The 2014 World Congress on
Advances in Civil, Environmental, and Materials Research (ACEM14)
Busan, Korea, August 24-28, 2014

Assessing Goodness-of-fit of Weibull Distributions for Wind Resource Prediction

*Ho-Sung Song[1)] and Soon-Duck Kwon[2)]

[1),2)] KOCED Wind Tunnel Center, Department of Civil Engineering
Chonbuk National University, Chonju, Chonbuk 561-756, Korea
[2)] sdkwon@chonbuk.ac.kr

ABSTRACT

This paper presents a model selecting procedure for proper probability distribution of wind speed. In order to assess the goodness-of-fit of various Weibull distributions, this study introduced Akaike information criterion (AIC) and Bayesian information criterion (BIC) in addition to the conventional Chi-square test (χ^2 test) and Kolmogorov–Smirnov test (K-S test). The results show that 3-parameter Weibull is good for modeling ordinary wind data, yet 5-parameter mixture Weibull reveals more suitable probability model for that representing two peaks.

Keyword: Weibull distribution, Goodness-of-fit test, Probability model selection, Akaike Information criterion, Bayesian Information criterion.

1. INTRODUCTION

Wind speed is the most important factor for wind energy prediction. The annual energy production (AEP) from wind turbine can be estimated by the following integration, which is power curve of wind turbine $P_{WT}(V)$ multiplied by the probability density function of mean wind speed.

$$E_{AEP} = T \int_0^\infty P_{WT}(V) \cdot f(V)\, dV \tag{1}$$

where, *T* is the annual hours.

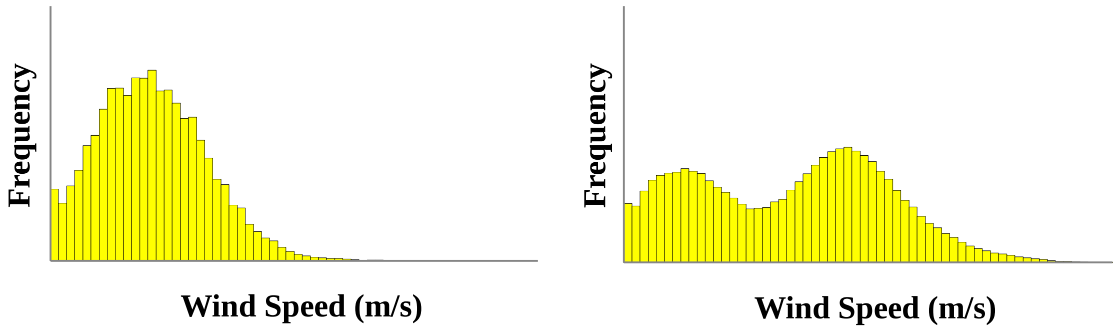
In Eq. (1), power curve is generally provided by wind turbine makers after field tests. Therefore the remaining uncertainty to evaluate the annual energy production is wind

---

[1)] Senior Researcher
[2)] Professor

probability model. It is generally believed that the most significant error in prediction of wind power output is improper estimation of wind probability.

In order to model the temporal repeated feature of wind speed during the days or the years, 2-parameter Weibull distribution or Rayleigh distribution mostly used in the previous studies and applications. However, it is sometimes necessary to use more flexible type of probability distributions in order to fit two peaks (Fig. 1b) which observed in the terrains that have two or more sources of wind.



(a) One peak (b) Two peaks

Fig. 1 Two types of wind speed distributions.

In this study, we have tested the following four different Weibull distributions of wind speed in order to assess wind resource prediction reasonably.

a) 1-parameter Weibull distribution (known as Rayleigh distribution)
b) 2-parameter Weibull distribution
c) 3-parameter Weibull distribution
d) 5-parameter mixture Weibull distribution

The goodness-of-fit tests were performed in order to verify the effectiveness of the linear relationship between observed wind data and the assumed probability models. In addition, we introduce Akaike Information Criterion (Akaike, 1974) and Bayesian Information Criterion (Schwarz, 1978) which are widely used as the criterion of model selection for number of estimated parameters.

