Wind Data Analyzer

Ziad Hatahet

March 2006

Contents

List of Figures

1 Introduction

Wind power refers to extracting useful energy from wind. The cost of electrical power generated by wind is continually dropping, although it remains higher than the conventional fuel-generated counterpart. The cost of wind power in the late 1990s was about five times the cost of what it was in 2005; and as larger multi-megawatt turbines are mass-produced, that trend is expected to continue[1].

Finding a model to describe wind speed variation is of paramount importance, as it used to estimate the amount of energy earnings, as well as to optimize the design of wind turbines. The Weibull and Rayleigh distribution models are commonly used for this purpose[2] – the latter being a special case of the former.

The aim of this project is to calculate the Weibull distribution parameters for wind speed variations spanning several days, and subsequently use these parameters to estimate the wind power earnings.

2 Theory

This section highlights the theoretical concepts used in this project.

2.1 The Weibull Distribution

The Weibull distribution has many applications in survival analysis, and reliability and biomedical engineering. Furthermore, it is widely used to model variations in wind speed. This distribution exists in two main forms: the two-parameter, and three-parameter Weibull distribution.

2.1.1 The Two-Parameter Weibull Distribution

The two-parameter Weibull distribution has the following density and distribution functions respectively:

$$f(x) = \left(\frac{b}{a}\right)\left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^b}$$
$$F(x) = 1 - e^{-\left(\frac{x}{a}\right)^b}$$

where $x \geq 0$, and $a > 0$ is the *scale parameter* and $b > 0$ is the *shape parameter* of the distribution. Both the exponential and Rayleigh distributions are special cases of the Weibull distribution. Setting $b = 1$ in the Weibull density function results in an exponential density function with parameter $\lambda = \frac{1}{a}$. The Rayliegh distribution is a special case of the Weibull distribution with $a = 2$.

The n^{th} raw moment for the Weibull distribution is given by the following moment generating function:

$$m_n = a^n \Gamma\left(1 + \frac{n}{b}\right)$$

From the above equation, the mean and variance respectively are:

$$E[X] = a\Gamma\left(1 + \frac{1}{b}\right)$$
$$Var[X] = a^2\left\{\Gamma\left(1 + \frac{2}{b}\right) - \Gamma^2\left(1 + \frac{1}{b}\right)\right\}$$

2.1.2 The Three-Parameter Weibull Distribution

The density and distribution functions of the three-parameter Weibull distribution respectively are:

$$f(x) = \left(\frac{b}{a}\right)\left(\frac{x-c}{a}\right)^{b-1} e^{-\left(\frac{x-c}{a}\right)^b}$$
$$F(x) = 1 - e^{-\left(\frac{x-c}{a}\right)^b}$$

where $x \geq 0$, $a > 0$, $b > 0$, and $-\infty < c < \infty$. The difference between this and the two-parameter Weibull distribution is the introduction of the *location parameter* c. The mean and variance respectively are:

$$E[X] = c + a\Gamma\left(1 + \frac{1}{b}\right)$$
$$Var[X] = a^2\left\{\Gamma\left(1 + \frac{2}{b}\right) - \Gamma^2\left(1 + \frac{1}{b}\right)\right\}$$

The three-parameter Weibull distribution is more general than its two-parameter counterpart. It is effective for modeling days with higher wind speeds as the location parameter is used to *shift* the distribution to the higher speeds. However, one main drawback is that it is more difficult (maybe even impossible) to determine the parameters for this distribution using the technique described in section 2.3.1.

2.2 Histogram

A histogram of the empirical data is used to visualize the general distribution of these data. A relative histogram is constructed in order to compare the distribution of the wind speed variation with the computed Weibull distribution. An important concept in this process is choosing an appropriate bin width to realize the histogram, as the former determines the general shape of the latter.

In our case, we found that a bin width equal 0.5 m/s or 1 m/s produces good results to compare the relative histogram to the computed fit. It should be noted that the width of the bins will not have any effect on the shape of resulting fit. The sole purpose of the histogram – in this project – is to use it to visually compare the fit with the distribution of the wind speed.

2.3 Parameter Estimation

This is the *heart* of the procedure. This section describes the techniques used to estimate the parameters of the Weibull distribution function. The techniques used herein operate directly on the collected data, and hence are not effected by the bin width used in the histogram.

