Ph. D. Candidate and ASME, CRES, ASES Member.

PII: S0167-6105(99)00122-1

important to note that the analysis presented here does not consider extreme wind speed analysis, for this see [2].

Before sufficiently powerful computers were widely available, the preferred method of calculating the Weibull parameters was a graphical technique which entailed generating the cumulative wind speed distribution, plotting it on special Weibull graph paper, and drawing a line of best fit. This procedure is now commonly implemented by performing a linear regression on a computer. A more accurate and robust approach, however, is given by the maximum likelihood method. The purpose of this paper is to demonstrate that the maximum likelihood method is a more suitable computer-based method for estimating the Weibull parameters. Both methods are presented and demonstrated in this paper, as are a variation of the maximum likelihood method and a qualitative method for estimating the parameters using the average wind speed.

2. The Weibull distribution

This family of curves is widely used in statistical analysis. In wind energy analysis it is used to represent the wind speed probability density function, commonly referred to as the wind speed distribution. The Weibull distribution function is given by

$$P(v < v_i < v + \mathrm{d}v) = P(v > 0)\left(\frac{k}{c}\right)\left(\frac{v_i}{c}\right)^{k-1}\exp\left[-\left(\frac{v_i}{c}\right)^{k}\right]\mathrm{d}v, \tag{1}$$

where c is the Weibull scale parameter, with units equal to the wind speed units, k is the unitless Weibull shape parameter, v is wind speed, v_i is a particular wind speed, $\mathrm{d}v$ is an incremental wind speed, $P(v < v_i < v + \mathrm{d}v)$ is the probability that the wind speed is between v and $v + \mathrm{d}v$ and, $P(v > 0)$ is the probability that the wind speed exceeds zero.

Eq. (1) and the other equations in this paper that refer to probability can be applied equally well whether probability is interpreted as relative (fractional or percent) or absolute (number of data points). For example, $P(v > 0)$ in Eq. (1) can be interpreted as the fractional probability that the wind speed exceeds zero or the number of hours per year that the wind speed exceeds zero.

The cumulative distribution function is given by

$$P(v < v_i) = P(v \geqslant 0)\left\{1 - \exp\left[-\left(\frac{v_i}{c}\right)^{k}\right]\right\}, \tag{2}$$

where $P(v < v_i)$ is the probability that the wind speed is less than v_i, and $P(v \geqslant 0)$ is the probability that the wind speed equals or exceeds zero.

The two Weibull parameters and the average wind speed are related by

$$\bar{v} = c \cdot \Gamma\left(1 + \frac{1}{k}\right), \tag{3}$$

where $\bar{v}$ is the average wind speed and $\Gamma(\,)$ is the gamma function.

3. Wind speed data

Measured wind speed data are commonly available in time-series format, in which each data point represents either an instantaneous sample wind speed or an average wind speed over some time period. An example of such data (giving hourly averages over a 24 h period) is given in Table 1. In some instances, wind speed data may instead be available in frequency distribution format. In this format, the frequency with which the wind speed falls within various ranges (bins) is given. An example of such data is given in Table 2. The methods described in the following section can be used to estimate the Weibull parameters given wind speed in either time-series or frequency distribution format.

4. Determination of Weibull parameters

Three methods of estimating the parameters of the Weibull wind speed distribution are presented: two variations of the maximum likelihood method as well as the popular graphical method.

4.1. The maximum likelihood method

The Weibull distribution can be fitted to time-series wind data using the maximum likelihood method as suggested by Stevens and Smulders [3]. The shape factor k and

Table 1
Wind speed data in time-series format

Hour	Speed (m/s)	Hour	Speed (m/s)	Hour	Speed (m/s)
1	3.3	9	5.2	17	6.5
2	3.8	10	6.7	18	4.2
3	4.2	11	6.8	19	4.3
4	3.3	12	6.8	20	3.7
5	2.8	13	5.7	21	4.0
6	3.0	14	8.3	22	2.8
7	4.0	15	9.2	23	3.7
8	2.7	16	9.3	24	3.3

Table 2
Wind speed data in frequency distribution format

Wind Speed (m/s)	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10
Frequency (%)	2	7	9	15	20	17	8	1	0	0

the scale factor c are estimated using the following two equations:

$$k = \left(\frac{\sum_{i=1}^{n} v_i^k \ln(v_i)}{\sum_{i=1}^{n} v_i^k} - \frac{\sum_{i=1}^{n} \ln(v_i)}{n} \right)^{-1}, \tag{4}$$

$$c = \left(\frac{1}{n} \sum_{i=1}^{n} v_i^k \right)^{1/k}, \tag{5}$$

where v_i is the wind speed in timestep i and n is the number of nonzero wind speed data points.

