Uncertainty in the application of the Measure-Correlate-Predict(MCP) method in wind resource assessment

Sundus Cordelia Ramli[*][1] and Mikkel Haugaard Windolf[1]
[1] DONG Energy Renewables. Nesa Allé 1, 2820 Gentofte, Denmark
[*] Corresponding author, sunra@dongenergy.dk

SUMMARY

This paper presents a study of the uncertainty that arises when the measured wind resource data are used to estimate the long-term wind resource. A process called Measure-Correlate-Predict(MCP) uses the measured data along with long-term data from a reference site to estimate the long-term wind resource at a site.

Long term data from the Netherlands and Denmark were used whereby 11 offshore sites and 44 onshore sites data sets were available for analyses which results in 1,892 onshore site pairs and 50 offshore site pairs.

The MCP algorithms used in this study for comparison are the Variance Ratio, Matrix Method and Linear Regression. The Variance Ratio and the Linear Regression with Residual Resampling were found to be superior to the other methods, however, the Variance Ratio was used almost exclusively in the rest of the study. Several possible influences on uncertainty were used to determine the overall uncertainty in the MCP proces.

It was observed that by using high quality reference stations uncertainty can be reduced from 6% to 4% for onshore sites and from 4% to 3% for offshore sites.

A method for checking the consistency of data sets is also introduced by performing "MCP cross predictions(MCPCP)". Several possible influences on uncertainty was investigated whereby uncertainty can be reduced which includes the length of concurrent data sets and the number of reference stations used. A combined parameter estimation of both of these influences on uncertainty is also defined.

1. INTRODUCTION

Uncertainties can arise in the assessment of wind resource hence it is important to account for the various individual sources of uncertainty in order to provide with the best estimate of the wind resource. This uncertainty from the wind resource can be combined with the other uncertainties that arises from other factors such as the power curve and the energy losses to contribute to an overall annual energy production(AEP) uncertainty. Since the ability to secure a wind development project is closely tied to the perceived risk, it is critical to estimate the uncertainty to be able to capture the risk associated with the potential venture.

The uncertainty of interest for the scope of this paper is the uncertainty that arises by using the measure-correlate-predict(MCP) methodology in the estimation of long term wind resource. The MCP is a statistical technique used for predicting the local long-term wind resource based on a series of short term measurements of wind speed at a prospective target site and correlating them with concurrent wind speed measurements made at a nearby long-term reference meteorological site.

The long-term reference site data provide information about variations in the wind resource at time scales longer than the length of the target site data. MCP has been shown to provide more accurate estimates of the long-term wind resource at a site than simply assuming that the measured data alone is representative of the long-term resource[2], [3].

Wind speed and direction data are measured simultaneously at both wind site for a recommended period of at least 1 year, to avoid seasonal fluctuations in wind speeds [4], [5], [6].

In summary, MCP uses the measured data with long-term data from a reference site to estimate the long-term wind resource at a site.

2. METHODOLOGY

The goal of any MCP method is to closely predict the predicted values to the true values of the target site wind resource. The availability of long-term data sets are critical to any test involving MCP predictions as it offers a means of objectively evaluating the MCP method. Being able to compare the predictions to the true values offers a mean of objectively evaluating the MCP method. Furthermore, multiple MCP predictions can be made with a single pair of long-term data sets, which allows for the variability of the predictions to be assessed.

The steps below are a general method of investigation used to quantify uncertainty in this study :

1. Two sets of long-term data are used at a time whereby one is treated as the target site data.
2. The long term target site data can be divided into smaller segments and each segment can be evaluated as the measured data set.
3. The measured data set is used together with a the long term reference data to predict the long term target site wind resource.
4. This results in multiple MCP predictions .
5. The value from each MCP prediction is compared to the true target site values.
6. The error of the multiple predictions indicate the uncertainty whereby the coefficient of variation (COV=σ/μ) of the errors equals to the standard uncertainty of the process.

Additionally, the error values can be compared to available statistics and fit to a normal distribution.

2.1.1 MCP Algorithms

a) *Linear Regression:* The most basic tool to create a functional relation between two stochastic variables, i.e. wind speed data of prospective and reference sites, is the simple linear regression. It can be expressed as follows :

$$\hat{y} = \beta_0 + \beta_1 x$$

where x is wind speed at the reference site, $\hat{y}$ is predicted wind speed at the target site, and β_0 and β_1 are the estimated intercept and slope of the linear relationship. This will be labelled as LR hereafter.