2. PROBABILITY MODELS

2.1 Weibull distribution

Weibull distribution is generally used to represent the occurrence frequency of wind speed (V). Probability density function (PDF) of Weibull distribution is given by

$$f(V;C,k,\tau) = \frac{k}{C}\left(\frac{V-\tau}{C}\right)^{k-1} \exp\left\{-\left(\frac{V-\tau}{C}\right)^{k}\right\}, \qquad (V>0,\ C>0,\ k>0,\ \tau \le V)$$
$$f(V) = 0, \qquad (V \le 0) \tag{2}$$

where C , k and τ are scale, shape and location parameters respectively. The generalized form of Weibull distribution is made up of three parameters. If the location parameter τ is 0, the distribution is the 2-parameter Weibull distribution which is generally used in wind power industry. The Rayleigh distribution is a special form with that shape parameter equal to 2 from 2- parameter Weibull distribution.

Wind speed at well-behaved area can be properly expressed using 2-parameter Weibull and Rayleigh distribution. However, it is not easy to use the 2-parameter Weibull in some specific regions that have a least two different characteristics of wind speed. Jaramillo and Borja (2004), Carta and Ramirez (2007) have investigated various probability density function of wind speed using mixture Weibull distribution consisted of a linear combination of two 2-parameter Weibull distributions.

2.2 Mixture Weibull distribution

A mixture distribution is a convex linear combination of each other distributions. Suppose that wind data sets, V_i $(i = 1{,}2)$ are independently distributed as 2-parameter Weibull, $f(V; C, k)$. Then V is distributed as V_i with mixing parameter ω is said to have a 2-component mixture Weibull distribution. The probability density function of V, which depends on five parameters $(C_1, k_1, C_2, k_2, \omega)$, is given by

$$f(V;\ C_1, k_1\ C_2, k_2, \omega\ ) \\ = \omega\left[\frac{k_1}{C_1}\left(\frac{V}{C_1}\right)^{k_1-1}\exp\left\{-\left(\frac{V}{C_1}\right)^{k_1}\right\}\right] + (1-\omega)\left[\frac{k_2}{C_2}\left(\frac{V}{C_2}\right)^{k_2-1}\exp\left\{-\left(\frac{V}{C_2}\right)^{k_2}\right\}\right] \tag{3}$$

and the cumulative distribution function is given by

$$F(V;\ C_1, k_1\ C_2, k_2, \omega\ ) = \omega\left[1-\exp\left\{-\left(\frac{V}{C_1}\right)^{k_1}\right\}\right] + (1-\omega)\left[1-\exp\left\{-\left(\frac{V}{C_2}\right)^{k_2}\right\}\right] \tag{4}$$

In the Eqs. (3) and (4), the probability parameters, C_i, k_i and ω, can be estimated using moment method, maximum likelihood method and least squares method from wind speed data.

3. GOODNESS-OF-FIT TEST AND MODEL SELECTION CRITERIA

3.1 Goodness-of-fit test

If the collected wind data is assumed as an arbitrary probability model, the effectiveness of the assumed probability model is verified by goodness-of-fit test. In order to compare the performance of each probability distributions, we use five different measures; correlation coefficient, root mean square error (RMSE), R-square test, Chi-Square test and Kolmogorov-Smirnov (K-S) test.

3.2 Model selection criteria

The number of estimated parameters consisting the probability model is increased, the minimum value of the error converge to zero. However, this is likely to be over-fitting. While the opposite case, the over-fitting is prevented, but the minimum value of error is increased and we cannot represent appropriate probability model. More objective and quantitative criteria is needed, because the method using error directly can be adjusted by adding or deleting the independent variable arbitrarily.

To prevent over-fitting problem, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are introduced in this study. AIC and BIC impose the penalty on the conduct to reduce the error increasing independent random variables. The difference between the two methods is that BIC imposes the heavier penalty on conduct to increasing independent random variables than AIC. AIC and BIC are given by

$$\begin{aligned} \mathrm{AIC} &= -2\ \ln L + 2\mathrm{p} \\ \mathrm{BIC} &= -2\ \ln L + \mathrm{p} \ln n \end{aligned} \tag{5}$$

where L, p and n are value of the maximum likelihood, the number of parameters and the number of all observed data respectively. The model is better when there are smaller value of AIC and BIC both.

4. NUMERICAL EXAMPLES

4.1 Wind data

Table 1 shows information about wind data sets used in testing of probability models of wind speed. Total ten sites are classified into four categories according to the shape of wind speed distribution as shown Fig. 2.

Table 1 Information of wind data sets.

