Comparative Study of Five Methods to Estimate Weibull Parameters for Wind Speed on Phangan Island, Thailand

Warit Werapun*, Yutthana Tirawanichakul, Jompob Waewsak

*Faculty of Science and Industrial Technology, Prince of Songkla University, Suratthani Campus, Suratthani, Thailand 84100

Abstract

The Weibull distribution has been widely used to represent the time distribution of wind speed. The two parameters to fit are k for shape and c for scale, and five methods were compared to estimate these Weibull parameters. Wind speed data from January 2012 to December 2014 from Phangan island were used with the fitting methods, namely empirical, energy pattern factor, maximum likelihood, modified maximum likelihood, and graphical method. The goodness of fit was quantified using Kolmogorov-Smirnov test, $R^2$, RMSE and the percent error of wind power density. With these data, the energy pattern factor method performed the best in terms of percent error of power density, and had the highest $R^2$. The graphical method performed the worst among these methods, in this case study.

Keywords: Weibull distribution; wind energy; power density; estimating Weibull parameters

1. Introduction

The assessment of wind energy potential is an important step affecting decisions about wind turbines or wind farms. Generally, long-term meteorological wind speed data in time series reveal the frequency distribution of wind speed. The Weibull distribution has been used to fit these data, and it has two adjustable parameters, namely k for shape and c for scale. The Weibull fits are widely used in Wind
Atlas Analysis and Application Program (WAsP) to estimate the wind power density or the wind energy. However, the Weibull distribution is not suitable to represent the wind distributions across all geographical locations in the world [1]. Normally the wind power correlates with the time average of cubed wind speed, so the frequency distribution wind speed allows the computation of wind energy. Many numerical methods for fitting the Weibull distribution to data are available [2-4]. The maximum likelihood method is recommended for use with time series wind speed data, whereas a modified maximum likelihood is suggested for use with wind speed data in frequency format, and a graphical method is recommended for small data samples [5]. The power density method is sometimes called the energy pattern factor method, and was proposed by Akdag and Dinler [6] as a direct method that does not require iterations like the maximum likelihood fitting does. However, Chang [7] compared six numerical methods for estimating the Weibull parameters for three areas in Taiwan, and found that the maximum likelihood method performs best, with the modified maximum likelihood the next best. To assess how suitable the various methods are for fitting wind speed frequency distributions by the Weibull function, the results can be statistically evaluated using the coefficient of determination ($R^2$), the chi-square ($\chi^2$), and the Root Mean Square Error (RMSE). Rocha et al. [8] introduced the equivalent energy method for Weibull fitting with data from a Northeastern region of Brazil, and demonstrated better performance than with the alternative methods.

The aim of this work is to evaluate and compare five numerical methods for fitting Weibull distributions to wind speed data from January 2012 to December 2014, on Phangan Island, Thailand.

2. Methods of Estimating the parameters $k$ and $c$

The Weibull probability density for fitting the wind speeds is shown in equation 1.

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

(1)

Here, $C$ is a scale parameter and $k$ is a shape factor. The cumulative distribution is

$$F(V) = 1 - \exp \left( -\frac{V}{C} \right)^k$$

(2)

The methods selected for comparative assessment were the empirical method, the energy pattern factor method, the maximum likelihood method, the modified maximum likelihood method, and the graphical method. The values of $k$ and $c$ are then determined by equations 3 to 14.

2.1 Empirical Method

For the empirical method $k$ and $c$ are defined by equations 1 and 2. The gamma function is defined by equation 5, and is closely related to the factorial.

$$k = \left( \frac{\sigma}{\bar{v}} \right)^{1.086}$$

(3)
2.2 Energy Pattern Factor Method

This method was suggested by Akdag Ali [6], and is defined by equations 6 to 8.

\[ E_{pf} = \left( \frac{\overline{V}^3}{\overline{V}} \right) \]  \hfill (6)

\[ k = 1 + \left( \frac{3.69}{E_{pf}^2} \right) \]  \hfill (7)

\[ \overline{V} = C\Gamma\left( 1 + \frac{1}{k} \right) \]  \hfill (8)

2.3 Maximum likelihood method

The fitting of parameters requires numerical iterations. The shape parameter \( k \) and the scale parameter \( c \) are estimated with the following two equations, suggested by Steven and Smulders [9].

\[ k = \left( \frac{\sum_{i=1}^{n} v_i^k \ln(v_i)}{\sum_{i=1}^{n} v_i^k} \right)^{-1} \] \hfill (9)

\[ c = \left( \frac{1}{n} \sum_{i=1}^{n} v_i^k \right) \frac{1}{k} \] \hfill (10)

where \( v_i \) is the wind speed in time step \( i \) and \( n \) is the number of nonzero wind speed data points. Iterations are required to satisfy equation 9, which determines \( k \)

2.4 Modified maximum likelihood method

If wind speed data in frequency distribution format are available, this method may be appropriate. The Weibull parameters are calculated as follows.
Here $v_i$ is the wind speed central to bin $i$, $n$ the number of bins, $f(v_i)$ is the frequency for wind speeds within bin $i$, and $f(v \geq 0)$ is the probability for wind speed equal to or exceeding zero. Again, "k" needs to fitted by iterative solution of equation 11.