2.3.1 Maximum Likelihood

The first method used is the Maximum Likelihood Estimation (MLE). In general, this technique is preferred to the alternative method implemented in the program as MLE is usually considered to be more robust and produces more accurate results. The basic idea behind MLE is to find the most likely values of the parameters for a given distribution that would best describe the data. Given a probability density function (PDF):

$$f(x; \theta_1, \theta_2, \ldots, \theta_n)$$

where $\theta_1, \theta_2, \ldots, \theta_n$ are the distribution parameters required to be estimated. After drawing k independent samples out of the distribution, we could seek an estimate of the values of $\theta_1, \theta_2, \ldots, \theta_n$. Let the likelihood function be defined as:

$$lik(\theta_1, \theta_2, \ldots, \theta_n | x_1, x_2, \ldots, x_k) = \prod_{i=1}^{k} f(x_i, \theta_1, \theta_2, \ldots, \theta_n)$$

Maximizing this function over all possible values of θ_j will enable us to obtain the *maximum likelihood estimators* $\hat{\theta_1}, \hat{\theta_2}, \ldots, \hat{\theta_n}$ of parameters $\theta_1, \theta_2, \ldots, \theta_n$.

A number of properties make MLE attractive for use: it is *asymptotically consistent* – as the sample size increases, the parameter estimates converge to the right values. It is *asymptotically efficient* – it produces the most precise estimates for large samples[3]. Furthermore, it is *asymptotically unbiased* – one expects to get the right values on average for large samples. It should be noted however,that MLE may not be precise with small sample sizes. Fortunately, this should not be an issue in our case.

The MLE for the Weibull distribution cannot be expressed in closed form[4]; hence, iterative computational methods are used to determine the parameters. Furthermore, the MLE for the three-parameter Weibull distribution is not always stable, and may collapse if $b \sim 1$. To overcome this issue, we shall use a separate procedure to estimate the location parameter before using MLE to determine a and b.

2.3.2 Least Squares

Although the term *least squares* is usually used to refer to fitting a regression line, it is going to be used in a different context at this juncture. The method herein describes a different application of least squares: univariate distribution fitting[5].

The Weibull cumulative distribution function (CDF) is going to be transformed such that it will have a form resembling that of the equation of the straight line. Then, the empirical CDF (ECDF) values will be plugged into the resulting equation, producing a (mostly) straight line. The "straighter" the line, the more do the data tend to follow the Weibull distribution.

After that, we use least squares to fit a straight line that passes through these points. This line represents the Weibull distribution that is "closest" to the data. From the resulting fit, we can extract the Weibull parameters by performing the inverse of the transformation. The transformation is illustrated below:

$$\ln(x) = \frac{1}{b}\ln(-\ln(1 - F(x))) + \ln(a)$$

Comparing the transformation with the equation of the straight line, we can notice that the slope of the line is $\frac{1}{b}$ and the y-intercept is $\ln(a)$. Hence, the reciprocal of the slope and the exponential of the y-intercept are the b and a parameters of the Weibull distribution respectively.

2.3.3 The Location Parameter

In the case we want to fit the data to the three-parameter Weibull distribution, an additional procedure must be added to estimate the location parameter, as is difficult to use MLE for this case. The procedure will be used regardless of whether we intend to use MLE or least squares to fit our data. As previously mentioned, the purpose of the location parameter is to shift the distribution on the x-axis. Given the fact that we don't know the value of this parameter, we will use linear regression to estimate it. Our purpose is to try to find a value c such that the relation between $ln(x - c)$ and $\ln(-\ln(1 - F(x)))$ is a linear one. Once this parameter is found, it will be subtracted from the data, and subsequently MLE or least squares will be used to estimate the a and b parameters.

2.4 Goodness-of-Fit

Due to the general consensus that the Weibull distribution provides a good model for wind speed variation, no goodness-of-fit tests (such as Chi-square or Kolmogorov-Smirnov tests) were implemented in the program. The program does however provide the ability to view the empirical histogram and the computed Weibull fit superimposed on the same graph for comparison purposes. The same functionality is provided for the empirical and computed CDF. Furthermore, the mean and variance of the empirical data and the resulting fit are displayed. For the case of the least squares method, it is possible to view the plot of the empirical data and the straight line fit to visualize to what extent the data follow the Weibull distribution.

3 Energy and Power Calculation

Once the Weibull parameters have been estimated, we are able to proceed with calculating the energy and power gains.

