Long-term correction of wind observations by diffusion-based transformation

Nielsen, Morten

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

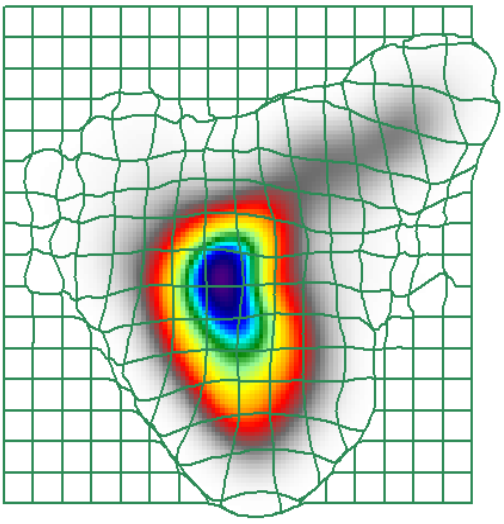
Long-term correction of wind observations by diffusion-based transformation

Department of Wind Energy
E Report 2019

Morten Nielsen

DTU Wind Energy E-0183

April 2019



DTU Wind Energy
Department of Wind Energy

Author: Morten Nielsen

Title: Long-term correction of wind observations by diffusion-based transformation

Department: DTU Wind Energy

Summary (max 2000 characters):

A wind climate based on local measurements is typically limited to a few years and it needs correction to become representative for the long-term wind climate. This is usually done by the measure-correlate-predict (MCP) strategy, which often relies on linear regression of local observations and reference time series. We propose to transform measurements and reference data to Gaussian variables before establishing the statistical correlation. A new transformation for this purpose was inspired by the cartogram method of Gastner and Newman (2004) originally developed for displaying thematic maps in geographical information systems. Transforming wind data to Gaussian variables further allows conditional simulation of time series by Fourier transformation.

The aim of the RECAST project is to reduce the time for wind resource assessments by careful planning and advanced measurement systems. This report investigates the uncertainty of long-term correction based on a relatively short measurement campaign. Three years of data from the ten masts used as reference stations for the South African Wind Atlas and simulations by the WRF model are used for this purpose. Measured time series from very short campaigns sometimes lack observations for certain wind directions, and this leads to imprecise long-term correction. Problems with lack of data for certain wind conditions both seem to depend on season and location, so it is difficult to recommend a standard campaign duration. However, based on periods with successful long-term correction, the root-mean-square error of the wind power density was found to be on the order of 10%, 5% and 2% for 30, 90 and 360 days, respectively. The performance of the new method is comparable to the best MCP methods when the measurement campaign is on 360 days, but it is better to use a robust method, like MCP by orthogonal linear regression, for very short campaigns, say 30 days.

For practical application it is of interest to estimate the uncertainty of long-term corrections based on the complete time series rather than multiple short samples. This is facilitated by the jackknife resampling method, which provides uncertainty of variables like Weibull parameters, mean wind speed or power density, either for all wind directions or for individual wind sectors.

DTU Wind Energy E-0183

April 2019

ISBN

978-87-93549-51-7

Contract no.:

7046-00021B

Project no.:

43356 4630-IN-RECAST

Sponsorship:

Innovationsfonden

Front page:

Two-dimensional wind distribution warping into a Gaussian distribution

Pages: 39

Tables: 3

Figures: 32

References: 12

Technical University of Denmark
Department of Wind Energy
Frederiksborgvej 399
Building 118
4000 Roskilde
Denmark
Telephone

www.vindenergi.dtu.dk

Preface

This work started in the DTU Wind Energy *Mesoscale generalization and data* internal development project of 2017 and continued in the *Reduced Assessment Time (RECAST)* project sponsored by the *Danish Innovation Fund, contract no. 7046-00021B*. The present report is a delivery of work package *WP1: RECAST concept*.

Contents

Summary

A wind climate based on local measurements is typically limited to a few years and it needs correction to become representative for the long-term wind climate. This is usually done by the measure-correlate-predict (MCP) strategy, which often rely on linear regression of local observations and reference time series. We propose to transform measurements and reference data to Gaussian variables before establishing the statistical correlation. A new transformation for this purpose was inspired by the cartogram method of Gastner and Newman (2004) originally developed for displaying thematic maps in geographical information systems. Transforming wind data to Gaussian variables further allows conditional simulation of time series by Fourier transformation.

The aim of the RECAST project is to reduce the time for wind resource assessments by careful planning and advanced measurement systems. This report investigates the uncertainty of long-term correction based on a relatively short measurement campaign. Three years of data from the ten masts used as reference stations for the South African Wind Atlas and simulations by the WRF model are used for this purpose. Measured time series from very short campaigns sometimes lack observations for certain wind directions, and this leads to imprecise long-term correction. Problems with lack of data for certain wind conditions both seem to depend on season and location, so it is difficult to recommend a standard campaign duration. However, based on periods with successful long-term correction, the root-mean-square error of the wind power density was found to be on the order of 10%, 5% and 2% for 30, 90 and 360 days, respectively. The performance of the new method is comparable to the best MCP methods when the measurement campaign is on 360 days, but it is better to use a robust method, like MCP by orthogonal linear regression, for very short campaigns, say 30 days.

For practical application it is of interest to estimate the uncertainty of long-term corrections based on the complete time series rather than multiple short samples. This is facilitated by the jackknife resampling method, which provides uncertainty of variables like Weibull parameters, mean wind speed or power density, either for all wind directions or for individual wind sectors.

1 Introduction

Wind resource assessment is usually based on observations from a meteorological mast installed at a potential wind-farm site. For practical reasons, the measurements campaign is typically limited to a few years, and the wind climate in this period may deviate from the long-term wind climate. The long-term wind climate could be derived from observations at a not too distant meteorological mast or by a numerical weather model, although it would not be fully representable for the local wind conditions.

MCP methods

The aim of measure-correlate-predict (MCP) methods is to combine local and long-term information. The general idea is to model the relation between local observation and reference data statistically using data from a period of concurrent data. This statistical model is then used for prediction of local wind conditions by reference data from a longer more representable period.

wind-climate presentations

Figure 1 compares different presentations of the 2D probability density function of the variable wind vector. The first sub-figure shows the raw distribution of 10 min average winds sampled in a bit map with a resolution on the order of 0.1 m/s. Distributions like this will be noisy when the observation period is of limited duration. The next sub-figure shows the probability distribution when the data are sampled in twelve wind sectors and wind-speed bins with 1 m/s discretization, a method which is used in wind resource assessment. This distribution is smoother than the raw distribution but it has discontinuities between sectors and wind-speed bins. The third method shows sector-wise Weibull distributions, which is the parameterization used in WAsP software. This distribution has a smooth dependency on wind speed, but it is still discontinuous between wind sectors. The Weibull distributions are either fitted by raw data or bin-statistics. The fourth distribution is smooth everywhere. It is calculated by convolution of the raw distribution in the first subfigure and a 2D Gaussian kernel with a standard deviation of 1 m/s.

data for this study

The wind time series in this study were observed at ten meteorological stations with similar instrumentation during phase one of the Wind Atlas for South Africa (Mortensen, Hansen, Kelly, Prinsloo, Mabille & Szewczuk 2014). The data set also contains reference time series extracted from wind fields modelled by the weather research and forecasting model WRF (Hahmann, Lennard, Badger, Vincent, Kelly, Volker, Argent & Refslund 2015, Hahmann, Vincent, Pea, Lange & Hasager 2015). The sample period is three years although it would actually be possible to extend the period with continued measurement as these masts are still in operation. Figures 2 and 3 show the observed and modelled wind climates for a three year period. The wind climates are quite different and fairly well predicted by the model.

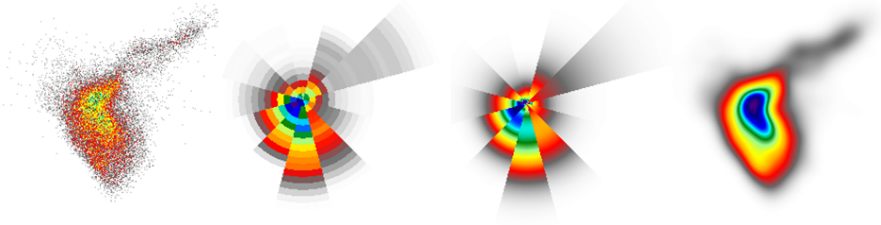


Figure 1: Different presentations of the wind-vector probability distribution - from left to right: raw distribution, bin sampling, sector-wise Weibull distributions and smooth distribution.

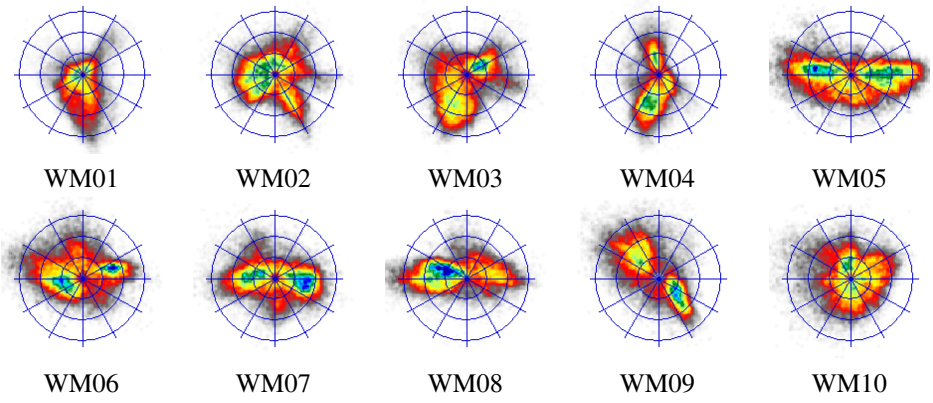


Figure 2: Observed wind distributions

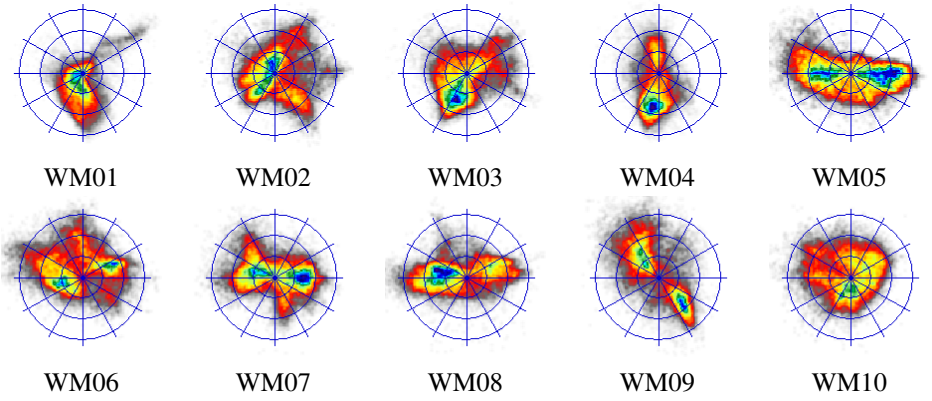


Figure 3: Wind distributions predicted by WRF

2 A new long-term correction method

inspiration

A cartogram is a graphical representation of statistical data with geographical variation. The variation is illustrated by a transformation, which inflates or deflates areas until the density of the data becomes uniform. Gastner & Newman (2004) developed a diffusion-based cartogram, where the initial condition is a probability density map with geographical areas defined by coastlines and political borders, e.g. national states, each assigned the thematic data of interest. This initial density map is diffused by convolution with a Gaussian kernel corresponding to an imaginary diffusion process. The resulting density map becomes increasingly uniform during the diffusion process, and movements of the probability mass pushes area boundaries into new shapes. Figure 4 shows a cartogram of water resources for all countries of the world calculated by this method.

diffusion-based transformation

Our aim is to correlate reference-data wind vectors with wind vectors observed at the potential wind energy site, but the first step is to transform the wind data to variables with Gaussian distribution. The transformation is similar to that of the diffusion cartogram method of Gastner & Newman (2004), but there are subtle differences. The original method assigned an average value to areas with no data, like the sea or unpopulated land, in order to preserve the total area of land with known statistics. We skip this map pre-processing and let the probability distributions of reference data and observations diffuse to Gaussian distributions. The final state of the density map for the cartogram was a uniform distribution, whereas we stop when the processed probability map is sufficiently close to a 2D Gaussian distribution. The diffusion cartogram tracked the boundaries of political areas, but we track the translation of nodes in a reference grid, either for for-

