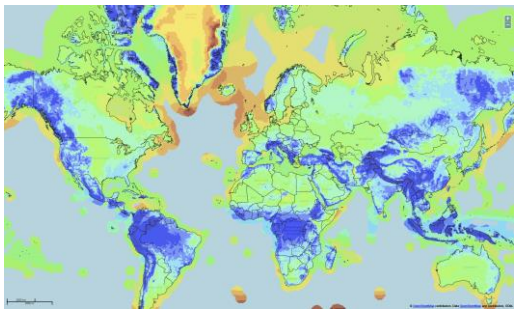
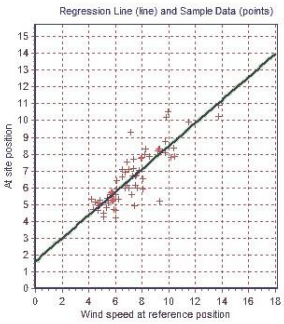
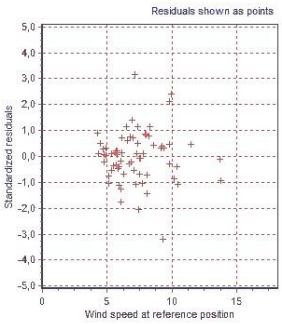
windPRO / MCP

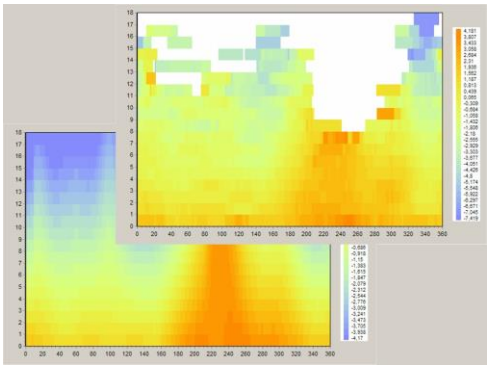
MCP - Measure-Correlate-Predict

An Introduction to the MCP Facilities in windPRO









EMD International A/S

Niels Jernes Vej 10, DK-9220 Aalborg, Denmark
www.emd.dk or www.windpro.com
phone +45 6916 4850
e-mail: windpro@emd.dk

Published by:

EMD International A/S
Niels Jernesvej 10
9220 Aalborg Ø
Tel: +45 6916 4850
Email: emd@emd.dk
web: www.emd.dk

Author
Morten Lybech Thøgersen, M.Sc.,

Co-authors
Per Nielsen, M.Sc.
Thomas Sørensen, M.Sc.
Lasse Ulrik Svenningsen, Ph.D.
Mads V. Sørensen, M.Sc.

Document history

18-08-05	1st draft – Morten Lybech Thøgersen, M.Sc., mt@emd.dk (NCAR/NCEP data extraction only)
20-11-05	2nd draft – Added description of Regression MCP and added more lengthy introduction
28-11-05	3rd draft – Matrix Method MCP description, Weibull Scale Description, Wind Index Description
29-01-07	1st released version
14-06-07	Added description on R^2 (coefficient of determination) in the Regression MCP chapter
16-06-10	Updated to reflect the windPRO 2.7 release. Removed section NCEP/NCAR data (now obsolete as this is now found in the online-services manual).
25-06-10	Added description of ‘r’ and ‘s’ in MCP-Predict tab (Section 6-4)
28-01-13	Added description on comparing 'r' and 's' between different methods (Section 6-4-1)
26-08-21	Updated chapter 7 and 8 and removal of outdated content.

Front cover

The front cover shows windPRO logo. Upper right picture shows the coverage of the ERA5 dataset. Lower left figure shows linear regression MCP analysis, with fit and residuals from the analysis. Lower right figure shows sample surface and model surface for the Matrix MCP service.

Table of Contents

1. Introduction

This document gives an introduction to the Measure-Correlate-Predict tools (MCP) that are included with windPRO. The MCP module in windPRO consists of the following sub-modules that were first introduced with the release of windPRO 2.5 in 2005:

1. Linear Regression MCP
2. Matrix method MCP
3. Neural Network MCP
4. Weibull Scale MCP (MCP2005)
5. Wind Index MCP (MCP2005)

Semi-offshore turbines at Frederikshavn, Denmark.

The Measure-Correlate-Predict toolbox in windPRO (MCP) enables the user to calculate long term corrected wind data directly in windPRO. The MCP module provides not only a direct access to different MCP-methods, but also provides reporting through overview-reports and detailed reports for each of the methods available.

Data that is used by the MCP-methods are contained in the windPRO Meteo object(s). This object is the data container for wind data saved as either time series data, table data or Weibull distribution parameters. Most MCP-methods require two overlapping time series, each holding a concurrent time series for the reference position and site position respectively. The long-term reference data could be either Weibull data, table data or time series data.

The result from the MCP analysis is – typically – one or more new long term corrected Meteo-Object(s). This meteo-object is located at the exact same position as the one holding the site data – and the intension is, that this long term corrected object should be used in further analysis, e.g. in WAsP, PARK or other windPRO modules.

Long term reference wind data is available directly from within windPRO as a part of the windPRO online services, see more in the windPRO documentation available from the windPRO knowledge base at http://help.emd.dk/knowledgebase [14].

1.1 About the Application of MCP

MCP is the abbreviation for Measure-Correlate-Predict techniques, which is widely in use for establishing a long-term wind statistic using limited wind data from the current site and long-term data from a more-or-less nearby site.

The goal of any estimation of a long-term wind statistic is to establish a transfer model between the available short-term or long-term wind data and the long-term statistic on the prediction site. The transfer model can be grouped into (at least) four different types:

1. Physical models (e.g. CFD flow models)
2. Statistical models
3. Empirical models
4. Other (combinations of the above, e.g like WAsP)

MCP-models may belong to any one of the categories or a combination hereof, indicating that the application of MCP-models has a very wide scope:

Some MCP-models operate on large timescales – like index correction methods, where monthly data typically are used. Other MCP-models tries to decode a one-to-one relationship between wind speeds and wind

directions on site and on the reference mast, calling for high quality measurements. In some situations, MCP-methods are applied in order to correct the lack of ability for a model to take long-term variation into account. This is the case when using local short-term site data in a WAsP analysis. Unfortunately, not all methods will perform equally well in all situations, calling for the user to get acquainted to the performance and limitations of the individual methods.

In general, the applied MCP method modifies one or more of the following descriptive data:

1. Wind energy index
2. WTG energy index (measured production from the WTGs)
3. Weibull A-parameter + Weibull k-parameter
4. Wind speed and possibly also wind direction

In windPRO methods working on (1), (3) and (4) are implemented.

1.2 A Conceptual Model for the MCP-Framework

A conceptual model for the measure-correlate-predict model is shown in Figure 1 below. From this figure it is seen that three potential different measured datasets are input to the MCP-model:

- Reference data: Long data
- Reference data: Short term data (concurrent data)
- Site data: Short term data (concurrent data)

The long term data could be any type of distribution, e.g. Weibull data, table data (joint wind distribution) or time series data. The concurrent datasets are typically required to be time series data. Not only measurements on wind speeds and wind direction could be used as input to advanced MCP models but also temperature differences could prove useful. However none of the models in windPRO currently uses other input than wind speed and wind direction.

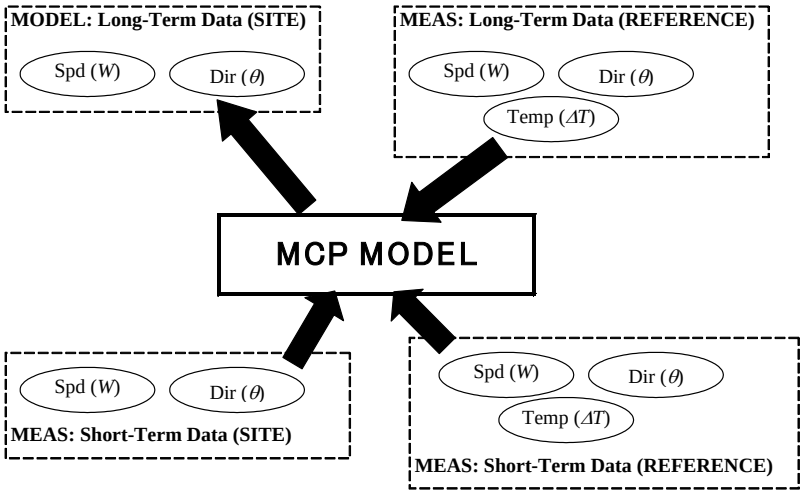


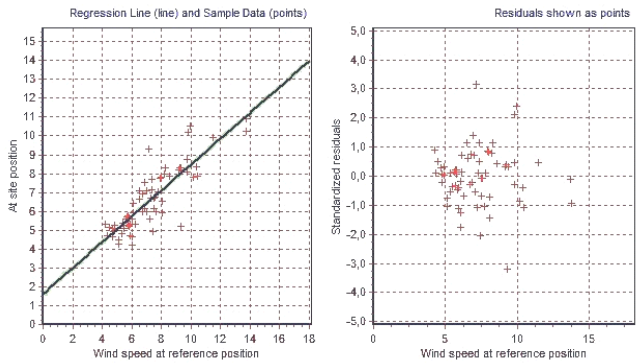
Figure 1: Conceptual Model for Measure-Correlate-Predict Models.

Note: When doing MCP calculations in windPRO, the input data are always internally treated as three different input sources. However, in many cases the long term reference data and the concurrent reference data series may come from the same time series data source. In such a situation, windPRO will automatically select the same series for both input sources.

1.3 The MCP facilities in windPRO

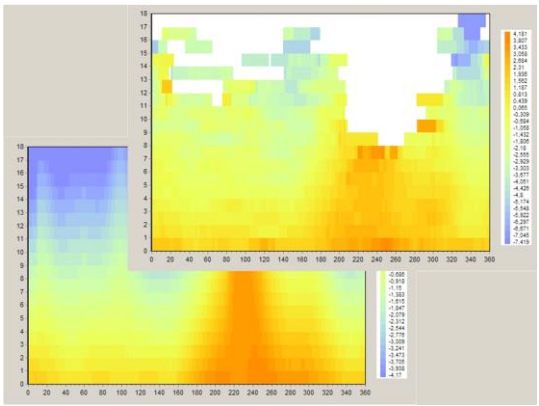
Below is a list of sub-sections briefly outlining the MCP facilities, which was introduced from windPRO 2.5 and improved with later windPRO releases.

1.3.1 Linear Regression MCP



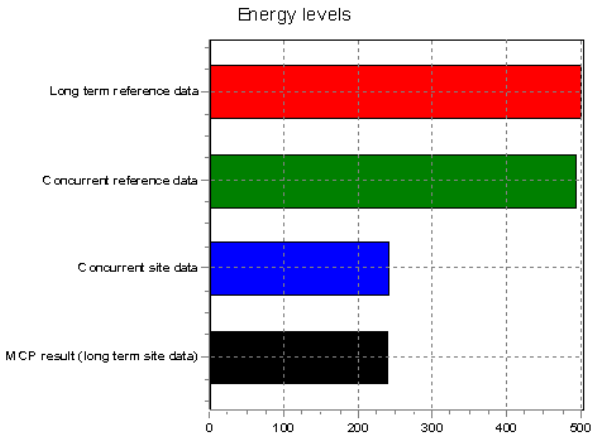
The (Linear) regression tool enables the user to inspect the fit directly through an animated graph. If the fit is not satisfactory, a wide range of parameters may be fine-tuned to provide a better fit. The regression tool is not limited to linear regression, but also higher order polynomials may be used in modelling wind speeds and wind veer.

1.3.2 Matrix Method MCP



The matrix method in windPRO models the changes in wind speed and wind direction through a joint distribution fitted on the 'matrix' of wind speed bins and wind direction bins. The user may choose to either use polynomials fitted to the data statistics or – where appropriate - to use the measured samples directly when doing the matrix MCP.

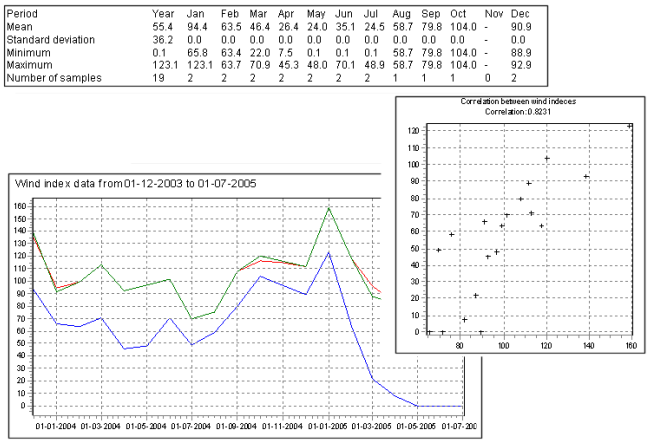
1.3.3 Weibull Scale MCP (only in MCP2005)



The Weibull Scale method is a very simple empirical method, which does its manipulation directly on the Weibull form and scale parameters (A,k) as well as on the frequency distribution.

The Weibull method has the advantage, that it will match the nature of the wind at most places, but beware that application of this method should be done with caution on locations with significant non-Weibull distributions.

1.3.4 Wind Index MCP (Only in MCP2005)

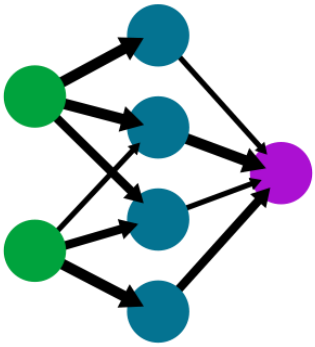


Period	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	55.4	94.4	63.5	46.4	26.4	24.0	35.1	24.5	58.7	79.8	104.0	-	90.9
Standard deviation	36.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0
Minimum	0.1	65.8	63.4	22.0	7.5	0.1	0.1	0.1	58.7	79.8	104.0	-	88.9
Maximum	123.1	123.1	63.7	70.9	45.3	48.0	70.1	48.9	58.7	79.8	104.0	-	92.9
Number of samples	19	2	2	2	2	2	2	2	1	1	1	0	2

The index correlation method is a method typically making the MCP analysis by using monthly averages of the energy yield, thus disregarding the directional distribution of the winds. Even though this method may seem rather crude and primitive when comparing to other more advanced MCP methods, it has its advantages in stability and performance – even in the cases where other MCP methods seem to fail.

The Wind Index MCP method in windPRO offers the opportunity to calculate the wind indexes using real power curves from the wind turbines included in the wind turbine catalogue in windPRO. Also a generic power curve may be chosen.