3.1 Energy

The effective energy (in $\frac{W}{m^2}$) is given by:

$$E_{effective} = \frac{1}{2}\rho\, m_3 = \frac{1}{2}\rho \left\{ a^3 \Gamma \left(1 + \frac{3}{b} \right) \right\}$$

where ρ is the air density constant (varies with temperature), and m_3 is the third raw Weibull distribution moment. This equation provides the energy earnings at any one specific wind turbine location.

3.2 The Machine Curve and Power

The power gain depends on the properties of the turbine used to "extract" it. These are depicted by what is known as the *machine curve*. This curve shows the amount of power generated at a specific wind speed. The one used in this project is shown below[6].

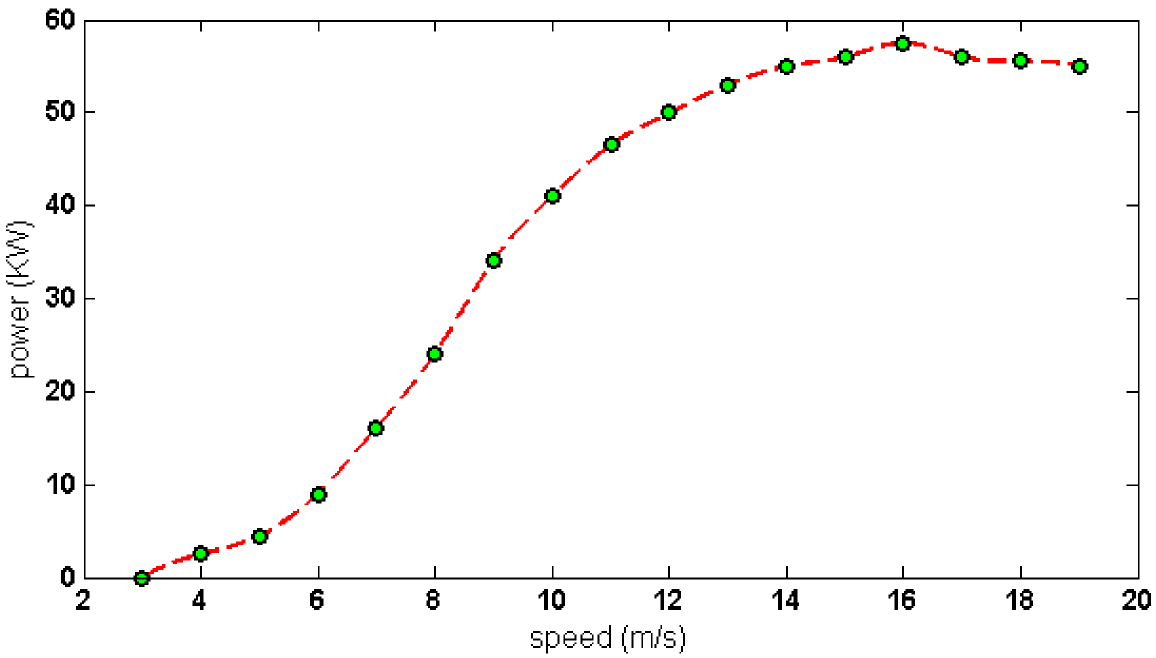

Figure 1: Machine Curve $\Phi(u)$

The resulting power (in W) is given by:

$$P = \int_0^\infty f(u) \cdot \Phi(u)\, du$$

where $f(u)$ is the computed Weibull fit PDF, and $\Phi(u)$ is the machine curve at a given wind speed u.

4 The Program

This section describes the features of the program implemented in this project. The program was written using MATLAB® 7 R14 SP2 with GUIDE v2.5 and uses the features of Statistics Toolbox 5.0.2.

4.1 The User Interface

The program incorporates a user interface to simplify interaction with its features for the general user. When the program is launched, the user is presented with the following window:

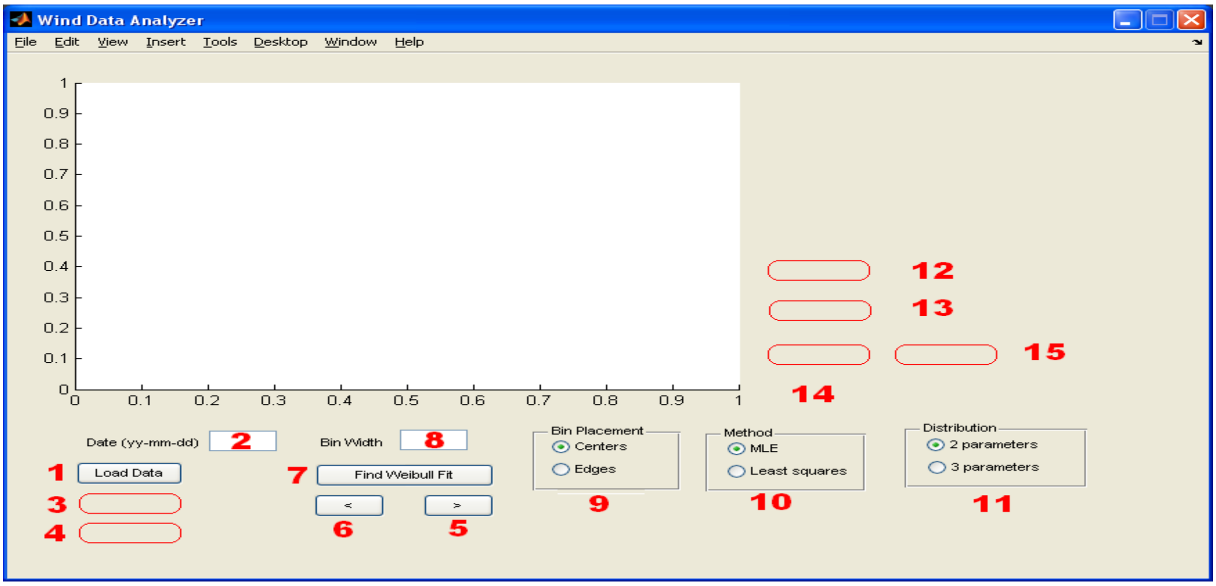

Figure 2: The User Interface

1 – Load Data pushbutton Opens a dialog box to allow the user to select the input file. Once the file is loaded, the program lists the range of dates included in the file.

2 – Date text field After loading the file, the user can select a specific date to work on.

3 – Wind Rose pushbutton (*not shown*) Displays a plot of the wind rose for the day specified in the *Date* text field.

4 – Speed vs. Time pushbutton (*not shown*) Displays a plot showing the variation of wind speed according to time for the day specified in the *Date* text field.

5 – (>) pushbutton Steps to the next date in the file (if possible).

6 – (<) pushbutton Steps to the previous date in the file (if possible).

7 – Find Weibull Fit pushbutton Calculates the Weibull distribution parameter estimates for the day specified in the *Date* text field. Furthermore, it displays the resulting fit superimposed on the histogram of the empirical data. The results depend on the options set in the *Bin Width* text field, and the *Bin Placement*, *Method*, and *Distribution* radio buttons.

8 – Bin Width text field Sets the bin width of the histogram of the wind speed.

9 – Bin Placement radio buttons Select whether the bins are going to contain events centered on the bin width (and its multiples), or events occurring between consecutive bin width multiples.

10 – Method radio buttons Select what parameter estimation method to use.

11 – Distribution radio buttons Select whether to use the two-parameter or three-parameter Weibull distribution to fit the data.

12 – Show CDF pushbutton (*not shown*) Displays a plot of the ECDF. If a fit was calculated (*Weibull Fit* pushbutton), the CDF of the resulting fit is also displayed.

13 – Show straight line fit pushbutton (*not shown*) In case the *Least Squares* method was used, this pushbutton displays a plot of the empirical data and resulting straight line fit.

14 – Generate Table pushbutton (*not shown*) This makes the program go through the selected input file, computing the Weibull parameters, and the energy and power gains for each day. The results are displayed in the prompt as well as being written to a file.

15 – Energy Variation pushbutton (*not shown*) This button plots the energy gain for each day in the input file. It is visible after the *Generate Table* pushbutton had been selected.

4.2 The Input File

The program was written to work with ASCII files with a specific layout (comma separated values) containing nine columns. However, only three fields are used: date (2$^{\text{nd}}$ column), speed (6$^{\text{th}}$ column), and direction (8$^{\text{th}}$ column). The code was written in a modular way such that it should be easily adaptable to work with files of other formats/layouts if necessary, as string manipulation is done in separate functions.

4.3 Sample Session

Note: Clicking on any figure that appears on the user interface will make it display in a new window, providing the ability to export the plot, change its size, etc...

Clicking on the *Load Data* pushbutton will display a typical dialog box allowing the user to select the input file.