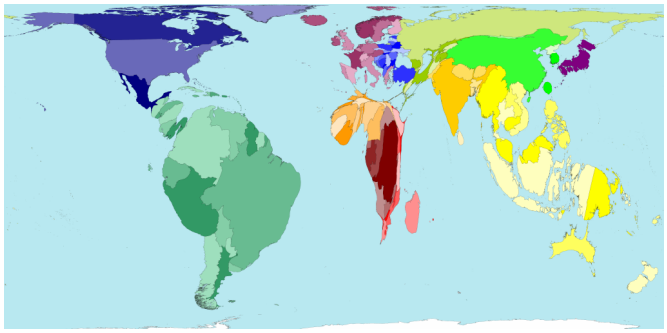


Figure 4: Diffusion cartogram of the water reserve per country using the method of Gastner and Newman (2004). © www.worldmapper.org

ward transformation from original to Gaussian distributions or backward from Gaussian variables to wind vectors. Translations for arbitrary points in the map are found by interpolation between these grid nodes.

velocity field

The wind vectors are converted from speed and direction to orthogonal velocity components and the initial wind-vector probability distribution is sampled in a grid map. The evolution of the density map is modelled by a diffusion equation

$$\frac{\partial \rho(\mathbf{r},t)}{\partial t} = K\nabla^2\rho(\mathbf{r},t) \tag{1}$$

where ρ is the probability density depending on a 2D position vector $\mathbf{r}$ and time t. The map for a given time is calculated by convolution of the original density map and a Gaussian kernel with a width determined by $\sigma(t) = \sqrt{2Kt}$. Following Gastner & Newman (2004), we set the diffusion coefficient equal to one $K = 1$ since the time scale of the diffusion process is of no significance for the solution. The velocity field at a given time during the diffusion process is found by gradients of the diffused density map.

$$v(\mathbf{r},t) = -\frac{\nabla\rho(\mathbf{r},t)}{\rho(\mathbf{r},t)} \tag{2}$$

efficiency by FFT

The 2D Fourier transformation of the density map and the inverse Fourier transformation are defined by

$$\begin{aligned} \widetilde{\rho}(k,l) &= \iint \rho(x,y)\exp\left[-\imath 2\pi(kx+ly)\right]\,dxdy \\ \rho(x,y) &= \iint \widetilde{\rho}(k,l)\exp\left[\,\imath 2\pi(kx+ly)\right]\,dkdl \end{aligned} \tag{3}$$

where x, y are velocity components and k, l are corresponding wave numbers in the Fourier decomposition. We also use the short-hand notation

$$\rho(x,y) \Leftrightarrow \widetilde{\rho}(k,l) \tag{4}$$

An efficient way to calculate the derivatives of the map is to multiply the Fourier components by the wave-number vector and then make an inverse transformation to physical space.

$$\begin{aligned} \imath 2\pi k\,\widetilde{\rho}(k,l) &\Leftrightarrow \rho'_x(x,y) \\ \imath 2\pi l\,\widetilde{\rho}(k,l) &\Leftrightarrow \rho'_y(x,y) \end{aligned} \tag{5}$$

The initial density map $\rho(x,y,0)$ is found by bin sampling of the empirical distribution of wind vectors, and the map at arbitrary time $\rho(x,y,t)$ is calculated by convolution of the initial map and a 2D Gaussian kernel with a spreading defined by the diffusion time.

$$\rho(x,y,t) = \rho(x,y,0) * g(x,y,t) \Leftrightarrow \widetilde{\rho}(k,l,0)\widetilde{g}(k,l,t) \tag{6}$$

The density gradient maps are calculated in a similar way. The Gaussian kernel at a given stage in the diffusion process is defined by

$$g(x,y,t) = \frac{1}{2\pi\sigma^2(t)} \exp\left[-\frac{x^2+y^2}{2\sigma^2(t)}\right] \quad \text{with } \sigma(t) = \sqrt{2t} \tag{7}$$

We do not need to calculate the Fourier transform of the Gaussian kernel numerically, as this is known theoretically.

$$\widetilde{g}(k,l,t) = \exp\left[-2\pi\sigma^2(t)\left(k^2+l^2\right)\right] \tag{8}$$

The calculations are actually done by discrete Fourier transformations using the FFTW library (Frigo & Johnson 2005). The size of the density map includes a hidden buffer zone to avoid wrap-around effects in the convolution calculus associated with cyclic boundary conditions. In order to avoid redefining the FFTW plans for the Fourier transformations, the size of the density map remains the same for every time step of the integration. The map scale is, however, optimized for each step in the integration to ensure that the diffused probability map is covered efficiently. Each Fourier transformation handles two real-valued maps simultaneously, using the data-packing method described by Press, Flannery, Teukolsky & Vetterling (1992).

grid deformation

The nodes of the grid describing the forward transformation are calculated by a fourth order Runge-Kutta (RK) integration scheme with adaptive time steps (Press et al. 1992). Density and gradients at grid node positions are found by interpolation in maps calculated by FFT techniques for the diffusion times required by the RK algorithm. Grid node movements for the backward transformation are also calculated by fourth order RK integration. This integration applies the same time steps as the forward transformation using the pre-calculated density and velocity maps for the forward transformation, so we avoid additional FFT transformations.

forward and backward transformation

Figure 5 shows the forward transformation of the wind distribution of the reference data to a Gaussian distribution. Figure 6 shows the backward transformation from a Gaussian distribution (rightmost figure) to the wind distribution at the potential wind-energy site. We will refer to the forward transformation from wind vector (ξ_1, ξ_2) to corresponding Gaussian vector (x_1, x_2) by

$$(x_1, x_2) = \mathcal{F}(\xi_1, \xi_2)$$

and the backward transformation by

$$(\xi_1, \xi_2) = \mathcal{F}^{-1}(x_1, x_2)$$

Both functions are defined by interpolation in the transformed grids. According to equation 2, it is impossible to calculate velocity vectors for zero density and inaccurate for near-zero density. We consequently ignore grid-node movements when the density is below a small threshold. The density of a grid node may be zero at first and later get a positive value during the forward diffusion process. As in Figure 5, we often see a front near the outer area of the distribution during the forward diffusion and the solution beyond this front is unreliable. During the backward diffusion process it seems like the density will stay at zero once it has this value, so we remove grid nodes once this condition is met.

The integration is stopped when the distribution is nearly Gaussian, i.e. at $t_2 = 0.5\sigma_2^2$ with σ_2 much larger than the spread of the raw distribution. The start of the integration is set to $t_1 = 0.5\sigma_1^2$ where a non-zero value of σ_1 will avoid too fine details in the backward transformation. Figures 5 and 6 were calculated with $\sigma_1 = 0.1$ m/s and $\sigma_2 = 50$ m/s.

summary of method

Figure 7 illustrates the long-term correcting method using a diffusion-based transformation derived from concurrent data in a short period. The top-centre frame shows a transformation of the pdf of the short-term reference distribution to a 2D Gaussian distribution and the associated grid deformation. The transformation of the pdf at the site of observation is shown in the frame below. Here we both see grid deformations for forward and backward transformation.

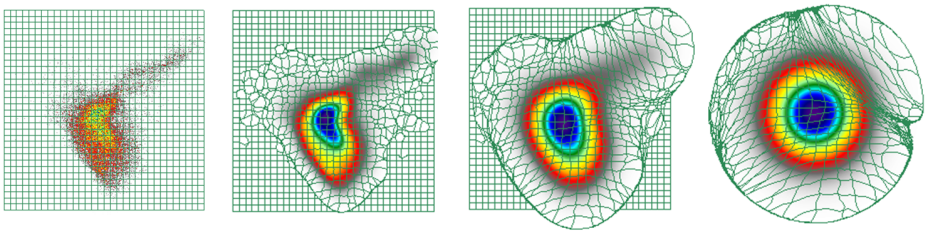


Figure 5: Development of density map and grid during the forward diffusion process.

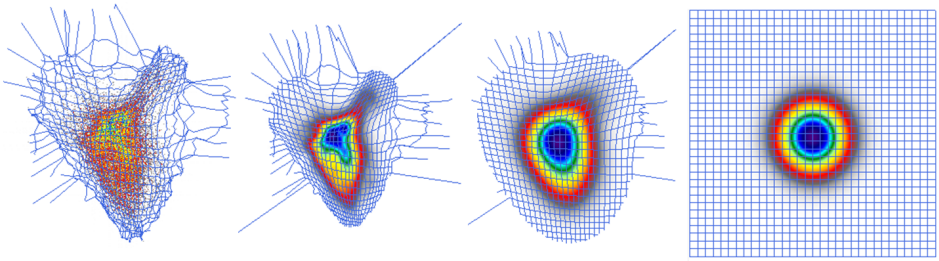


Figure 6: Development of density map and grid during the backward diffusion process.

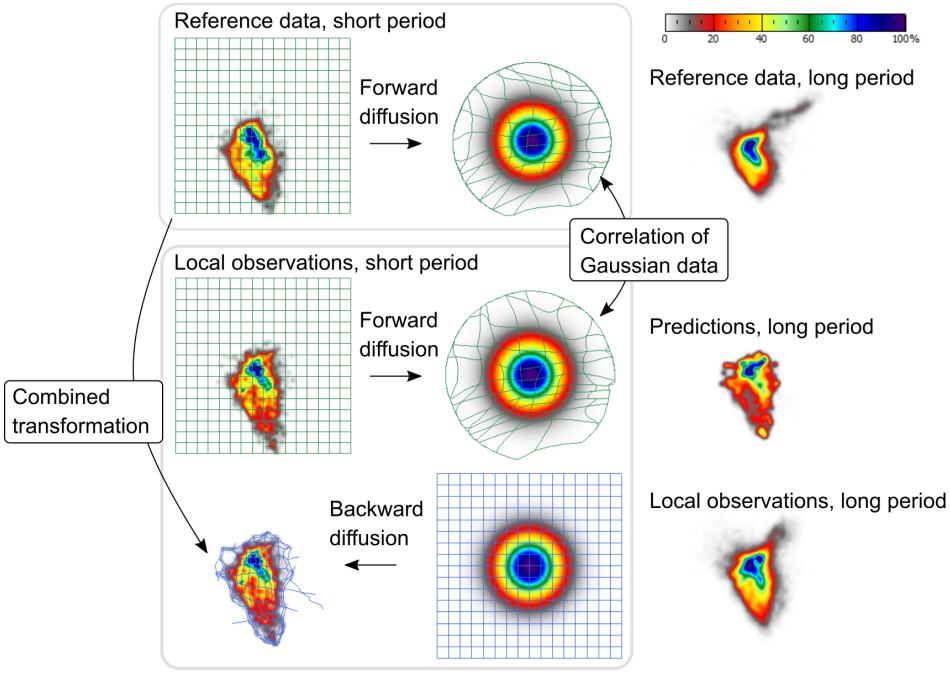


Figure 7: Illustration of the long-term correction

The combined transformation of the distributions for reference data to observations via intermediate Gaussian values is shown at the left. We note that transformations are small in the central area, i.e. for modest wind speeds, but often large in areas with low probability density and different for the two distributions. Furthermore, the transformation is only defined in areas with data. This sometimes become a problem for short periods of concurrent data.

The plots on the right-hand side show pdfs for a longer period. The plot at the top shows the long-term distribution of reference data and the plot at the bottom shows the long-term distribution of local observations. The plot in the middle is a prediction of the pdf at the local site based on the transformation derived from the period of concurrent data and the long-term reference pdf above. The ideal is a predicted pdf which is close to the long-term distribution of local data. The shortcomings in the present example are due

to the short duration of concurrent data, chosen to be three months only.

transformation vs simulation

We may transform the probability distributions directly, but when time series are needed it will be more realistic to make a stochastic simulation, which takes the lack of correlation between the two data sets into account. This is best done for the transformed Gaussian time series.

