PERFORMANCE OF LONG TERM WIND ESTIMATION METHOD AT WIND POWER DEVELOPMENT

AN ANALYSIS OF RESULTS FROM DIFFERENT WIND ESTIMATION METHODS

Master of Science Thesis in the Master Degree Program Industrial Ecology for a Sustainable Society

In collaboration with Triventus Consulting AB

EMIL SVENSSON

Department of Energy and Environment

Division of Electrical Power Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden, 2012

PERFORMANCE OF LONG TERM WIND ESTIMATION METHOD AT WIND POWER DEVELOPMENT

AN ANALYSIS OF RESULTS FROM WIND ESTIMATION METHODS

EMIL SVENSSON

Department of Energy and Environment

Division of Electric Power Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden, 2012

Examiner: Ola Carlson, Electrical Power Engineering

PERFORMANCE OF LONG TERM WIND ESTIMATION METHOD AT WIND POWER DEVELOPMENT

AN ANALYSIS OF RESULTS FROM WIND ESTIMATION METHODS

EMIL SVENSSON

Department of Energy and Environment

Division of Electrical Power Engineering

Chalmers University of Technology

SE-41296 Göteborg

Sweden

Telephone: +46 (0) 31-772 1000

Göteborg, Sweden 2012

Abstract

The purpose of this thesis is to compare wind estimation methods such as Wind Index Linear Regression and Weibull in order to minimize the uncertainties. This is done by perform wind estimations using all methods at different locations and compare the errors and finally determine the magnitude of correlation between the errors.

The main findings were that Linear Regression was the most accurate method and that Wind Index was the least accurate. Linear Regression correlated with Wind Index on all investigated locations and it is of great importance to check correlation between measured and reference data.

Acknowledgement

I would like to express my gratitude to the following

Triventus Consulting: For giving me the opportunity to do this thesis and helping me with guidance along the way.

Jakob Sondegger: For good times in the office and

Fredrik Pettersson: For guidance through the databases.

Göteborg, November 2012

Emil

Table of content

1 Introduction

Global warming is on the agenda in almost every country around the world and there are and the ambition is to develop technologies that reduce our environmental impact. An available and rather mature technology is wind power, and countries such as Denmark and Germany has a significant part of the electricity produced from wind power. This is due to political decision that makes it favorable to produce energy from wind power in these countries. However, the development has not increased in the same rate in Sweden, although, the Swedish government decided in June 2009 to state a plan to increase the energy produced from wind power with 30 TWh[1]. This was a further development of a decision taken in 2002 that stated an expansion of 10 TWh by 2015[1].

In order to reach these goals the Swedish government need to develop an energy market that attract investors into the wind power industry. However, it is also of importance that the industry itself is eager to develop the technology, and with that increase establishment of wind power projects.

As most industries the wind power industry face a number of technical and political issues. A crucial issue that will be discussed in this thesis is the prediction of wind. The reason for why it is so important is that the wind forecast is used in order to determine whether a project is economically beneficial or not. Improved forecast could be a parameter that could increase investments since it would minimize the economical risk. This task involves both technical challenges such as development of equipment used to measure winds and mathematical issues such as improvement of statistical methods.

As mention earlier, an increase of wind power in the Swedish electricity system depend on a number of factors but this study will focus on statistical methods that are used to estimate wind resources. The objective of this thesis is to compare methods that are currently used to estimate wind resources. Methods that will be compared are Weibull, Linear Regression and Wind Index, and the ambition is to find relations between them.

1.1 Background

When planning to erect a wind turbine at a specific location it is crucial to determine wind resource of the location. This is done with wind measured on the site. However, measurements can only be carried out for a shorter time, typically one year, in comparison to the life time of a wind turbine, which usually is twenty years. This implies that a forecast of future wind resource is required in order to determine whether a location is suitable or not.

The most common method used is Linear Regression, but other methods such as Weibull, Wind Index and Matrix method is also used. All methods are rather simple statistical tools, which more or less are based on a linear relationship, and since winds are rather complex and hard to understand the models involves significant uncertainties. In order to reduce these uncertainties the result of these methods can be compared and result in statistical relation between independent methods can be identified. It should also be said that a forecast is not only dependent on the statistical method, both measured and reference data is also influencing the estimation.

1.2 Aim

The aim of this thesis is to find correlation coefficient between errors calculated from different wind estimation methods.

2 Analytical framework

In order to make a comprehensive comparison of the methods used for long term forecasting one need deep understanding of the methods and how they are applied. This section will first discuss available reanalysis data used as reference for estimation of wind resources. Later will it describe the main process of long term estimation of wind, and also deal with the theory associated with Weibull distribution, Linear Regression and Wind Index. The section will end up with description of other necessary theoretical terms such as; U-wind, V-wind and interpolation, in order to give the reader a complete view of wind resource estimations.

2.1 Atmospheric reanalysis data and measured data

Atmospheric reanalysis data and data from actual measurements are the foundation of wind resource estimations. The reason for that is that all wind estimations are based on a relationship between measured data and reference data.

2.1.1 Atmospheric reanalysis

Atmospheric reanalysis/reference data is based on a synthesis of worldwide observations that is combined in an atmospheric model and results in a global three dimensional grid of wind conditions. Reanalysis data covers a long time interval and is therefore suitable as long-term reference data for predicting future wind climate[2].

There are a number different Reanalysis data available but the most frequently used are NCAR (National Center for Atmospheric Research), NCEP (National Center for Environmental Prediction), MERRA (Modern Era Retrospective-analysis for Research Applications) and CSFR (Climate Forecast System Reanalysis).

For a person that is not familiar with wind estimations, the differences between these data sets can be explained as a variety in time resolution, how measurements are carried out and number of grid points around the globe. However, there could also be great differences in how data are collected and in the construction of the models[2].

2.1.2 Measured data

The measured data, or site target data, is wind speed measures at a specific location and height. These measures have much shorter time perspective than reference data and have also generally a higher time resolution.

Wind measurements are often carried out by putting up a mast at a location that corresponds to the conditions where the wind turbine/turbines are planned. Anemometers are positioned at different heights on the mast, the most common timeframe is one year and the time resolution is usually hours.

Further is measured data considered as more accurate than atmospheric reanalysis data and therefore is always all measured data points used. However, it is necessary to check measured data in order make sure error is left out of the estimation.

2.2 Main procedure

The general process of estimating wind is almost the same for all three methods. Figure 1 illustrates a simplified process chain of the considered MCP-methods. The process surrounded by a blue frame is linked to the Linear Regression model and the entire process chain is linked to both Weibull and Wind Index model. The major difference that can be distinguished is that Weibull and Wind Index involves the long term reference data, which not is involved in the Linear Regression model. This project will focus on the whole process for all the different methods.