Eq. (4) must be solved using an iterative procedure ($k = 2$ is a suitable initial guess), after which Eq. (5) can be solved explicitly. Care must be taken to apply Eq. (4) only to the nonzero wind speed data points.

4.2. The modified maximum likelihood method

When wind speed data are available in frequency distribution format, a variation of the maximum likelihood method can be applied. The Weibull parameters are estimated using the following two equations:

$$k = \left(\frac{\sum_{i=1}^{n} v_i^k \ln(v_i) P(v_i)}{\sum_{i=1}^{n} v_i^k P(v_i)} - \frac{\sum_{i=1}^{n} \ln(v_i) P(v_i)}{P(v \geqslant 0)} \right)^{-1}, \tag{6}$$

$$c = \left(\frac{1}{P(v \geqslant 0)} \sum_{i=1}^{n} v_i^k P(v_i) \right)^{1/k}, \tag{7}$$

where v_i is the wind speed central to bin i, n is the number of bins, $P(v_i)$ is the frequency with which the wind speed falls within bin i, $P(v \geqslant 0)$ is the probability that the wind speed equals or exceeds zero.

Eq. (6) must be solved iteratively, after which Eq. (7) can be solved explicitly.

4.3. The graphical method

From Eq. (2), the following relation can be developed:

$$\ln\{-\ln[1 - P(v < v_i)]\} = k \ln v_i - k \ln c. \tag{8}$$

So a plot of $\ln\{-\ln[1 - P(v < v_i)]\}$ versus $\ln v_i$ presents a straight line with a slope of k and a y-intercept of $-k \ln c$. This logarithmic transformation is the basis of the graphical method.

The application of the graphical method requires that the wind speed data be in cumulative frequency distribution format. Time-series data must therefore first be sorted into bins. The line of best fit can be drawn by hand or determined using a least-squares regression. This method will be referred to in this article as the "graphical method" even though the least-squares regression can be performed without producing a graph of the data.

4.4. Qualitative estimation of the Weibull parameters using average wind speeds

This subsection is included because it is sometimes necessary to estimate the Weibull parameters in the absence of any information about the distribution of wind speeds. For example, only annual or monthly averages may be available. In such a situation, the value of k must be estimated. The value of k is usually between 1.5 and 3, depending on the variability of the wind. Smaller k values correspond to more variable (more gusty) winds. A distribution commonly used in wind engineering is the Rayleigh distribution. The Rayleigh distribution is equivalent to a Weibull distribution with $k = 2$, which corresponds to moderately gusty winds. If no information about the variability of the wind is available, a k value of 2 is often assumed. With the estimated value of k and the average wind speed, the value of c can be obtained using Eq. (2).

5. Demonstration of the methods

In this section, each method is applied to the same sample data set. The sample time-series data set is shown in Table 3. Table 4 gives the frequency distribution and the cumulative frequency distribution of this data set. For clarity, the sample data set consists of only three days of hourly wind speed data. It is important to note that a true evaluation would require many months or years of measured wind speed data.

5.1. The Maximum likelihood method

Using $n = 72$ and an initial guess of $k = 2$, successive applications of Eq. (4) to the data in Table 3 give $k = 4.07$, 2.30, and 3.60, converging after several iterations to 2.93. Eq. (5) then gives $c = 5.75$ m/s.

5.2. The modified maximum likelihood method

Using $n = 12$ and an initial guess of $k = 2$, successive applications of Eq. (6) to the data in Table 4 give $k = 4.20$, 2.31, and 3.70, converging after several iterations to 2.99. Eq. (7) then gives $c = 5.77$ m/s.

5.3. The graphical method

Using a linear least-squares regression to compute the line of best fit, the graphical method applied to the data in Table 4 gives a slope of 2.62 and an intercept of -4.35, corresponding to $k = 2.62$ and $c = 5.27$ m/s. The final bin (10–11 m/s) was excluded from the regression in order to avoid taking the logarithm of zero (see Fig. 1).

Table 3
Sample time series data set

	Wind speed (m/s)		
Hour	Day 1	Day 2	Day 3
1	3.3	4.0	4.7
2	3.8	4.0	4.5
3	4.2	2.0	4.2
4	3.3	2.7	5.7
5	2.8	2.7	2.7
6	3.0	3.3	4.3
7	4.0	2.7	4.3
8	2.7	2.7	4.5
9	5.2	5.8	4.5
10	6.7	5.7	6.0
11	6.8	6.2	10.4
12	6.8	6.5	6.7
13	5.7	4.5	6.3
14	8.5	5.8	9.4
15	8.9	4.8	7.7
16	9.3	4.8	6.0
17	6.5	5.5	8.9
18	4.2	5.7	7.7
19	4.3	5.0	6.2
20	3.7	4.3	5.7
21	4.0	4.0	5.7
22	2.8	3.5	7.5
23	3.7	5.0	7.5
24	3.3	3.7	5.3