3 Stochastic simulation

Gaussian variables are easier to model statistically than the original wind data. These variables are described by a multivariate normal distribution with the probability density function

$$f\left(x_1, x_2, \ldots, x_n\right) = \frac{1}{\sqrt{(2\pi)^n \left|\mathbf{\Sigma}\right|}} \exp\left(-0.5(\mathbf{x}-\mu)^{\mathrm{T}} \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right) \tag{9}$$

Here, $\mathbf{x}$ is a vector with n stochastic variables x_i, μ is the mean vector and Σ is the covariance matrix with determinant $|\Sigma|$. The mean vector and covariance matrix are estimated by data from a period of concurrent observations at the met mast and reference site or model. When the model is calibrated we can derive a conditional distribution for Gaussian parameters for the met-mast site and use this for prediction of long-term wind conditions.

3.1 Time-independent simulation

Let $\boldsymbol{x}$ be a vector with Gaussian components translated from the velocity components by the diffusion-based mapping function. There will be two components for each data set, e.g. four components for the basic case of one mast and one reference time series. We define the mean vector $\boldsymbol{\mu} = \langle \boldsymbol{x} \rangle$ and covariance matrix $\boldsymbol{\Sigma} = \langle (\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T \rangle$. A covariance matrix is by definition non-negative definite allowing for the Cholesky 'square-root' decomposition $\boldsymbol{LL^T} = \boldsymbol{\Sigma}$. However, to improve numerical stability, we apply the closely related $\boldsymbol{LDL^T} = \boldsymbol{\Sigma}$ decomposition, where $\boldsymbol{D}$ is a diagonal matrix, $\boldsymbol{L}$ is a lower-left matrix with unit-valued elements at the diagonal, and $\boldsymbol{L^T}$ is the transposed matrix. The $\boldsymbol{LDL^T}$ decomposition allow us to simulate Gaussian variables with realistic mean and covariance by

simulation

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{LD^{1/2}\varepsilon} \tag{10}$$

where ε is a vector with independent random elements of standardized normal distribution. Now, let's arrange the $\boldsymbol{x}$ vector with known components $\boldsymbol{x_1}$ at the start and unknown components $\boldsymbol{x_2}$ at the end. We then write the system as

$$\begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{L_{11}} & \boldsymbol{0} \\ \boldsymbol{L_{21}} & \boldsymbol{L_{22}} \end{bmatrix} \begin{bmatrix} \boldsymbol{D_1^{1/2}} \\ \boldsymbol{D_2^{1/2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon_1} \\ \boldsymbol{\varepsilon_2} \end{bmatrix} \tag{11}$$

where the sub-matrix $\boldsymbol{L_{11}}$ contains covariances of known components, $\boldsymbol{L_{22}}$ contains covariances of unknown components, and $\boldsymbol{L_{21}}$ contains cross-covariances between known and unknown components. We now eliminate the 'random input' $\boldsymbol{\varepsilon_1}$ for the known components $\boldsymbol{x_1}$ and deduce a conditional simulation of the unknown components.

conditional simulation

$$\boldsymbol{x_2} = \overbrace{\boldsymbol{\mu_2} + \boldsymbol{L_{21}L_{11}^{-1}}\left(\boldsymbol{x_1} - \boldsymbol{\mu_1}\right)}^{\text{conditional mean}} + \overbrace{\boldsymbol{L_{22}D_2^{1/2}\varepsilon_2}}^{\text{random component}} \tag{12}$$

The factor $\boldsymbol{L_{21}L_{11}^{-1}}$ in the conditional mean term will converge to an identity matrix and the residual covariance $\boldsymbol{L_{22}D_2L_{22}^T}$ for the unknown components will vanish when $\boldsymbol{x_2}$ is perfectly correlated with $\boldsymbol{x_1}$. Conversely, the bias term will disappear and the residual

covariance will remain equal to the original covariance of the unknown components $\mathbf{\Sigma_{22}}$ when $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are perfectly uncorrelated.

In general we will have two 'known' variables for the reference time series and two 'unknown' variables for the site of prediction, i.e. a $\mathbf{4} \times \mathbf{4}$ covariance matrix $\mathbf{\Sigma}$. It will be possible to extend the method for multiple reference time series and multiple sites of prediction and include their correlation.

3.2 Time series simulation

discrete Fourier transformation

Consider a Gaussian process $x_i(t)$ and its Fourier transformation $X_i(f)$. Discrete sampling results in the times series $x_i[k] = x_i(k\Delta t)$ with M steps of fixed length Δt, and the discrete Fourier transformation of this becomes $X_i[m] = \Delta t^{-1} X_i(m/(M\Delta t))$. The time series and its Fourier transform relate to each other by

$$\begin{aligned} X_i[m] &= \sum_{k=0}^{M-1} x_i[k] \exp(\mathrm{i}\, 2\pi km/M) \\ x_i[k] &= \frac{1}{M} \sum_{m=0}^{M-1} X_i[m] \exp(-\mathrm{i}\, 2\pi km/M) \end{aligned} \tag{13}$$

Here, we use a wrap-around index style and place spectral values for negative frequencies in the upper half of the array, as allowed because of the cyclic variation of the exponential function with imaginary argument (Press et al. 1992). For real time series, Fourier values for positive and negative frequencies become complex conjugates, i.e. $X_i[m] = X_i^*[M-m]$ for $m > M/2$.

Fourier simulation

The general idea of Fourier simulation is to generate Fourier components with proper statistics and subsequently convert them into time series by inverse FFT. This procedure is actually a linear combination of stochastic variables, so the result will be Gaussian if the input is. For each frequency, a set of N correlated variables is constructed from uncorrelated input ε_k by

$$X_i[m] = \begin{cases} \mu_i, & m = 0 \\ \sum_{k=1}^{N} l_{ik}[m]\, d_k^{1/2}[m]\, \varepsilon_k[m], & m = 1 \ldots M/2 \end{cases} \tag{14}$$

This is similar to the model for the time-independent case except that the lower-left $\boldsymbol{L}$ matrix now is complex and all components depend on a frequency index m. The cross-spectral correlation between two time series $x_i[k]$ and $x_j[k]$ becomes

$$\chi_{ij}[m] = \left\langle X_i[m]\, X_j^*[m] \right\rangle = \frac{1}{2} \sum_{k=1}^{N} l_{ik}[m]\, d_k[m]\, l_{jk}^*[m] \tag{15}$$

A time-lagged correlation function has the symmetry $\rho_{ij}(\tau) = \rho_{ji}(-\tau)$, which implies that the cross spectral matrix is Hermitian symmetric $\chi_{ij}[m] = \chi_{ji}^*[m]$. We use a variant of the LDLT decomposition applicable for this symmetry.

$$\begin{aligned} d_j[m] &= \chi_{jj}[m] - \sum_{k=1}^{j-1} l_{jk}[m]\, l_{jk}^*[m]\, d_k[m] \\ l_{ij}[m] &= \begin{cases} 1 \text{ for } i = j \\ d_j^{-1}[m] \left(\chi_{ij}[m] - \sum_{k=1}^{j-1} l_{ik}[m]\, l_{jk}^*[m]\, d_k[m] \right) \text{ for } i > j \end{cases} \end{aligned}$$

The result is a lower-left triangular matrix, so we can replace the upper limit of the sum in equation 14 by index i.

conditional Fourier simulation

Similar to the time-independent case we construct a conditional simulation of Fourier

components for the unknown processes by

$$\boldsymbol{X_2}[m] = \overbrace{\boldsymbol{L_{21}}[m]\boldsymbol{L_{11}^{-1}}[m]\boldsymbol{X_1}[m]}^{\text{conditional mean}} + \overbrace{\boldsymbol{L_{22}}[m]\boldsymbol{D_2^{1/2}}[m]\boldsymbol{\varepsilon_2}[m]}^{\text{random component}} \tag{16}$$

where, again, indices 1 and 2 refer to vectors of known and unknown variables. The main difference compared to equation 12 is that the equation applies to all frequencies $\boldsymbol{m}$. Independent random variables $\boldsymbol{\varepsilon_2}[\boldsymbol{m}]$ are generated by the Box-Müller algorithm (Press et al. 1992), and then time series $\boldsymbol{x_2}[\boldsymbol{k}]$ with the prescribed cross-correlations are found by inverse FFT. If we ignore the random contributions in equation 16 we get the time-dependent conditional mean of an ensemble of simulations. The residual covariance matrix for the conditional simulation is a sum of the covariances for all frequencies

residual covariance

$$\boldsymbol{\Sigma_{x_2}} = \frac{\mathbf{1}}{\boldsymbol{M}} \sum_{\boldsymbol{m=1}}^{\boldsymbol{M-1}} \boldsymbol{L_{22}}[\boldsymbol{m}]\boldsymbol{D_2}[\boldsymbol{m}]\boldsymbol{L_{22}^*}[\boldsymbol{m}] \tag{17}$$

This residual covariance matrix is time independent and uniform for the entire Gaussian domain. However, due to divergence in the backward transformation $\mathcal{F}^{-1}$ and transformation from cartesian to polar coordinates, we cannot assume uniform residual covariance in the domain of wind conditions, see section 3.4.

3.3 Time series analysis

We need cross spectra of the transformed Gaussian variable as input for stochastic simulation of time series. This is obtained by discrete Fourier transformation of each time series x_i followed by multiplication of all Fourier components of one signal X_i and complex conjugate of Fourier components of another signal X_j^*.

$$\chi_{ij}\left[m\right] = X_i\left[m\right] X_j^*\left[m\right] \tag{18}$$

The result is real for the power spectrum $i = j$ but complex for cross-spectra between distinct signals $i \neq j$. False correlations due to wrap-around effects are avoided by zero padding (Press et al. 1992).

We have several options for improving the spectral analysis; we may choose to

options

- smooth spectral results by a band-width averaging filter, e.g. applying a $\pm 10\%$ frequency range,
- divide signals into a number of shorter periods, analyse each period and average spectral results,
- remove predictable cycles before spectral analysis,
- remove linear trends or high-pass filter the data,
- minimize end effects by a Hamming window filter.

The first two procedures will smooth the spectra and make them more representable for the long-term climate. Splitting the time series into blocks is also convenient if we want to avoid episodes of missing or corrupted data. Abrupt ends of the signal in block time series may introduce noise, but that disturbance may be mitigated by multiplication of a tapering window such as a the Hamming window.

Figure 8 shows the power spectrum and autocorrelation function of one of the transformed Gaussian series with and without diurnal variation. The uncorrected spectrum have strong peaks for frequencies corresponding to 24 and 12 hours and this is also seen in the autocorrelation function. These variations are removed in time domain by calculating the average daily variation and subtracting it from the signal before spectral analysis.

diurnal variation

It is often a good idea to remove linear trends in signals, especially when operating with small blocks. The typical reason for trends is low-frequency variations which are

signal trend

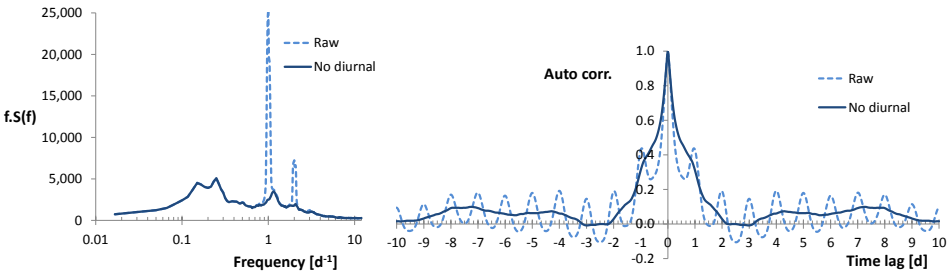


Figure 8: Power spectrum and autocorrelation function of a transformed Gaussian series with and without diurnal variation.

impossible to resolve with a data window of limited duration. The correlation function of a signal with a trend would, however, not approach zero for large time lags if the signal had a linear trend. Tests with the present data indicate that the variance associated with trends becomes sufficiently small for block lengths of 2-4 weeks. As an alternative to linear trend correction we can apply a high-pass filter, and to avoid phase-shift problems we apply a forward-backward first-order autoregressive filter.

expanded frequency range

The frequencies of the discrete spectra are

$$f = \frac{m}{M\Delta t} \quad \text{for } m = 0 \ldots M/2 \tag{19}$$

where $T = M\Delta t$ is the time series duration and $f_{\text{Nyquist}} = (2\Delta t)^{-1}$ is the highest frequency resolved. The duration of time series needed for conditional simulation $T_2 = M_2\Delta t$ is generally longer than the duration of the analysed time series $T_1 = M_1\Delta t$, and the indices in the two spectra giving the same frequency relate by

$$m_1^* = m_2 \cdot M_1/M_2 \tag{20}$$

This is generally a non-integer value so we must interpolate in the analysed spectra. The lowest frequency needed for simulation usually drops below the lowest analysed frequency $(M_1\Delta t)^{-1} < (M_2\Delta t)^{-1}$ and it is not obvious how to extrapolate beyond the resolved frequencies. However, we chose to extrapolate to zero spectral energy for zero frequency, hoping that a high correlation of the low-frequency variations will reduce low-frequency residual variance in the conditional simulation and not affect the simulations significantly.