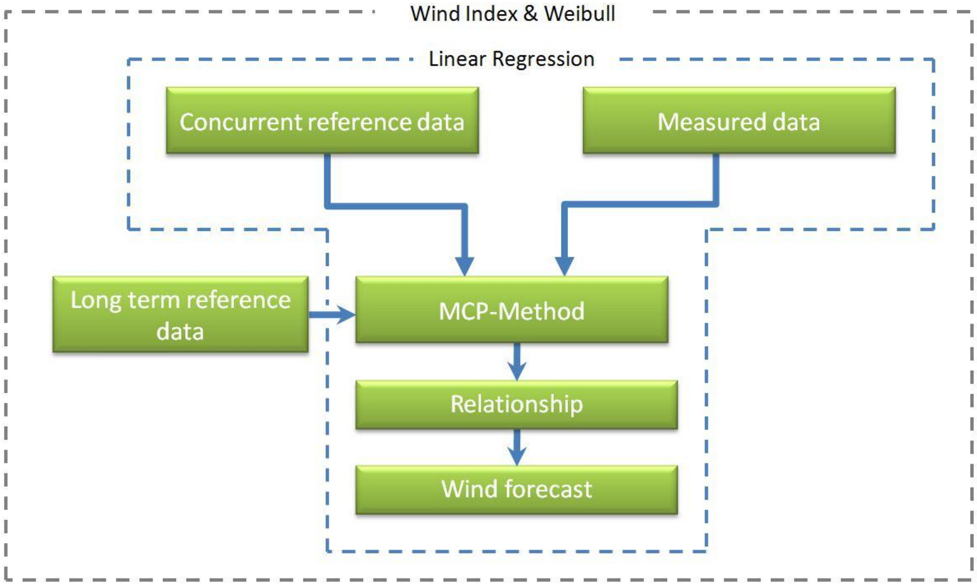


Figure 2.2 Main procedure for wind estimations

2.2.1 Linear Regression

Linear Regression is commonly used as a statistical tool in a number of industries, and it is because of the fact that it is a rather simple and relievable method. It is a method where a linear function is fitted to the available data according to equation (1)[3].

$$Y = \alpha + \beta x \ (1)$$

When using Linear Regression the objective is, as mentioned earlier, to fit line between available data. This is done by using least square method, equation (2). Both parameter α and β can be determine by setting wind speed from reference data as x-values and measured data as y-values[3].

$$\beta = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \ (2)$$

Fig. 2.2.1 illustrates an example of a plot of reference wind speed versus measured wind speed. The line is a fitted line calculated by using equation (1) and (2).

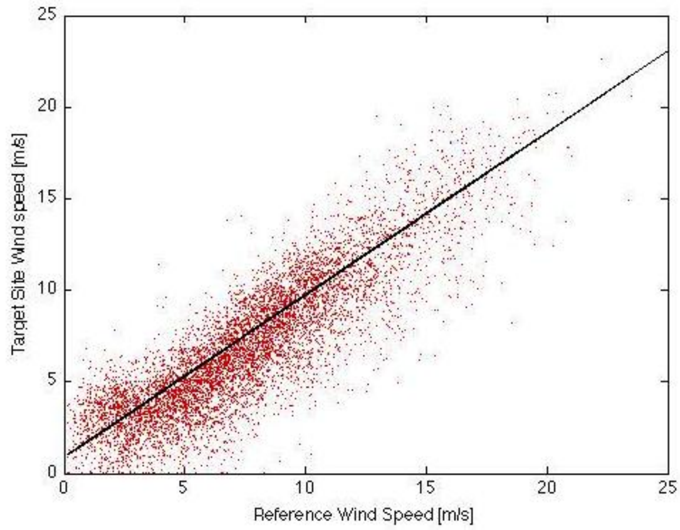


Figure 2.2.1 Example of Linear Regression

2.2.1.1 *Residual analysis*

After a linear function is fitted it is necessary to look at residuals in order to validate the model. The residual from a regression model are $e_i = y_i - \hat{y}_i$, where y_i is the actual measured value and $\hat{y}_i$ is the corresponding fitted value from the regression model. By plotting residuals in time sequence it is possible to evaluate the model. These graphs will usually look like one of the four general patterns shown in figure 2.2.1.1. Pattern (b) in figure 1 illustrates the ideal situation, while patterns (a), (c) and (d) represent deviation from randomized pattern illustrated in (b). If residuals appears as in (a), the variance of the observations may be increasing with time or magnitude. Residual plots that look like (c) indicate model inadequacy; that is, higher order terms should be added to the model. Finally is pattern that looks like (d) in figure indicates inequity of residuals[4].

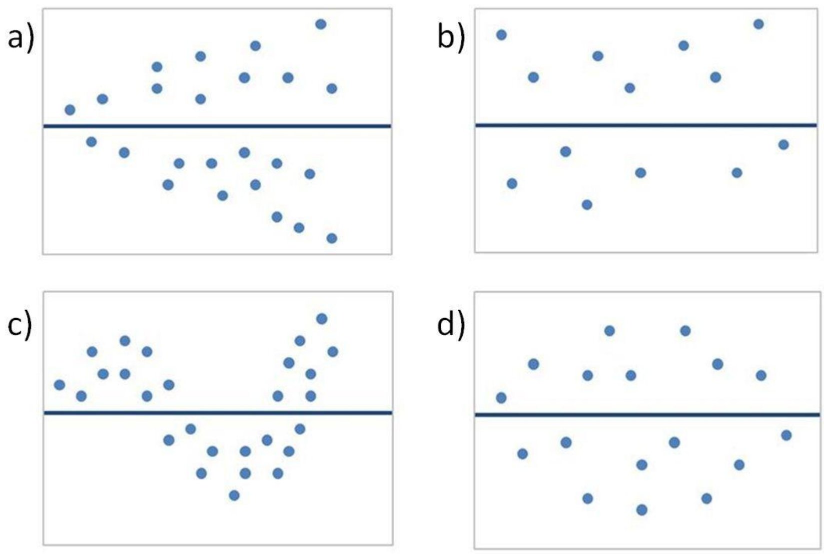


***Figure* 2.2.1.1 *Patterns for residual plots. (a) Funnel, (b) Satisfactory, (c) Nonlinear and (d) double bow*[4].**

If desirable pattern of residuals are obtain, a residual parameter is added to the regression function, equation (3)[5].

$$Y = \alpha + \beta x + e \quad (3)$$

Where e are randomized zero-mean Gaussian distributed residuals[5].

2.2.2 Correlation

When α, β and residuals are checked it is possible to determine the correlation coefficient. The correlation coefficient is determined in order to explain if the estimated line describes the input data. A value of 0 describes that there is no correlation between x and y and 1 or -1 describes that there is a linear or negative linear, correlation, and it is determined by using the following equation[3].