1.3.5 Neural Network MCP



The Neural network model is a machine learning algorithm which takes in the concurrent period of measured wind data and uses it to train a neural network. This neural network detects a pattern between the reference wind conditions and the measured wind conditions

2. Regression Measure-Correlate-Predict Methods

2.1 Introduction

The regression MCP method holds the traditional linear regression MCP analysis as well as other - more general - regression models using polynomials of other orders. Specializations of the polynomial-fitting methods are also included, provided in order to comply with methods used or suggested by other companies. One of the specialization is here an MCP method forcing the regression line through the origin (0,0). However, this option should be used only with caution as it typically provides a significantly poorer fit to the data than the methods where a non-zero intersection with the y-axis is allowed.

Wind Farm in Southern Europe

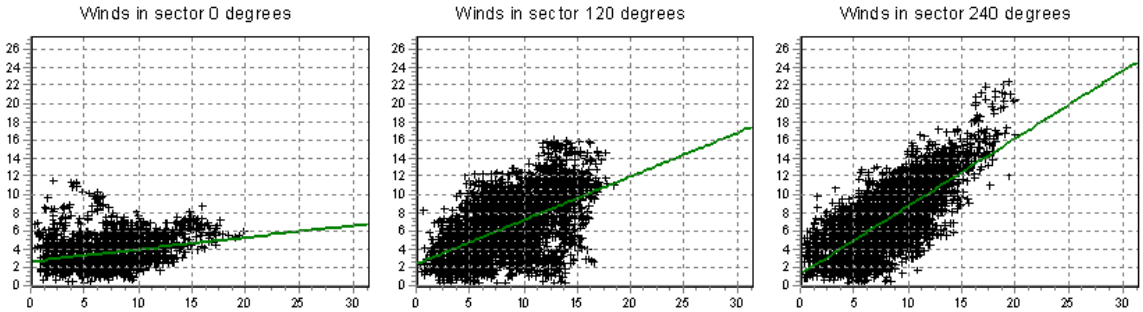


Figure 1: Linear fitting of wind speeds (x: winds at reference position, y: winds at site position)

The regression MCP methods in windPRO are improved over a traditional linear regression analysis, as a model for the distribution of the residuals is also included. This model allows the regression MCP method to capture the energy content in the MCP corrected site wind distribution much better than regression models without this option. Experience has shown, as much as 10% energy can be erroneously lost in the long-term correction if the model is run without this option.

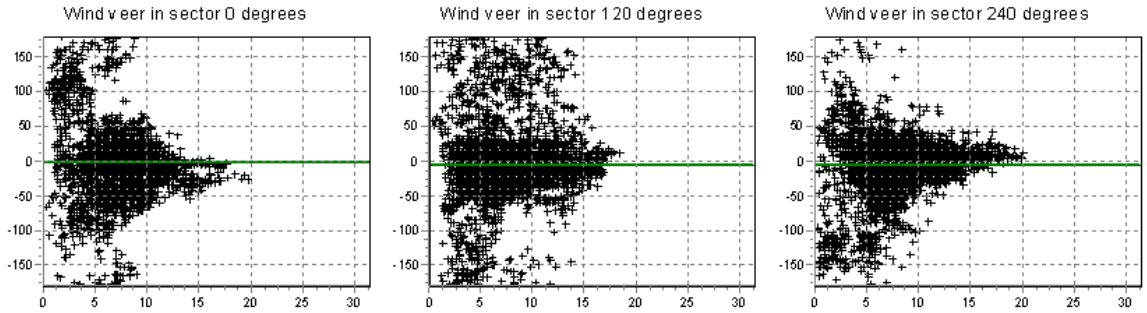


Figure 2: 'Constant' fitting of the wind veer (x: winds at reference position, y: veer from reference to site).

The Figure 1 and Figure 2 show examples on the fitting of wind speed and wind direction respectively. Note, that a very significant scatter is present. windPRO models this scatter through the residuals in the linear model, which are included as zero-mean Gaussian random variables conditioned on the wind direction and also on wind speed. Especially for the wind speeds, it may be relevant to include a model of the scatter (residuals). This is due to the fact, that the measured points above the regression line represent significantly more energy than the points below the line. This is due to the third power relationship between wind speed and energy and the non-linearity of the WTG power curves. Failure in including the scatter will

erroneously remove as much as 5-10% of the measured energy, though applying the scatter where a Gaussian model does not apply may add too much energy.

2.2 The Regression Model

Regression modeling, where only one independent (x) and one dependent (Y) variable is present, is based on the following equation:

$$Y = f(x) + e$$

where Y is the dependent variable
x is the independent variable
$f(x)$ is the regression model
e is a random error (residual)

The regression model could be polynomials of any order or other models, but traditionally a linear model is assumed, as this model has been found to give reasonable fits for wind energy estimation. In the case of a regression MCP analysis, the independent variable could be the wind speed measured at the reference position. The dependent variable (Y) is then the wind speed at the local WTG site position.

3.2.1 Wind Speed and Wind Veer Regression Model

In windPRO, the regression models are made as models for wind speed and wind veer. The wind veer is taken as the difference in wind direction at the site mast and the wind direction at the reference mast. All analyses are made conditioned on the wind direction on the reference mast, i.e. the fitted regression parameters are functions of reference site wind direction. The binning is made with one-degree resolution in the angular variable and a 1.0 m/s resolution in the wind speed variable. However, data are loaded into these bins from larger window - typically a 30 degree window is used – otherwise some bins may suffer from missing data.

The following regression models are currently available in windPRO:

Model description	*Model equation, f(x)*
No model	$Y = x$
Constant	$Y = \beta_0$
Linear - 1st order polynomial	$Y = \beta_1 \cdot x + \beta_0$
Linear regression through (0,0)	$Y = \beta_1 \cdot x$
Second order polynomial	$Y = \beta_2 \cdot x^2 + \beta_1 \cdot x + \beta_0$
Second order polynomial through (0,0)	$Y = \beta_2 \cdot x^2 + \beta_1 \cdot x$

The regression model named ‘No model’ actually just transfers the measured data from the reference position to the site position, i.e. if no model is chosen for the wind veer model, then the directional distribution for the site will be the same as for the reference site. Please note that the modelling of the residuals should be disabled if such a transfer is desired.

2.2.2 Estimation of the Regression Parameters

The regression parameters are estimated through a least squares algorithm. The least squares method calculates the summed squared difference between the estimated response and the actual measured response:

$$SS_R = \sum_{i=1}^{n} \left[Y_i - f(x_i) \right]^2$$

The least squares method then chooses the set of regression parameters in f(x), that minimizes the squared sum, SS_R. Note, that for the linear model, analytical expressions are available for calculating the optimal parameters enabling a quite fast calculation with this model, see S.M. Ross [11]. However for polynomials of other orders, a general optimization algorithm is used instead, resulting in longer calculation times. The optimization algorithm is the Amoeba method, e.g. found in Press et al [13].

2.2.3 The Distribution of the Residuals, *e*

As expressed earlier, the regression model yields,

$$Y = f(x) + e$$

where *e* is a random error.

The distributions of the random errors may, according to Ross [11], reasonably be assumed to follow a zero mean Gaussian distribution, $e \sim N(0,\sigma)$. However, in many cases, data shows that a better assumption is to model the residuals conditioned on both the wind direction and the mean wind speed. This model is also available in windPRO, where the residuals are still modelled as Gaussian, but with the mean and standard deviation conditioned on both wind speed and wind direction. Note that caution is needed when choosing a residual model: such a model is needed in the MCP-modelling in order to give the right energy levels in the new MCP-corrected time series. Currently (2010), the default model in windPRO 2.7 is to model residuals for wind speeds conditioned on both wind direction and wind speed, see the examples in Figure 5.

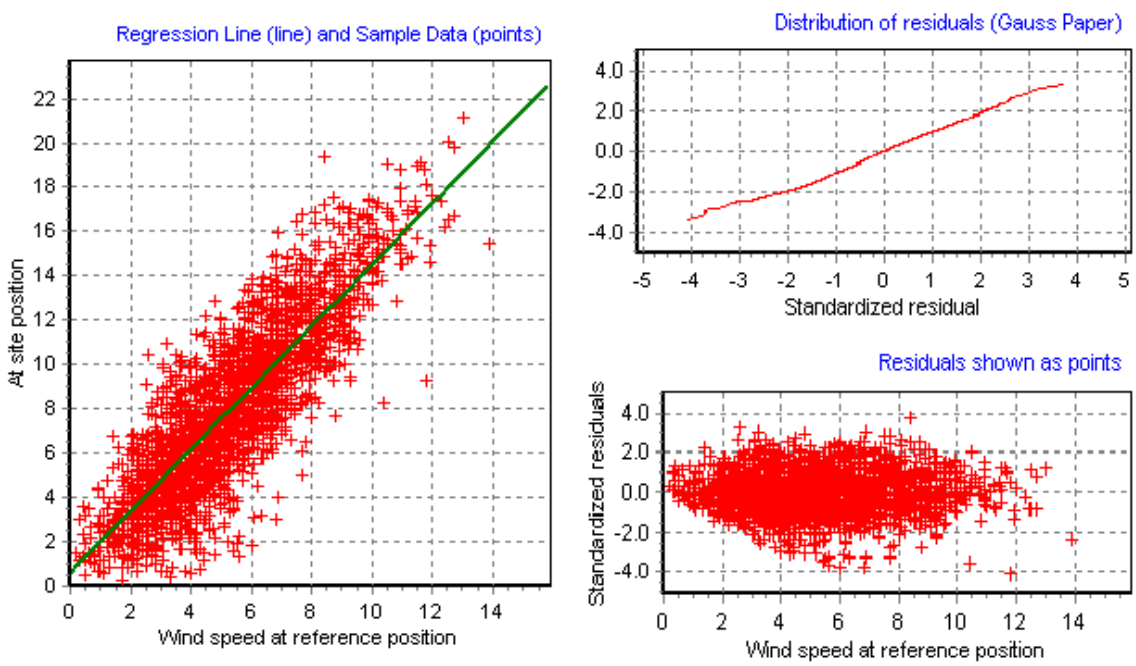


Figure 3: Wind Speed Fit (Left) with the Distribution of the Residuals (plot captured from windPRO). To the right is shown the modelling of the mean and standard deviation of the residuals.

When comparing the distribution of the residuals, it is convenient to consider the Standardized Residuals. This is a normalization procedure made to bring the residuals on a common scale:

$$\frac{Y_i - f(x_i)}{\sqrt{SS_R / (n-2)}}$$

where

Y_i is the measured response for the i-th sample (x_i)
$f(x_i)$ is the model prediction for the x_i sample
n is the total number of samples
SS_R is the sum of squares of the residuals

Inspecting the Figure 3 where the wind speeds is plotted, it is seen, that the distribution of the residuals follows a Gaussian distribution quite well. This is easiest seen from the top-right figure showing the distribution of the residuals plotted on Gauss paper, i.e. a straight line in this graph and the distribution is a Gaussian distribution. The lower right figure shows, that the residuals seem to be reasonably independent on the reference wind speed (the standard deviation seems quite constant regardless of the reference wind speed).

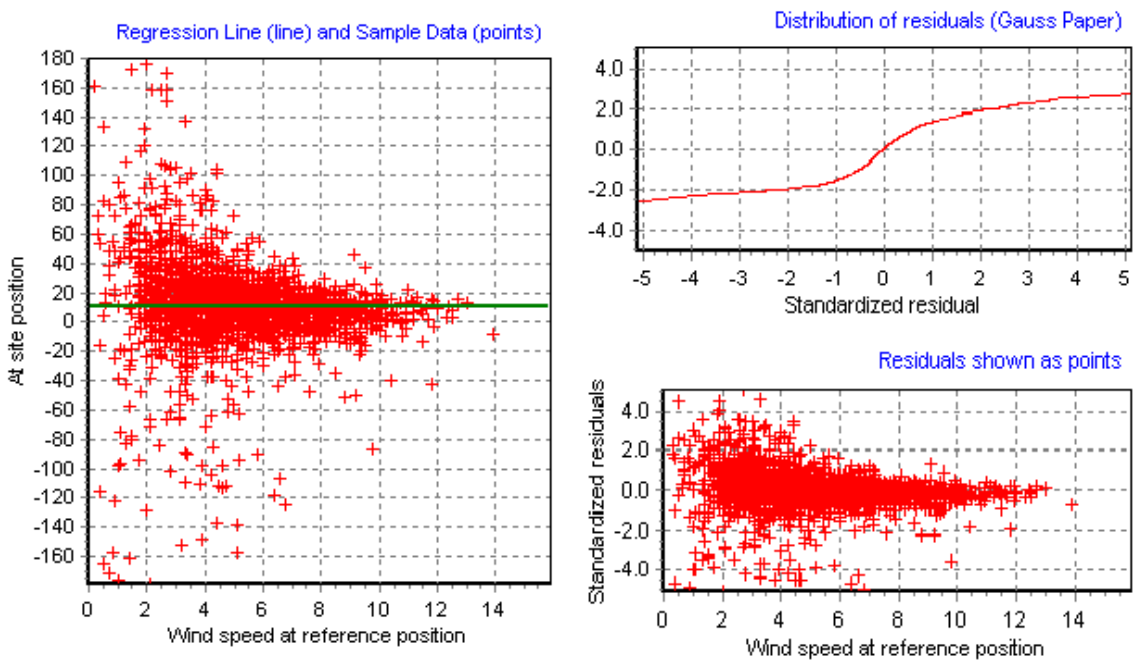


Figure 4: Wind Veer Fit with the Distribution of the Residuals.

The Figure 4 shows the fit to the wind veer distribution. By inspecting the Gaussian paper plot it is seen that a Gaussian distribution (straight line) is a quite raw approximation in this case. Also, when inspecting the lower right graph with the plot of the residuals, it is also noticed that the wind veer values at the low reference site wind speeds has a significant higher standard deviation than the ones at high wind speeds. This suggests that a revised model should be refined to include a residual random variable with standard deviation conditioned not only on the reference site wind direction but also the reference site wind speed. This model is actually also included with windPRO as an 'Advanced Gaussian' model. An example of this modelling is shown in the Figure 5 where the blue line is the conditioned mean value and the green lines are the mean plus/minus one standard deviation. As an alternative, one could investigate the modelling options in the Matrix Method MCP; these are also able to catch such effects.