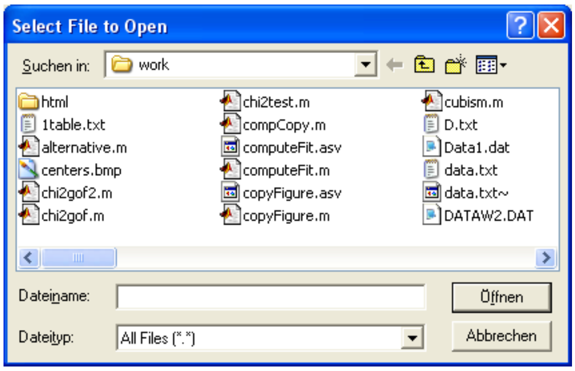

Figure 3: Select Input File

Once the file is selected, three events occur: the first date in the file is automatically loaded into the *Date* text field; the range of the dates contained in the file is displayed in the prompt; and four pushbuttons appear – *Wind Rose*, *Speed vs. Time*, *Show CDF*, and *Generate Table*.

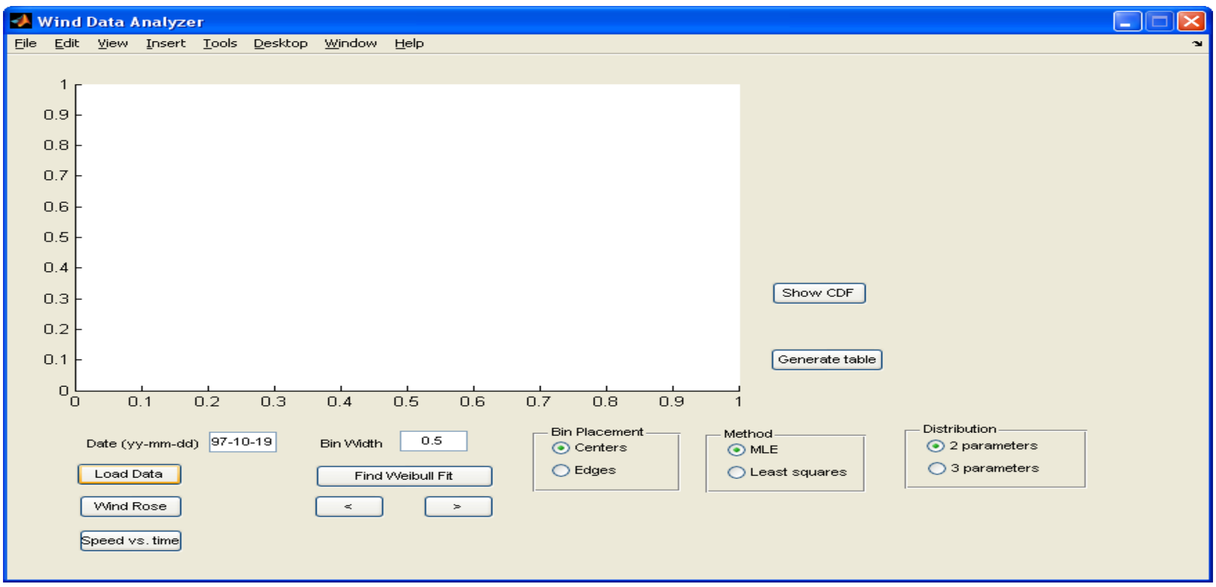

Figure 4: Post-file Loading

It is now also possible to step through the dates in the file using the (<) and (>) pushbuttons. The *Wind Rose* pushbutton will display a rose plot of the wind speed for the day specified in the *Date* text field. Similarly, the *Speed vs. Time* pushbutton will display a plot showing the speed variation with respect to time for the specified day.