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right)\left(\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right)}} \quad (4)$$

2.2.2.1 Applying Linear Regression on wind estimations

When using Linear Regression in wind estimations it is necessary to divide both measured and reference wind speed into wind directions. Usually this is done by a 30 degrees window, starting at 345 degrees, and constantly moving 1 degree clockwise, see figure 2.2.2.1, until it eventually end up where it started. In total 360 sectors is created and a Linear Regression is done for all 360 sectors[5].

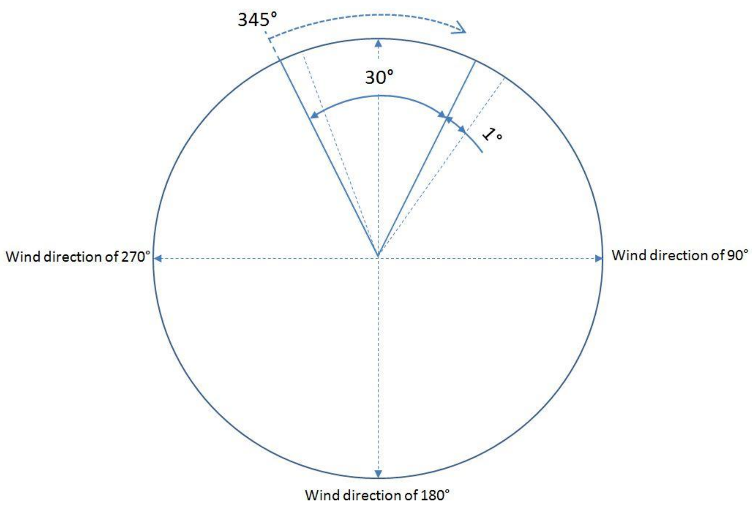


Figure 2.2.2.1, Illustration of how wind speed is divided into wind direction

The artificial series for target site is then created by putting in long-term reference data into equation (3) for every sector, and the residuals (e) are randomized from a zero mean normal distribution[5].

2.2.3 Weibull

Weibull is an empirical method that includes a linear manipulation in form of Weibull scale and shape parameters (A, k) as well as adjustment on the frequency distribution. Both shape and scale factor are important parameters since they got a great impact on the frequency and cumulative Weibull distribution.

The shape factor can simply be set to a specific value if the distribution is known, but in wind estimations it is usually estimated. There are a number of methods used to estimate this parameter, but the method used in this thesis is called the empirical method, which incorporate an evaluation of measured wind according to formula (5). However, when Weibull is used in a wind prediction analysis the k-value in most cases is very close to 2, which is known as the Raleigh distribution[6]. When shape factor is calculated it is possible to determine scale factor by using equation (6).

$$k = \left(\frac{\sigma}{v_{mean}}\right)^{-1{,}086} \quad (5)$$

$$A = \frac{v_{mean}}{\Gamma(1 + 1/\mathrm{k})} \quad (6)$$

Where: σ is the standard deviation and v_{mean} is the average wind speed.

After the shape factor and scale factor are estimated it is possible to calculate the Weibull distribution. It is described through the probability density f(v) or/and cumulative distribution function F(v) given as:

$$f(v) = \frac{k}{A}\left(\frac{v}{A}\right)^{k-1} exp\left[\left(-\frac{v}{A}\right)^{k}\right] (7)$$

$$F(v) = 1 - exp\left[\left(-\frac{v}{A}\right)^{k}\right] (8)$$

Where v is the wind speed, k the dimensionless shape parameter, and A is the scale parameter having the same unit as v[6].

2.2.3.1 Applying Weibull distribution in wind estimations

When Weibull is used to estimate wind resources for a site it is necessary to sort out both reference and target wind speeds in wind direction. It is almost the same procedure as in Linear Regression but instead of divide data into 360 sectors, 12 sectors are created[5] (See *Figure 2.2.3.1*).

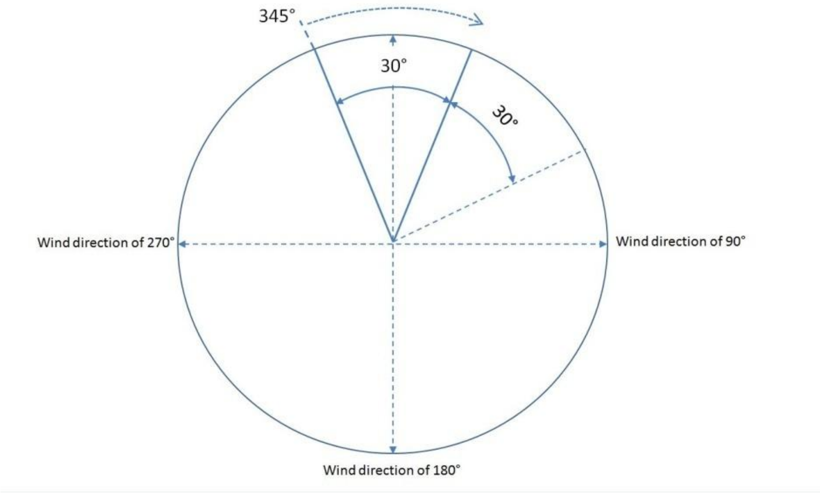


Figure 2.2.3.1 Illustration of how wind speed is divided into wind direction

When Weibull is used it is of great importance to consider the defined underlying assumptions. The main assumption is, that it exist a monotonic relationship between the concurrent reference and target wind speeds, which means that you assume a function between the data that is consistently increasing and never decreasing. This means also that it is possible say that the values of their cumulative distribution are the same for concurrent reference and site data.

If site target wind speeds is well correlating with wind speeds from reference data a reasonable approximation is that; the cumulative distribution of site data comports with the cumulative reference data. Thereby is it possible to state the following equations:

$$F_x(x) = F_x(x) <=> 1 - e^{\left(\frac{x}{A_x}\right)^{k_x}} = 1 - e^{\left(\frac{y}{A_y}\right)^{k_y}} => y = A_y \left(\frac{x}{A_x}\right)^{\frac{k_x}{k_y}} \quad (9)$$ [6]

Where

k_x and k_y are the reference and target site shape factors

A_x and A_y are the reference and target site scale factors

According to equation (9) it is possible to state a linear relationship between reference and target wind speed if shape factor k_x and k_y are equal. However, the relation is non-linear in cases where k_x *and* k_y differ from eachother.