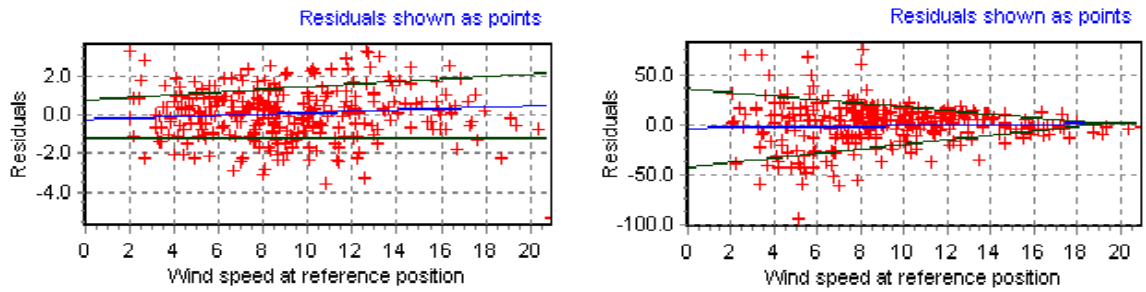


Figure 5: Example of the Advanced Gaussian Residual Model. Left: Residuals of Wind Speed. Right: Residuals of Wind Veer.

2.2.4 R^2: Coefficient of Determination

The coefficient of determination, R^2, is a statistic used for interpreting the proportion of variability in the data set that is accounted for by the selected statistical model. The R^2 gives some information about the goodness of fit of the chosen model. The R^2 is defined as

$$R^2 = 1 - \frac{SS_E}{SS_T}$$

$$SS_T = \sum_i \left(Y_i - \overline{Y}\right)^2 \qquad SS_E = \sum_i \left(Y_i - \hat{Y}_i\right)^2$$

where SS_T is the total sum of squares
SS_E is the sum of squared errors
Y_i is the sample considered
$\overline{Y}$ is the mean of the samples
$\hat{Y}_i$ is the response value for the sample i - calculated using the chosen regression model

When considering a linear regression model, $Y=A\cdot x+B+e$, then the R^2 value is actually an indication of whether the linear model is an improvement in the fit rather than the model $Y=A+e$ (i.e. using a constant model only). Again, e is a random error (residual). Please note, that by definition, the R^2 always relates the chosen model to the constant model.

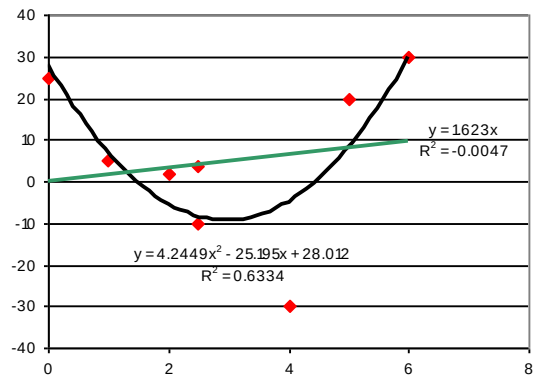
Typically, the value of R^2 is in the range from 0 to 1. However - in the case where a regression model with a fixed intercept is used (e.g. when forcing a line through (0.0)) - then the R^2 value may in fact become negative. This happens when SSE is larger than SST. This is only an indication that the chosen model does not fit the data as well as the 'constant' model.

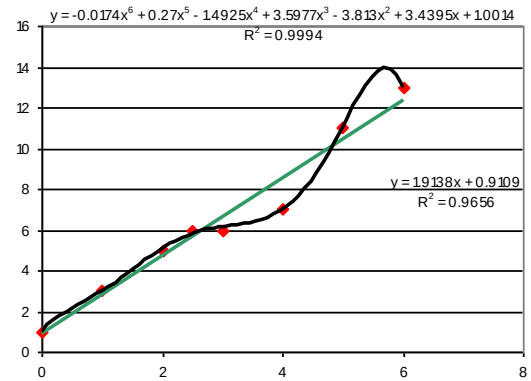
The R^2 values are often assessed like this:

R^2 = 1.00:	Indicates that all response variation is explained by model chosen (because of a prefect fit, so SS_E =0)
R^2 = 0.00:	The model chosen performs like the 'Constant' model
R^2 = 0.75:	About 75% of the variation in the response variable is explained by the model. The 25% remaining is explained by inherent variability, unknown or hidden variables.

The value $R = \sqrt{R^2}$ is called the index of fit. It commonly used as an indication of how the model fits the data. The index of fit is also named the sample correlation coefficient, because – in the case of a linear model – it is a natural estimator of the correlation coefficient, ρ. If not using the linear model, then this relationship does not hold.

In the two graphs below different models have been fitted to sample data, using the Microsoft Excel spreadsheet tool. In the graph to the left, a linear curve with intercept in (0,0) is fitted. Note that the R^2 is negative, implying that a constant model would be a better choice. Also shown is a 2 order polynomial fit with R^2 = 0.63. In the graph to the right, a linear regression fit with a high R^2 is shown (R^2 = 0.97). Also shown is a 6-th order polynomial fit with R^2 = 0.999. This high R^2 does not necessarily mean that the 6-th order polynomial is a (physical) better model, because the variability in the response may actually be caused by some random component inherent in the sample data.





2.3 Calculating the Long Term Corrected Data

The long term corrected meteorological data are calculated using Bootstrap and Monte-Carlo simulation techniques, i.e. probabilistic methods enabling generation of the long term corrected wind distribution through an 'artificial' time series. For details on Monte-Carlo simulation, see [15]. For an introduction to the Bootstrap, see [12].

An outline of the calculation and simulation procedure is as follows:

1. Fit the wind speed model with the corresponding residuals (conditioned on the reference wind direction, i.e. actually in 360 bins from 0 to 359 degrees).
2. Fit the wind veer model with the corresponding residuals (also conditioned on the reference wind direction, binned from 0 to 359 degrees)
3. Given the long term reference wind distribution, simulate out a random sample of wind direction and wind speed, (W_{ref}, θ_{ref}).
4. Using the wind speed model in (1), calculate a sample of the on site wind speed data: $W_{site}=f(W_{ref})+e$. Note that f is the regression model and e is a realization of the residual random variable.
5. Using the wind veer model in (2), calculate a sample of the on site wind direction: $\theta_{site}=\theta_{ref}+f(\theta_{ref})+e$. Note that f is the (wind veer) regression model and e is a realization of the residual random variable. Also note, that the mathematical formulation is slightly different from the wind speed model, as we are here modelling the wind veer and thus need to add the reference wind direction.
6. Repeat 2-5 until the number of samples is 'sufficiently' large. This is typically the case when the number of 'artificial' samples has reached the number of samples expected in the long term reference period. If a time series is available for the entire reference period it is preferable to pick the actual measurements sequentially from the reference in (3). The artificial time series is then complete when all reference measurements have been transformed.
7. From the sample distribution generate table data and fit Weibull table data.

The simulation of samples from the long term reference distribution may (step 3) may be done using data from either the fitted Weibull data, the distribution table or from a long term reference time series. It is recommended to use the data from the distribution table or time series, as the Weibull data are subjected to modelling errors.

When modelling from Weibull or table data, then no correlation structure in the generated time series is preserved. Thus only the resulting distribution tables should be compared to the actual measurements and used in further analysis.

If however the reference data input is a time series a correlation between measured site data and predicted site data is a measure of the success of the prediction.

3. Matrix Method MCP

3.1 Introduction

The matrix method in windPRO models the changes in wind speed (speed-up) and wind direction (wind veer) through a joint distribution fitted on the 'matrix' of wind speed bins and wind direction bins.

Wind Turbines at Göteborg, Sweden

The parallel period of measured wind data is used to calculate the set of the transfer functions, used for transferring wind speeds and wind directions from the reference site to the site position. Since real measurements will suffer from data missing in bins in the dataset, this method needs a way to fill out missing input bins. In windPRO, polynomials are fitted to the statistics of the sample data enabling this interpolation/extrapolation. The user may choose to either use polynomials fitted to the data statistics or, where appropriate, to use the measured raw samples directly when doing the matrix MCP.

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. Actually this relationship is basically modeled through a joint distribution between the two variables wind speed-up and wind veer. How this joint distribution is modeled should actually depend on the data in question, but experience using the windPRO MCP, suggests that a combination of binned sample distributions and a modeled joint Gaussian distribution seem to work quite well.

The transfer model, given as a conditional distribution, is actually the key distribution in the generalized matrix method. The distribution gives the 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*.

3.2 Matrix MCP Modelling

The Matrix MCP in windPRO models the wind speed-up and the wind veer as functions of the wind speed and wind direction on the reference site, see Figure 1 and Figure 2. Two modelling options are available in windPRO, either to use the measured samples themselves (through a resampling technique) or alternatively to use the fitted polynomial model (actually through a joint Gaussian random variable). If no sample data is available in a particular bin, then the polynomial model is always used.

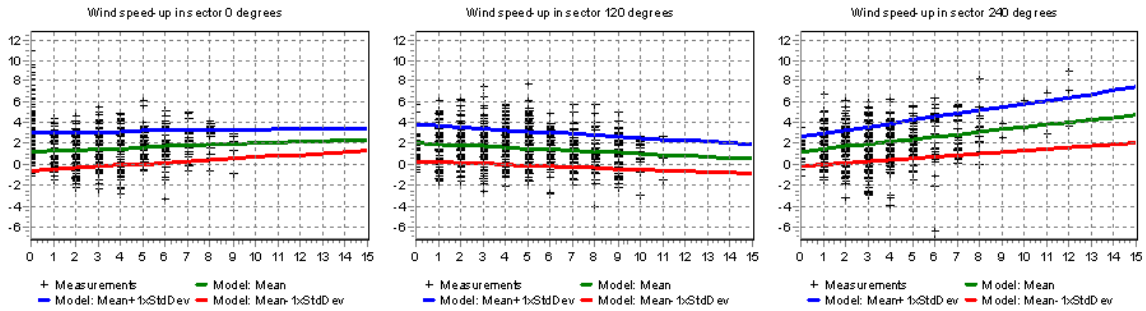


Figure 1: Sample data and first order model for the wind speed-up (x: wind at reference, y=speed-up).

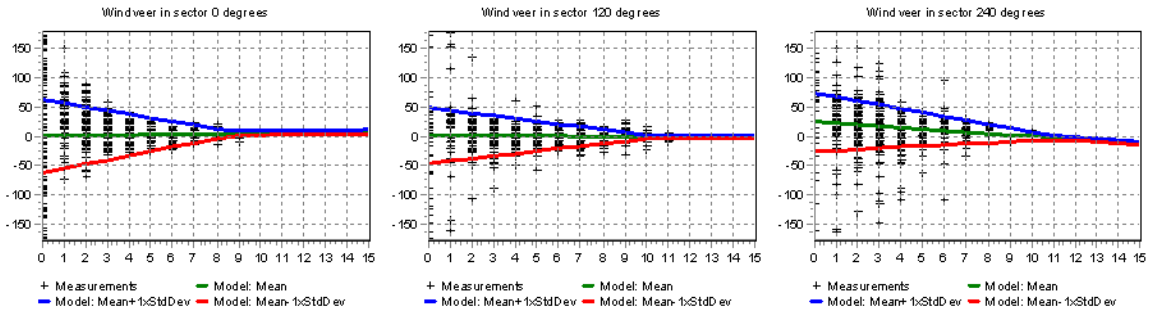


Figure 2: Sample data and first order model for the wind veer (x: wind at reference, y=wind veer).

The model is based on the joint distribution of the measured wind speed-ups and wind veers. Thus, for each measured sample we have to calculate/measure pairs of the two quantities (a pair is data with identical timestamps):

$$\Delta u = u_{site} - u_{reference}$$
$$\Delta\theta = \theta_{site} - \theta_{reference}$$

where Δu is the wind speed-up
u_{site} the wind speed at the site position
$u_{reference}$ the wind speed at the reference position
$\Delta\theta$ is the wind veer
θ_{site} the wind veer at the site position
$\theta_{reference}$ the wind veer at the reference position

The joint distribution of $f_{\Delta u,\Delta\theta}$ is then modelled conditioned on the wind speed and the wind direction on the reference site. This joint distribution is represented as either through the samples (bootstrap model) or through a joint Gaussian distribution. In the case of the joint Gaussian distribution, the distribution parameters – mean, standard deviation and correlation - are modelled through polynomials of any order.

3.2.1 Sample Data, Sample Distributions and Statistical Moments

When the data has been measured and a match between the short-term site data and the short-term reference data has been established, then the samples are sorted into bins with the resolution 1 m/s and 1 degree. Since a 1-degree angular resolution is too small in most cases, it is possible to feed in data from a larger window, typically pre-set to around 30 degrees. The result from this binning is a set of joint *sample distributions* of wind veer and wind speed-up. Since the data are binned with wind speed and wind direction, these sample distributions are said to be conditioned on the mean wind speed at the reference position and the wind direction on the reference position. The sample distributions calculated are used directly in a bootstrapping technique (see Efron & Tibshirani [1]) when doing the Matrix MCP calculation.

Based on the sample distributions, the following sample statistics are calculated for the wind veer and the wind speed:

- Mean value
- Standard deviation
- Skewness
- Kurtosis
- Correlation

An example is shown in Figure 3 (left column).

3.2.2 Polynomial Model of the Statistical Moments

In order to enable interpolations and extrapolations into bins where no data are present, we choose to parameterize a model fitted to the sample distribution statistics. This parametric distribution is represented by the two first statistical moments and the correlation, and it is assumed that a joint Gaussian distribution is a reasonable distribution assumption (see the section 4.2.3 below).