Setting $\beta_0 = 0$ will result in a linear regression through (0,0). This will be labelled as LR(0,0) hereafter.

b) *Residual Resampling:* An improved way over the traditional linear regression study is to model the distribution of the residuals. The long term corrected data is calculated using the regression model(transfer function). All of the samples in the long term reference series are transferred and then the residuals are calculated using a Monte-Carlo Simulation technique. The result is a long term corrected wind distribution through an 'artificial' time series.

c) *Variance Ratio:* In this model, it is assumed that the variance of the predicted wind speed at the target site is expected to have the same overall variance as observed at the target site. For a linear model of the form $\hat{y} = mx + b$, the variance of the predicted target site wind speeds is $\sigma^2(\hat{y}) = \sigma^2(mx+b) = m^2\sigma^2(x)$.
Setting m^2 equal to $\sigma^2(y)/\sigma^2(x)$ ensures that $\sigma^2(\hat{y}) = \sigma^2(y)$. Thus, the variance ratio method uses the following equation :

$$\hat{y} = \left[\bar{U}y - \frac{\sigma y}{\sigma x}\bar{U}x\right] + \left[\frac{\sigma y}{\sigma x}\right]x$$

where x is the wind speed at the reference site, $\hat{y}$ is the wind speed at the target site, $\bar{U}x$ and $\bar{U}y$ are the sample mean wind speeds at the reference and target sites, respectively, and σx and σy are the corresponding sample standard deviations.

d) *Matrix Method (as implemented in WindPRO):* Instead of prescribing a linear relationship between two variables this techniques models the changes in wind speed(speed-up) and wind direction(wind veer) through joint distributions fitted on the 'matrix' of wind speed bins and wind direction bins. A basic assumption of the matrix method is that the long- term site data (wind speed and direction) can be expressed through the simultaneous measurements of on-site data and reference site data. The relationship is modelled through a joint distribution between the two variables wind speed-up and wind veer.

In this study, the matrix method used is the as the one implemented in WindPRO. The transfer model, given as a conditional distribution is actually the key distribution in the generalized matrix method. The distribution gives the reference relationship between the site wind climate and the reference wind climate. When applying the matrix method this conditional distribution is stipulated to hold regardless of the time frame considered. For more details , refer to Thørgersen et. al[8]

2.1.2 MCP parameters Selection

When implementing the MCP methods, several parameters can be adjusted which can enhance the performance of a particular MCP method which includes the number of direction sectors and the cutoff wind speed. In this research, both one direction sectors and twelve direction sectors are performed and a cutoff wind speed of <2m/s is used to avoid the large and unrepresentative scatter in wind direction at low wind speed.

2.1.3 Data Sets Available for Analysis

For this study, long-term data sets were taken from the Netherlands, for which the Royal Netherlands Meteorological Institute(KNMI) makes hourly wind speeds available for all meteorological stations and from Denmark. A total of 44 onshore sites and 11 offshore sites were available for analysis.

2.2 RESULTS OF ANALYSIS

An initial analysis was performed for 1,892 onshore site pairs and 50 offshore site pairs using the Variance Ratio method; one direction sector of wind speeds was used.

The coefficient of variation on predicted wind speed/actual wind speed ratios is V = 0.04 for offshore sites and V = 0.06 for onshore sites as presented in Table I.

When compared to a normal distribution, the distribution of ratios fits well. This indicates that 90% of all predictions are in the interval 0.9-1.1. A typical example is shown in Figure 1.

Inconsistent Data	*Offshore*	*Onshore*
No. of Station Pairs	50	1892
No. of Predictions	754	37,612
Mean ratio pred/meas	1.00	1.00
COV	0.04	0.06

Table I: Results from initial analysis using MCP method Variance Ratio. One direction sector of wind speed was used

<u>Improving the accuracy of the MCP process</u>

2.2.1 Importance of consistent data :

It is obvious that in order to obtain accurate estimates of the target site wind resource, it is important to use a reference site data set that has high quality, consistently measured wind speed data. This part of the analysis investigates the influence of having consistent data on the uncertainty of the MCP process. The Variance Ratio method will be used in this part of the analysis. A similar study on has also been done by Siddabanthani et. al[9]