Further could this be extended by introducing location parameters; θ_x , θ_y for reference and target site wind speeds. This is explained by equation (10) below.

$$1 - e^{\left(\frac{x}{A_x}\right)^{k_x}} = 1 - e^{\left(\frac{y}{A_y}\right)^{k_y}} <=> y = A_y \left(\frac{x-\theta_x}{A_x}\right)^{\frac{k_x}{k_y}} + \theta_y \ (10)^{6}$$

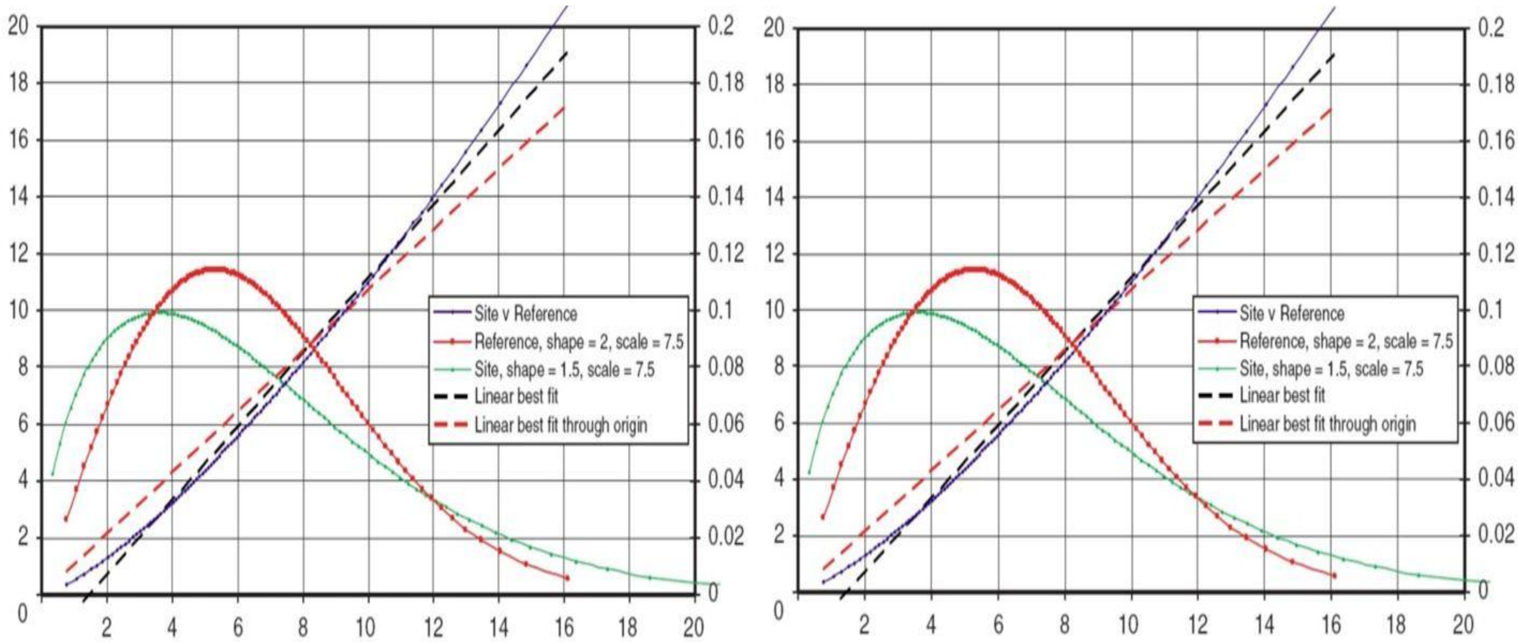
By setting $\theta_x = 0$, such that θ_y becomes the wind speed at target site when wind speed at reference site is zero, results in the following equation:

$$y = mx^{\alpha} + c \ (11)^{6}$$

Where, $m = A_y A_x^{-\alpha}$ and $\alpha = \frac{k_x}{k_y}$

Thus becomes the co-efficient *m* the slope of a linear relationship when shape parameters are equal.

In order to explain equation (11), figure *2.2.3.2* illustrates two examples of a linear fit for different shape factors.



2.2.3.2 Reference site shape parameter greater than target site (left) and Target site shape parameter greater than reference(right)

2.2.3.2 *Predicting wind speed*

In order to predict wind speed distribution, equation (5) is used. By calculating shape factor and scale factor,$k_x\ k_y, and\ A_x, A_y$, for the short term target site and for concurrent reference wind speeds. And determine shape and scale factor for long term reference data,$A_\delta\ and\ k_\delta$, it is possible to calculate shape and scale factor for long term target site,$A_\varphi\ and\ k_\varphi$.

$$\frac{k_x}{k_y} = \alpha = \frac{k_\delta}{k_\varphi} <=> k_\varphi = \frac{k_\delta k_y}{k_x} \quad (12)$$

$$A_y A_x^{-\alpha} = m = A_\varphi A_\delta^{-\alpha} <=> A_\varphi = \frac{A_y A_\delta^{\alpha}}{A_x} \quad (13)$$ [6]

After scale and shape factor is determined for the long term target site, the long reference data is used to create artificial Weibull distribution data.

2.2.4 Wind Index

The index correlation method is a method creating the MCP analysis by means of monthly average of energy yield, without taking into account the directional distribution of the wind. The relationship between wind speed v [m/s] and energy is shown in equation (14), where ρ[kg/m^3] is the density of air, A [m^2] is the area swiped by the rotor. Another limiting factor is how much energy a wind turbine can absorb and this parameter is determined by Betz's law. According Betz's law is 59% ($\frac{16}{27}$) of the kinetic energy can be absorbed by the rotor.

$$P_{actual} = \left(\frac{16}{27}\right) * \left(\frac{1}{2}\right) * \rho * A * v^3 \quad (14)$$

The Wind Index does not estimate a mean wind speed as Linear Regression and Weibull, instead the Wind Index concern about the power output, and the generic equation for calculating Wind Index is:

$$I = \left(\frac{Average\ power\ output\ in\ some\ period, T}{Avergae\ power\ output\ in\ reference\ period, T_{ref}}\right) * 100 \quad (15)$$

By calculating the Wind Index for site data and reference data and combine these it is possible to calculate a correction factor. The correction indicates whether the calculated power production on measured data should be seen as high or low.

In order to calculate power production of a wind power plant it is necessary to specify the stall speed, which is the highest rotational speed of the wind turbine. This information provided for all wind power plants, and an example is illustrated by the red line in *Figure 2.2.4.* It is also possible to use a generic power curve, green line in *Figure 2.2.4*, when calculating power production. If the type of wind turbine not is specified, a generic power curve is commonly used, and it is also what will be used in this thesis.