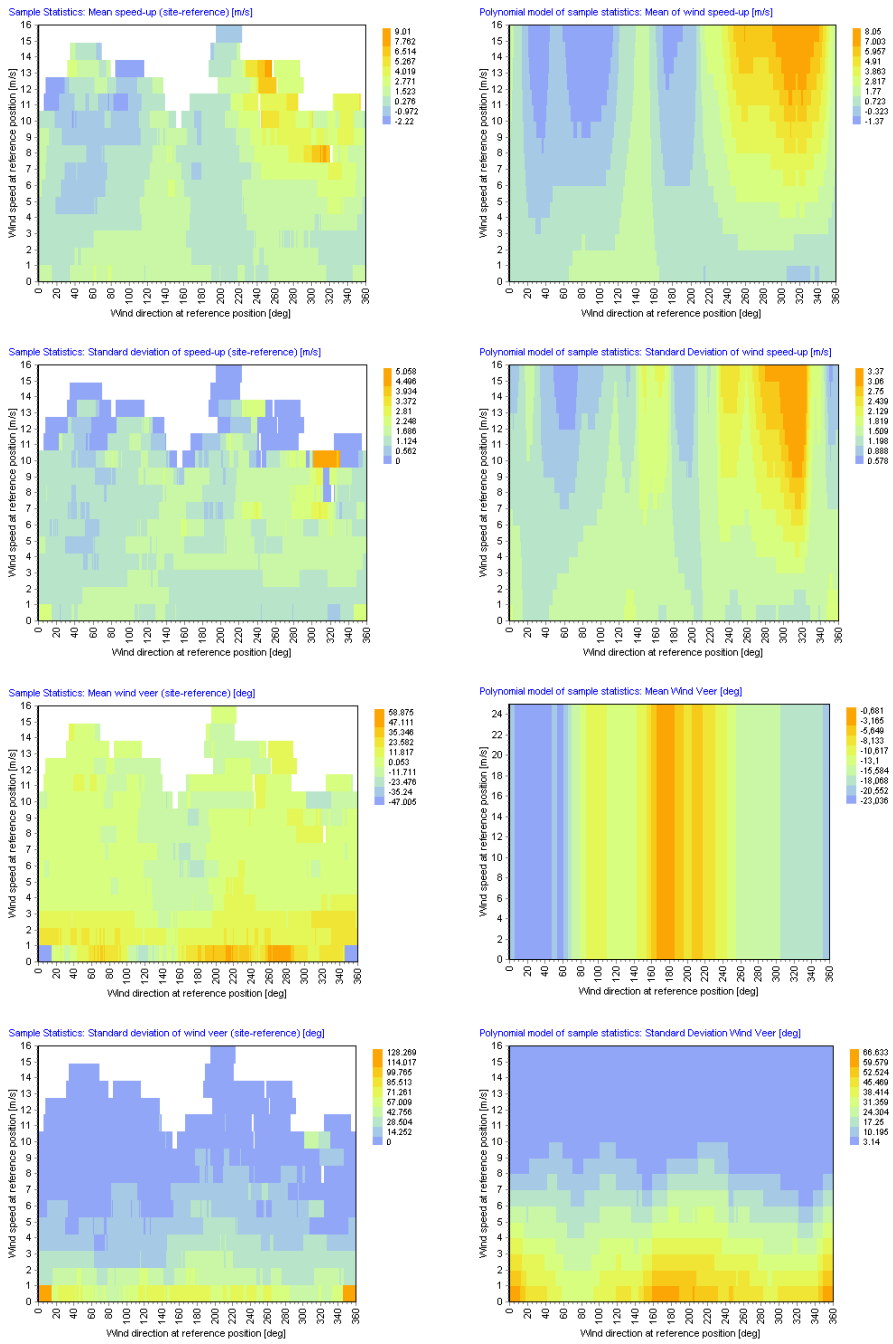


Figure 3: Left: Sample Statistics, Right: Model of the Statistics Moments.

The mean, standard deviation and correlation are now modelled as ‘slices’ of polynomial surfaces:

$$P \mid (U_{reference}, \theta_{reference}) = \sum_{i=0}^{n} a_i(\theta_{reference}) \cdot U_{reference}^{\ i}$$

where P denotes the sample statistical moment (or correlation) considered
n is the order of the polynomial
a_i is the polynomial coefficients (which are also functions of $\theta_{reference}$)

An example of this modeling is shown in Figure 3 (right side). It is seen, that the surface model seems to capture the trends in the sample data quite well (left side of the same figure). In this case 1st order polynomials are used in all cases except Mean wind veer where a 0th order polynomial is used.

3.2.3 Parametric Model for the Joint Wind Speed-up and Wind Veer Distribution

As mentioned earlier, it is assumed that the joint distribution of wind veer and wind speed is adequately modelled by a joint Gaussian distribution. This bivariate probability density function is given by:

$$f_{\Delta u,\Delta\theta}(\Delta u, \Delta\theta) =$$

$$\frac{1}{2\pi\sigma_{\Delta u}\sigma_{\Delta\theta}\sqrt{1-\rho^2}} \cdot \exp\left[-\frac{(\Delta u - \mu_{\Delta u})^2 / \sigma_{\Delta u}^2 - 2\rho(\Delta u - \mu_{\Delta u})(\Delta\theta - \mu_{\Delta\theta}) / \sigma_{\Delta u}\sigma_{\Delta\theta} + (\Delta\theta - \mu_{\Delta\theta})^2 / \sigma_{\Delta\theta}^2}{2 \cdot \sqrt{1-\rho^2}} \right]$$

In the Figure 4 below, two examples of the joint wind veer and speed-up distribution are shown. This particular example was calculated using the following mean, standard deviation and correlation:

Wind speed-up: $(\mu_{\Delta U}, \sigma_{\Delta U})$ = (0.3 m/s, 0.5 m/s)
Wind veer: $(\mu_{\Delta\theta}, \sigma_{\Delta\theta})$ = (5.0 deg, 7.0 deg)
Correlation $\rho_{\Delta U \Delta\theta}$ = 0.1 (left figure), $\rho_{\Delta U \Delta\theta}$ = 0.9 (right figure)

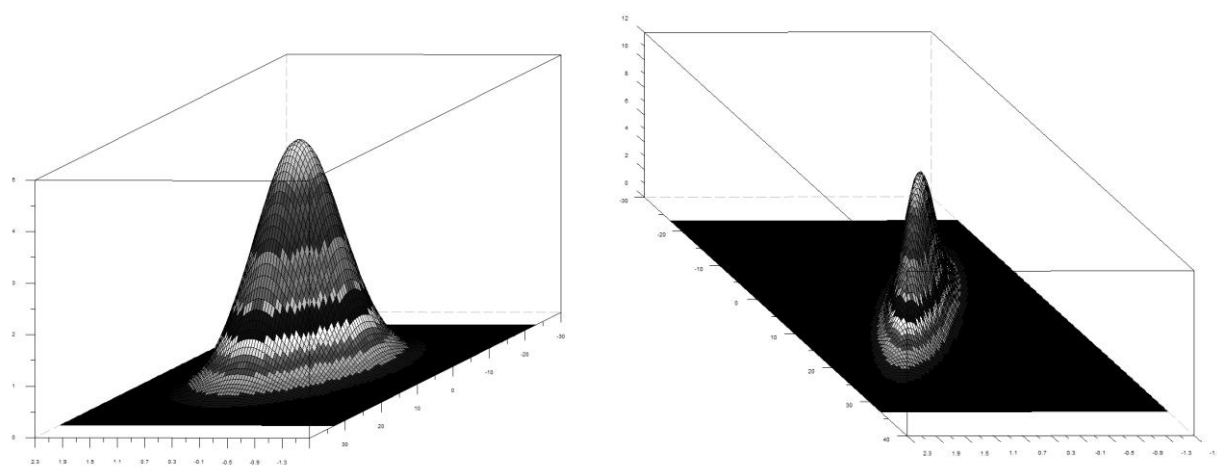


Figure 4: Bivariate Gaussian Distribution: Example of the Joint Speed-Up and Wind Veer Distribution.

The parametric model for the veer and speed-up distribution is typically only used in the cases where only few samples are available in the bin. In that case, the polynomials mentioned in Section 4.2.2 are used to find the distribution parameters. Then a realization of the wind veer and the wind speed-up are made using the appropriate techniques.

3.3 Calculating the Long Term Corrected Data

As in the case of regression MCP (see Chapter 3), the long term corrected meteorological data are calculated using Bootstrap and Monte-Carlo simulation techniques, i.e. probabilistic methods enabling generation of the long term corrected wind distribution through an 'artificial' time series. For details on Monte-Carlo simulation, see [15]. For an introduction to the Bootstrap, see [12].

An outline of the Matrix MCP calculation and simulation procedure is as follows:

1. In the concurrent period, the wind veer and wind speedup are calculated as the site data minus reference data.
2. All sample sets (ΔU, $\Delta\theta$) are sorted/grouped into the appropriate bins.
3. From the binned sample distributions, the statistical moments are now calculated (mean, standard deviation and correlation).
4. The polynomial surfaces used to describe the statistical moments are fitted.
5. Given the long term reference wind distribution a random sample of wind direction and wind speed, (W_{ref}, θ_{ref}) is simulated.
6. Knowing the wind speed and wind direction on the reference site, the appropriate bin is now found. Using the data connected to this bin, a realization of the wind veer and wind distribution is now simulated from either the real sample data (by a bootstrapping technique) or using the modelled polynomials and the joint Gaussian model (by a Monte Carlo simulation technique).
7. Repeat 5-6 until the number of samples is 'sufficiently' large. This is typically the case when the number of 'artificial' samples has reached the number of samples expected in the long term reference period. If a time series is available for the entire reference period it is preferable to pick the actual measurements sequentially from the reference in (5). The artificial time series is then complete when all reference measurements have been transformed.
8. From the sample distribution generate table data and fit Weibull table data.

The simulation of samples from the long term reference distribution (step 3) may be done using data from either the fitted Weibull data, the distribution table or from a long term reference time series. It is recommended to use the data from the distribution table or time series, as the Weibull data are subjected to modelling bias (as the Weibull distribution fit used is an 'Energy' fit).

When modelling from Weibull or table data, then no correlation structure in the generated time series is preserved. Thus only the resulting distribution tables should be compared to the actual measurements and used in further analysis.

If however the reference data input is a time series a correlation between measured site data and predicted site data is a measure of the success of the prediction.

4. Wind Index MCP

4.1 Introduction


Brokilde Wind Farm

The wind energy index of a particular period is defined as ratio between the energy in the considered period (such as a particular month) and the average (monthly) energy taken over a longer reference period. For instance: a monthly wind index in January 2010 of 115 means that this particular month has a 15% higher wind energy than the average month in the reference period – which is typically taken as 20 or 30 years. The formulation used in windPRO for the Wind Index is:

$$I = \left[\frac{\text{Average power output in some period}, T}{\text{Average power output in reference period}, T_{ref}} \right] \cdot 100$$

The wind energy index could be derived from wind data (as in the windPRO MCP module) or as production indexes (derived from databases on wind turbine performance (such as the Danish Wind Index available from www.vindstat.dk).

Thus, the index correlation method is a method for making the MCP analysis by using averages of the energy yield (typically monthly averages), thus disregarding the directional distribution of the winds. Even though this method may seem rather crude when comparing to other more advanced MCP methods, which takes the wind veer into account; this method has its advantages in stability and performance as it may even succeed in the cases where other MCP methods seem to fail.

This is due to the fact, that the wind indexes are related directly to WTG energy yield. The Wind Index MCP method in windPRO offers the opportunity to calculate the wind indexes using real power curves from the wind turbines included in the wind turbine catalogue in windPRO. Also a generic power curve based on a truncated squared wind speed approach may be chosen. When the wind indexes has been calculated, the MCP correction is done on the estimated WTG energy yield, i.e. by multiplying the production estimated with a correction factor based on the difference in the wind index from the short term site data to the long term site data estimate.

4.2 The Power Curve used in the Wind Index Calculation

The energy level in the wind is proportional to the third power of the wind speed. However, since the power curve of a WTG is a non-linear function of the wind speed, then the wind index is typically calculated as either

1. Through a generic power curve, e.g. the square of the wind speed for wind below the stall onset and a constant above the stall onset:

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

2. From a real power curve, $P(u)$, see an example from a Vestas V44 in the Figure 1.

The windPRO Wind Index MCP tool offers a possibility to calculate the wind index using both types of power curve methods.

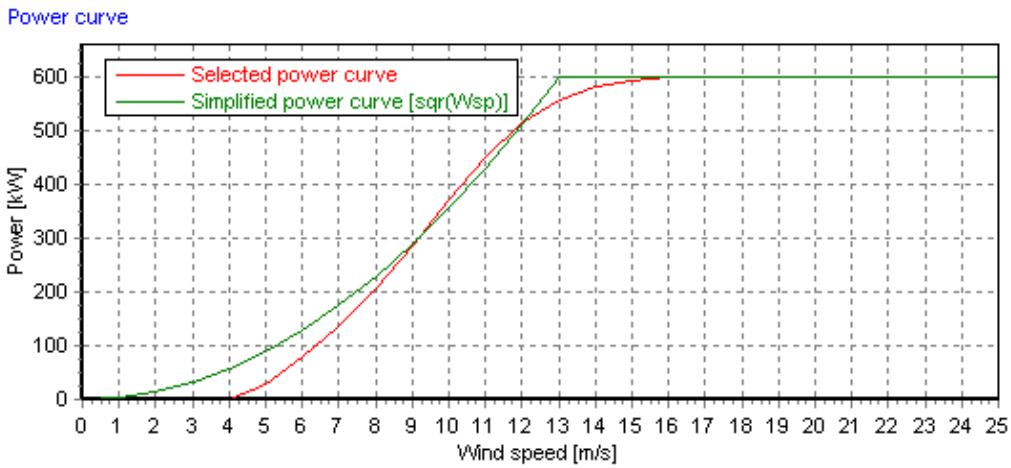


Figure 1: Power Curve from Vestas V44 and Simple Power Curve (scaled to fit the maximum power of the Vestas WTG).

4.3 Wind Speeds in the Wind Index Calculation

In order for the power output - calculated for local site and reference site - to be comparable; they must be based on a similar mean wind speed. This is done by assuming a sector uniform shear. This shear is then applied so that both concurrent mean when speeds are set to a fixed user-inferred wind speed, typically the expected mean wind speed at hub height.

The individual wind speed measurements are thus multiplied with the relevant factor. Both full time series wind speeds will be adjusted with the same ratio as was applied to the respective concurrent time series. The argument for this operation is that the variations in wind speed will only be interpreted correctly in terms of wind energy if a comparable section of the power curve is considered.

4.3 Calculating the Wind Index

The average power output in some period, W_T, is calculated using the modelled or measured power curve, $P(u)$:

$$W_T = \frac{\sum_{i=1}^{N} P(u_i)}{N}$$

Where N is the number of measured wind speeds within the period considered
u_i is the i-th wind speed measurement (typically 10 minute mean wind speeds)

In windPRO Wind Index MCP four different average power outputs are calculated or modelled. These are:

W_{Rf}, Average power output the full reference period of the reference wind dataset
W_{Rc} Average power output for the concurrent part of the reference wind data set

W_{Sf}. Average power output for the full reference period of the on-site wind dataset
W_{Sc} Average power output for the concurrent part of the on-site wind dataset

The power output for the full period on the reference site, W_{Rf}, is set to index =100, unless specifically stated otherwise, and the ratio in power output between W_{Rf} and W_{Rc} then gives the index of the concurrent period so that

$$I_{Rf} = 100$$
$$I_{Rc} = \frac{W_{Rc} \cdot 100}{W_{Rf}}$$

Where I is the wind index of the relevant period.

The assumption is now that the index for the concurrent period at the reference site is the same as the index for the concurrent period at the on-site, thus

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

That this is a valid assumption needs to be validated through correlation, see next section on Wind Index Correlation. Now knowing the index of the concurrent period on the site means that it possible to find the expected average power output for the entire site measurement period. This is done simply by multiplying the average power output of the concurrent period for the on-site measurements with the wind index. The relation is:

$$I_{Rc} = \frac{W_{Sc} \cdot 100}{W_{Sf}}$$
$$W_{sf} = (W_{Sc} \cdot 100)/I_{Rc}$$

This is then the wind index for the measurement period. Please note that in this method the entire measured site data set is retained with original frequency and period length.

