Available online at www.iseec2012.com

I-SEEC 2012

Proceeding - Science and Engineering (2013) 377–382

Proceeding
Science and Engineering

www.iseec2012.com

Science and Engineering Symposium
4th International Science, Social Science, Engineering and Energy Conference 2012

The Method of Medians to Estimate Weibull Parameters for Wind Energy Applications

Y. Mert Kantar[a,*] , I. Usta[b] , I. Arik[c]

[a]*Department of Statistics, Faculty of Science, Anadolu University, Eskisehir 26470, Turkey*

Abstract

The Weibull family of distributions is frequently used in modeling wind speed data and calculating wind power density. Thus, its parameters have to be estimated accurately. The maximum likelihood is a popular method due to statistical properties and also available in various programs. However, this method is very sensitive to outliers, which are generally results of either mistake in measuring and recording data. In this study, we introduce the method of medians, proposed by He and Fung 2009, for the two-parameter Weibull wind speed distribution. Analyses are carried out to assess the performance of the considered method for Weibull distribution on wind speed data with/without outliers.

Selection and/or peer-review under responsibility of Faculty of Science and Technology, Kasem Bundit University, Bangkok.

Keywords: Weibull distribution, wind speed, wind power, outlier, robust, median

1. Introduction

The wind speed distribution is very important information needed in the assessment of wind energy potential. In the specialized literature on wind energy and other renewable energy sources, the Weibull distribution has been commonly used, accepted and recommended distribution to model the wind speed [1-3]. Therefore, its parameters have to be estimated accurately. In order to find the best method for Weibull, various estimation methods have been introduced and widely discussed in the literature [4-7]. In literature, graphic, maximum likelihood, least squares and moment estimators are the most frequently used as estimation methods. These estimators are introduced and some comparisons between these methods are carried out by using different criteria. It is generally found that the maximum likelihood estimator (MLE) is the recommended method due to statistical properties [4, 5]. However; it is known that if there are outliers in data, the MLE as well as the other mentioned estimators can be very unreliable [8-11]. Outliers generally arise either mistakes in measuring and recording the data. These outliers affect the estimates dramatically. Unfortunately, engineers working wind speed data often face the problem of outliers. This problem may lead to estimates of parameters of the Weibull

* Corresponding author. *E-mail address:* ymert@anadolu.edu.tr

distribution obtained by MLE and the other mentioned estimators, to be significantly biased [10]. For such cases, robust estimators may be good alternatives. It is well-known that robust estimators aim to reject or reduce effect of outliers in order to provide a better fit to the majority of data [10].

In this study, a robust estimator based on medians is introduced to estimate Weibull distribution parameters for wind energy applications. This estimator is called as the method of medians (MM) as in [11]. Also, to demonstrate performance of MM, some comparisons are carried out based on wind speed data. Results of this study show that the MM is an effective method to estimate Weibull parameters when wind speed observations as outliers exist in the data.

2. Methods for estimating Weibull parameters

Two parameter Weibull probability density function (pdf) is given as follows:

$$f_W(x) = ck(cx)^{k-1}\exp(-(cx)^k) \tag{1}$$

where k is the shape parameter and c is the scale parameter. The k values range from 1.5 to 3 for most wind conditions [12-14].

The cumulative distribution function of Weibull distribution is given as follows:

$$F_W(x) = 1 - \exp(-(cx)^k) \tag{2}$$

2.1 Maximum likelihood estimator

The log-likelihood function of a random sample from the Weibull distribution can be expressed as follows:

$$L(c,k|x_1,\ldots,x_n) = nlogk + nlogc + (k-1)\sum_{i=1}^{n}(\log(cx_i)) - \sum_{i=1}^{n}(cx_i)^k$$

By differentiating the log-likelihood function with respect to $\boldsymbol{\theta} = (c,k)$ and equating them to zero, the estimated parameters are obtained from equations (3)-(4) as follows:

$$\frac{n}{k} - \sum_{i=1}^{n}(log\,(cx_i)) - \sum_{i=1}^{n}(cx_i)^k \ln(cx_i) = 0 \tag{3}$$

$$\frac{n}{c} - (k-1)\frac{n}{c} - k\sum_{i=1}^{n}x_i(cx_i)^{k-1} = 0 \tag{4}$$

The equations (3) – (4) can be solved by iterative methods.

2.2 Method of Medians for estimating Weibull parameters

The logarithm of the Weibull distribution given in equation (1) is presented as follows:

$$log\,f_W(x) = logk - logc + (k-1)(c\,logx) - (cx)^k \tag{5}$$

The score function of the Weibull distribution is obtained as:

$$\varphi(x) = \frac{\partial \log f_W(x)}{\partial \beta}, \tag{6}$$

where $\boldsymbol{\beta} = (c,k)$ is the vector of parameters of the Weibull distribution.