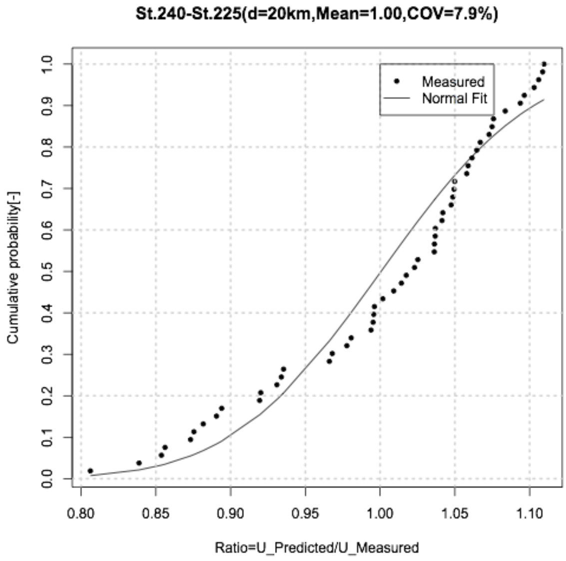

Figure 1 : Ratio of predicted/actual wind speed compared to a normal distribution of MCP predictions of long term station 225 with station 240 data for inconsistent data period.

In Figure 1, it can be observed that the majority of the distribution of ratios fits a normal distribution fairly well. However, it can also be observed that the distribution of ratios are skewed which indicates that there are sections of data which do not fit the normal distribition.

It is necessary to investigate the cause of this skewness in the graph. At a closer inspection of the tabular values, the clusters of data that do not fit the normal distribution can be seen to represent a certain period of time. For the particular example discussed, this period of inconsistency is for the years 1951 – 1983.

Following this observation, the MCP predictions was repeated by excluding time series data from the years previous to 1984 hence this time 22 predictions was made. Figure 2 shows the cumulative distribution of the ratio of predicted/actual wind speeds and again is compared to a normal distribution. It can be observered that the value of COV is reduced more than half of its original value to 0.034.

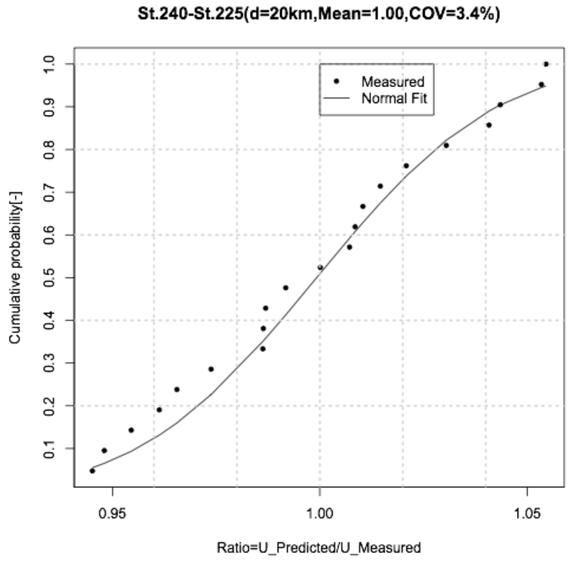

Figure 2 : Ratio of predicted/actual wind speed compared to a normal distribution of MCP predictions of long term station 225 with station 240 data for consistent period of data

By excluding inconsistent data, the coefficient of variation on predicted wind speed/actual wind speed ratios is COV = 0.032 for all 110 pairs of offshore sites and COV = 0.042 for ll 1,892 pairs of onshore sites; as shown on Table II. The value for the offshore sites did not reduce too dramatically due to the majority of the data sets were already within the consistent periods.

Consistent data	*Offshore*	*Onshore*
No. of Station Pairs	110	1892
Mean ratio pred/meas	1.00	1.00
COV	0.032	0.042

Table II : Results from using only consistent data sets by using the Variance Ratio Method.

2.2.2 Performance comparison of MCP methods :

This part of the analysis seeks to find if the uncertainty is influenced by different MCP methods. In order to do this, it is necessary to evaluate the performance of several MCP methods. The data sets used for this investigation are from the 110 pairs of offshore long-term data sets.

The predictions made via the MCP method is validated by comparing the predicted time series of data and $\{x_i\}$ and the measured wind speed time series $\{y_i\}$ at the prospective site. Table III shows the results for each of the 6 methods discussed earlier.