Table 4
Sample data frequency distribution and cumulative frequency distribution

Speed (m/s)	Frequency (%)	Cum. freq. (%)
2–3	12.50	12.50
3–4	13.89	26.39
4–5	26.39	52.78
5–6	18.06	70.83
6–7	15.28	86.11
7–8	5.56	91.67
8–9	4.17	95.83
9–10	2.78	98.61
10–11	1.39	100.00

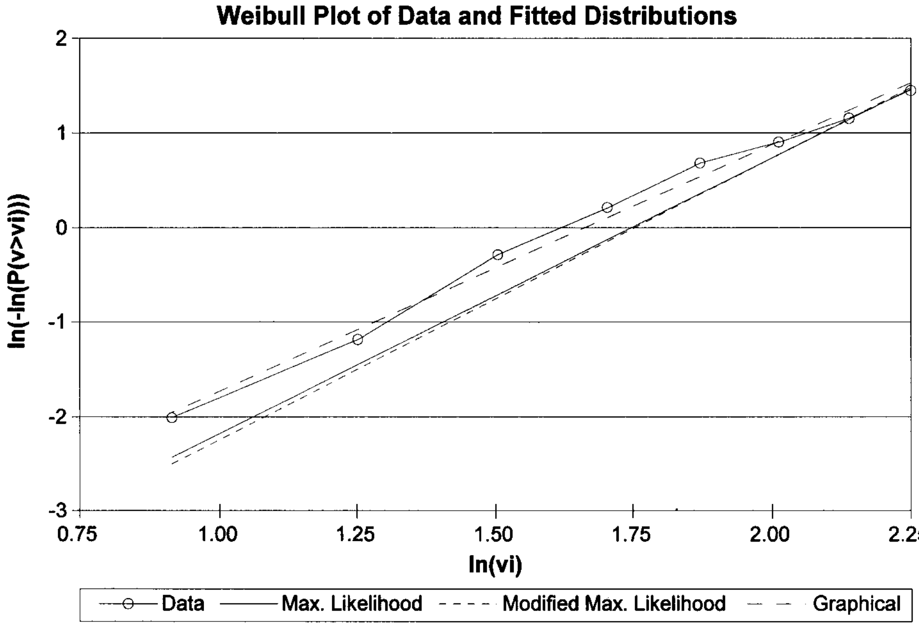

Fig. 1. The "Weibull plot" of the sample data set along with the Weibull distributions fitted by each method.

6. Accuracy of the Methods

Two tests were employed to determine the accuracy of the three methods given in this article. In the first test, each method was applied to sample sets of wind speed data drawn from known Weibull distributions, and the estimated Weibull parameters were compared to the known values. In the second test, a wind turbine power curve was used to translate wind speeds into energy outputs. The "reference energy output" was calculated using the sample time-series data set shown in Table 3. This value was then compared to the energy output corresponding to each of the three estimated Weibull distributions.

6.1. Accuracy test using known Weibull distributions

In this experiment, each method was applied to 20 data sets drawn from known Weibull distributions. Each data set was generated by randomly drawing 8760 values from a Weibull distribution with known values of k and c. The modified maximum likelihood method and the graphical method were applied using two bin sizes to show how accuracy improved with the smaller bin size. The results of this experiment are shown in Table 5. The known values of k and c are shown in the first two columns of