No.	*Site*	*Country*	*Terrain*		*Data acquisition period*
1	Blyth	UK	offshore	flat	2001-08-16~2003-01-24
2	Borglum	Denmark	pastoral	flat	1997-12-11~2001-12-31
3	Capel	UK	pastoral	flat	1989-05-01~1995-08-31
4	Equinox	US	scrub	hill	1991-05-16~1991-11-10
5	Gorgonio	US	scrub	hill	1991-02-09~1992-05-10
6	Hidalgo	Mexico	scrub	hill	2002-01-01~2006-12-31
7	Nasudden	Sweden	coastal	flat	1992-01-09~1995-07-21
8	Oakcreek	US	scrub	hill	1998-05-18~2000-07-26

No.	*Site*	*Country*	*Terrain*		*Data acquisition period*
9	Ventosa	Mexico	scrub	hill	2000-05-01~2006-08-01
10	Windland	US	scrub	hill	1990-12-08~1993-11-09

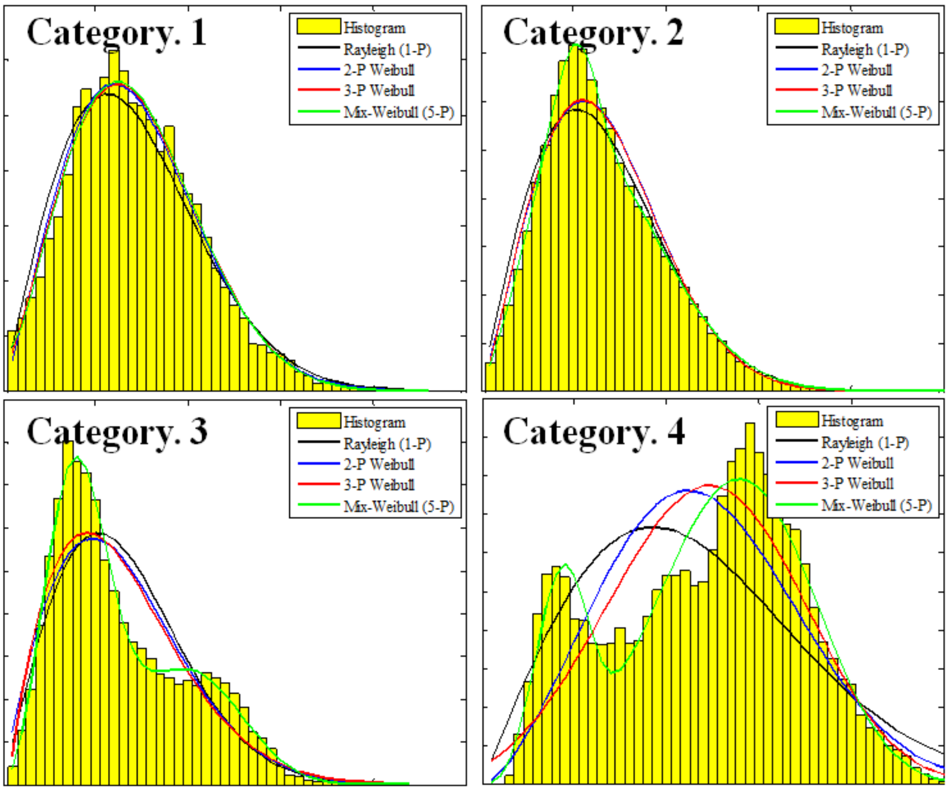
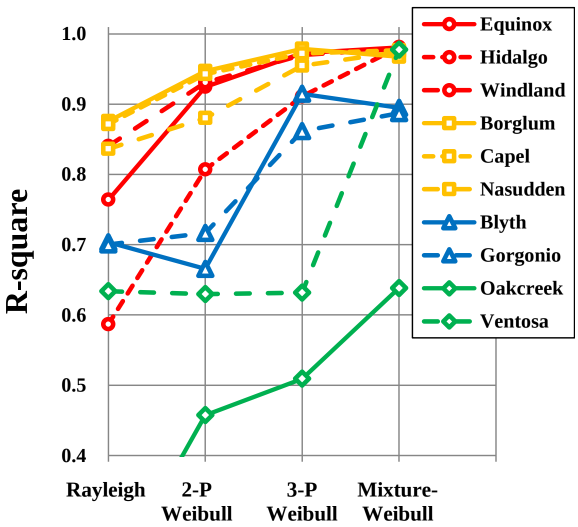