3.4 Uncertainty band of simulated wind time series

wind vector

The diffusion-based transformation was formulated in a cartesian coordinate system, so we transform wind speed u_n and direction θ_n to eastward ξ_{2n-1} and northward ξ_{2n} components of the wind vector.

$$\begin{aligned} \xi_{2n-1} &= u_n \sin\theta_n \\ \xi_{2n} &= u_n \cos\theta_n \end{aligned} \tag{21}$$

Here, the index n indicate time series pairs from reference data and observations. The forward diffusion process in Figure 5 defines a transformation from the wind vector ξ to the Gaussian variables x

$$(x_{2n-1}, x_{2n}) = \mathcal{F}_n\left(\xi_{2n-1}, \xi_{2n}\right) \tag{22}$$

and the backward diffusion process in Figure 6 defines the reverse transformation from Gaussian variables to wind speed components.

$$(\xi_{2n-1}, \xi_{2n}) = \mathcal{F}_n^{-1}(x_{2n-1}, x_{2n}) \tag{23}$$

covariance of wind components

The uncertainty of the Gaussian variables x_{2n-1} and x_{2n} has a two-dimensional normal

distribution Φ_{2D} with the covariance matrix from equation 17. According to equation 23, the velocity components are functions of the Gaussian variables, so we may calculate moments of wind component errors numerically by

$$m_{pq} = \langle \xi_{2n-1}^p \xi_{2n}^q \rangle = \iint \xi_{2n-1}^p \xi_{2n}^q \Phi_{2D}\left(x_{2n-1}, x_{2n}\right) dx_{2n-1} dx_{2n} \tag{24}$$

and find the uncertainties

$$\begin{aligned} \sigma_{\xi_{2n}}^2 &= m_{20}/m_{00} - m_{10}^2/m_{00}^2 \\ \sigma_{\xi_{2n-1}}^2 &= m_{02}/m_{00} - m_{01}^2/m_{00}^2 \\ \mathrm{cov}\left(\xi_{2n-1}, \xi_{2n}\right) &= m_{11}/m_{00} - m_{10}m_{01}/m_{00}^2 \end{aligned} \tag{25}$$

Wind speed u_n and direction θ_n are calculated by a polar transformation of the orthogonal wind speed components ξ_{2n-1}, ξ_{2n}.

$$u_n = \sqrt{\xi_{2n-1}^2 + {\xi_{2n}}^2} \quad \text{and} \quad \theta_n = \arctan(\xi_{2n-1}, \xi_{2n}) \tag{26}$$

variance of speed and direction

Thus, we may calculate the uncertainties of wind speed and direction by the uncertainties of the wind speed components[1].

$$\begin{aligned} \sigma_{u_n}^2 &= \frac{\xi_{2n-1}^2 \sigma_{\xi_{2n-1}}^2 + \xi_{2n}^2 \sigma_{\xi_{2n}}^2 + 2\xi_{2n-1}\xi_{2n}\mathrm{cov}\left(\xi_{2n-1}, \xi_{2n}\right)}{u_n^2} \\ \sigma_{\theta_n}^2 &= \frac{\xi_{2n}^2 \sigma_{\xi_{2n-1}}^2 + \xi_{2n-1}^2 \sigma_{\xi_{2n}}^2 - 2\xi_{2n-1}\xi_{2n}\mathrm{cov}\left(\xi_{2n-1}, \xi_{2n}\right)}{u^4} \end{aligned} \tag{27}$$

These wind uncertainties depend on the divergence of the backward transformation $\mathcal{F}^{-1}$ and thereby on the combination of wind speed and direction. The uncertainty of the wind direction also tends to increase for low wind speeds.

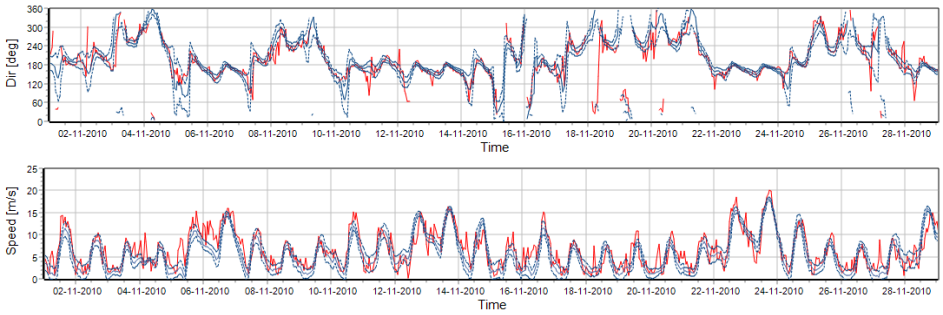


Figure 9: Time series simulation (red line) for station WM01 with conditional mean (blue line) plus/minus one standard deviation (dashed blue lines).

Equation 16 splits Fourier modes into conditional means and random components. The time dependent ensemble-average of multiple simulations is found by ignoring the random component, inverse Fourier transformation to Gaussian time series, diffusion-based mapping into orthogonal velocity components, and finally a polar transformation to wind speed and direction. The uncertainty of the simulated Gaussian variables is described by the covariance matrix in equation 17. This matrix is constant for the entire time series but the conditional mean vector $\mu_2(t)$ will vary in time. Knowing the ensemble mean and covariance of the Gaussian variables at a given time $\boldsymbol{\Phi_{2D}}\left(\boldsymbol{x_2} \,|\boldsymbol{\mu_{x_2}}, \boldsymbol{\Sigma_{x_2}}\right)$ we can calculate a time-dependent uncertainty of the orthogonal velocity components by equation 24 and finally the uncertainty of wind speed and direction by equation 27. In practice we calculate a table of the integral in equation 24 for all combinations of the Gaussian variables and interpolate in this when predicting the uncertainty of the time series.

simulation error band

Figure 9 shows part of a conditional simulation of speed and direction together with

[1] By a change of notation $u = \sqrt{x^2 + y^2}$ and $\theta = \arctan(x, y)$ we get the perhaps more readable expressions $\sigma_u^2 = \frac{x^2\sigma_x^2 + y^2\sigma_y^2 + 2xy\mathrm{cov}_{xy}}{u^2}$ and $\sigma_\theta^2 = \frac{y^2\sigma_x^2 + x^2\sigma_y^2 - 2xy\mathrm{cov}_{xy}}{u^4}$

the conditional mean values and an error band of plus/minus one standard deviation, and we note that the error bands vary in time. This is both due to nonlinear variations in the diffusion-based mapping functions and the polar transformation, which enhances the wind direction uncertainty when the wind speed is low. We see that the simulated time series sometimes exceeds the error band, as we expect them to do for about 33% of the time.

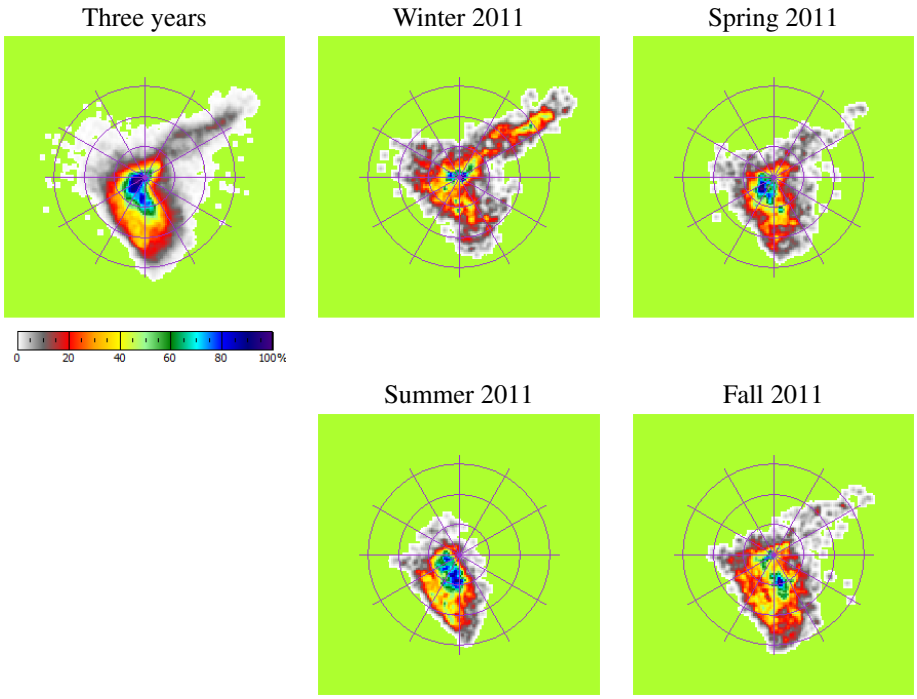


Figure 10: Probability distributions of the reference wind data for WM01 Alexander Bay sampled with different durations. The lime-green colour indicates conditions with no observations.

3.5 Problems due to limited data coverage

Figure 10 illustrates how data coverage depends on sample period. The top-left sub-figure shows the probability distribution for the long-term period, and the following sub-figures show distributions for four seasons in year 2011[2]. The lime-green colour shows wind conditions with no observations. All long-term reference data need to be corrected, but due to insufficient data coverage, this may not be possible with the diffusion-based transformation. There were no observations of high wind speeds from northeast during the summer of 2011, so a model based on these data would not know how to correct such wind conditions. Even a MCP method based on sector-wise linear regression would be inaccurate, since during the summer we only have observation of NE winds up to 2 m/s so extrapolation to 20 m/s would be rather uncertain.

Figure 11 is similar to Figure 9, except that here we compare the reference wind and local observations to simulation error bands calculated by one year of concurrent data. The simulation error bands follow the observations for most of the time although there are episodes with deviations.

Figure 12 makes the same comparison, except now only for the wind speed signals and with simulation error bands based on six months of concurrent observations. The differences are very small. Figures 13 and 14 show comparisons for periods of three

[2]Summer represented by Dec 2010 to Feb 2011, fall represented by Mar 2011 to May 2011, etc.

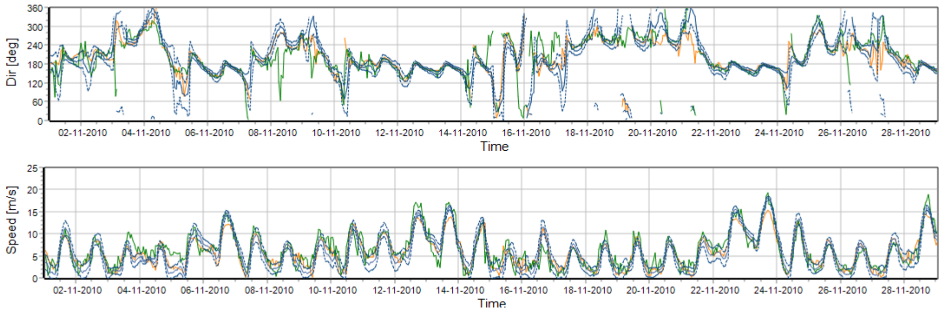


Figure 11: Observed wind (green line) and reference signal (orange line) with mean (blue line) plus/minus one standard deviation (dashed blue lines) of a simulation target based on observation in year 2011.

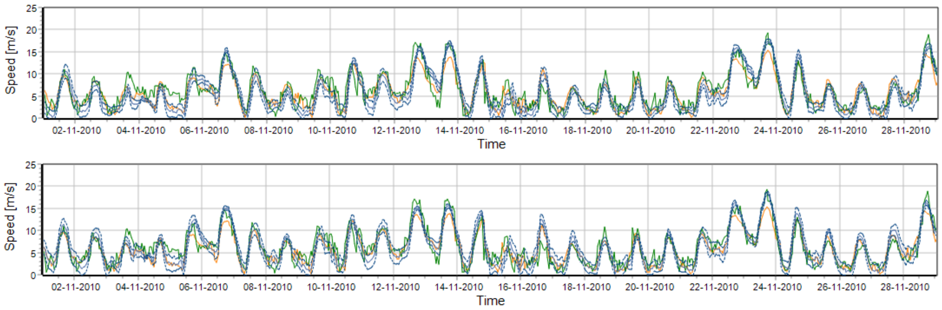


Figure 12: As previous figure with simulation targets based on observations in first (top) and second (bottom) half of year 2011.

and one months of concurrent data, respectively. We see differences among the seasons and gaps in the signals where we cannot map the simulated Gaussian signals into a wind vector. The narrower error bands indicate that we did not catch the full variation of the wind climate.