Thus, the score function of the Weibull distribution is calculated as follows:

$$\varphi(x) = \begin{cases} \frac{k}{c}(1-(cx)^k) \\ (1+(1-(cx)^k)\log((cx)^k)/k \end{cases} \tag{7}$$

The estimates based on the method of median (MM) are obtained by equating the sample median of $\varphi(x)$ to corresponding population median. It is known that $Y = (cX)^k$ has exponential distribution with mean one [11]. By using this relationship, the following two equations are obtained.

$$\text{Median}_i \{(cx_i)^k\} = log2 \tag{8}$$

$$\text{Median}_i \{(1-(cx_i)^k) log(cx_i)^k\} = a, \tag{9}$$

where $a = Median\left((1-Y)logY\right) \approx -0.51$.

From equation (8), c parameter is estimated as:

$$\hat{c} = (log2)^{1/k} / \text{Median}_i\{x_i\} \tag{10}$$

Thus, equation (9) is iteratively solved for shape parameter k.

3. Comparison of methods

In order to show the performance of the MM for wind speed data, the comparison between MM and MLE based on wind speed data taken from Amasya, Turkey was conducted. The considered wind speed data are artificially contaminated with outliers. Some criteria such as root mean square errors (RMSE) and chi-square are used for comparison.

The hourly wind speed data in empirical frequency distribution (ED) format and the frequency distributions calculated by the Weibull distribution estimated from MLE and MM are given in Table 1. For the monthly wind speed data measured in February, 2004, the first column show empirical frequency distribution, second two columns provide frequency distributions calculated from the Weibull distribution estimated by MLE and MM, respectively. The same monthly wind speed data is contaminated with one outlier (large wind speed observation), thus, for new data set with outlier, the last two columns are calculated from MLE and MM. It is seen that changing the value of the single wind speed observation (outlier) has not influence on the frequency distributions obtained from MM, but, change dramatically the frequency distributions obtained from MLE.

Table 1. Frequency distributions

	Wind speed data without outlier		**Wind data with one outlier**	
ED	MLE	MM	MLE	MM
0.0172	0.0167	0.0181	0.0357	0.0181
0.0517	0.0492	0.0504	0.0707	0.0504
0.0977	0.0771	0.0768	0.0893	0.0768
0.0920	0.0979	0.0958	0.0981	0.0958
0.1020	0.1101	0.1066	0.0999	0.1066
0.0948	0.1134	0.1093	0.0963	0.1093
0.0991	0.1090	0.1052	0.0892	0.1052
0.1121	0.0987	0.0958	0.0798	0.0958
0.0805	0.0847	0.0831	0.0694	0.0831
0.0690	0.0692	0.0690	0.0588	0.0690
0.0704	0.0539	0.0549	0.0487	0.0549
0.0330	0.0401	0.0419	0.0395	0.0419
0.0302	0.0286	0.0309	0.0314	0.0309
0.0201	0.0195	0.0219	0.0245	0.0219
0.0129	0.0128	0.0150	0.0188	0.0150
0.0115	0.0081	0.0099	0.0141	0.0099
0.0057	0.0049	0.0063	0.0105	0.0063

Also, by using criteria such as RMSE and Chi-square, the results of comparisons of MM and MLE for wind speed data without/with outliers are shown in Table 2-4, respectively.

The estimated Weibull parameters, RMSE and Chi-square values according to the MM and MLE have been shown Table 2 for measured in Amasra for January, February of 2004 and whole year 2004. It is seen from Table 2 that MM gives satisfactory results according to RMSE and Chi-square, when even there is no outlier in sample.

Table 2. MM and MLE comparison for wind speed data measured in Amasya, Turkey

		Estimation Methods	
	Parameters	**MLE**	**MM**
January	*c*	0.153970	0.148437
	k	1.603295	1.527869
	RMSE	0.010742	0.010146
	Chi-square	0.000129	0.000115
February	*c*	0.131946	0.129593
	k	2.016635	1.959312
	MSE	0.009421	0.008821
	Chi-square	0.000101	8.82E-05
Whole year	*c*	0.17942	0.181613
	k	1.497446	1.353483
	RMSE	0.008131	0.00885
	Chi-square	7.24E-05	8.58E-05

To observe performance of method in terms of robustness property, we consider wind speed data samples with outliers. Therefore, wind speed data samples are artificially contaminated with outliers.
In this study, the outlier data configuration is considered as follows:

(a) One outlier is generated from $U(\bar{X} + 5s, \bar{X} + 10s)$, where $\bar{X}$ is the mean wind speed and s is the standard deviation and U show uniform distribution.