When a generic curve is considered equation (16) and (17) can be stated in order to calculate power production at a certain wind speed[5].

$$P(u) = \begin{cases} u^2, & for\ u < u_{stall}\ (16) \\ {u_{stall}}^2, & for\ u \geq u_{stall}\ (17) \end{cases}$$

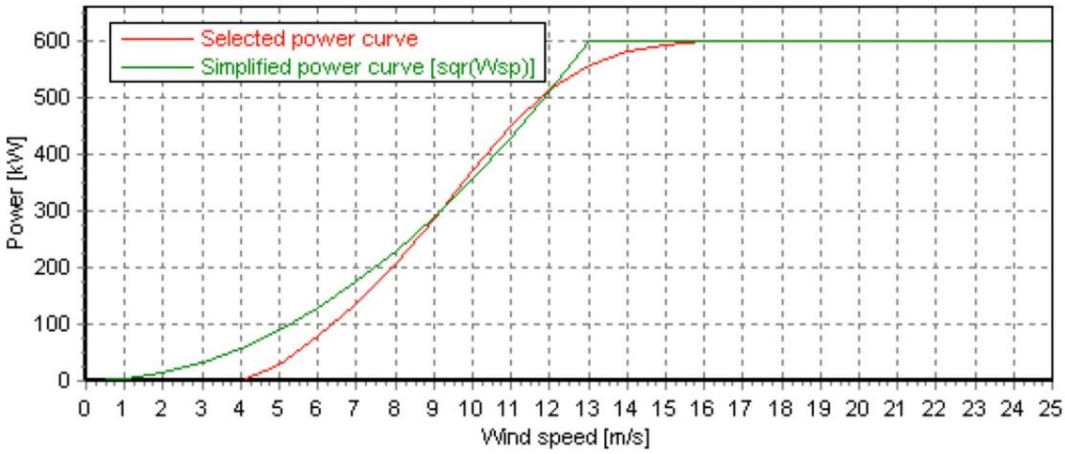


Figure 2.2.4 Illustrates a generic (green) and real (red) power curve[5]

2.2.4.1 Calculating Wind Index

In order to obtain a correction factor between the reference wind data and the measured wind data, four different averages of power output is calculated. It is; Wrf, which is the average power from the full reference series, Wrc, which corresponds to the average power output, calculated of the concurrent reference wind data, Wsf which is the full site power output and finally Wsc which is the power output for the concurrent period.

When all monthly average power production is calculated the Wind Index for every period is calculated, starting with the full reference power production and the concurrent reference series (equations below).

$$I_{Rc} = \frac{W_{Rc} * I_{Rf}}{W_{Rf}} \quad (18)$$

Where:

$$I_{rf} = 100$$

In order to calculate the Wind Index for the measured data one need to assume that the Index for the concurrent reference Index is equal to concurrent measured Index:

$$I_{Rc} = I_{Sc}$$

Since Index of the concurrent site is known it possible to calculate index for entire site measurement. The average power production is calculated with the original time resolution and the Index is determined by the following equation[5]:

$$I_{Sf} = \frac{W_{sc} * I_{Sc}}{W_{Rc}} \quad (19)$$

2.2.4.2 Wind Index Correction

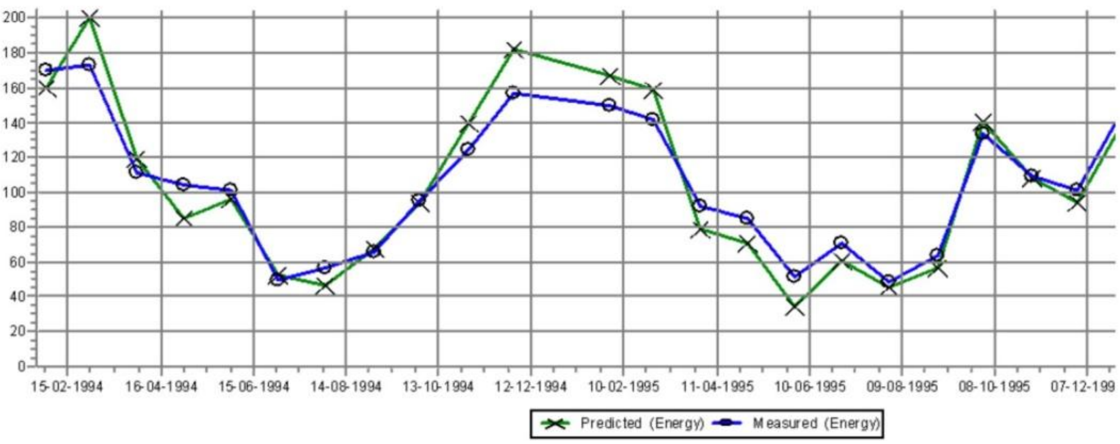
Wind Index for the measured site data is then converted into a correction factor in equation 20.

$$C_{Sf} = \frac{100}{I_{Sf}} \quad (20)$$

The correction factor is then used to create the artificial data. This is done by using the reference data, which cover a long period, and multiply that data with the correction factor, which results in a predicted long term artificial power production series for the site.

2.2.4.3 Validating the Wind Index correlation

In order to validate that the Wind Index for the concurrent period of the reference series is identical to the Index of concurrent period of the site, it is necessary check the degree of correlation between the data sets. This can be done in two different ways. One approach is to plot monthly Wind Index for both data sets in time sequence (Figure 2.2.4.3) in order to see if the course of the Index is similar.



Figure, 2.2.4.3 Monthly Wind Index for reference and measured data[5]

Another method is to plot Wind Index at site versus Wind Index at site (Figure *2.2.4.4*) to see whether is correlating or not. If the plots not is illustrating desirable correlation that is a sign that the climate at the two locations are different and the assumption of equal Index not is the case.

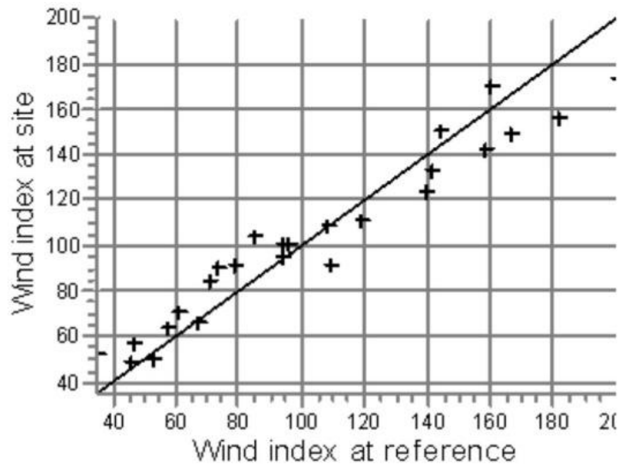


Figure 2.2.4.4 Linear fitted line between Wind Index at site and reference[5]

2.2.5 Bilinear interpolation

Bilinear interpolation is a method used in order to estimate data for a specific point when no data is available other than for surrounding points. This is a common issue within the area of wind estimation since the exact location of the measurement rarely is found in the reference database. Therefore it becomes necessary to pick the closest points and interpolate in order to obtain the wind conditions at a specific site.

There are many different interpolating methods, but bilinear interpolation is a rather simple and accurate method where surrounding data points are used in order to estimate the data at a specific location.