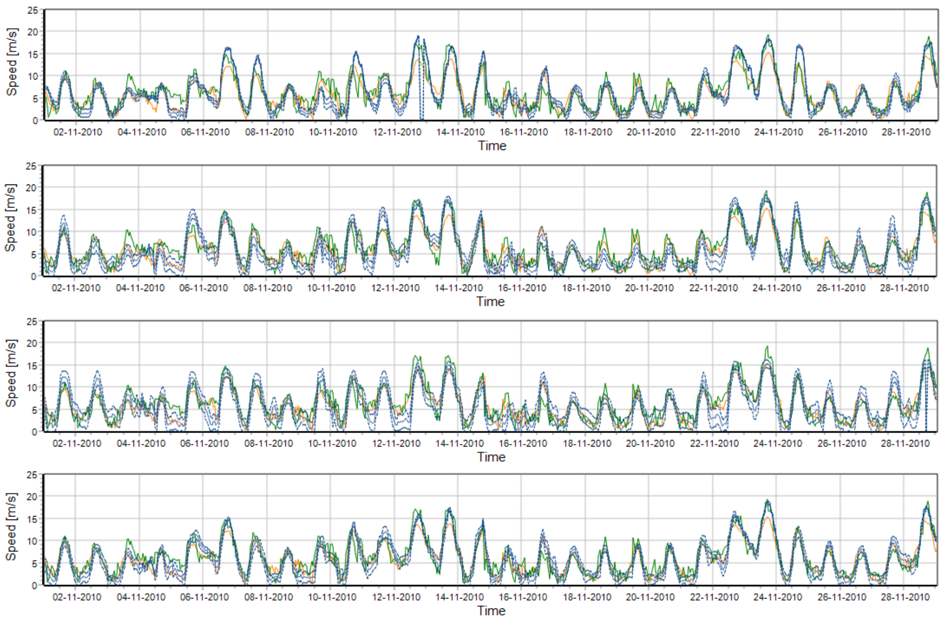


Figure 13: As previous figures with simulation targets based on observations in the three-month periods: Summer (top), fall, winter and spring (bottom) of year 2011.

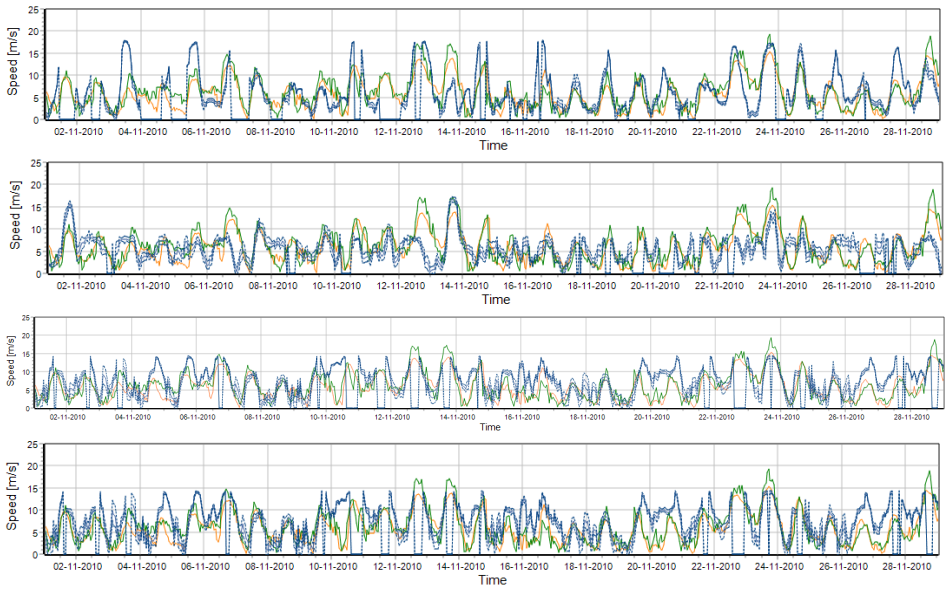


Figure 14: As previous figures with simulation targets based on observations in January (top), April, July and October (bottom) of year 2011.

4 Traditional MCP methods

Many MCP methods have been suggested in the literature, see Carta, Velázquez & Cabrera (2013) for a comprehensive review. These methods are often based on a linear relation $\hat{u}_p = a + b\,u_r$ predicting the wind at a potential wind energy site by a reference wind signal u_r. The offset parameter a is determined by the mean values and the slope parameter b.

$$a = \langle u_p \rangle - b\,\langle u_r \rangle \tag{28}$$

simple linear regression

The most obvious regression method is to consider reference observations as independent variables and observations at the predicted site as dependent stochastic variables. Minimization of squared model errors $(\hat{u}_p - u_p)^2$ leads to a prediction of the slope parameter.

$$b = \rho\sigma_p/\sigma_r \qquad (u_r \text{ independent}) \tag{29}$$

where ρ is the correlation coefficient. On the other hand, if u_p is considered independent and we minimize $(\hat{u}_r - u_r)^2$ the slope becomes

$$b = \sigma_p/(\rho\sigma_r) \qquad (u_p \text{ independent}) \tag{30}$$

and the offset will change accordingly.

orthogonal linear regression

York (1966) discussed methods based on linear regression where both observation and reference data have measurement errors. One of these methods (Worthing & Geffner 1946) normalises the data (u_r, u_p) by their expected measurement uncertainties $(\sigma_{\varepsilon,r}, \sigma_{\varepsilon,p})$, which for simplicity are assumed constant for all wind conditions. The slope of the regression line then becomes

$$b = \frac{1}{2\rho}\left(\frac{\sigma_p}{\sigma_r} - c\frac{\sigma_r}{\sigma_p}\right) + \sqrt{c + \left[\frac{1}{2\rho}\left(\frac{\sigma_p}{\sigma_r} - c\frac{\sigma_r}{\sigma_p}\right)\right]^2} \tag{31}$$

where $c = \sigma^2_{\varepsilon,r}/\sigma^2_{\varepsilon,p}$. If the uncertainties of the two signals are equal $c = 1$ then the regression line is termed the major axis.

variance ratio method

A very simple approach (Teisier 1948) is to define a slope by the ratio of the standard

deviation of the two distributions.

$$b = \sigma_p / \sigma_r \tag{32}$$

The associated regression line is called the reduced major axis (York 1966) and the method is also known as the variance ratio method (Rogers, Rogers & Manwell 2005). Equation 31 reduces to equations 29, 30, and 32 for $c = 0$, $c \to \infty$ and $c = \sigma_p^2/\sigma_r^2$, respectively. Thus, the reduced major axis approaches the major axis if the ratio of the measurement uncertainties in the signals is equal to the ratio of the variances of the signals.

Weibull stretching

As an alternative to linear regression, we can match the probability distributions of the two data sets. Assuming Weibull distributions with shape parameters (k_r, k_p) and scale parameters (A_r, A_p) for the reference data and predicted site, respectively, leads to this non-linear transformation.

$$u_p = A_p \left(u_r / A_r\right)^{k_r/k_p} \tag{33}$$

speedup factor

If we divide the data for a short measurement period into samples for each wind sector, we risk having very few data points in some samples. In this case it might be more robust simply to model the slope by the ratio of the mean wind speeds.

$$b = \frac{\langle u_p \rangle}{\langle u_r \rangle} \tag{34}$$

comparison

Figure 15 compares the above-mentioned fitted models to a joint wind speed distribution with marginal Weibull distribution and a correlation of $\rho = 0.8$, using the model in Appendix A. The definition of the offset parameter makes every linear regression line cross the centre of gravity. Linear regression with u_r as the independent parameter gives a moderate slope and high offset parameter, while the converse is true when u_p is considered independent. The slopes of the major axis and reduced major axis are in the middle of the range and close to each other. The regression lines will collapse on the reduced major axis in the limit of maximum correlation $\rho \to 1$. The Weibull stretching function is curved in this example, but it becomes a straight line when the Weibull shape parameters of the two functions are equal $k_r = k_p$.

negative wind-speed

It is well known that prediction by linear regression sometimes leads to negative wind-speed predictions. Rogers et al. (2005) decided to adjust such false predictions to zero speed. This introduces a small bias of the mean wind speed, whereas the effect on energy density or power production seems to be small. An alternative option could be to reject these false predictions from the data set, but that would lead to positive bias of mean wind speed, energy density and production.

alternative methods

There are alternative methods for linear regression, such as maximum likelihood or minimisation of the absolute error, which are less sensitive to data outliers than least-square-error minimisation.

data quality and filtering

Quality control and data filtering will also improve the modelling, e.g. we could remove data with obvious measurement errors or even data for very low wind speeds more affected by stability effects than wind speeds relevant for wind resource estimates. Some measurement stations have more than one anemometer and it may be a good idea to fill in gaps of missing data from the preferred instrument by scaled measurements from a nearby anemometer.

correction of direction

Some MCP methods correct the wind direction and thereby the frequency of occurrence in wind sectors. Simple ways to do this is first to calculate the mean difference in wind directions in the period of concurrent data for all sectors, and then apply this in the long-term correction. In order to avoid voids or overlaps in the corrected wind direction we can interpolate between known corrections at sector mean directions, eg. by piece-wise linear interpolation or by a FFT-based spline function.

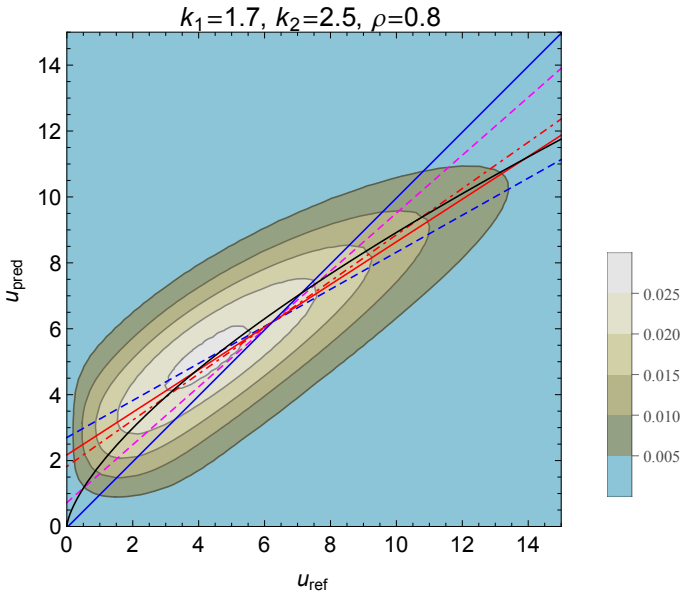


Figure 15: Example of a joint wind-speed probability distribution with marginal Weibull distributions with fitted linear regression models based on u_r as independent variable (dashed blue line), u_p as independent variable (dashed magenta line), major axis (red line) reduced major axis (dashed red line), linear regression forced through zero (blue line) and the Weibull stretching transformation (black curve).

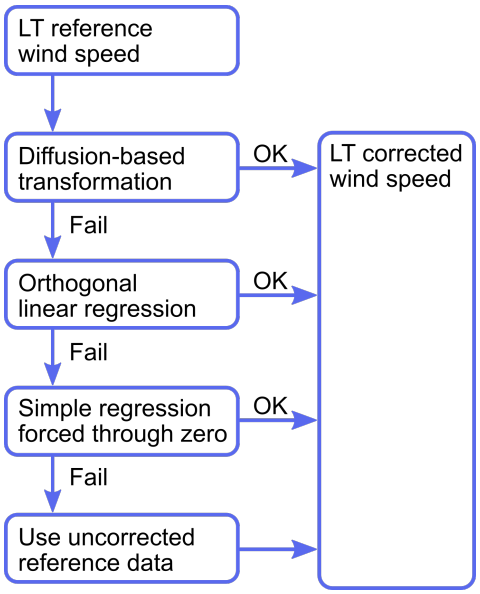


Figure 16: Contingency strategy for failing transformations.