Method	*Mean(abs)*	*St. Devn*
LR(0,0)	0.73	0.59
LR	0.08	0.32
LR(0,0) with RR*	0.12	0.41
LR with RR*	0.035	0.29
Matrix Method	0.50	0.56
Variance Ratio	0.008	0.28

Table III : Wind Speed Time Series : Performance comparison of different MCP methods. Mean and St. Devn of the bias error is presented. *RR refers to Residual Resampling.

In order to see whether one method is significantly better than the others it is necessary to use a standard statistical test, the F-test. The F-test (Fischer's test) enables tests for comparing the standard deviation of two samples to distinguish whether they come from the same population or from different population. Taking the 95% confidence level, the F-test can be performed.

When using the F-test on the 3 least biased methods the ratios do not exceed the F-statistic as demonstrated in Table IV. It can be concluded that there is no statistical evidence for rejecting the hypothesis that the two samples for each pair of tests were drawn from Gaussian distributions with a common variance.

Another test, the Student's t-test was also performed. In this test, a test value is calculated from the absolute difference of the means of the two data populations and their variances.

.

Method 1	*Method 2*	*F-test*	*t-test*
LR with RR*	LR	Same population	Same population
LR with RR**	VR**	Same population	Same population
LR	VR**	Same population	Same population

Table IV : Wind speed time series : Comparison of least biased MCP methods using the F-test and the t-test. **RR refers to Residual Resampling, **VR refers to Variance Ratio

When using the t-test on the Variance Ratio and LR with Residual Resampling and with LR, the t-value is not within the rejection region and therefore we can accept the hypothesis that $\mu_{VR}=\mu_{LR(residual)}=\mu_{LR}$. From this hypothesis, we can clearly state that there is no reason to prefer one method over the other. Hence it is necessary to see the effects of the MCP accuracy on the power

time series.

For each of the least biased methods, the time series of hourly averages of wind speed at the predictor site

were converted to average power using a power curve for a wind farm parks.

The methods were then compared on the basis of real and predicted time series power data. Table V shows the result for the three methodologies.

Method	*Mean(Abs)*	*St. Devn*
LR	3.16	7.24
LR with RR	0.981	6.39
Variance ratio	0.187	6.174

Table V : Power time Series : Performance comparison of different MCP methods. Mean and St. Devn of the bias error is presented. *RR refers to Residual Resampling.

As previously, it is necessary to use 2 standard statistical test, the F-test and the t-test so see
whether one method is significantly better than the other.

Table VI indicate similar results to the wind speed prediction time series in that the methods all share a common variance but only the Variance Ratio and the LR with Residual Resampling can accept the hypothesis that $\mu_{VR}=\mu_{LR(residual)}=\mu_{LR}$.

This hypothesis clearly states that the Variance Ratio and the LR with Residual Resampling are superior to the methods.

Method 1	*Method 2*	*F-test*	*t-test*
LR with RR*	LR	Same population	Different population
LR with RR*	VR**	Same population	Same population
LR	VR**	Same population	Different population

Table VI: Power Time Series : Comparison of least biased MCP methods using the F-Test (Fisher's Test) and the Student's t-test. *RR refers to Residual Resampling; **VR refers to Variance Ratio.

However, these results do suggest that the success of any of the prediction methodologies is less to do with the mechanics of the methodology itself, and more to do with the facets of the data being analysed. This is pursued further in the next following subsections.

2.2.3 Variation with Concurrent Data Length

For this part of the analysis, the Variance ratio was used on the 50 pairs of offshore long-term data sets. The lengths of concurrent data investigated were 6 months, 1 year, 2 years, 3 years, and 4 years.

Figure 3 shows the relationship between the normalized bias error of wind speeds and the amount of concurrent data length. A clear relationship can be observed whereby the longer the concurrent data is, the more accurate the predictions are; the spread of data decreases the longer the concurrent data.