An example of how bilinear interpolation can be applied is illustrated in Figure *2.2.5* where data for the specific location P is but data is only available for the surrounding location P1, P2, P3 and P4. By fitting a bilinear scheme of the four known data points the value of the unknown point P could be estimated. The interpolation is performed in three stages. The first stage involve a interpolation between point P1 and P2 which is described by equation (21), and the second stage P3 and P4 are interpolated and results in point P34.

$$P_{12} = dx \cdot P_1 + (1 - dx) * P_2 \quad (21)$$

$$P_{34} = dx \cdot P_3 + (1 - dx) * P_4 \quad (22)$$

The third step consists of an interpolation between the interpolated points and could be described with equation (3).

$$P = dy \cdot P_{12} + (1 - dy) * P_{34} \quad (23)$$

By combining equation (21), (22) and (23), the equation for the entire interpolation is described in equation (24).

$$P = dx \cdot dy \cdot P_1 + (1 - dx) * dy \cdot P_2 + dx * (1 - dy) * P_3 + (1 - dx) * (1 - dy) * P_4 \quad (24)$$

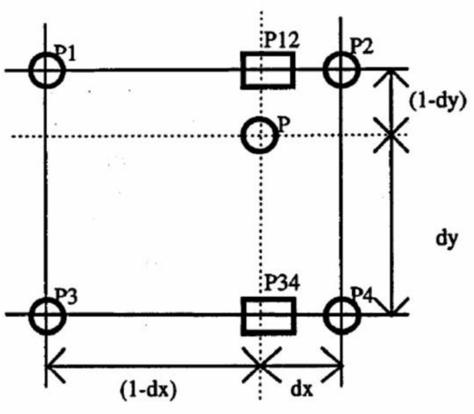


Figure 2.2.5 Illustration of interpolation

2.2.6 U- and V-wind

In order to calculate the wind directions from the u and v components it is necessary to understand how the u- and v-wind are defined, and also how the actual wind vector is calculated from these two components. As figure 2.2.6 illustrates, it is possible to calculate the actual wind vector if both the U and V-wind s known by using Pythagoras[7]. (25).

$Actual\ wind\ vector = \sqrt{v^2 + u^2}$ (25)

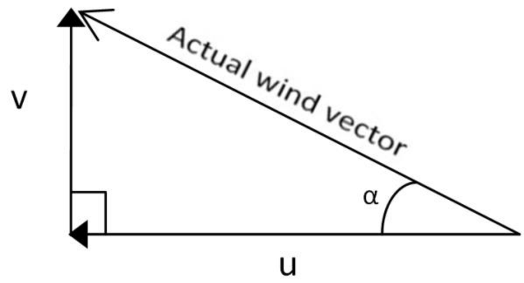


Figure 2.2.6, Relationship between U-and V-wind

As figure 2.2.7 illustrates; is the direction of the wind defined as the angle from where the wind is blowing.

Wind direction $= \arctan\left(\frac{u}{v}\right) + \theta$ (26)

Where:

$\theta = 180° \ if\ v \geq 0,$

$\theta = 360° \ if\ u \geq 0\ and\ v < 0$

$\theta = 0° \ if\ u < 0\ and\ v < 0$

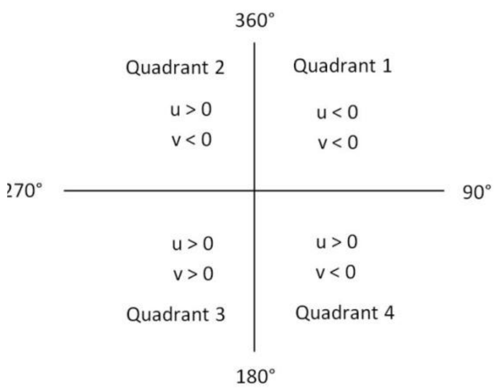


Figure 2.2.7 Illustartion of how actual wind direction can be estimated

By considering the definition of the wind direction and at the same time involve figure 2.2.7 it is possible to determine the direction of the actual wind.

3 Evaluation of the methods

This chapter will discuss the procedure of how the considered methods were evaluated. This is done by giving a brief description of the modeling and comparison process in Matlab. However, the first part of this chapter will provide information about which reference and target site data that was chosen and also justify the choice. There will also be a discussion about how data are sorted and recalculated in order to agree with the theoretical model described in an earlier chapter. Finally will this section concern about the procedure of comparison.

3.1 Input data and validation

In order to obtain a reliable result it is crucial to analyze and chose data that is consistent. Therefore this chapter will discuss how measured data was sorted, how choice of reference data was justified and how reference data points can be combined to correspond to conditions at a certain location. It will also discuss possibility of using different time intervals of measured data in the different analysis.

3.1.1 Measured data

This study is based on local measurements at Näsudden (Gotland), Oskarshamn and Älvsborgsbron. The timeframe of these measurements is different but they have the same time resolution.

Table 3.1.1 Available Measured site data

Site	Latitude	Longitude	Mean speed	Height	No. of Years
Älvsborgsbron	57.69 N	11.90 E	6.04 m/s	100m	9 (2002-2010)
Näsudden	57.07 N	18.20 E	7.86 m/s	90m	14 (1981-1997, except 1984, 1990, 1991)
Oskarshamn	57.27 N	16.43 E	6.85 m/s	110m	9 (2001-2009)
Ringhals	57.26 N	12.11 E	8.00 m/s	120m	9 (1999-2007)

3.1.1.1 Validation and sorting of measured data

When measurements are carried out, failure on the equipment could occur, which results in incorrect data. Therefore is it crucial to check all measured data and clear those points where errors can be distinguished. This was done visually by plotting data against time.

After the data was checked the concurrent reference data for every location was sorted. This was done by creating a filter that compared year, month, day and time and deletes values in the reference data if a point in the measured data is missing.

3.1.2 Reference data

As discussed earlier, there are a number of available reference data sets and it is hard to determine which of them that is most accurate and reliable. According to a study made by Lilèo and Olga Petrik, MERRA and NCEP/CFSR correlate best with local measurements, and therefore was MERRA used as reference data.

3.1.2.1 Translate reference data to local conditions

Since location of the measured data not has an exact corresponding grid point in the MERRA data set, the reference data had to be interpolated between four of the closest grid points. The coordinates of the considered location was known and the four closest point of MERRA was downloaded in WindPro. Then these points were interpolated according to the bi-linear procedure and the definition of U- and V-wind described in the theoretical framework.

The estimation was rather straight forward and consisted of an interpolation of U- and V-wind in order to determine the “new” wind speed and direction of considered locations.

3.2 Performing Regression, Weibull and Wind Index analysis in Matlab

This section will describe the procedure and give information about how the methods were modeled and assumptions associated with the modeling. It will also discuss possibility of using different time intervals of measured data.

Since all methods are well described in the analytical framework, no basic equations will appear in this section.

3.2.1 Time interval of measured site data

In order to compare the different methods it is crucial to determine the time interval of involved measured data. As specified earlier, the different measured data series; Näsudden, Oskarshamn and Älvsborgsbron, has different historical time interval. However, the historical time interval does not influence the analysis itself; it only implies a difference in number of ways that the analysis can be compared. Anyway, the time interval in this case was set to one year, since this is the typical measured timeframe in wind resource estimations.