5 Contingency strategy

As discussed in section 3.5, we expect problems for the diffusion-based long-term correction if the measurement campaign did not include all the conditions which occurs in the long-term wind climate. This is of particular concern for situations with high wind speed, so simply rejecting the reference data which are impossible to map to local winds would lead to a bias in the predicted distribution. A possible mitigation strategy would be to correct the problematic data points by more robust methods as illustrated in Figure 16. The preferred method in the figure is diffusion-based mapping and if this is not

possible we try the method of orthogonal linear regression. If that regression model also becomes unreliable we try correction by a scaling factor and, as a last resort, simply use the reference data with no correction.

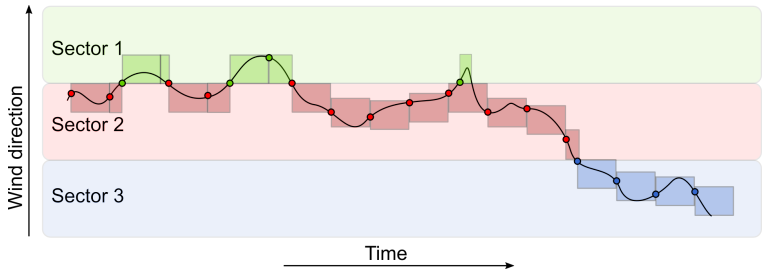


Figure 17: Sketch of independent samples per sector.

failure criteria

We need failure criteria to implement the contingency strategy. The diffusion-based transformation is considered to fail when the backward transformation from Gaussian variables to wind components $\mathcal{F}^{-1}$ is undefined. For the MCP methods we reject linear regression lines of individual wind sectors when we have

- too few independent measurements linear regression, $N < N_{\min}$
- too low coefficient of determination, $R^2 < R^2_{\min}$, in the linear regression
- regression lines with very unexpected slope, $b \notin [b_{\min} \dots b_{\max}]$

independent data

Linear regression assumes observations to be independent, so the effective sample size N is often lower than the number of available records. As an approximate criteria we assume consecutive data points x_i and x_{i+1} to be independent if $t_{i+1} - t_i > \mathcal{T}$ where $\mathcal{T}$ is the integral length scale of the wind speed autocorrelation function $\rho_x(\tau)$

$$\mathcal{T} = \int_0^\infty \rho_x(\tau)\, d\tau \tag{35}$$

This time scale is typically on the order of one day, see Figure 8, although it will vary for different sites. MCP models apply regression lines for a number of wind sectors N_{sec}. Figure 17 shows a sketch with the wind direction as a continuous curve and dots indicating sample points independent of one another according to the quarantine periods $\mathcal{T}$ indicated by the squares. The quarantine periods are independent for each sector so sometimes we are allowed to sample even though we made a recent sample in a neighbouring sector. The resulting estimate of the number of independent observations in each sector is used for failure criteria and uncertainty estimates, but for convenience we include all observations in each sector for fitting regression lines.

6 Model performance

RECAST

The aim of the RECAST project is to reduce the time for wind resource assessments by careful planning and advanced measurement systems. As a contribution to this project we will investigate long-term corrections based on short periods with concurrent data. The objective is to answer questions like

- Will uncertainty and bias depend on the duration of the measurement campaign?
- Will uncertainty and bias depend on the local wind climate?
- Which correlation method is best suited for a short measurement campaign?

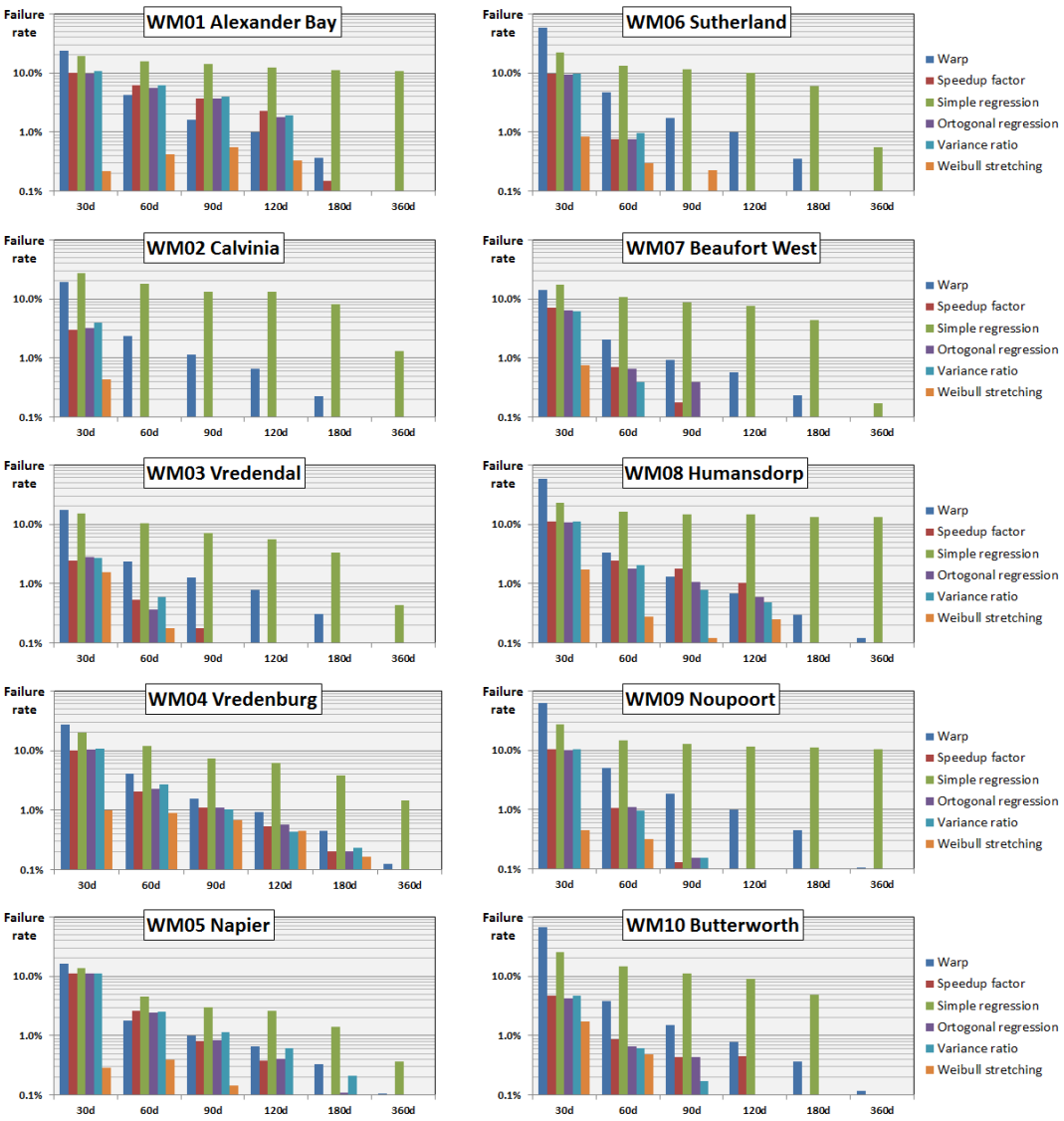


Figure 18: Failure rate for different correlation models as a function of the length of the correlation period.

This section 6 examines long-term corrections by multiple time-shifted extracts from the available observations and compare with long-term corrections based on the complete time series. The following section 7 will focus on the perhaps more practical problem of how to estimate the uncertainty of a long-term correction based on a single time series.

failure rate

Figure 18 displays percentages of the long-term data where long-term correction is impossible. The failure rates are calculated for correlation periods of 30, 60, 90, 120, 180 or 360 days and to avoid seasonal bias, they are averages of multiple samples distributed over three years of data. The failure rates of the diffusion-based method, named *warper* in the figure, are defined by the percentage of the long-term reference data where the backward mapping function $\mathcal{F}^{-1}(\mathbf{x}_1)$ is undefined. For each regression line used in a MCP method we required a reasonable amount of independent observations $N_{\mathrm{eff,s}} \geq 5$, a reasonably high coefficient of determination $R^2 > 0.5$ and a realistic regression line slope $0.5 < b < 2$. For the Weibull stretching method we require a similar number of independent observations $N_{\mathrm{eff,s}} \geq 5$ and reasonable similarity of Weibull shape factors $0.5 < k_{\mathrm{obs}}/k_{\mathrm{ref}} < 2$. The overall failure rate is estimated by the frequency of the long-term reference data in sectors with unsuccessful regression models. Although these requirements are somewhat

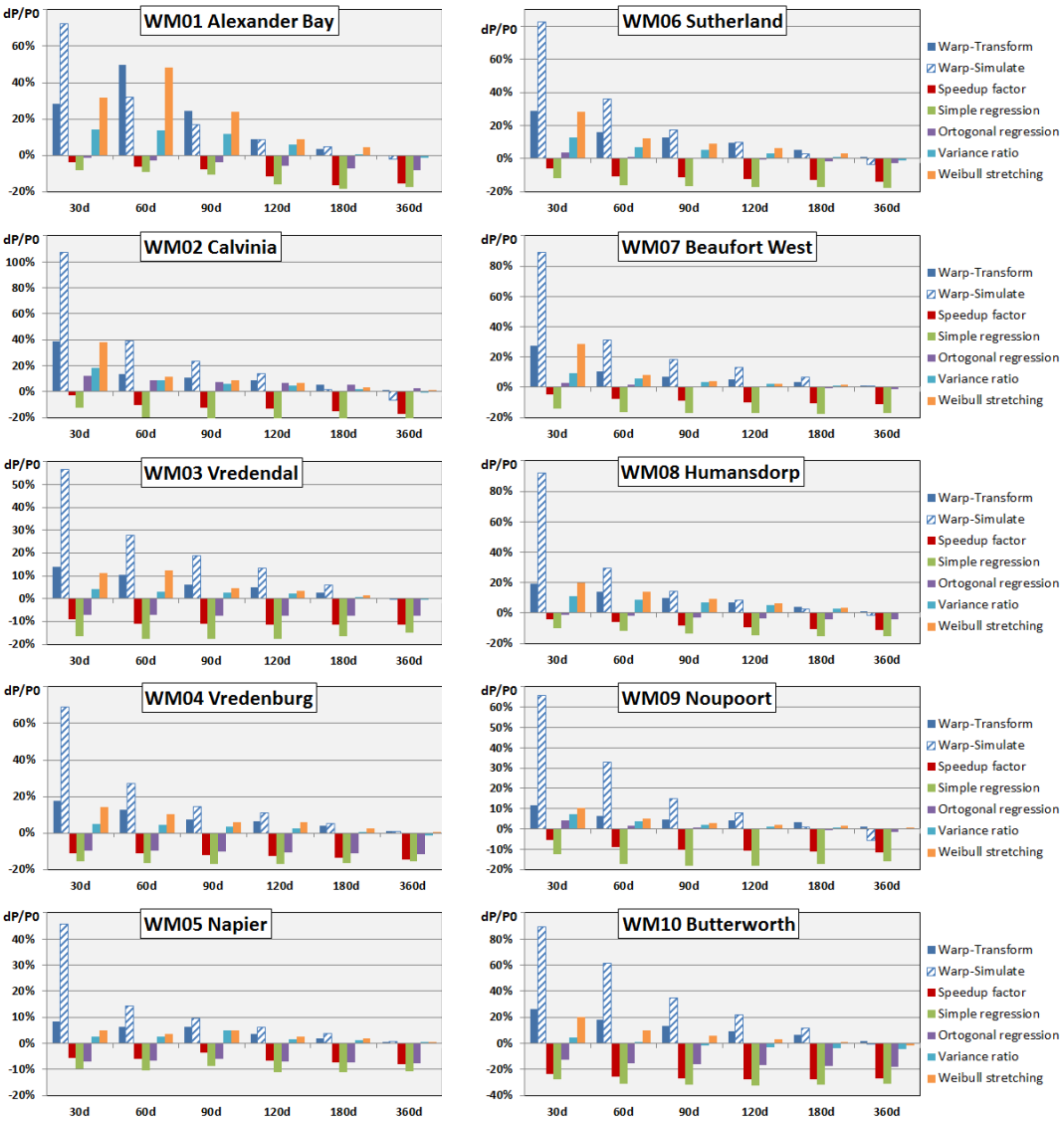


Figure 19: Bias of predicted mean wind power density for different correlation models as a function of the length of the correlation period. Note the variable scale of the vertical axis.

subjective, we will draw the following conclusions:

- The failure rate depends on the duration of the correlation period.
- The failure rate varies for different sites.
- The failure rate of diffusion-based mapping is higher than for MCP methods when the correlation period is short. This is due to problems with extrapolation from low to high wind speeds.
- The failure rate for MCP by simple linear regression is generally higher than for other MCP methods. This is because simple linear regression tends to predict lower slope parameters when data lack correlation.

bias

The long observed time series enable us to calculate the bias of the predictions. Figure 19 shows[3] relative biases of the mean power density $\langle \Delta u^3 \rangle / \langle u^3_{\mathrm{ref}} \rangle$ and similar to figure 18 we display average values for all seasons. We both include statistics for direct

[3] Appendix B shows similar plots for mean wind speed bias.