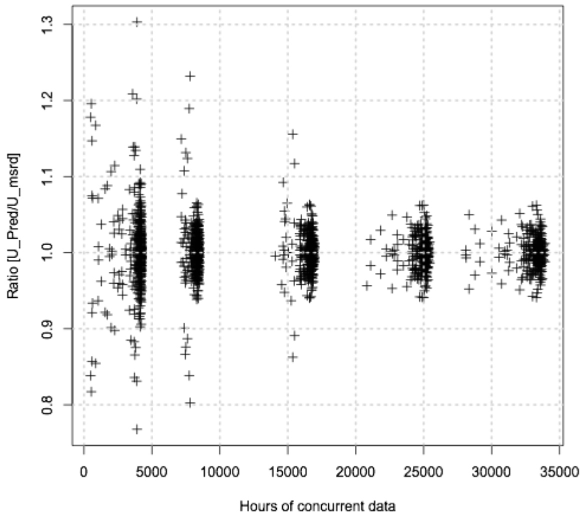

Figure 3 : Variation with concurrent length of data

The uncertainty of the MCP process is defined as the coefficient of variation of the ratio predicted/measured wind speed. The uncertainty for all predictions made by all 50 site pairs variation with hours of concurrent data is demonstrated in Table VII.

Concurrent Data Length, Hours	*COV*
4380	0.057
8760	0.036
17520	0.027
26280	0.021
35040	0.020

Table VII : Variation of uncertainty with concurrent data length

It can be confirmed that the %COV decreases as the square root of the concurrent data length :

$$Uncertainty \approx \frac{\%COV}{\sqrt{N}}$$

where N is the number of hours of concurrent data.

2.2.4 Variation with Correlation Coefficient :

Presented in Figure 4 is the normalized bias error wind speed plotted against the product moment correlation coefficient.

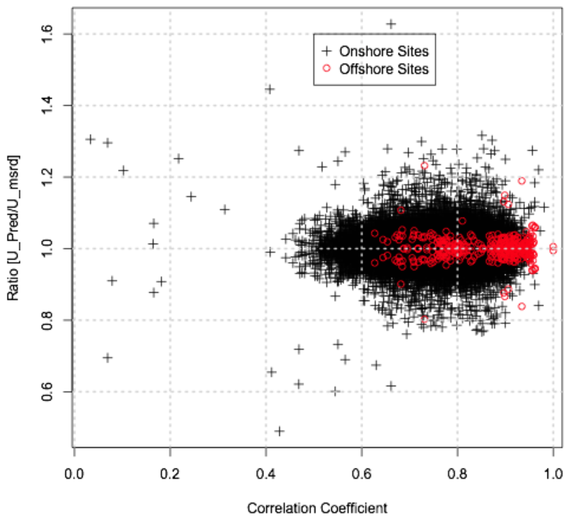

Figure 4 Ratio [Upred/Umeasured] as a function of correlation coefficient. Note that (1-Ratio) is the normalized bias error.

The data sets presented are from the 1,892 pairs of onshore and 50 pairs of offshore sites. The majority of the correlation coefficient lies in the region of 0.6-0.98. It can be observed that the scatter of the results suggests a weak tendency for the normalized bias error to decrease with increasing correlation.

2.2.5 Variation with Distances :

This part of the analysis investigates whether prediction improves with smaller distance. Figure 5 presents the coefficients of variations of all predictions for all of the onshore and offshore station pairs in the Netherlands.

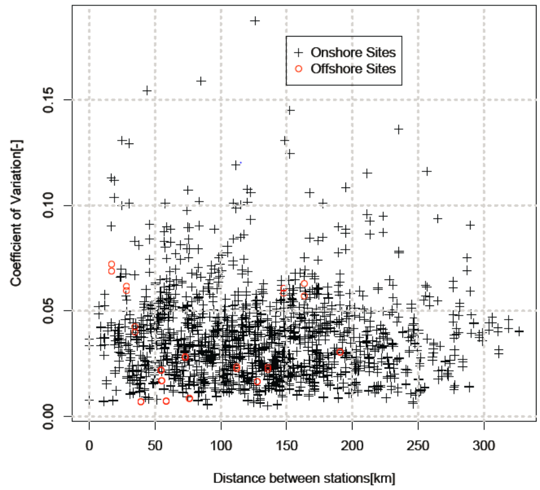

Figure 5 : Variation of COV with distance(km)

It can be observed that the scatter of the results suggest that the predictions do not improve with smaller distances and that there is zero correlation with distance. This results suggests that all local wind speeds are correlated with the same geostrophic wind, and that wind speed variations are due to local terrain effects. It is important to highlight that the topography of Netherlands is similar throughout the region and hence, this may not be applied to other regions.