Table 5
Calculated Weibull parameters for each method

True Weibull parameters		Maximum likelihood method		Modified maximum likelihood method				Graphical method			
				1 m/s bins		0.1 m/s bins		1 m/s bins		0.1 m/s bins	
k	c (m/s)	*k*	*c* (m/s)	*k*	*c* (m/s)	*k*	*c* (m/s)	*k*	*c* (m/s)	*k*	*c* (m/s)
1.50	4.00	1.48	4.01	1.52	4.01	1.48	4.01	1.28	3.18	1.41	3.88
1.50	5.00	1.51	4.93	1.51	4.90	1.50	4.93	1.30	4.00	1.45	4.76
1.50	6.00	1.51	5.98	1.50	6.03	1.51	5.98	1.30	5.11	1.47	5.78
1.50	7.00	1.50	7.02	1.49	6.93	1.50	7.02	1.30	5.98	1.45	6.88
2.00	4.00	1.98	3.96	1.95	4.02	1.98	3.96	1.60	3.23	1.88	3.83
2.00	5.00	2.03	5.00	1.95	4.95	2.03	5.00	1.64	4.12	1.97	4.89
2.00	6.00	1.98	5.98	1.99	6.01	1.98	5.98	1.68	5.14	1.89	5.87
2.00	7.00	2.00	6.96	1.99	7.03	2.00	6.96	1.74	6.16	1.94	6.83
2.50	4.00	2.54	4.00	2.42	4.01	2.54	4.00	1.93	3.26	2.31	3.89
2.50	5.00	2.49	4.97	2.43	4.96	2.49	4.97	1.98	4.25	2.70	4.97
2.50	6.00	2.50	5.97	2.47	5.96	2.50	5.97	2.08	5.17	2.54	5.89
2.50	7.00	2.55	7.05	2.45	6.98	2.54	7.05	2.08	6.22	2.55	6.97
3.00	4.00	2.95	4.01	2.99	3.98	2.95	4.01	2.42	3.30	2.85	3.97
3.00	5.00	2.98	5.00	2.95	5.03	2.98	5.00	2.37	4.32	2.91	4.91
3.00	6.00	3.02	6.03	2.96	5.99	3.01	6.03	2.48	5.26	3.00	5.90
3.00	7.00	3.00	7.02	2.99	6.99	3.00	7.02	2.55	6.31	3.11	7.04
3.50	4.00	3.52	3.99	3.39	3.97	3.52	3.99	2.74	3.34	3.77	3.95
3.50	5.00	3.54	5.01	3.40	4.97	3.54	5.01	2.79	4.33	3.52	4.97
3.50	6.00	3.49	6.01	3.47	5.98	3.49	6.01	2.82	5.36	3.59	5.97
3.50	7.00	3.49	6.99	3.42	7.01	3.49	6.98	2.86	6.41	3.56	6.97
RMS error:		0.0096	0.0052	0.0183	0.0071	0.0095	0.0053	0.1780	0.1505	0.0440	0.0229

Table 6
Calculated energy output

	Actual data	M.L. method	Mod. M.L. method	Graphical method
k	N/A	2.93	2.99	2.62
c (m/s)	N/A	5.75	5.77	5.27
AEO (kWh)	203,114	198,792	199,582	156,695
Error (%)	N/A	− 2.13	− 1.74	− 22.85

Table 5. The bottom row shows the root mean square error in the estimates of k and c for each method.

The final row of Table 5 reveals several important findings. First, the reduced bin size resulted in improved accuracy for both the modified maximum likelihood method and the graphical method. This result is expected, since the larger bin sizes result in less detailed information being available to the algorithms. Second, the maximum likelihood method is considerably more accurate than the traditional graphical method even when the smaller bin size is used. Third, the modified maximum likelihood method is much more accurate than the graphical method regardless of bin size. In this experiment, in fact, the modified maximum likelihood method when used with the smaller bin size had virtually the same error as the original maximum likelihood method.

The fundamental reason for the inaccuracy of the graphical method is that the least-squares regression is performed not on the actual wind speed data, but on its cumulative frequency distribution. Each point is weighted equally even though some of the bins represent many more data points than others. This problem is somewhat diminished as the bin size is reduced, but the bin size is limited by the resolution of the wind speed data.

6.2. *Accuracy test using total energy output*

In this experiment, a particular wind turbine power curve was used to convert wind speeds into expected energy outputs. This procedure can be performed using either time series data or a Weibull distribution. Total energy outputs were calculated for the sample wind data set shown in Table 3 and for each of the three Weibull distributions fitted to the sample data set. Since the value calculated using the actual time-series data is most accurate, it was used as the "reference energy output". The results of this experiment are shown in Table 6.

Once again, the maximum likelihood method was most accurate, followed by the modified maximum likelihood method. It is interesting to note that for this particular data set, if a Weibull k factor of 2 had been assumed, the resulting error in annual energy output would have been + 19.7%.

7. Conclusions

When wind speed data is available in time-series format, the maximum likelihood method is the recommended method for estimating the parameters of the Weibull distribution for wind energy analysis. The graphical method is not only less accurate than the maximum likelihood method, it is also less robust since its accuracy is affected by external variables such as the bin size in the cumulative frequency distribution. When wind speed data is available in frequency distribution format, the modified maximum likelihood method is the recommended method since its accuracy is superior to the graphical method regardless of the bin size.

When computers were less powerful and less readily available than they are today, the graphical method was the method of choice because it could be performed by hand with minimal computation. Because of its familiarity, the graphical method is still commonly used even though it is typically implemented on a computer. The intent of this article has been to demonstrate that the maximum likelihood method is the more appropriate computer-based method.

References

[1] C.G. Justus, W.R. Hargraves, A. Yalcin, Nationwide assessment of potential output from wind-powered generators, J. Appl. Meteorol. 15 (1976) 673–678.

[2] E. Simiu, N.A. Heckert, T.M. Whalen, Mean recurrence intervals of ultimate wind loads, Proceedings of the 17th International Conference on Offshore Mechanics and Architectural Engineering, Lisbon, Portugal, 1998.

[3] M.J. Stevens, P.T. Smulders, The estimation of the parameters of the Weibull wind speed distribution for wind energy utilization purposes, Wind Eng. 3 (2) (1979) 132–145.