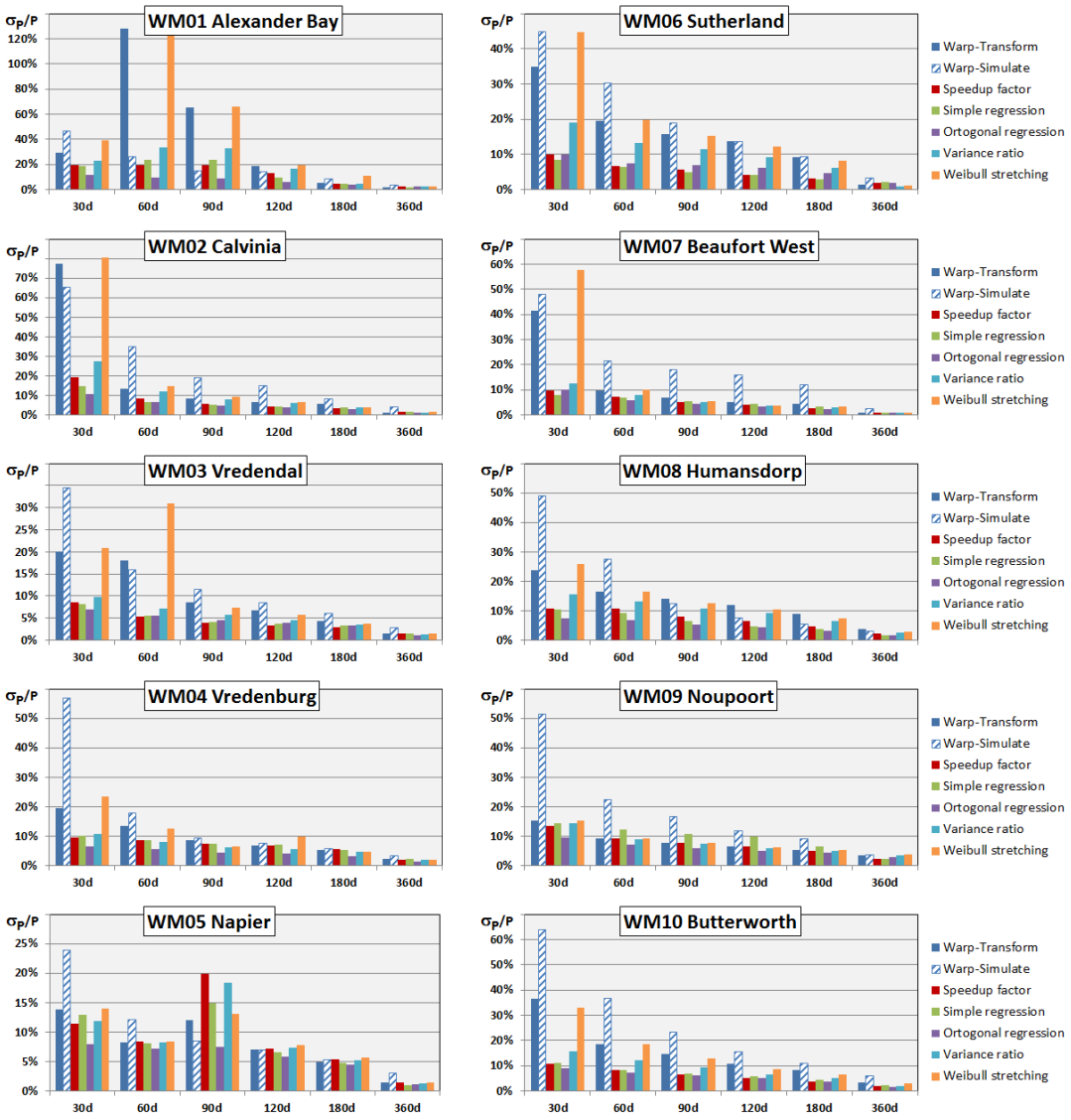


Figure 20: Root-mean-square error of predicted mean wind power density for different correlation models as a function of the length of the correlation period. Note the variable scale of the vertical axis.

diffusion-based transformation and the conditional time series simulation described in section 3.2.

We conclude that

- The bias increases for short correlation periods.
- The bias for short correlation periods depends on the site.
- Diffusion-based mapping has a problem with short correlation periods.
- Conditional simulation has more bias than direct transformation. This is probably due to inaccuracies in spectral model defined by short time series.
- Simple linear regression often results in negative bias and this does not improve much for long correlation periods.

RMS error

Figure 20 is based on the same calculations as figure 19; the only difference is that we now display the root-mean-square error (RMSE) of the predictions instead of their average bias. We conclude that

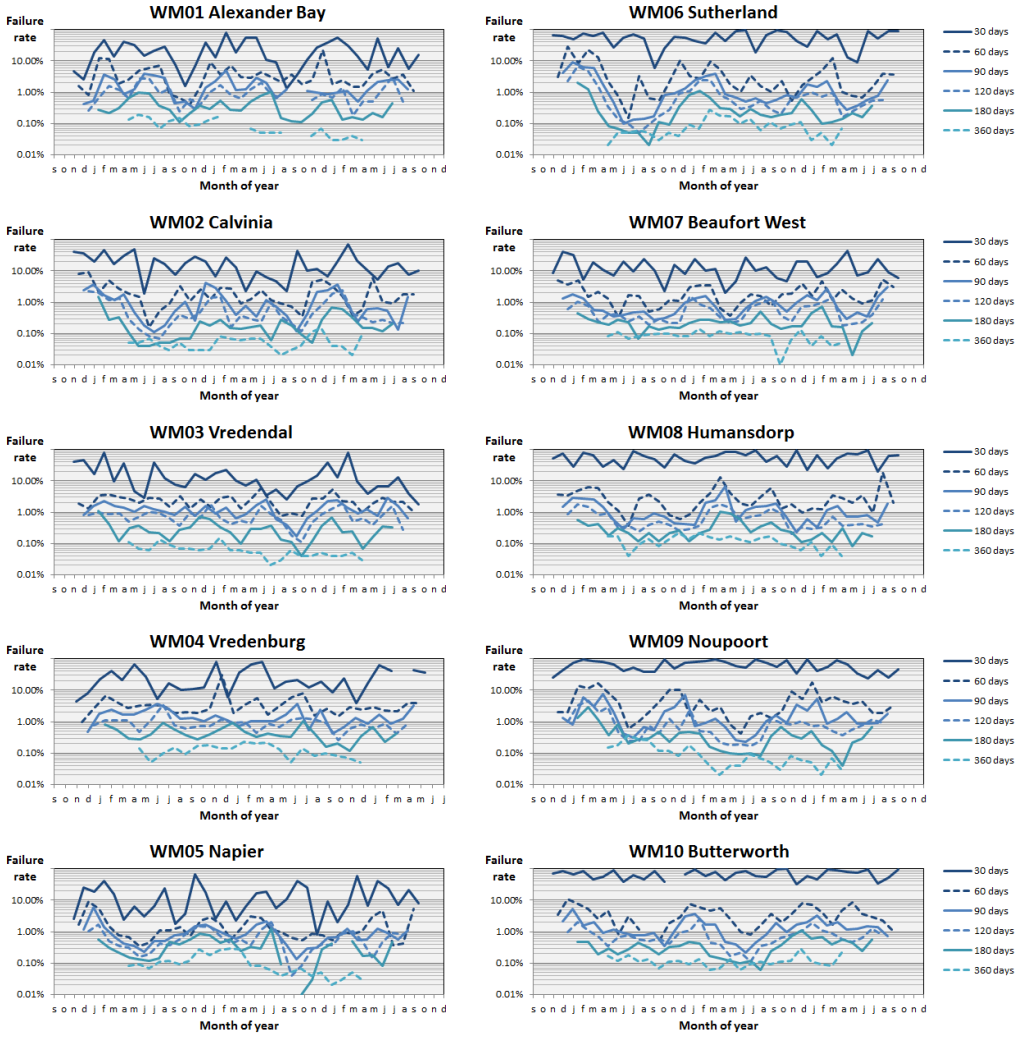


Figure 21: Seasonal variation of failure rate for diffusion-based mapping with correlation periods of variable duration at ten stations.

- Power density RMSE decreases for long correlation periods. Some results for short periods are affected by the lack of data, but it seems like we can expect relative accuracy on the order of $\sigma_p/p \simeq 10, 5, 2\%$ for 30, 90 and 360 days, respectively.
- Power density RMSE depends on the local wind climate.
- Power density RMSE by different methods approach the same levels for long correlation periods.
- MCP methods result in lower RMSE than diffusion-based mapping and Weibull stretching for short correlation periods.
- Conditional simulation generally results in higher RMSE than direct transformation. This effect is expected since the simulation has a stochastic component, but the dependence on correlation period suggests problems when fitting the spectral method to short time series.

seasonal failure rate?

Figure 21 shows the failure rate of the diffusion-based transformation as function of time. We recognize the dependence on the duration of the concurrent data period and differences among stations. At some stations we see better performance in winter periods, but there is no clear seasonal variation and it is not universal for all stations. Figure 22 shows similar plots of the prediction error of the wind power density as function of time as

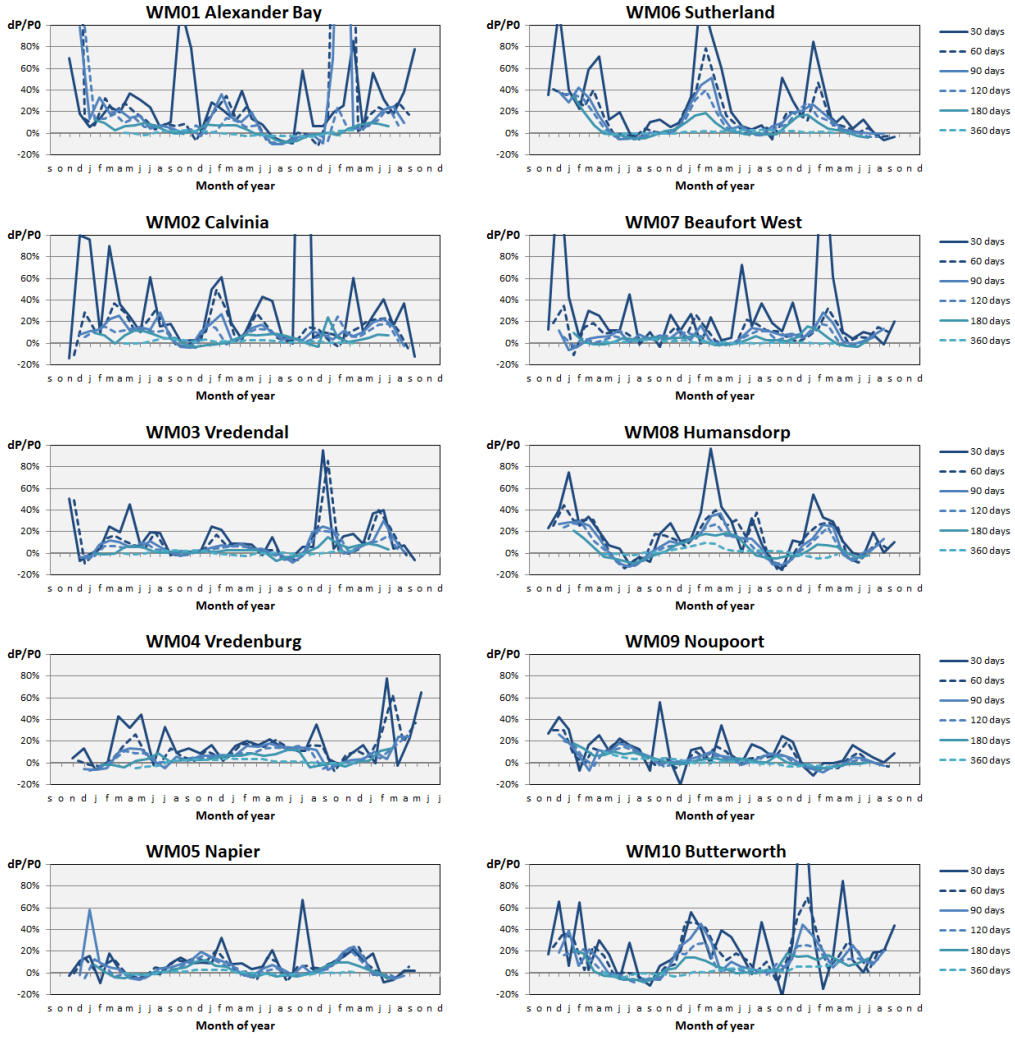


Figure 22: Seasonal variation of the error of the predicted mean wind power density for diffusion-based mapping with correlation periods of variable duration at ten stations.