3.2.2 Linear Regression

The first task when using a Linear Regression model for wind estimation is to sort data. The concurrent periods have been sorted but the filtration by wind direction is also crucial to consider when using Linear Regression. The measured site data was sorted by wind direction, according to earlier chapter, were concurrent MERRA series is used as a base. This means that measured data is sorted by the wind direction of the concurrent MERRA data point. When all essential data for the regression analysis was sorted the next task is to create the regression model in Matlab.

3.2.3 Performing the analysis

The regression analysis was done for all years, location and sectors. From these regression models an artificial series was produced, by putting in long-term reference data for every year. This artificial series was then used to calculate the predicted average wind speed. Since the regression model involves 360 different wind sector, the predicted average speed for every year was calculated by taking the mean value of all 360 sectors.

Further was a residual analysis carried out for every regression analysis and the residuals was added, in terms of a randomized mean zero Gaussian distribution, to the regression model.

3.2.4 Weibull

As input data in Linear Regression, Input data in Weibull is sorted out by wind direction. This was done according to the procedure explained in section 3.2.3. The only difference in this filtration process, in comparison to Linear Regression, is that Weibull uses 12 wind sectors instead of 360. When all data was sorted the model was created in Matlab.

3.2.5 Performing the analysis

A Weibull analysis, as Linear Regression, was done for every year, location and sector. Then the mean wind speed was calculated for every sector, and eventually a long term predicted wind speed was calculated by take a mean value for all 12 sectors.

There is no simple method used to validate Weibull. Therefore, was the distribution of the predicted wind speed checked visually in terms of plots, in order to see if the result was reasonable.

3.2.6 Wind Index

Wind Index uses input data without any concern about the wind direction, and is therefore not in need of any filter when created in Matlab. The necessary data is the wind speed for the long term data and concurrent reference and measured data. Although, no filter is necessary, the Wind Index involves a power curve, and in this study, a generic power curve was used. This means that if a wind speed is above a generic stall speed the power output is constant. In order incorporate a generic power curve it is necessary to determine stall speed. The generic stall speed was set to 13 m/s[5]. Wind speeds above stall in reference and measured data series was therefore changed to 13 m/s.

3.2.7 Performing the analysis

The analysis was carried out by calculate monthly average power output, corresponding monthly correction factor and Wind Index for every year. The correction factor was then used to correct the measured monthly power. Further an average of the monthly artificial power output was calculated.

The validation of a Wind Index was done according to 3.2.4.3 and correlation factor was estimated.

3.3 Estimation of Error

The procedure of estimating error for all method was the same. The only difference was that the result of a Wind Index analysis gives an error in power output, whereas Linear Regression and Weibull results in an error in average wind speed. However, the final error will be presented in percentage and is therefore comparable.

The average wind speed of complete measured series was used as a reference for Weibull and Linear Regression, and power output from the same series was used as reference for Wind Index.

3.4 Comparison of methods

The comparison was carried out by first plot the errors from the different models against one other in order to interpret the result visually. In a second stage a regression analysis was done to see whether they correlate or not. Finally the error and absolute error was determined, since that could provide information about an underlying pattern between errors at different locations.

4 Results

In this chapter the result will be presented in terms of error, absolute error and correlation factor for each location and method.

4.1 Oskarshamn

The table below shows the mean error, absolute mean error and correlation factors estimated from wind measurements done in Oskarshamn. The last column shows the correlation between the different estimation models.

Further was the correlation between reference and measured data 0,65 for this location.

Oskarshamn				
Method	**Mean Error (%)**	**Absolute Mean Error (%)**	**Correlation**	
Linear Regression	-2,25	2,74	Regression/Wind Index	r = 0,81
Weibull	1,9	5,26	Regression/Weibull	r = 0,60
Wind Index	23,42	23,4	Wind Index/Weibull	r =0,25

Table 4.1 Results from estimations done by using measurements from Oskarshamn

4.2 Näsudden

The table below shows the mean error, absolute mean error and correlation factors estimated from wind measurements from Näsudden (Gotland). The last column shows the correlation between the different estimation models.

Further was the correlation between reference and measured data 0,82 for this location.

Näsudden				
Method	**Mean Error (%)**	**Absolute Mean Error (%)**	**Correlation**	
Linear Regression	– 4,51	4,51	Regression/Wind Index	r = 0,81
Weibull	– 2,43	4,8	Regression/Weibull	r = 0,43
Wind Index	– 7,09	8,1	Wind Index/Weibull	r =0,24

4.2 *Results from estimations done by using measurements from Näsudden*

4.3 Älsvborgsbron

The table below shows the mean error, absolute mean error and correlation factors estimated by using wind measurements from Älvsborgsbron (Gothenburg). The last column shows the correlation between the different estimation models.

Further was the correlation between reference and measured data 0,83 for this location.

Älvsborgsbron				
Method	**Mean Error (%)**	**Absolute Mean Error (%)**	**Correlation**	
Linear Regression	-5,24	5,24	Regression/Wind Index	r = 0,66
Weibull	– 6,71	9,30	Regression/Weibull	r = 0,90
Wind Index	2,12	4,72	Wind Index/Weibull	r =0,76

4.3 ***Results from estimations done by using measurements from Älvsborgsbron***

4.4 Ringhals

The table below shows the mean error, absolute mean error and correlation factors estimated by using wind measurements from Ringhals (Varberg). The last column shows the correlation between the different estimation models.

Further was the correlation between reference and measured data 0,85 for this location.

Ringhals				
Method	**Mean Error (%)**	**Absolute Mean Error (%)**	**Correlation**	
Linear Regression	0,07	1,01	Regression/Wind Index	r = 0,85
Weibull	– 0,62	5,20	Regression/Weibull	r = -0,16
Wind Index	2,04	3,02	Wind Index/Weibull	r =-0,44

4.4 ***Results from estimations done by using measurements from Ringhals***

4.5 Combining all locations

When errors from all locations were used in a regression model the following correlation factors and mean errors were determined.