2.5 Graphical Method

The cumulative distribution is shown again in equation 13. Taking natural logarithms twice gives equation 14, which is linear in the parameters to be fit. They can be identified from a plot of $ln(v)$ versus $ln[-ln[1-F(V_i)]]$, where the $k$ shape parameter equals the slope, and the scale parameter is obtained from the intercept with y-axis.

$$F(V_i) = 1 - e^{\left(\frac{v_i}{C}\right)^k}$$  \hspace{1cm} (13)\\
$$ln[-ln[1-F(V_i)]] = -klnC + klnV_i$$  \hspace{1cm} (14)

3. Goodness of fit and data used

3.1 Goodness of fit

The Kolmogorov-Smirnov test enables comparing cumulative distributions, and is appropriate for checking goodness of fit to the distribution in the data. It uses as quantifier of discrepancy the maximal absolute error between the two cumulative distributions. Another statistic describing goodness of fit is $R^2$. Also the root mean square error is common as a descriptor of the goodness of fit. Finally, the percent error of power density between the Weibull fit and the actual data is calculated in equation 15. The power density based on actual time series data can be calculated using equation 16, and the power density from the Weibull fit is given by equation 17.

$$\text{Error\%} = \left| \frac{P_u - P_a}{P_a} \right| \times 100\%$$  \hspace{1cm} (15)\\
$$P_a = \frac{1}{2} \rho V^2$$  \hspace{1cm} (16)\\
$$P_u = \frac{E_u}{T} = \frac{1}{2} \rho C^2 T (1 + \frac{3}{k})$$  \hspace{1cm} (17)
3.2 Training and validation data

The annual numbers of wind speed data from 2012 until 2014 were 51289, 27751 and 52413, respectively. In the year 2013, the wind speed data for the months May to October was missing due to sensor problems. The data in 2012 was used as the training set, while the data for 2013 and 2014 were used for validation, to assess the methods for their suitability to fit and predict wind speed frequencies on the Phangan island.

4. Results and Discussion

The probability densities for the actual wind speed data and the fits by the five methods, from 2012 to 2014, are shown in Figures 1 to 3. The graphical method had subjectively the worst performance in fitting the actual wind speed data. The Weibull parameters k, c and percent error of power density are shown in Table 1. The maximum likelihood method gave a c value nearly equal to the average wind speed, and was followed by the modified maximum likelihood, the empirical method, the energy pattern factor, and the graphical method in this order. However, the percent error of power density was least (best) with the energy pattern factor method, followed by the empirical, the maximum likelihood, the modified maximum likelihood, and the graphical method in this order. These rankings match the $R^2$ values. The Kolmogorov-Sminov test was in the range of 0.036-0.223. By the root mean square error the graphical method was the worst.

![Histogram of actual wind speed distribution and Weibull fits from various methods, for the year 2012-2014](image)

Table 1. The parameter estimates k, c and percent error of power density (PE) from various methods, for each year separately.

<table>
<thead>
<tr>
<th>Method of Estimation</th>
<th>2012</th>
<th></th>
<th>2013</th>
<th></th>
<th>2014</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k</td>
<td>C</td>
<td>PE</td>
<td>k</td>
<td>C</td>
<td>PE</td>
</tr>
<tr>
<td>Energy pattern factor</td>
<td>1.891</td>
<td>4.470</td>
<td>0.420</td>
<td>1.724</td>
<td>4.150</td>
<td>0.143</td>
</tr>
<tr>
<td>Average wind speed from actual</td>
<td>3.964</td>
<td></td>
<td>3.697</td>
<td></td>
<td>3.934</td>
<td></td>
</tr>
</tbody>
</table>
5. Conclusions

Based on wind speed data from 120 m above sea level on Phangan island, Thailand, from January 2012 to December 2014, the maximum likelihood method performed best in conserving the average wind speed from the data to the fit. The energy pattern factor method had the lowest percent error in power density and the highest $R^2$. The scope of this study was limited to using the Weibull distribution, while further assessments with other types of distribution could be considered.

Acknowledgements

The authors would like to thank the National Research Council of Thailand and RDO, Prince of Songkla University, and also research funding from the Prince of Songkla University, Suratthani Campus. Assoc. Prof. Dr. Seppo Karrila provided helpful suggestions to improve a draft manuscript.

References