4.4 Wind Index Correction

When the expected long-term average power for measured site data has been found, W_{sf}, it may be converted to a correction factor that may be used on other calculations, such as the energy yield based on a wind statistic from the local meteorological mast.

$$Wsf = C_{sf} \cdot W_{sc}$$
$$C_{Sf} = \frac{W_{sf}}{W_{sc}} = \frac{100}{I_{Rc}}$$

Where C_{Sf} is the correction factor for the full site data set.

This correction factor needs to be applied to the final result of the energy calculation. If a wind statistic is created in relation to the long term correction calculation the correction factor is automatically embedded in the resulting wind statistic and will be applied whenever the wind statistic is used. The term used for the correction factor in windPRO is the *Regional Correction Factor, RCF*.

4.5 An Example

On a particular site, the following average power outputs have been measured (or calculated via windPRO):

Reference site: W_{Rf}, = 450 kW; W_{Rc} = 350 kW

On-site: W_{Sc} = 200 kW

The wind index for the concurrent period on the reference site is calculated below to 77.8. Thus, this period of consideration holds only 77.8% of the energy as the full long term reference period.

$$I_{Rc} = \frac{350 \cdot 100}{450} = 77.8$$

Now the Wind Index Correction factor can be calculated 1.285, i.e. the expected long term yield is 28.5% higher than the local measurements show:

$$C_{Sf} = \frac{100}{77.8} = 1.285$$

4.6 Validating the Wind Index Correlation

In order to make the crucial validation that the wind index for the concurrent period of the reference station is identical to the index of the concurrent period of the site; then it is necessary to establish whether there is a fair correlation between the two data sets. This can be established by establishing a set of monthly wind index using the same method as described above.

Wind indexes are calculated for each month during the concurrent period comparing the monthly average power output to that of the entire concurrent period. This is done for both reference and site data.

When the monthly wind indexes are plotted against each other like shown in Figure 2 and Figure 3, then the course of the indexes must be similar. If they are divergent it is a sign that the climates at the two locations are different and the assumption of similar index for the same period grows weak. Similarly does the assumption that the reference data are representative for long term conditions at the site. The correlation is measured in either correlation value of the monthly indexes or as a standard error of the difference between the index graphs.

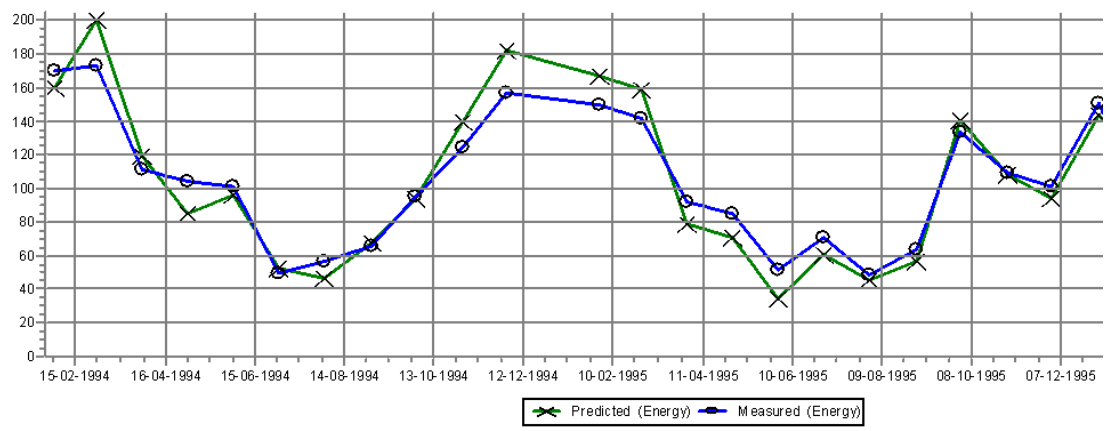


Figure 2: Monthly wind index for reference (green) and concurrent data sets (blue). The correlation is a test of how well the two graphs fit together.

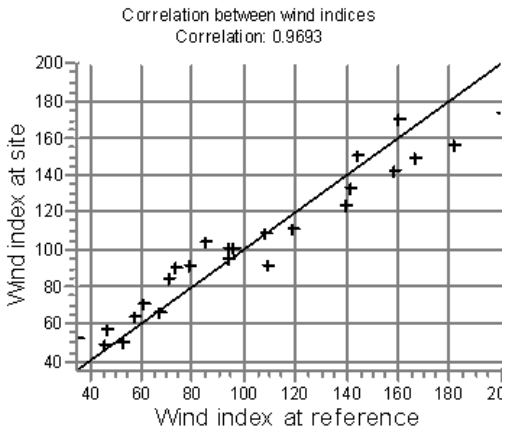


Figure 3: Correlation between wind indexes on the site and the reference station.

4.7 Calculating a LTC Wind Statistic based on a Wind Energy Index

All MCP methods in windPRO have the opportunity to output a wind statistic where the wind speeds are corrected into the proper long term corrected (LTC) reference level. This MCP corrected wind statistic is primarily used for energy yield estimations such as with a windPRO/PARK calculation.

While the calculation of a long term corrected wind statistic is fairly straightforward for MCP-methods based on modelling of the wind speeds; then this is not the case when a wind energy index is involved. Since the Wind Energy Index relates to the power yield from an actual turbine, then this power curve must also be taken into consideration when using the wind index with the wind statistic. The correction method as implemented in windPRO is based on a numerical optimization scheme, where the Weibull form and scale parameters are fitted to match the increase in energy as calculated from the Wind Index MCP method. An outline of the procedure is given in the following pseudo-code:

```
For each pair of A and k in the Wind Statistic do
  1. Calculate original AEP based on original A and k as well as turbine power curve
  2. Calculate LTC-AEP based on original AEP multiplied with Wind Energy Index
  3. Find the LTC Weibull parameters by numerical optimization to match LTC-AEP in (2)
  4. Save the LTC Weibull parameters in new LTC corrected wind statistic
End For
```

5. Weibull Scale MCP

5.1 Introduction

The Weibull Scale method is a very simple empirical method, which does its linear manipulation directly on the Weibull form and scale parameters (*A*,*k*) as well as adjustments on the frequency distribution.

The Weibull method has the advantage, that it will match the nature of the wind at most places, but beware that application of this method should be done with caution on locations with significant non-Weibull distributions.

Also, the scaling which is linear, is a quite simple and radical assumption. A very good directional distribution correlation is needed for the calculation to make sense. The method works best when only small corrections are needed.

Enercon E33 WTGs at Fehmarn, Germany.

5.2 The Weibull Scale Method

The Weibull Scale method presumes that the relationship between the Weibull distribution parameters and the frequency follow the general relation:

$$\lambda_{site}^{long} = \left\lfloor \lambda_{site}^{short} \Big/ \lambda_{reference}^{short} \right\rfloor \cdot \lambda_{reference}^{long}$$

where λ is the distribution parameter under consideration (Weibull *A*, *k*)

In the case of considering the frequencies, the modified long term frequency distribution must be normalized to 100%, i.e. for the *N* sectors under consideration:

$$f_{site,i}^{long} = \left[\left[\frac{f_{site,i}^{short}}{f_{reference,i}^{short}} \right] \cdot f_{reference,i}^{long} \right] \Bigg/ \sum_{i=1}^{N} \left[\left[\frac{f_{site,i}^{short}}{f_{reference,i}^{short}} \right] \cdot f_{reference,i}^{long} \right]$$

where *f* is the frequency
N is the number of sectors (typically 12)
i is the sector under consideration

5.3 Doing the Weibull Scale MCP

Weibull Scale MCP primarily requires look-up in the appropriate Weibull distributions, calculating the correction table and finally doing the calculation of the long term distribution. An example of the application of a Weibull Scale Method is shown in the Figure 1 below.

Reference full: Alstedgårde 3yr

Height: 10.00

Sector		0-N	1-NNE	2-ENE	3-E	4-ESE	5-SSE	6-S	7-SSW	8-WSW	9-W	10-WNW	11-NNW	Mean
A- parameter	[m/s]	4.12	4.57	4.86	5.61	5.86	4.44	4.59	5.54	5.49	5.95	4.56	4.11	5.14
Mean wind speed	[m/s]	3.69	4.06	4.31	4.97	5.19	3.93	4.07	4.90	4.87	5.27	4.05	3.66	4.55
k- parameter		1.655	1.821	2.001	2.016	2.276	1.949	1.919	2.176	2.415	2.428	1.916	1.715	2.013
Frequency	[%]	7.10	3.38	4.82	9.40	8.61	5.12	8.30	13.27	12.57	11.67	9.56	6.21	100.00

Reference Concurrent: Alstedgårde 3yr

Height: 10.00

Sector		0-N	1-NNE	2-ENE	3-E	4-ESE	5-SSE	6-S	7-SSW	8-WSW	9-W	10-WNW	11-NNW	Mean
A- parameter	[m/s]	4.15	5.40	5.55	6.26	6.40	4.32	4.30	4.74	4.84	5.15	4.42	3.74	5.02
Mean wind speed	[m/s]	3.70	4.78	4.91	5.55	5.68	3.83	3.82	4.21	4.28	4.57	3.91	3.33	4.45
k- parameter		1.738	1.985	2.025	2.360	2.603	2.073	1.877	1.837	2.172	2.241	2.109	1.757	1.971
Frequency	[%]	8.75	4.45	6.39	12.13	9.98	4.71	8.21	11.21	9.96	9.29	8.55	6.36	100.00

Local Site Concurrent: Risø mast96-97(1y)

Height: 44.00

Sector		0-N	1-NNE	2-ENE	3-E	4-ESE	5-SSE	6-S	7-SSW	8-WSW	9-W	10-WNW	11-NNW	Mean
A- parameter	[m/s]	6.50	5.37	6.17	6.96	7.91	6.33	5.16	6.11	6.80	7.12	7.17	5.74	6.65
Mean wind speed	[m/s]	5.76	4.77	5.47	6.19	7.05	5.61	4.57	5.42	6.03	6.31	6.35	5.22	5.89
k- parameter		2.109	1.930	2.184	2.769	2.900	2.454	2.015	1.940	2.151	2.087	2.137	1.433	2.125
Frequency	[%]	7.35	5.24	7.60	12.06	11.45	7.51	5.61	7.34	9.62	10.03	9.75	6.44	100.00

Correction Factors

Sector	0-N	1-NNE	2-ENE	3-E	4-ESE	5-SSE	6-S	7-SSW	8-WSW	9-W	10-WNW	11-NNW
A- parameter	0.993	0.848	0.877	0.896	0.916	1.027	1.067	1.167	1.136	1.154	1.032	1.098
k- parameter	0.952	0.917	0.988	0.854	0.874	0.940	1.022	1.185	1.112	1.083	0.909	0.976
Frequency	0.811	0.760	0.754	0.775	0.863	1.086	1.010	1.184	1.262	1.256	1.117	0.977

Local MCP-Corrected Data: Risø mast96-97(1y)

Height: 44.00

Sector		0-N	1-NNE	2-ENE	3-E	4-ESE	5-SSE	6-S	7-SSW	8-WSW	9-W	10-WNW	11-NNW	Mean
A- parameter	[m/s]	6.46	4.55	5.41	6.24	7.25	6.50	5.50	7.13	7.73	8.21	7.40	6.30	6.87
Mean wind speed	[m/s]	5.72	4.05	4.79	5.53	6.44	5.75	4.87	6.32	6.85	7.28	6.56	5.75	6.09
k- parameter		2.008	1.770	2.158	2.366	2.536	2.307	2.060	2.299	2.392	2.261	1.942	1.399	2.035
Frequency	[%]	6.00	4.01	5.77	9.41	9.95	8.21	5.71	8.74	12.23	12.68	10.97	6.33	100.00

Figure 1: Sample Calculation with the Weibull Scale Method.

Note that required input to the Weibull Scale Method is distributions from three different sources:

- long term reference data (named ‘Reference Full’ above)
- the overlapping parts of the reference data (named ‘Reference Concurrent’), and finally
- the site data (named ‘Local Site Concurrent’ in the Figure 1)

The calculation of the concurrent datasets is made directly in windPRO MCP using the data from the correlation table. These data are then fitted to a distribution table and finally the Weibull parameters are fitted from the distribution table.

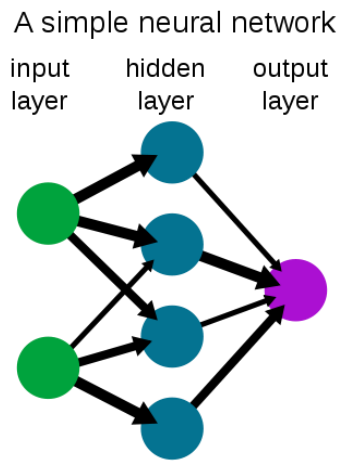
Please note that as this method does not involve the transformation of a time series it is not possible to compare a measured site time series with a predicted concurrent time series.

6. Neural Network Method MCP

6.1 Introduction

The neural network method in windPRO models the changes in wind speed (speed-up) and wind direction (wind veer) through a trained neural network.

The concurrent period of measured wind data is used to train a neural network to detect a pattern between the provided input (reference climate conditions) and the output (measured wind conditions). This is done by applying input to the input layer, and propagate the result through the neural network – following the given output is compared to the expected output (the measured wind condition), and the weights of the neural network is adjusted to provide a more accurate output on the given input. Generally neural networks are universal function approximators, which tries to find a function in a multidimensional space, which maps the provided input to the expected output. But whereas the regression method, and the matrix method uses a fixed set of assumptions of the mapping between the input and the output (linear, polynomial, etc.), the neural network is not limited to by these assumptions.



An example of a simple neural network.

Neural networks are good at interpolating between input to provide an output to input never seen during the training of the input, provided the function between the input and the output of the domain is expected to describe by a (multidimensional) smooth surface. As long as extrapolations are close to the trained input, the neural networks also perform reasonably – but the longer the input travels from the trained data, the worse the neural network is expected to perform.

Thus, a basic assumption of the neural network, as for the other models, 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.

As the neural networks tries to model a *smooth* multidimensional function from the input to the output, neural networks cannot have a built-in residuals model. Therefore, an external residuals model can be applied after the output of the neural network have been determined.

6.2 Neural Network MCP Modelling

The Neural Network MCP in windPRO models the wind speed-up and the wind veer in two different neural networks.

6.2.1 Modelling Wind Speed

A model of the neural network used for modelling the wind speed is shown in **Error! Reference source not found.**, the two sets of 2 input neurons to the right are optional, and only enabled of the user select these.