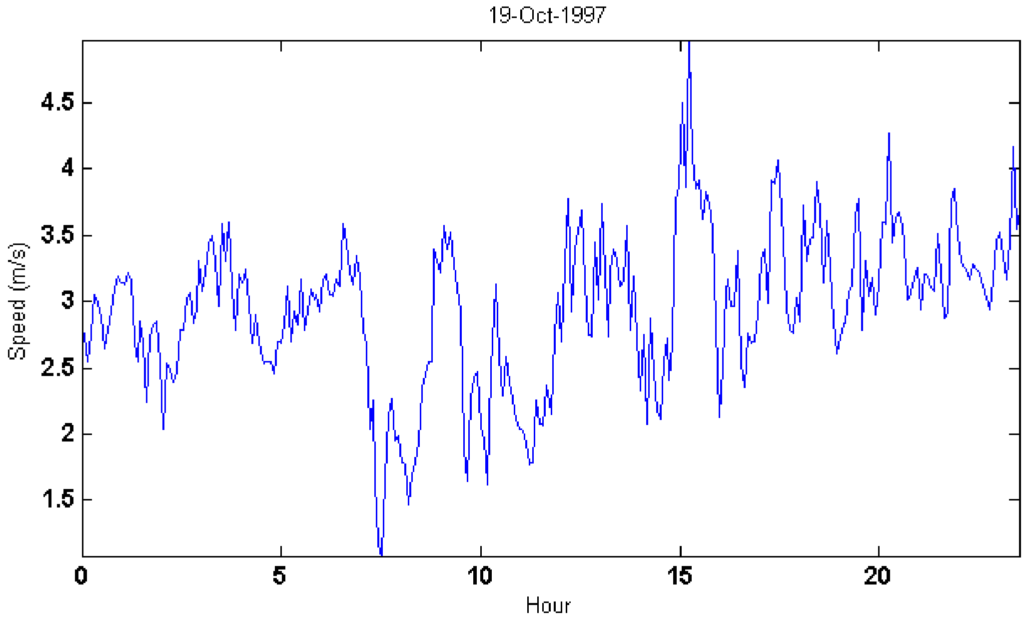

Figure 5: Speed vs. Time (24 hours)

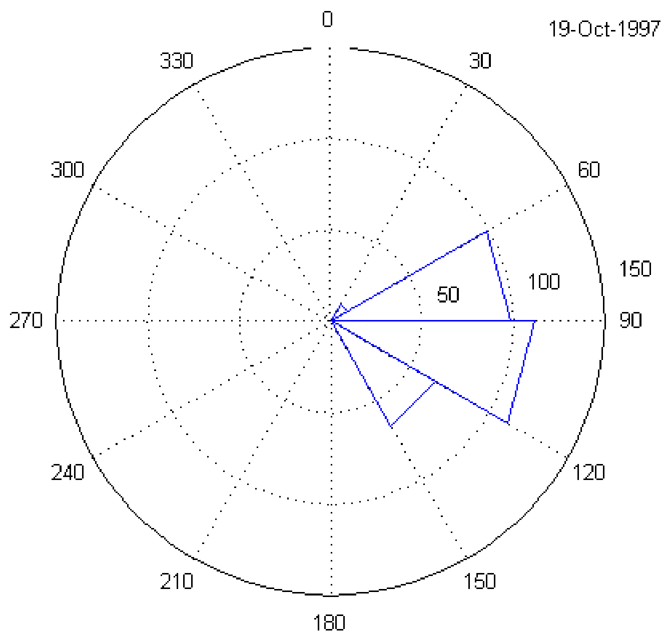

Figure 6: Wind Rose

Since the Weibull fit is not yet computed, the *Show CDF* button will only display the ECDF. Clicking on the *Generate Table* pushbutton will cause the program to pass through the whole input file, generating a table containing the estimated Weibull parameters, the empirical mean, the mean from the fit, the energy gain, and the power gain for each day. The table is displayed on the prompt as well as stored in a file – table.txt – in the present working directory. The generated table is affected by the selections in the *Method* and *Distribution* radio button groups. For example, the default selection of *MLE* and *2 parameters* will generate a table with the two parameters of the Weibull distribution estimated using MLE.

```
Command Window
Generating...
--------------------------------
Parameter estimation method: MLE
--------------------------------

Day            a          b          Fit mean(m/s)  Emprcl mean  Energy(W/m^2)  Power(W)  Samples
-----------------------------------------------------------------------------------------------------
19-Oct-1997    3.1768     5.7002     2.9387         2.9457       18               611     288
20-Oct-1997    5.0899     2.4780     4.5152         4.5643       91              6055     288
21-Oct-1997    1.7282     1.3120     1.5932         1.6157       9                365     288
22-Oct-1997    4.1524     3.9559     3.7614         3.7585       41              2294     288
23-Oct-1997    7.8042     3.4039     7.0115         7.0192       283            18877     288
-----------------------------------------------------------------------------------------------------
Done
>> |
```

Figure 7: Generated Table (5 days)

Now, the *Energy Variation* pushbutton becomes visible. Clicking on it will plot the energy gain for each day in the input file.