(b) One outlier is generated such that the last wind speed observation (v_{max}) in the original sample are shifted by $v_{max} \times 4s$.

Table 3. MM and MLE comparison for monthly wind speed data with one outlier (a)

		Estimation Methods	
	Parameters	**MLE**	**MM**
January	*c*	0.1536045	0.148437
	k	1.570418	1.527869
	RMSE	0.010575	0.010411
	Chi-square	0.000126	0.000122
February	*c*	0.131635	0.129593
	k	1.961918	1.959312
	RMSE	0.008961	0.008833
	Chi-square	9.10E-05	8.84E-05
Whole year	*c*	0.179397	0.181613
	k	1.495113	1.353483
	RMSE	0.008106	0.00885
	Chi-square	7.20E-05	8.58E-05

Table 4. MM and MLE comparison for monthly wind speed data with one outlier (b)

		Estimation Methods	
	Parameters	**MLE**	**MM**
January	*c*	0.152556	0.148437
	k	1.162153	1.527869
	RMSE	0.019053	0.010411
	Chi-square	0.000408	0.000122
February	*c*	0.130089	0.129593
	k	1.298289	1.959312
	RMSE	0.024045	0.008833
	Chi-square	0.000655	8.84E-05
Whole year	*c*	0.180548	0.181613
	k	1.357692	1.353483
	RMSE	0.008755	0.00885
	Chi-square	8.40E-05	8.58E-05

From Table 3-4, we see that changing the value of the single wind speed observation (outlier) has little influence on the MM, but, induces a large variation in the maximum likelihood estimates. These analyses show that the MM has a built-in protection against a certain amount of recording or measuring errors. Also, it should be emphasized that MM may provide better estimates than MLE and other non-robust estimators when there are outliers in the sample, however, MM may not yield satisfactory results every time for the case of no outliers.

4. Conclusions

In this study, we introduce the method of medians (MM) for the two-parameter Weibull wind speed distribution. Analysis is carried out to assess the performance of MM for Weibull distribution on wind speed data with/without outliers. Analyses show that the MM has a smaller RMSE and chi-square values than the MLE for all the considered wind speed samples. It is emphasize that even if there is only one outlier in data, this outlier has little influence on the estimates obtained from MM, but induces a large variation in the maximum likelihood estimates.

Moreover, we will pursue this study with two goals in mind. The first one is to expand analyses and second goal is to investigate estimation of wind power density via the Weibull estimated by MM.

References

[1] Ahmed AS. Wind energy as a potential generation source at Ras Benas, Egypt. Renew Sustain Energy Rev 2010;14(8):2167-73.

[2] J Weisser D. A wind energy analysis of Grenada: an estimation using the'Weibull' density function. Renew Energy 2003;28:1803-12.

[3] Akdag S, Guler O. Wind characteristics analyses and determination of appropriate wind turbine for Amasra Black Sea region, Turkey. Int J Green Energy 2010;7(4):422-33.

[4] Kantar YM, Senoglu B. A comparative study for the location and scale parameters of the Weibull distribution with given shape parameter. Comput Geosci 2008;34:1900-9.

[5] Seguro JV, Lambert TW. Modern estimation of the parameters of the Weibull probability density distribution. J Wind Eng Ind Aerodyn 2000;85:75-84.

[6] Akdag S, Dinler A. A new method to estimate Weibull parameters for wind energy applications. Energy Convers Manage 2009;50:1761-6.

[7]Akdag S, Dinler A. A new method to estimate Weibull parameters for wind energy applications. Energy Convers Manage 2009;50:1761-6.

[8] D.R. Shier and K.D. Lawrence, A comparison of robust regression techniques for the estimation of weibull parameters, Comm. in Stat.-Sim. & Comp. 13(6) (1984), pp. 743-750.

[9] X. He and W.K. Fung, Method of medians for lifetime data with weibull models, Stat. in Med. 18 (1999), pp. 1993-2009.

[10] L.F. Zhang, M. Xie L.C. Tang, Robust regression using probability plots for estimating the weibull shape parameter, Qual. Reliab. Eng. Int. 22 (2006), pp. 905-917.

[11] X. He and W.K. Fung, Method of medians for lifetime data with weibull models, Stat. in Med. 18 (1999), pp. 1993-2009.
[12] Shamilov A, Kantar YM, Usta I. Use of MinMaxEnt distributions defined on basis of MaxEnt method in wind power study. Energy Convers Manage 2008;49:660-77.
[13] Kantar YM, Usta I. Analysis of wind speed distributions: wind distribution function derived from minimum cross entropy principles as better alternative to Weibull function. Energy Convers Manage 2008;49:962-73.
[14] Usta I, Kantar YM. Analysis of some flexible families of distributions for estimation of wind speed distributions. Applied Energy 2012;89: 355-367.