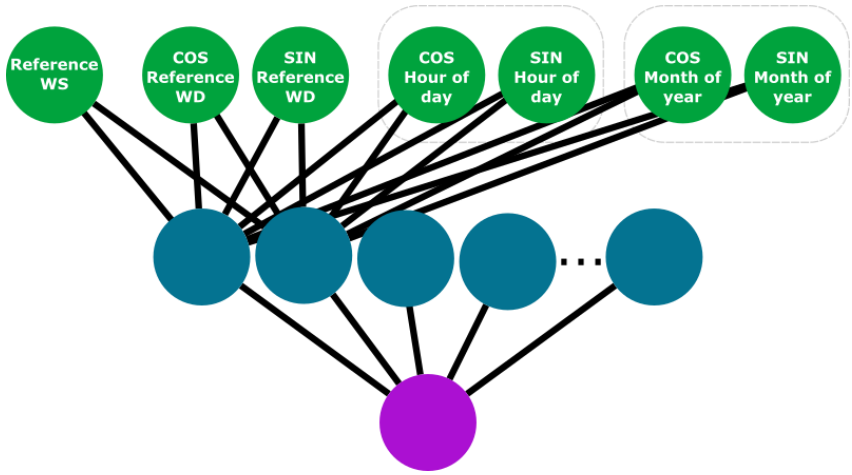


Figure 1: Neural network for modelling wind speed

Reference WS: The reference wind speed, is the input wind speed, scaled to the interval [-1;1], this is done by assuming input wind speeds in the interval [0;40], and scaling this interval to [-1;1].
COS Reference WD: Cosine of the input wind direction
SIN Reference WD: Sinus of the input wind direction
COS Hour of day: Cosine of the diurnal hour of the input (mapped to [0;2 PI]), this input is optional, and should only be used if the user believes that the diurnal hour is important for the modelling.
SIN Hour of day: Sinus of the diurnal hour of the input (mapped to [0;2 PI]), this input is optional, and should only be used if the user believes that the diurnal hour is important for the modelling.
COS Month of year: Cosine of the month of the input (mapped to [0;2 PI]), this input is optional, and should only be used if the user believe that the season is important for the modelling, and only if at least one is used for training data.
SIN Month of year: Sinus of the month of the input (mapped to [0;2 PI]), this input is optional, and should only be used if the user believe that the season is important for the modelling, and only if at least one is used for training data.

MCP always uses only a single hidden layer, and the number of neurons in the middle layer has twice the number of neurons compared to the input layer, that is between 6 and 14 depending on the user settings for input.

The output neuron models the speedup from the input wind speed (non scaled). Thus, to get the output wind speed, multiply the reference wind speed, with the output of the Wind Speed Neural Network, after applying the input wind condition to it.

6.1.2 Modelling Wind Direction

The neural network model for modelling wind direction is very similar to the network for modelling wind speed. A model of the neural network used for modelling the wind direction is shown in **Error! Reference source not found.**, the two sets of 2 input neurons to the right are optional, and only enabled of the user select these.

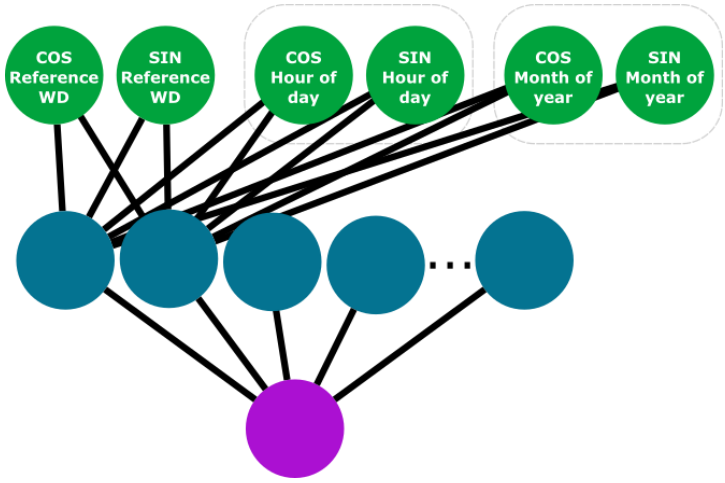


Figure 2: Neural network for modelling wind direction

The description of the input is the same as the description of the input neurons for the network modelling wind speeds.

When modelling wind direction MCP always uses only a single hidden layer, and the number of neurons in the middle layer has twice the number of neurons compared to the input layer, with a minimum of 6 hidden neurons. That is, between 6 and 12 depending on the user settings for input.

The output neuron models the veer of the wind direction. Thus, to get the output wind direction, just add the reference wind direction to the output of the Wind Direction Neural Network, after applying the input wind condition to it.

6.3 Residuals

As neural networks always tries to find a *smooth* multidimensional from the input to the output, neural networks cannot model residuals. But as some form of residuals is needed to capture the real energy content. MCP adds a residual model in the Neural Model by adding residuals after running the input through the neural network.

The Neural Network method in MCP, only offers to add residuals to the wind speed of the modelled data, that is, residuals are not added to the wind direction.
The residuals are added using the “Advanced Gaussian: Mean and std.dev. conditioned on wind speed modelled as polynomials (Of order: 1)”, for a more thorough explanation of residuals, please refer to section 2.2.3.

7. Measure & Correlate: Data Evaluation as part of the MCP-Analysis

7.1 Introduction

Performing a long-term correction is a lot more than applying a technique to two data sets. The choice of reference data and its quality has a large impact on the result as well as the actual MCP method.

This chapter outlines a guide for evaluating of input data, aiming for making a qualified decision procedure for doing MCP source data evaluation.

7.2 Consistency of Data

For both, short-term and long-term data, consistency has to be assured.

For short-term data the general quality criteria for energy yield prediction have to be respected:

- Measurements should be consistent through the measurement period. Aging instruments, drift or other circumstances that may change the measured wind speed should be avoided.
- Keep the local data to an integer number of years in order to avoid seasonal bias.
- Secure a high recovery rate of the data. Typically, 90% data recovery should be aimed for. However, the consistency depends not only on the recovery rate, but also on how the missing data is distributed: if data is missing systematically e.g. due to icing in winter, the result might be biased even when the recovery rate might have reached 90% .
- Data substitution on the local dataset need to respect the temporal variation. Data with different temporal resolution cannot be combined and noise from residual resampling will distort the temporal variation. The wind speed needs to be right at the right time.

As a general principle the longer the local data period, the better. Assuming statistical independence the uncertainty of using local data without long-term correction would be:

$$Wind\ speed\ variability\ uncertainty = \frac{Annual\ variability\ of\ wind\ speed}{\sqrt{Period\ length\ (years)}}$$

However, statistical independence is not always given, since the wind conditions at a specific location might be subject to longer oscillations, like NAO or ENSO.

The consistency of the reference data can be promoted by maintaining the same standards as mentioned above for short-term data.

The consistency of data over time can be tested using a Mann-Kendall (M-K) test. The M-K test is a trend test that is used to test if a sequence of event is trended, using a 95% probability threshold to accept the hypothesis that they are trended. A value of 0 in the M-K test is perfectly trended while a value of 1 is perfectly untrended with a threshold at 0.05.

The data values are evaluated as an ordered time series. Each data value is compared to all subsequent data values. The initial value of the Mann-Kendall statistic, S, is assumed to be 0 (e.g., no trend). If a data value from a later time period is higher than a data value from an earlier time period, S is incremented by 1. On the

other hand, if the data value from a later time period is lower than a data value sampled earlier, S is decremented by 1. The net result of all such increments and decrements yields the final value of S.

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(x_j - x_k)$$

With

$$sgn(x) = \begin{cases} 1 \; if \; x > 0 \\ 0 \; if \; x = 0 \\ -1 \; if x < 0 \end{cases}$$

A very high positive value of S is an indicator of an increasing trend, and a very low negative value indicates a decreasing trend. However, it is necessary to compute the probability associated with S and the sample size, n, to statistically quantify the significance of the trend. This is done in three steps.

Firstly the variance of S is found as

$$Var(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p-1}^{g} t_p(t_p - 1)(2t_p + 5)\right]$$

where g is the number of tied groups (a tied group is a set of sample data having the same value) and t_p is the number of observations in the p'th group. This means that each unique wind speed is grouped and the number of events in each group is counted.

The variance of S is used to calculate the normalized test statistics Z:

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} \; if \; S > 0 \\ 0 \; if \; S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} \; if \; S < 0 \end{cases}$$

If Z>0, it indicates an increasing trend, and vice versa.

Finally, the probability associated with this normalized test statistic is computed. The probability of Z is using a normal distribution with a mean of 0 and a standard deviation of 1:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{\frac{Z^2}{2}}$$

windPRO considers the trend significant for a probability of Z is less than 0.05 (corresponding to a probability level of significance of 95%).
The trend is said to be decreasing if Z is negative and the computed probability is greater than the level of significance. The trend is said to be increasing if the Z is positive and the computed probability is greater than the level of significance. If the computed probability is less than the level of significance, there is no trend.

A low test result may be indicative of a faulty dataset (like a change in source data for mesoscale data), but it may also be caused by long wave oscillations in the wind climate. In those cases, extending the reference dataset may make the dataset less trended.

In case a significant trend is detected, the slope of the trend is given. Since the least squares method for estimating the slope of a regression line is not valid when the data elements do not fit a straight line and is also sensitive to outliers, the more robust, nonparametric estimate of the slope, called Sen's slope is used to find the slope of the trend. The Sen's slope for the set of pairs (i, x_i), where x_i is a time series, is defined as

$$\beta = Median\left(\frac{x_j - x_i}{j - i}\right), j > i$$

β>0 indicates upward trend in a time series. Otherwise the data series presents downward trend during the time period.

While the test result is only significantly trended for values below 0.05, it appears from inspection of the time series that values just above 0.05 is not healthy either and that it is worthwhile aiming for reference data with as high a Mann-Kendall test value as possible, preferably above 0.4.

7.3 Representativeness

windPRO offers several options for comparison of local data with reference data:

- wind speed
- wind energy (production index)
- directional distribution

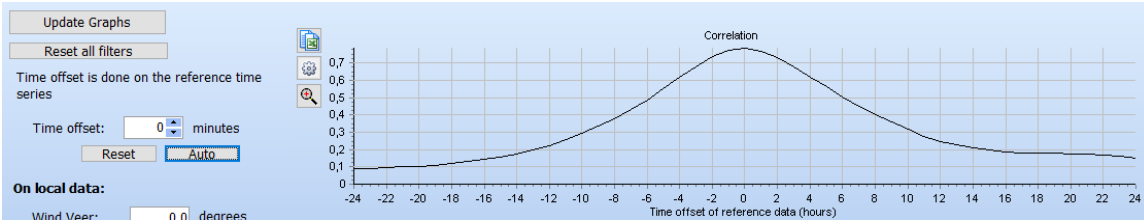
As default windPRO compares onsite and reference data without averaging. A 10-minute local dataset and an hourly reference dataset will therefore be correlated on the hourly value with the same time stamp. The general meteorological standard is that hourly data represents the last 10 minutes of the hour and the two datasets will therefore be compared on equal terms. Arguments have been made that mesoscale data represents a longer averaging period due to the nature of the input into the model, and consequently for the purpose of comparison the local data might have to be averaged similarly. This, however, tends to suppress the natural scatter in the local data, which the residual resampling technique (section 2) seeks to recover and therefore introduce an error. Default in windPRO is therefore to assume that the 10 minute and the hourly data can and should be compared without further averaging.

The correlation between the local and reference data decreases with distance. For large distances the introduction of averaging might be recommended.

7.3.1 Representativeness of Wind Speed

To establish the representativeness of wind speed windPRO correlates the wind speeds of the local and reference data using the Sample Correlation Coefficient (r). r is derived from r^2 as described in section 2 and assesses to what extend there is a relationship between concurrent data in the local and the reference datasets. An r of 1 is a perfect relationship, while 0 is no relationship. It is not possible to give an exact qualifier for how good a correlation coefficient is, as it depends on sample size and sample density. In the context of long-term correction, the correlation coefficient should preferably be better than 0.7

If the time stamps of the local and reference data are not synchronized, this can be corrected for. The "Auto" function on the Adjustment tab in the MCP tool will test in 10-minute increments, which time offset on the reference dataset results in the highest correlation of wind speed and suggest using this.



Automatic offset of reference data time stamp to improve the correlation coefficient

7.3.2 Representativeness of Wind Energy

The purpose of checking the representativeness on energy level is to test if a wind turbine will respond in the same way to the reference data as it would to the local data. In order to make the energy of the local and reference data comparable, a production index is introduced in two steps:

windPRO translates the wind speed into production using a power curve. To achieve comparability the wind speed from the reference time series is scaled to a user-defined, average wind speed, which should correspond to the expected hub height wind speed on site. Two options are available, a simplified or an actual power curve. The user can choose between the two options in the “Session setup” tab.

The default setting is to use a simplified power curve, squared wind speed up to 13 m/s, maintained at this level to 25 m/s, then 0 for wind speeds above. It is unnecessary to convert this to energy due to the normalization.

$$Yield_{time\ step}(u) = \begin{cases} u^2\ if\ 0 < u < 13m/s \\ 13^2\ if\ 13 < u < 25m/s \\ 0\ if\ u > 25m/s \end{cases}$$

Alternatively, a specific power curve can be used to calculate the yield in each time step. It is recommended to use the power curve of the specific project.

In a second step the resulting yield for each time step is normalized to the average production output across the period, creating a production index.

$$I_{production\ index} = \frac{Yield_{time\ step}}{Yield_{average}}$$

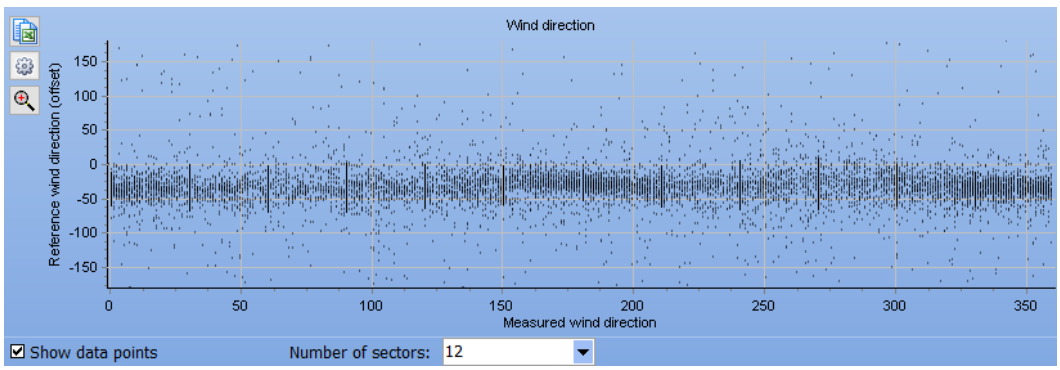
On a short time scale the result may be distorted by the nominal power and the cut-out periods so it is usually more meaningful to compare longer periods (days, weeks or month) to test if seasonal variations are captured correctly. Reducing the number of data point increases the sensitivity to individual faulty data points so the index setup includes a threshold for valid data. If data content in an averaged period is less than the threshold, this period is not included in the correlation calculation and will not be included in the graph. The data are not excluded from the MCP process itself.