All locations				
Method	**Mean Error (%)**	**Absolute Error (%)**	**Correlation**	
Linear Regression	-2,96	4,20	Regression/Wind Index	r = 0,69
Weibull	-2,07	6,68	Regression/Weibull	r = 0,39
Wind Index	5,73	12,10	Wind Index/Weibull	r =0,42

Table 4.5 ***A summary of the results when errors from all locations were used in Linear Regression***

In order to get a visual view of how the relation between the different methods a diagram for each pair will follow. The first diagram shows the relation between errors estimated when Weibull and Linear Regression were used

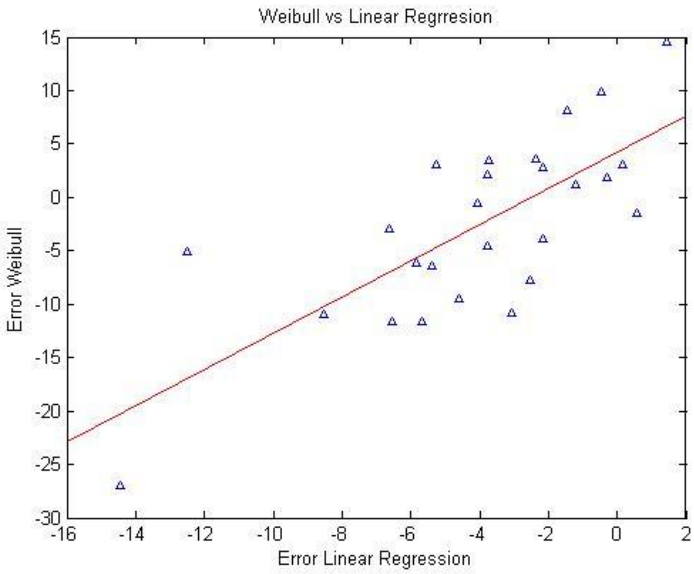


Diagram 4.5.1 Showing relation between Weibull and Linear Regression

Diagram 4.5.2 describes the relation between errors estimated using Weibull in relation to Wind Index

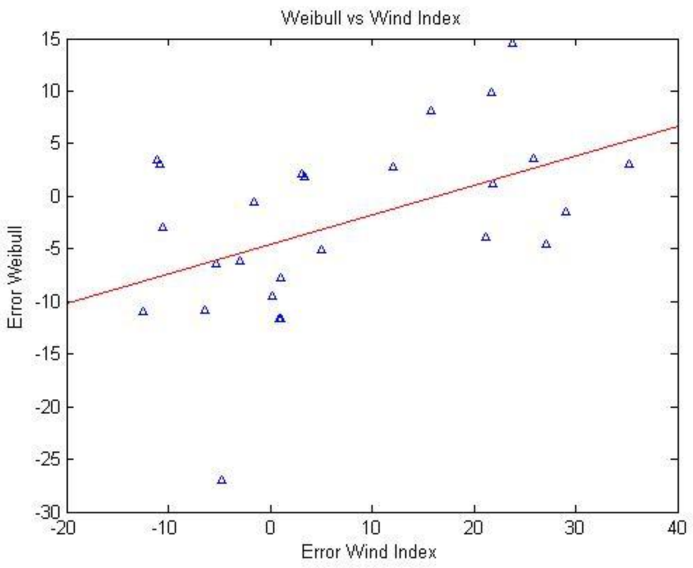


Diagram 4.5.2 Showing relation between Weibull and Wind Index for all locations

A visualization of the relation between Wind Index and Linear Regression is illustrated in diagram 4.5.3.

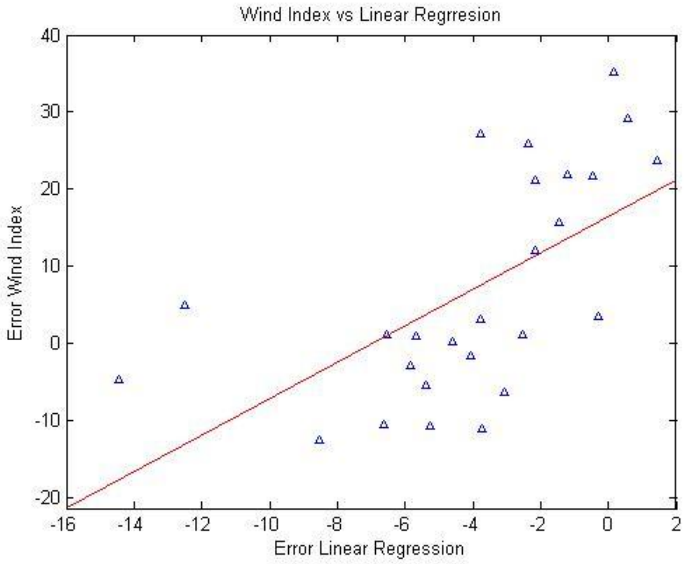


Diagram 4.5.3 Showing relation between Wind Index and Linear Regression for all locations

5 Discussion

This section will discuss the outcome of the result and also bring up underlying parameters that could have had an influence on the result. Further will also the local wind conditions at the considered sites be discussed.

5.1 Local conditions and measurement

As mentioned before is the timeframe of the measured site data different for all location. Oskarshamn has the shortest with 7 years available, whereas Näsudden has a time interval of 14 years. This resulted in that Näsudden gave additional comparison alternatives in comparison to Oskarshamn and should therefore be considered as more reliable from a statistical point of view.

The wind conditions at the different locations are shown in table *3.1* . The mean wind speed at Ringhals is higher than the other location. In general one can say that Ringhals has better wind conditions in comparison to Näsudden, Älvsborgsbron and Oskarshamn, and also a better correlation.

5.2 Corresponding reference data

When producing wind resource estimation the reference data and its correlation or non-correlation with measured data is of interest. In this case the correlation between given data varied from the different location. The correlation between measured data and reference data at Älvsborgsbron (0,83), Näsudden (0,82) and Ringhals (0,85) were rather similar. However, Oskarshamn (0,68) had, in comparison to the other locations, a rather low correlation between reference and measured data.

5.3 Estimated error

The resulting error and correlation vary to a relatively large extent between the different locations. It is interesting that the results from Wind Index varied very much and the highest number was much higher in comparison to the other methods. The aim of the work was to find correlation coefficients between the estimated errors form the studied methods, and the results shows that it is possible to state a correlation between these methods.

6 Conclusions

In general Linear Regression could be seen as the most accurate statistical method and Wind Index the least accurate. At all three locations the Linear Regression method had a relatively small deviation from the real average wind speed.

Correlation between the results from every method at the different locations was no stringent, but it was possible to distinguish a pattern; that Wind Index and Weibull had a weak correlation, Weibull and Linear Regression had a weak correlation and that Linear Regression correlated with Wind Index.

Further is it possible to state that the magnitude of correlation is dependent on the local conditions and correlation between measured and reference data, since the result from the different locations involved such great variety.

In order to state some further conclusion other reference data should be studied in order to see if they follow the same pattern as MERRA does.

7 References

[1] Swedish energy goals available at
http://www.sweden.gov.se, 2011-07-25

[2] Petrik O and Lileo S., 2010, Investigation on the use of NCEP/NCAR, MERRA and NCEP/CFSR reanalysis data in wind resource analysis

[3] Blomqvist U., 2003, Matematisk Statistik

[4] Montgomery D. and Runger G. 2011, Applied statistics and Probability for engineers, Fifth edition

[5] Thorgesen M. L. 2009, An introduction to the MCP Facilities in WindPro

[6] Thorgesen M., 2008, Measure-Correlate –Predict case study Implementation

[7] National weather forecast office, Seattle, available at
www.wrh.noaa.gov/sew/fire/olm/transport.htm, 2011-07-25