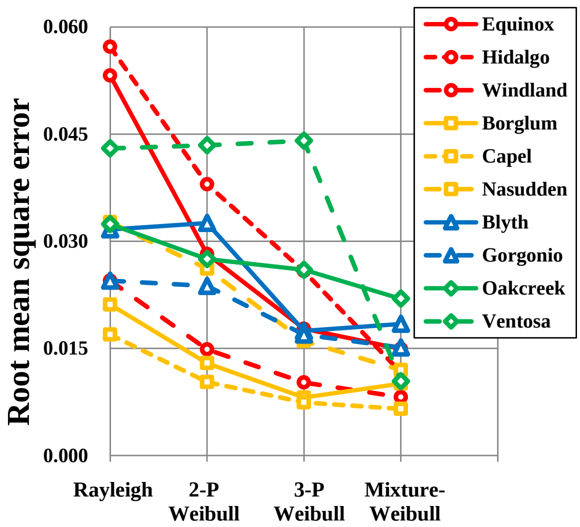
Fig. 2 Classification of wind speed distributions.

4.2 Results

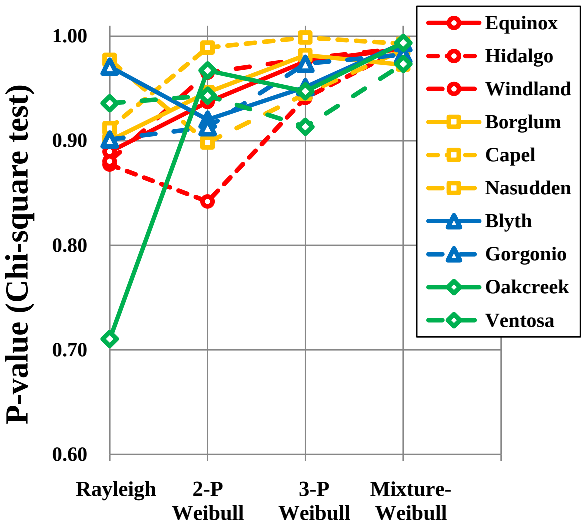
Fig. 3 shows the result of the goodness-of-fit tests. The 5-parameter mixture Weibull distribution shows the best performance in most cases. Table 2~3 and Fig. 4 shows result by AIC and BIC for model selection criteria. The 3-parameter Weibull distribution is found to be appropriate in case that probability density function of wind speed distribution has single peak. However the mixture Weibull distributions are more suitable for the probability distributions with two peaks.



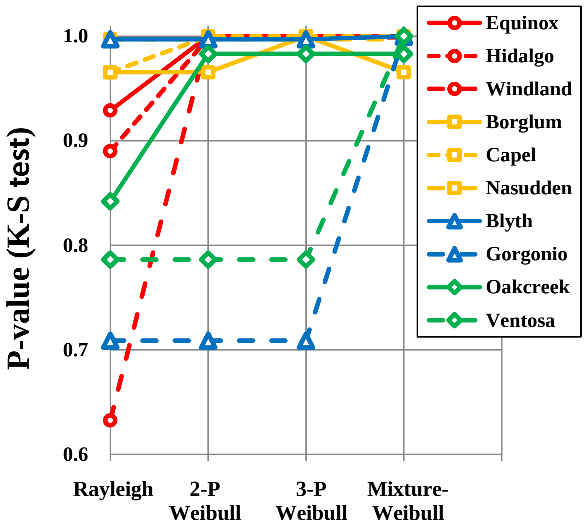
(a) R-square



(b) Root mean square error



(c) Chi-square test



(d) K-S test

Fig. 3 Result of goodness-of-fit tests.