2.2.6 Using Multi-Reference Stations :

After observing that there is zero correlation with distance, the motivation now it to make predictions using multiple reference points. Three studies were done :

i. 1 Reference Station
ii. 2 Reference Stations
iii. 3 Reference Stations

As previously, the uncertainty of the MCP process is defined as the coefficcnt of variation of the ratio predicted/measure wind speed. The uncertainty for all predictions by all 50 offshore site pairs variation with number of reference points is demonstrated in Table VIII and is shown graphically in Figure 6 where the error bars represents the standard deviation of the MCP process for the individual studies.

No. of Reference Stations	*1*	*2*	*3*
COV	0.032	0.023	0.018

Table VIII : Variation of uncertainty with concurrent data length

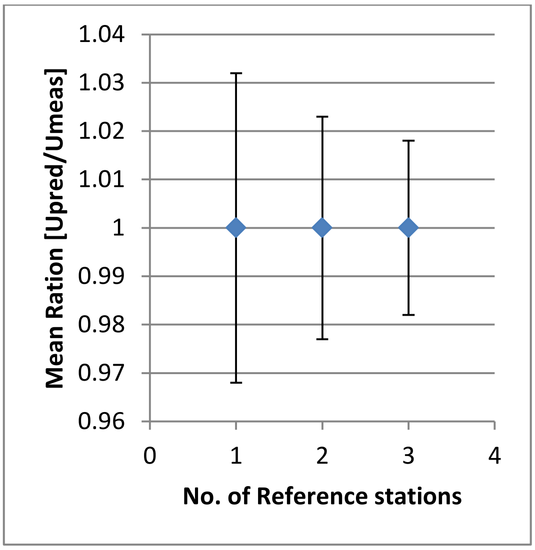

Figure 6 : Mean Ratio of [Upred/Umeasured] as a function of number of reference stations. Error bars are the standard deviation of the MCP predictions.

By parameter estimation, the uncertainty can be shown to decreases as the square root of the number of reference station used in the MCP predictions :

$$Uncertainty \approx \frac{\%COV}{\sqrt{R}}$$

where R is the number of reference stations.

2.4 The MCPCP(MCP Cross Predictions) method

It has been previously shown the importance of having consistent data when applying the MCP method to estimate the long term wind resource of a particular target site. In this section, a method for testing the consistency of a reference data set is introduced. This method is called the MCPCP (MCP cross Predictions) method.

I. All available reference stations and target site data are paired together in turn and steps 1-6 from the methodology described in this paper is performed.
II. A reference station with inconsistent data would appear to have a higher value of COV than the other paired stations.
III. The data set from this reference station can be discarded for inconsistency

Example : A real offshore project

A mast was erected and data was collected for a period of 5 years. There are four reference stations available each with long-term data sets of length 11 years.

The Variance Ratio MCP method was used to make predictions with each individual reference
stations and with concurrent data length of 5 years. The results are summarized below :

Reference station	1	2	3	4
Predicted Wind speed[m/s]	9.3	9.3	9.4	9.1

Table IX : Predicted wind speed as estimated when using the Variance Ratio MCP method for 4 available reference stations.

The estimated long term wind resource when using reference station 4 varies from the other reference stations. This could be due to the inconsistent data from reference station 4. To investigate this further the MCCP method is applied and the results are presented in Figure 7.

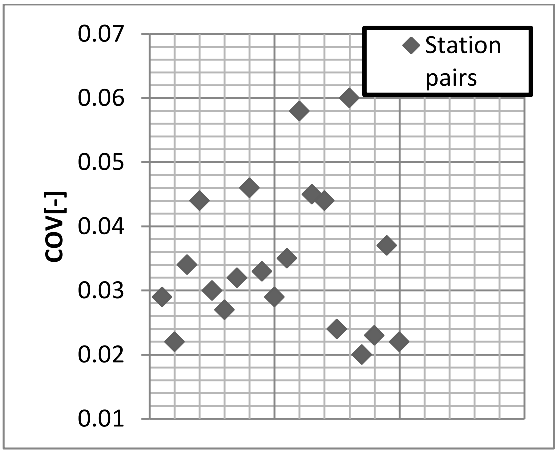

Figure 7 : Coefficient of variation of station pairs. Each point represents the station pairs results.

The points in Figure 7 that have a COV value above 0.04 are the results of the MCCP when using reference station 4 as one of the station pairs. However, it is interesting to note that when performing the MCCP prediction with the station pairs of reference station 4 and the "true" site data ,the COV value is low. This indicates that the consistent period for the reference station 4 is within the period of 5 years of data collected at the target site.