The actual correlation between local and reference data follows the procedure described in Chapter 7.3.1. (Representativeness of Wind Speed).

7.3.3 Representativeness of Wind Direction

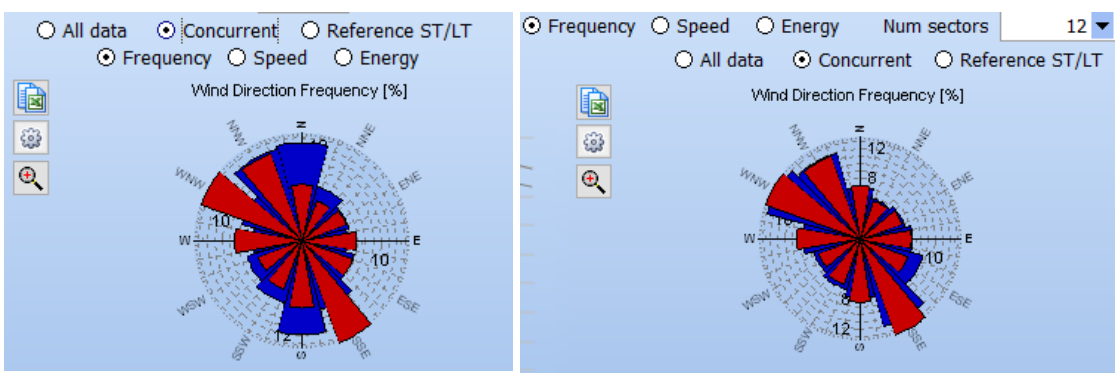
windPRO offers following qualitative methods to check the representativeness of reference directional data with local data. Please note that any temporal offset should be applied beforehand.

1. Visual inspection of time series of concurrent data
2. Visual inspection of systematic wind veer between local and possible correction of data
   The two datasets may have a direction offset. Under "Data Statistics" it can be tested if the wind veer is consistent across sectors. Varying veer from sector to sector indicates local terrain or meteorological effects rotating wind from specific directions. Depending on the chosen MCP concept and method the varying directional veer can be part of the transfer. A high standard deviation of the veer in some sectors may indicate a disconnection of the local and reference wind direction (or just very low wind speeds). For example, sea breezes can be measured locally but might be absent in reference data.
   A consistent wind veer across all sectors might point towards a misalignment of the local wind direction. It can be corrected for on the "Adjustment" tab. Careful attention should be paid to whether local or reference data are offset. On the "Adjustment" tab the veer correction is calculated as the average difference between wind direction on local and reference data.



Wind direction veer plotted for each data point with uncertainty bars.

3. Visual comparison of local versus reference data of frequency, mean wind speed and energy roses
   The direction distribution between local and reference data can be tested through a wind rose on either frequency or energy.



Wind direction distribution plotted before and after wind veer correction.

4. Visual comparison of concurrent (short-term) versus long-term period of frequency, mean wind speed and energy roses of reference data
   The comparison of the short-term and the long-term roses of the reference data tests if the short-term wind direction is long-term representative. If there is a sufficient match, it might not be necessary to apply directional correction to the local measurement, which is a prerequisite for the concept of "Local Scaling".

7.4 Evaluation of Reference Time Series

It is a healthy procedure to compare multiple reference data against the chosen reference data set, either with the objective to select the most appropriate reference data or in order to apply an ensemble result.

The above tests consider consistency of the reference data and representativeness between reference data and local data and this information can be used to discard reference data that fails the tests and to rank those that pass them. Often, no reference scores best in all tests, and among those that score high it is not obvious which is the best choice. It is therefore recommended to only make a final choice of reference data when also the tests are done on the different prediction models. Suitable reference data tends to converge on the resulting wind speed and production. In the end a combination of variation across reference data sets and correction methods may be a better indicator of uncertainty than the correlation between local and reference data.

7.5 Uncertainty of Long-term Correction using Linear Regression

windPRO offers an uncertainty calculation suggested by Klintø[16]. This uncertainty assessment algorithm is assuming that the linear regression method is used to create the long-term data and can therefore not necessarily be extended to other methods.

The uncertainty is calculated as

$$Unc = \sqrt{[A * |1 - WI|]^2 + (B * R^C)^2 + (D * V)^2} * Y^E * CF$$

Using the following constants:
A = 1.55, B = 0.06, C = -1.3, E = -0.3

WI is the wind index for the concurrent period using the full reference period as baseline. This is calculated according to the principle described in 7.3.2.
R is the correlation coefficient between local and reference data without averaging based on 7.3.1.
V is the annual variability of the wind speed for the reference series.
Y is the number of concurrent years between local and reference data
CF is the conversion from wind speed to production output. This sensitivity follows the same procedure as used for the Loss and Uncertainty calculation and is therefore dependent on the power curve used for the wind index (WI).

The uncertainty outcome of the algorithm is the uncertainty on production, not on wind speed.

An alternative approach is to follow 7.4 and consider the scatter of a host of results (ensemble method). This approach holds the advantage of offering uncertainty outside the linear regression method, but as this is largely a manual process care should be taken when considering which outcomes are dependent and which are independent.

[16] Klintø, F. (April 2015). Long-Term Correction - Uncertainty Model using different Long-Term data. London: Wind Power Conference 2015 - Wind Resource Assessment.

8. Measure & Correlate: Evaluating the result of the MCP-Analysis

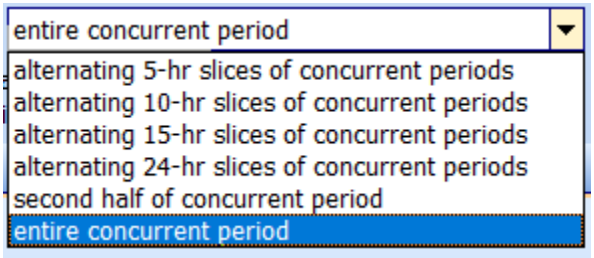
8.1 Introduction

windPRO offers a number of options for evaluation the long-term correction result, both quantitative in the form of statistical parameters and qualitative in the form of graphs and plots. Most of these are delivered through the “Model LT” tab in the MCP module and the background of the checks are described below.

These analyses are only available for MCP methods translating reference data to the site, not for translating local data to become long-term representative.

8.2 Slicing tests

All testing of the transfer of reference data onto the site are based on slicing tests. A slicing test is a test, where the concurrent dataset is divided into two even sized groups. One of the sets is used to find the transfer function (”training”) while the other set is used to test the quality of the transfer function. In the test the transfer function is applied to the reference data. The transferred data are then compared to the observed data. The success is the measured by how close the transferred data resemble the measured data.



The slicing options are listed to the right.
It is a concern that the slicing period should not be biased. This would be the case if the training period was systematically different from the testing period. Examples for biased slicing periods are day versus night or summer versus winter. The various slicing options try to avoid this, though the “Second half of concurrent period” should be treated carefully: If concurrent data covers one full year, there may be a seasonal bias (e.g. summer data used to predict winter data) and in that case this option should be avoided.

The option “Entire concurrent period” is essentially a self-test. The entire concurrent period is used to train the transfer function and it is tested on the entire period.

8.3 Slicing test results

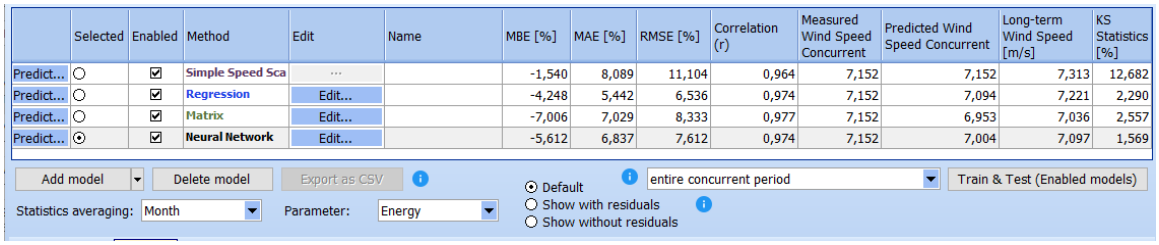
The results of the slicing test are tabulated in the top section of the Model LT tab (see figure). The categories are explained below.

The display can be managed according to averaging time and the parameter displayed

The averaging allows the results to be averaged over a time period before being compared. The error statistics for example will then be calculated based on the average value of the selected period, whether it be hour, day or month etc. If it is important to get a high precision in individual time steps you would aim for a good match with as little averaging as possible (such as for data substitution), while an annual yield assessment takes a broader view where the error has to be minimized on a larger time scale.

The results can be presented based on wind speed or energy. Wind speed is straight forward the wind speeds measured and calculated through the transfer. Energy is not the energy content of the wind speed but the

production output. The production output is established by applying the wind speed to a power curve selected in the Session Setup tab (see section 7.3.2). This also means that the error statistics is directly related to the error on the yield assessment. As opposed to section 7.3.2 there is no indexation of the production.



	Selected	Enabled	Method	Edit	Name	MBE [%]	MAE [%]	RMSE [%]	Correlation (r)	Measured Wind Speed Concurrent	Predicted Wind Speed Concurrent	Long-term Wind Speed [m/s]	KS Statistics [%]
Predict...	○	☑	Simple Speed Sca	...		-1,540	8,089	11,104	0,964	7,152	7,152	7,313	12,682
Predict...	○	☑	Regression	Edit...		-4,248	5,442	6,536	0,974	7,152	7,094	7,221	2,290
Predict...	○	☑	Matrix	Edit...		-7,006	7,029	8,333	0,977	7,152	6,953	7,036	2,557
Predict...	⊙	☑	Neural Network	Edit...		-5,612	6,837	7,612	0,974	7,152	7,004	7,097	1,569

The slicing test display. For each method transferred and observed datasets are compared on a number of criteria.

8.3.1 Residual resampling in slicing tests

Several of the transfer methods may include a random component. This corrects the energy content imbalance inherent in doing the transformation on wind speed (see section 2), the production content being non-linear with wind speed. The random component is distributed evenly over time and may disrupt the time domain fit of the transferred dataset to the measured dataset.

Both Linear Regression and the Neural Network method have an optional Residual resampling function. For the Matrix method the random element is inherent to the method.

Depending on the test criteria, the optional residual resampling may or may not be applied in order to get a meaningful result. In the dashboard, optional residuals can be applied or removed. From version 3.5 EMD has selected the option of using residual resampling as default where it is most meaningful.

Two examples, where resampling influences the test, are the MBE (Mean Bias Error) and the Correlation criteria. The Mean Bias Error tells you if there is an offset in the transferred dataset compared to the measured dataset. For wind speed the residuals are unnecessary, but for the production yield the MBE will only represent the correct production error if residuals are included. The Correlation criteria uses the correlation coefficient of the measured and transferred datasets. For this criteria the temporal aspect is important and a randomly applied residual component will disrupt it. Here, residuals should not be applied.

Below is a list of the parameters and settings with EMD's suggestion for residual setting. From version 3.5 these settings are set by default.

Parameter selection	Metric	Statistics averaging selection	Default residual setting		Parameter selection	Metric	Statistics averaging selection	Default residual setting
Wind	MBE	Hour	Residuals off		Energy	MBE	Hour	Residuals on
Wind	MBE	Daily	Residuals off		Energy	MBE	Daily	Residuals on
Wind	MBE	Week	Residuals off		Energy	MBE	Week	Residuals on
Wind	MBE	Month	Residuals off		Energy	MBE	Month	Residuals on
Wind	MAE	Hour	Residuals off		Energy	MAE	Hour	Residuals off
Wind	MAE	Daily	Residuals off		Energy	MAE	Daily	Residuals off
Wind	MAE	Week	Residuals off		Energy	MAE	Week	Residuals on
Wind	MAE	Month	Residuals off		Energy	MAE	Month	Residuals on
Wind	RMSE	Hour	Residuals off		Energy	RMSE	Hour	Residuals off
Wind	RMSE	Daily	Residuals off		Energy	RMSE	Daily	Residuals off
Wind	RMSE	Week	Residuals off		Energy	RMSE	Week	Residuals on
Wind	RMSE	Month	Residuals off		Energy	RMSE	Month	Residuals on
Wind	Correlation	Hour	Residuals off		Energy	Correlation	Hour	Residuals off
Wind	Correlation	Daily	Residuals off		Energy	Correlation	Daily	Residuals off
Wind	Correlation	Week	Residuals off		Energy	Correlation	Week	Residuals on
Wind	Correlation	Month	Residuals off		Energy	Correlation	Month	Residuals on
Wind	Wind speed	Hour	Residuals off		Energy	Wind speed	Hour	Residuals off
Wind	Wind speed	Daily	Residuals off		Energy	Wind speed	Daily	Residuals off
Wind	Wind speed	Week	Residuals off		Energy	Wind speed	Week	Residuals off
Wind	Wind speed	Month	Residuals off		Energy	Wind speed	Month	Residuals off
Wind	Kolmogorov-Smirnov	Hour	Residuals on		Energy	Kolmogorov-Smirnov	Hour	Residuals on
Wind	Kolmogorov-Smirnov	Daily	Residuals on		Energy	Kolmogorov-Smirnov	Daily	Residuals on
Wind	Kolmogorov-Smirnov	Week	Residuals on		Energy	Kolmogorov-Smirnov	Week	Residuals on
Wind	Kolmogorov-Smirnov	Month	Residuals on		Energy	Kolmogorov-Smirnov	Month	Residuals on

Default setting on residuals for each metric combination.

8.3.2 Mean Bias Error MBE

Mean Bias Error (MBE) is the single most important metric when it comes to evaluating the performance of the transfer function for the purpose of yield assessment.

It is the average of the difference between measured and predicted wind speed or respectively production output across the time steps, whether this being hourly or monthly or a selection in between.

MBE is calculated as

$$MBE = \frac{\frac{1}{N}\sum_{i=1}^{N} \overline{x_{p,i}} - \overline{x_{m,i}}}{\frac{1}{M}\sum_{i}^{M} x_{m,i}}$$

where N is the number of samples after averaging (e.g. using monthly averaging over a year gives 12 samples) in the testing period. x_p is the predicted value (wind speed or production) and x_m is the measured value, using averaged values in the numerator. M is the number of samples without averaging in the entire concurrent period (training and testing period)

MBE gives the offset of the value as a percentage. On energy an MBE of 1% means the production output is overestimated by 1% by the transfer function. The link between wind speed and production is established by assigning a power curve on the Setup tab.