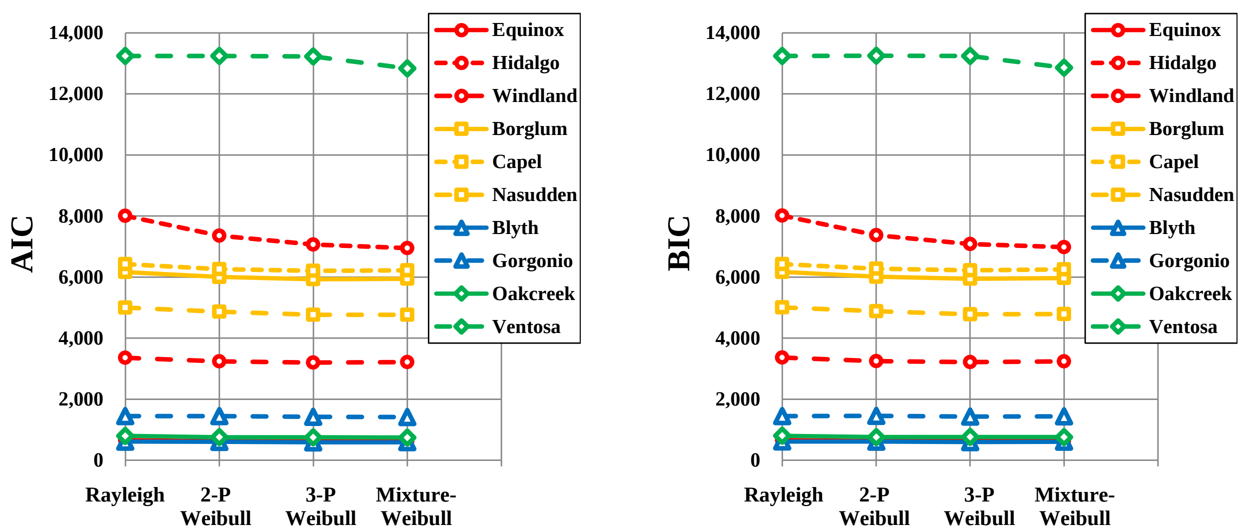
Table 2 Value of Akaike information criterion.

Site	*Category*	*Rayleigh (1P)*	*Weibull (2P)*	*Weibull (3P)*	*Mix-Weibull (5P)*
Equinox	1	658	622	**620**	623
Hidalgo	1	8,006	7,359	7,063	**6,950**
Windland	1	3,358	3,238	**3,199**	3,216
Borglum	2	6,163	6,002	**5,929**	5,944
Capel	2	6,421	6,262	**6,207**	6,223
Nasudden	2	5,000	4,869	4,762	**4,762**

Site	*Category*	*Rayleigh (1P)*	*Weibull (2P)*	*Weibull (3P)*	*Mix-Weibull (5P)*
Blyth	3	609	603	**585**	590
Gorgonio	3	1,444	1,446	1,420	**1,416**
Oakcreek	4	798	752	752	**743**
Ventosa	4	13,242	13,243	13,230	**12,834**

Table 3 Value of Bayesian information criterion

Site	*Category*	*Rayleigh (1P)*	*Weibull (2P)*	*Weibull (3P)*	*Mix-Weibull (5P)*
Equinox	1	661	**628**	629	637
Hidalgo	1	8,012	7,370	7,079	**6,977**
Windland	1	3,363	3,247	**3,213**	3,239
Borglum	2	6,169	6,012	**5,944**	5,970
Capel	2	6,426	6,272	**6,223**	6,248
Nasudden	2	5,005	4,879	**4,777**	4,787
Blyth	3	612	609	**593**	605
Gorgonio	3	1,448	1,453	**1,431**	1,434
Oakcreek	4	801	758	760	**757**
Ventosa	4	13,248	13,255	13,247	**12,863**



a) Akaike information criterion b) Bayesian information criterion

Fig. 4 Result of AIC and BIC.

5. CONCLUSIONS

In this study, we have tested four different Weibull probability distributions for predicting wind energy potential using various goodness-of-fit tests including Akaike information criterion (AIC) and Bayesian information criterion (BIC). The results show that the 3-parameter Weibull distribution is appropriate in case that probability density function of wind speed distribution has single peak. However the 5-parameter mixture Weibull distribution are more suitable for the probability distributions with two peaks.

ACKNOWLEDGEMENTS

This work is part of a research project supported by the Korea Ministry of Land, Infrastructure and Transport (MOLIT) through the Honam Regional Infrastructure Technology Development Center (14RDRP-B076564-01).

REFERENCES

Akaike, H. (1974). “A new look at the statistical model identification”, *Automatic Control, IEEE Transactions on*, **19**(6), 716-723.

Schwarz, G. (1978). “Estimating the dimension of a model”, *The annals of statistics*, **6**(2), 461-464.

Jaramillo, O. A., & Borja, M. A. (2004). “Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case”, *Renewable Energy*, **29**(10), 1613-1630.

Carta, J. A. & Ramirez, P. (2007). “Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions”, *Renewable energy*, **32**(3), 518-531.

WindData. Available at http://www.winddata.com. Department of Wind Energy Technical University of Denmark.