The points in Figure 7 that have a COV value between 0.03 and 0.04 are the results of the MCCP when using reference station 3 as one of the station pairs. When using reference stations 1 and 2 the COV values are the lowest and this is reflected in Table IX.

These results are illustrated in Table X.

Station Pairs	*COV*
S1&R2;S1&R5;S2&R1; S2&R5; S3&R2;S5&R1;S5&R2; S5&R4	>0.03
S1&R3;S2&R3;S3&R1; S3&R5;R5&S3	0.03<cov>0.04
S1&R4;S2&R4;S3&R4; S4&R1, S4&R2;S4&R3	<0.04

Table X : The MCCP results grouped into values of COV. S indicates when the station is used as a target site and R indicates when the station is used as a reference site. S5 and R5 refer to the real target site when used as target site and reference site respectively.

The usage of the MCCP method will give more confidence in choosing the right reference stations to utilise in the estimation of long term wind resource at a particular site hence reducing the risk in a potential venture.

2.5 Conclusions

Wind energy site assessment is a complex multistep process, with a high degree of uncertainty. This paper investigated further the uncertainty caused by the MCP process. A study has been conducted on 44 onshore and 11 offshore Dutch and Danish sites and uncertainty was quantified for both offshore and onshore sites. Parameter estimations of the uncertainty were also defined in this paper.

The following conclusions were reached :

- By using the F-test and the t-test it was found that the MCP methods Variance Ratio and Linear Regression with Residual Resampling are superior to the other methods.
- When using the consistent data method, uncertainty is improved and quantified at 4% for 3% for offshore sites - hence it is important to ensure that the long term reference data set has good high quality consistent data set.
- If multi-reference stations are available, the MCCP method can be used to check for inconsistency within the data.
- A clear relationship can be observed whereby the longer the concurrent data is, the more accurate the predictions are. Uncertainty decreases as the square root of the concurrent data length, N in hours:

$$Uncertainty = \frac{\%COV}{\sqrt{N}}$$

- There is no clear relationship between correlation coefficient and uncertainty.
- It is possible to improve predictions by simply using more reference stations. Using multi-reference stations is beneficial as it adds

certainty to the MCP process. Uncertainty decreases as the square root of the number of reference stations(R) used :

$$Uncertainty = \frac{\%COV}{\sqrt{R}}$$

- The above 2 equations can be combined to give an overall equation to define uncertainty :

$$Uncertainty = \frac{\%COV/\sqrt{N}}{\sqrt{R}}$$

REFERENCES

[1] Fontaine, Agnès, and Peter Armstrong. 2007. Uncertainty study in energy yield assessment. Paper presented at Poster Session, Milan,http://winddata.mek.dtu.dk/docsearch/pdf.php?id=468 (accessed Sept 2010).

[2] Landberg, L., and Mortensen, N.G., "A Comparison of Physical and Statistical Methods for Estimating the Wind Resource at a Site," 15th British Wind Energy Association Conference,York, UK, 1993, pp. 119-125.

[3] A. Derrick, "Development of the Measure-Correlate-Predict Strategy for Site Assessment,"14th BritishWind Energy Conference, Nottingham, UK, 1992, pp. 259-265.

[4] Rogers, A.L., Rogers, J.W., and Manwell, J.F., "Uncertainties in Results of Measure-Correlate-Predict Analyses," EWEC 2006.

[5] A.L. Rogers, J.W. Rogers, J.F. Manwell, "Comparison of the Performance of Four Measure-Correlate-Predict Algorithms," Journal of Wind Engineering and Industrial Aerodynamics, vol. 93, no.3, pp. 243-264, 2005.

[6] Anderson, M.B., 2004, "A Review of MCP Techniques", RES Report 01327R00022, Issue 03

[7] A.L. Rogers, J.W. Rogers, J.F. Manwell, "Comparison of the Performance of Four Measure-Correlate-Predict Algorithms," Journal of Wind Engineering and Industrial Aerodynamics, vol. 93, no.3, pp. 243-264, 2005.

[8] Thøgersen, M. L., and Per Nielsen. 2006. An introduction to the MCP facilities in WindPro. Aalborg: EMD International A/S, http://www.emd.dk/download/MCPforWindPRO25Beta.pdf

[9]Siddabanthi P.,Sørensen T., Uncertainty of long term correction of wind speed : The importance of consistency of data. Greentech Energy Systems A/S, EMD International

A/S