The MBE may not tell how good each sample match or whether the distribution is right, but it does tell if the model is able to predict the correct final wind speed and production.

8.3.3 Mean Absolute Error MAE

The Mean Absolute Error is very similar to the MBE, but consider the size of the average error rather than the sign. Negative or positive errors are combined for their absolute value rather than cancelling each other out.

The formula for MAE is described below.

$$MBE = \frac{\frac{1}{N}\sum_{i=1}^{N}\left|\overline{x_{p,i}} - \overline{x_{m,i}}\right|}{\frac{1}{M}\sum_{i}^{M} x_{m,i}}$$

where N is the number of samples after averaging (e.g. using monthly averaging over a year gives 12 samples) in the testing period, x_p is the predicted value (wind speed or production) and x_m is the measured value, using averaged values in the numerator. M is the number of samples without averaging in the entire concurrent period (training and testing period).

because the MAE describes absolute deviations, it is normal that the MAE is larger than the MBE. It is very rare that the prediction model is so precise that errors in the prediction disappear. Especially with residual models activated, the individual measurement point will have a natural scatter. This should however be reduced dramatically with longer averaging periods.

A relatively low MAE (compared to other methods or references) indicated that the model will normally do a good job at predicting the wind speed across the time series.

8.3.4 Root Mean Square Error RMSE

RMSE is another measure of prediction success. It is similar to MAE but squares the individual errors before summing them and then square root the sum.

$$RMSE = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_{p,i} - x_{m,i}\right)^2}{\frac{1}{M}\sum_{i}^{M} x_{m,i}}}$$

The consequence of doing this is that this becomes a measure of the standard deviation, σ, of the prediction errors and thus a potential input to an uncertainty estimation.

As with MAE, the RMSE decrease dramatically when the averaging period is longer. Monthly standard deviation is much smaller than hourly standard deviation and while it is interesting how good individual measurement are predicted, the monthly RMSE is of higher value in assessing the result. Residual resampling increases the RMSE on short averaging periods, but with longer averaging periods energy residual resampling will usually decrease this standard deviation.. The key figure on RMSE will therefore usually be obtained for the settings: monthly, energy and residuals on.

8.3.5 Correlation

The Correlation column will give the correlation coefficient, r, between the measured and predicted datasets.

Ideally each event in the dataset responds 1 to 1 between measured and predicted data, reflecting an unambiguous relation between the two datasets. This is a prerequisite between reference and measurements in the first place and a successful prediction model should maintain this unambiguous relation.

The formula for the correlation is described in section 2.

A poor correlation will be indicative that there are periods and sectors where the prediction model has failed to transfer the reference data to the site. A typical example is day/night differences. Local data may have a larger difference between day and night wind speed than reference data and often the prediction model will have no temporal sensitivity but use the same transfer function day and night. On average this may be correct, but days and nights are incorrectly predicted, and the correlation is poor (a solution to this may be to use a binning by time).

Correlation should be tested without residuals as this will add noise to the temporal element of the dataset. The residual resampling adds an error component without knowing when it should be added. Only to energy correlation with long averaging period will residual resampling be helpful.

The correlation between predicted and measured data is best evaluated using short averaging period and without residuals unless the purpose is to test the ability to predict monthly wind speed or production.

8.3.6 Measured and Predicted Concurrent Wind Speed

The slicing test has two columns for the concurrent wind speed, but as they are not entirely comparable, they should be handled with care.

Measured Wind Speed Concurrent is the arithmetic mean wind speed of the part of the local measurements which is concurrent with the reference data.

Predicted Wind Speed Concurrent is the arithmetic mean wind speed of the part of the prediction dataset that is concurrent with the local data.

If the slicing test is made for “Entire concurrent period” the Measured Wind Speed and the Predicted Wind Speed will be for the same period. The values will show how for a self-test the model was able to predict the mean wind speed. In this test Simple Scaling will always come out with a perfect match because matching the wind speed is the criteria for this model.

If any other slicing test is made the Predicted Wind Speed period will only cover half the dataset, the test period, and the period will therefore be different from that used for Measured Wind Speed Concurrent. The two wind speeds cannot be compared in this case.

8.3.6 Long-term Wind Speed

The Long-term Wind Speed is the predicted wind speed for the entire transferred reference dataset when the trained model is applied to the full reference dataset, not just the testing period.

If the slicing test is made for “Entire concurrent period” the Long-term Wind Speed will be identical to the long-term wind speed written to a meteo object when the final prediction is made. It therefore serves as a shortcut to the outcome of using various models on the dataset. In that single column the range of predicted long-term wind speed results is presented for an at-a-glance view.

If any other slicing test is made the trained model will be based on half the concurrent period and likely not as strong as based on the entire concurrent period. The Long-term Wind Speed will not be the same as the

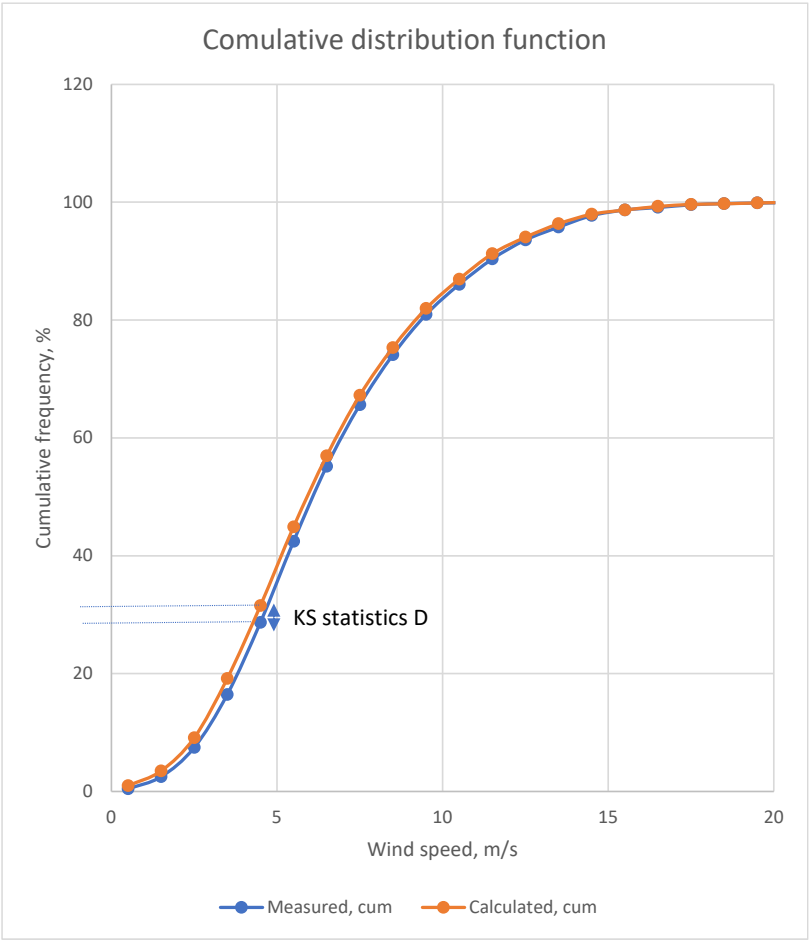
final predicted wind speed and the scatter of the wind speed results may not be representative of the scatter of the final wind speeds.

8.3.7 Kolmogorov-Smirnov (KS) Statistics

The slicing test includes a two-sample Kolmogorov-Smirnov test.

The objective of a two-sample Kolmogorov-Smirnov test is to test if two datasets follow the same distribution. This is done with the KS test statistics.

To illustrate the KS statistics imagine the frequency distribution of wind speed for both measured and predicted dataset, plotted as cumulative distribution functions (CDF). If they belong to the same distribution these two functions would be identical. The KS statistics (D) measures the largest vertical difference between the two datasets.



Kolmogorov-Smirnov statistics is the maximum vertical distance between the two cumulative frequency distributions.

Please note that windPRO does not perform a P test to confirm or reject that the two distributions are indeed identical, but rather, through the KS statistics gives the size of the difference.

The formula for the Kolmogorov-Smirnov statistics is

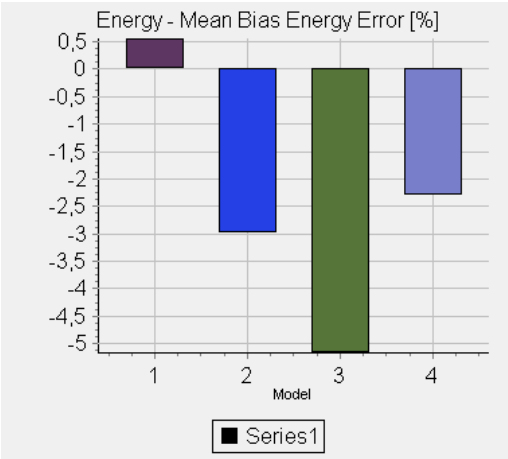
$$D = sup\left|F_{1,n}(x) - F_{2,n}(x)\right|$$

Where F_1 and F_2 are the measured and the predicted cumulative frequency distributions of wind speed (x) and sup is supremum (maximum) difference between the two.

8.3.8 Graphical result of tests

The slicing tests are accompanied with a series of graphical views.

Each test gets a diagram comparing the results of each prediction model. These are arranged so that the optimal result is "0". The best MBE result will approach 0, the best MAE result will approach 0 etc. To conform to this presentation correlation is presented as '1 – correlation'. In this way a superior correlation will approach 0. Numbers and colors refer to the list of models tested



Example of graphical view of the slicing test result. Each color refers to a transfer model. The ideal result is "0".

8.4 Testing Local Scaling result

While windPRO has included an extensive array of tests for transfer functions, there are none for the Local Scaling model. This is partly because the method is inherently different (the reference data are NOT being transferred to the site) and partly because the tests would leave meaningless results.

An MBE test would result is a systemic bias which is exactly what the model is trying to achieve. MAE would approach the same value as MBE as each local measurement point is adjusted the same way, Correlation is perfect and KS statistic result will simply reflect the offset in the adjustment.

The success of the Local Scaling method can only be tested as a comparison to the results using transfer function of reference data. windPRO offers a long-term wind speed which is comparable to the long-term wind speed of the slicing test. It is also recommended to compare production output using the resulting time series with those of the transfer functions. Such a test is not implemented in windPRO 3.4.

8.5 Session statistics

When a session is complete a number of statistics are calculated which can be used to compare with other statistics. This takes place on the Main tab in the MCP module.

There are four parameters on display.

Predicted Mean
This is the predicted long-term mean wind speed as calculated with the selected long-term correction method. It is the arithmetic mean of the time series so it may differ from the mean wind speed if the result is exported as a wind statistic (which will use a Weibull derived mean wind speed).

Concurrent Wind Index
This is the relative production output of a wind turbine in the concurrent period compared to the long-term corrected dataset, essentially comparing the production resulting from the measured, concurrent data with the resulting long-term dataset. It is used to test how large a change has been done to production output and how representative the measured period was in terms of production output.

The wind index is calculated as described in section 7.3.2 using the power curve selected in the session to translate wind speed to production.

AEP Uncertainty
This is the uncertainty calculated in the session as described is section 7.5. The unit is uncertainty in percentage on annual energy production (AEP) and it is important to note that it only relates to the linear regression method. If any other transfer function or Local Scaling was used, the uncertainty value will be invalid. It is included however in order to compare various selections of reference data in the sessions as it is calculated before the actual transformation.

Wind speed correlation
This is the correlation between local and reference data with no temporal averaging (simple match of time stamps). This is done before any transfer functions are applied and the result is therefore method independent. The unit is 1 – r, which means that a value of 0 means r = 1, perfect correlation.

9. References & Literature

Below is a list of selected literature related to the Measure-Correlate-Predict methods and also the windPRO implementation of the various MCP-methods. Additional literature and references are found at the end of the individual chapters which holds the detailed information of the implemented MCP-methods.

Enercon E40 and two Tackle Turbines at Schwarzenborn Knull, Germany

[1] Woods, J.C. & S.J. Watson: *A new matrix method of predicting long-term wind roses with MCP*, Journal of Wind Engineering and Industrial Aerodynamics 66 (1997), p. 85-94.
[2] Joensen, A., L. Landberg & H. Madsen: *A new measure-correlate-predict approach for resource assessment*, Proceedings of the 1999 European Wind Energy Conference, pp. 1157-1160.
[3] Dale Addison, J.F, A. Hunter, J. Bass, M. Rebbeck: *A neural network version of the measure correlate predict algorithm for estimating wind energy yield*,
[4] Gerdes, G. & M.Strack: *Long-term Correlation of Wind Measurement Data*, DEWI Magazin Nr. 15, August, 1999.
[5] Derrick, A.: *Investigations into the Measure-Correlate-Predict Strategy for Site Wind Assessment*, ETSU W/11/00236/REP.
[6] Nason, G.P. & T. Sapatinas: *Wavelet Packet Transfer Function Modelling of Nonstationary Time Series*
[7] Frandsen, S. et al: *Power Performance Assessment*, Risø National Laboratory, Final Report, June 1999, pp. 137.
[8] Riedel,Strack & Waldl: *Robust Approximation of Functional Relationship Between Meteorological data*: Alternative Measure-Correlate-Predict Algorithms, EWEA 2001.
[9] Zaphitopoulos, Dellaportas, Morfiadakis & Glinou: *Prediction of Wind Speed and Direction at a Potential Site*, Wind Engineering no 3, 1999
[10] Addison, Hunter, Bass & Rebbeck: *A neural network version of the measure correlate predict algorithm for estimating wind energy yield*
[11] Ross, S.M: *Introduction to Probability and Statistics for Engineers and Scientists*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, 1987
[12] Efron, B & R.J. Tibshirani: *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability, Chapman & Hall.
[13] Press, W.H; B.P. Flannery; S.A.Teukolsky, W.T.Vetterling, *Numerical Recipes in Pascal*, The Art of Scientific Computing, Cambridge University Press, 1989.
[14] Thøgersen M.L et al: w*indPRO Online: Remote Sensing Data in WindPRO* (WindPRO Reference manual on online services). EMD International A/S, 2010 (for WindPRO users: available from http://help.emd.dk/knowledgebase)
[15] Sørensen, J.D. & I. Enevoldsen: *Lecture Notes in Structural Reliability Theory*, Aalborg University, Spring 1996.
[16] Klintø, F. : *Long-Term Correction - Uncertainty Model using different Long-Term data*. London: Wind Power Conference 2015 - Wind Resource Assessment., April 2015