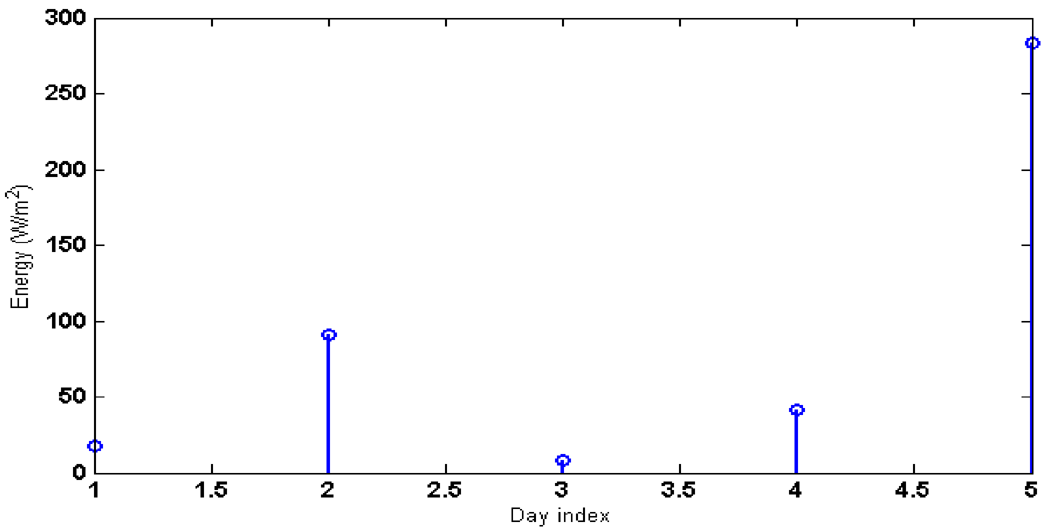

Figure 8: Energy Variation

The *Find Weibull Fit* pushbutton will calculate the Weibull parameter estimates for the day specified in the *Date* text field, taking into consideration the *Method* and *Distribution* radio button groups. Once the fit is found, the resulting PDF is displayed superimposed on a histogram of the empirical data (specified by the *Bin Width* text field). This allows the user to visualize how good (or bad) the data follows the Weibull distribution. The resulting parameter estimates, along with the empirical and computed mean and variance are displayed to the right of the plot.

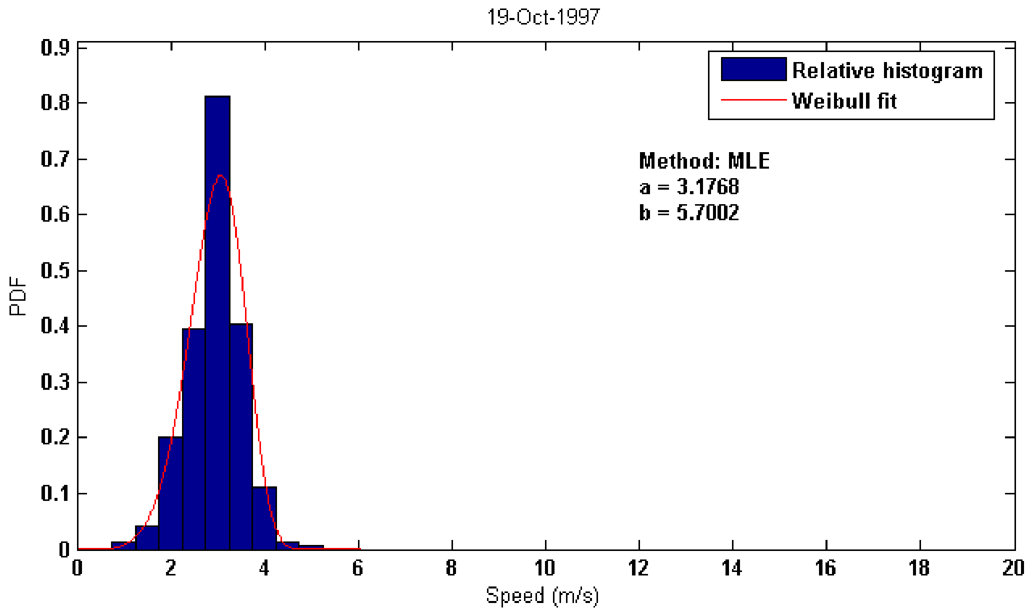

Figure 9: Histogram and Resulting Fit

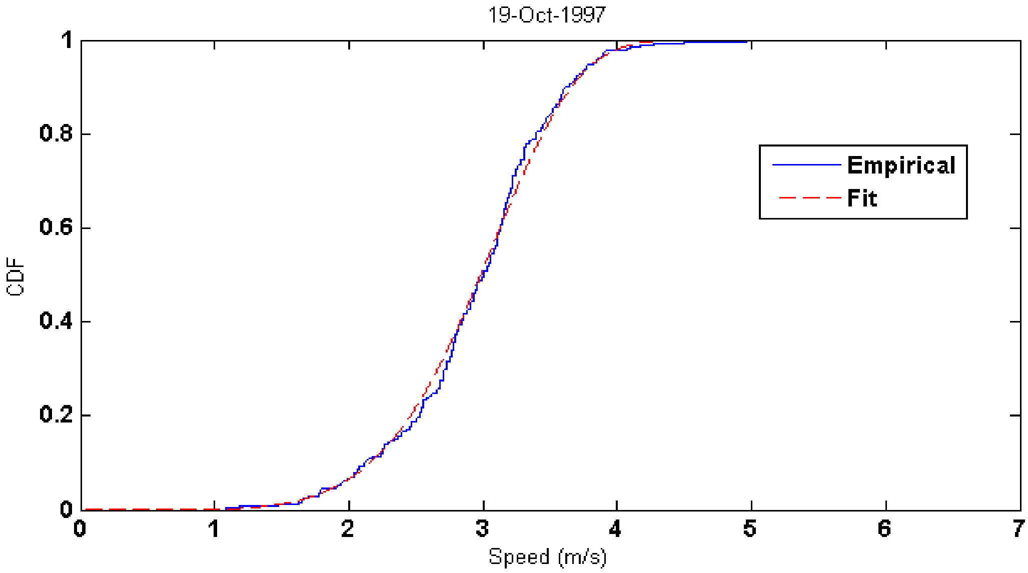

Figure 10: The ECDF and the CDF from Fit

In case the *Least Squares* method was chosen, the *Show straight line fit* pushbutton appears. This displays the empirical data and the Weibull fit after applying the transformation described in section 2.3.2. The degree the data points match the line signifies to what extent they follow the Weibull distribution.

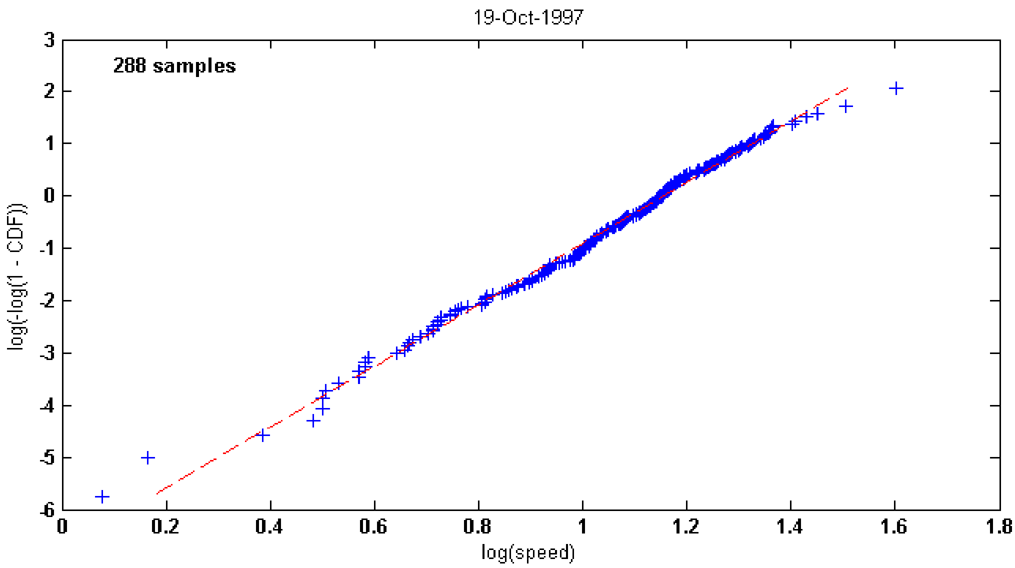

Figure 11: Least Squares Fit

Notes

[1] http://en.wikipedia.org/wiki/Wind_power

[2] http://www.windpower.org/en/tour/wres/weibull.htm

[3] http://www.weibull.com/LifeDataWeb/mle_for_complete_data.htm

[4] http://www.quantlet.com/mdstat/scripts/csa/html/node196.html

[5] http://www.mathworks.com/products/statistics/demos.html?file=/products/demos/shipping/stats/cdffitdemo.html

[6] Europaischer Windatlas, pg. 82