predicted by diffusion-based transformation. For many stations, e.g. WM06 Sutherland, the performance improves in winter periods, e.g. June to September, but it is difficult to prove a general seasonal dependence of the prediction error. Figure 23 shows similar prediction errors for MCP with orthogonal linear regression. This plot also suggests seasonal variations which, however, differ for different sites.

seasonal bias?

Figure 24 is based on the same calculations as the previous figures and compares bias and standard deviations for different methods and a long correlation period. The diffusion-based mapping is doing much better when calibrated by a long period of concurrent data, at least when using direct transformation. The methods of variance ratio and Weibull stretching are doing surprisingly well.

conclusions on performance

With reference to the questions at the beginning of this section we conclude that

- Uncertainty depends on the duration of the measurement campaign, e.g. the uncertainty of wind power density for all wind directions is on the order of $\sigma_p/p \simeq 10\%$, 5% and 2% for 30, 90 and 360 days, respectively. Uncertainty of mean wind speed is about a third of this, see Appendix B, and presumably the relative uncertainty of annual energy production will be somewhere between these estimates, depending on the wind turbine power curve and local wind distribution. The uncertainty will be higher for sector-wise wind speed distributions.
- Uncertainty depends on the local wind climate. Thus, uncertainty is not just a ques-

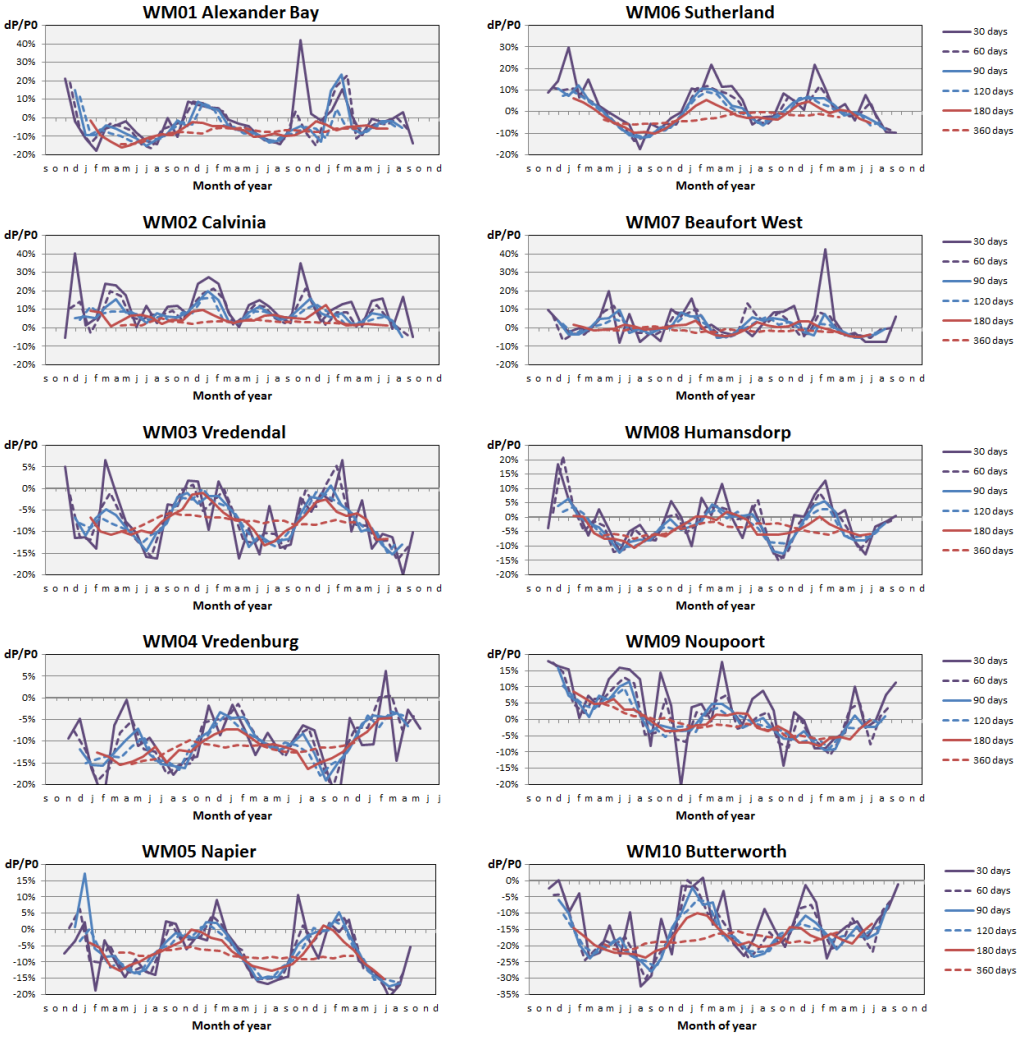


Figure 23: Same as previous figure, except for LT correction by MCP orthogonal linear regression. Note the variable scale of the vertical axis.

tion of the measurement accuracy or how well a model can predict local wind conditions; it is also important whether all relevant wind conditions were captured during the measurement campaign.

- Diffusion-based transformation is not ideal for short measurement campaigns as it cannot extrapolate from low to high wind speeds. MCP by orthogonal linear regression or the variance-ratio method seems more appropriate for short measurement campaigns.

7 Uncertainty estimates

We have seen that the accuracy of a long-term correction depends on the duration of the correlation period, but not in a simple way. Thus, we need a method for uncertainty assessment of long-term corrections based on specific single time series.

bootstrap method

The most popular resampling method is called the bootstrap method. The principle is to construct artificial data sets of the same size as the original time series by random sampling from the original data and, perhaps a little surprising, multiple samples of the same

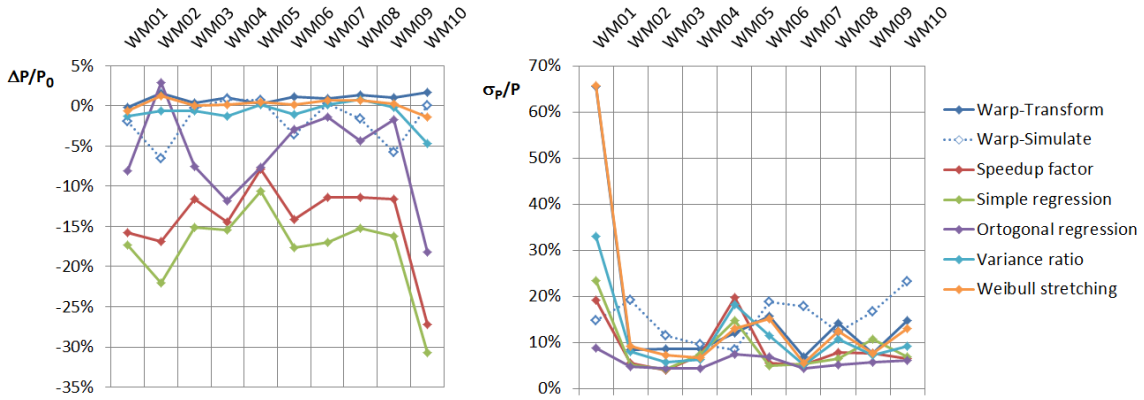


Figure 24: Bias and standard deviation of predicted mean power density for correlation period of 360 days.

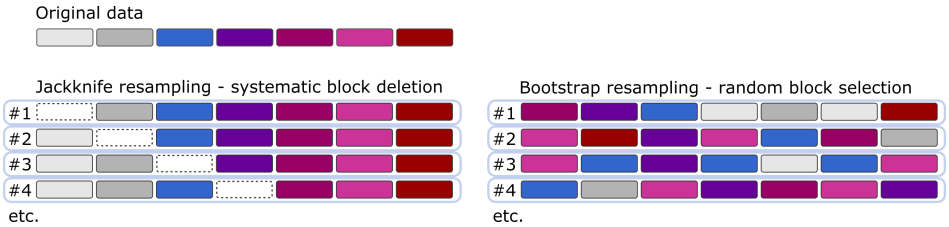


Figure 25: Sketch of the difference between jackknife and bootstrap resampling.

data point is allowed. The basic version of the method assumes independent data which is not true for time series. To minimize this error we divide the original time series into blocks, which are substantially longer than the autocorrelation time scale of the process. This procedure is repeated many times and uncertainty estimates are estimated by the variance between the resampled data sets.

jackknife method

The jackknife method is an earlier resampling method. Again, the original time series is divided into blocks, but this time the procedure is to include all but one block in a sample. The procedure is systematically repeated to generate as many variations as there are blocks to be left out. Figure 25 illustrates the difference between the two resampling strategies.

jackknife variance

The jackknife variance estimate is defined by

$$\sigma^2_{\mathrm{jk}}(x) = \frac{N_{\mathrm{jk}} - 1}{N_{\mathrm{jk}}} \sum_{i=1}^{N_{\mathrm{jk}}} (x_{i,\mathrm{jk}} - \mu_x)^2 \tag{36}$$

where N_{jk} is the number of data blocks in the original data. The variance estimate is relative to the mean of the original data set μ_x and the factor $N_{\mathrm{jk}} - 1$ compensates for the close similarity between each of the jackknife realizations and the original data set.

The jackknife method can estimate variances of Weibull parameters, mean wind speed, mean power density and similar. The required number of resampled data sets is lower with the jackknife method than with the bootstrap method. This is a considerable advantage in the present context, since we must establish computationally expensive transformation functions $\mathcal{F}$ and $\mathcal{F}^{-1}$ for each resampled data set.

number of jackknife blocks

Figure 26 shows jackknife uncertainty estimates of the long-term corrected wind power density at WM05 Napier. The data used for correlation were measured during the first three months of 2011, and uncertainty estimates were calculated by various correlation methods and repeated for 5 to 20 jackknife blocks, corresponding to block lengths of 18 to 4.5 days. The lowest uncertainties are found for the diffusion-based mapping, orthogonal linear regression and speedup factor methods. We see some dependence on the number of jackknife blocks and for certain block sizes, here 5 and 13, but probably random, some

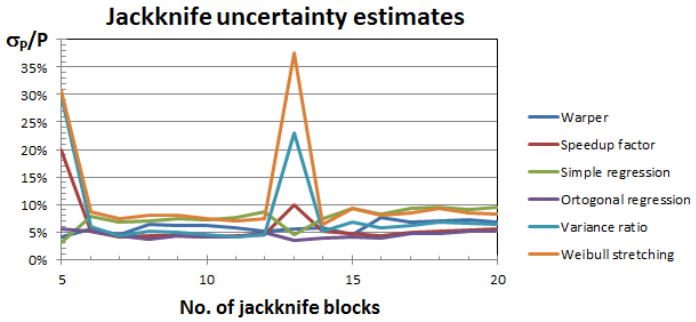


Figure 26: Comparison of jackknife uncertainty estimates for different correlation models and variable numbers of jackknife blocks.

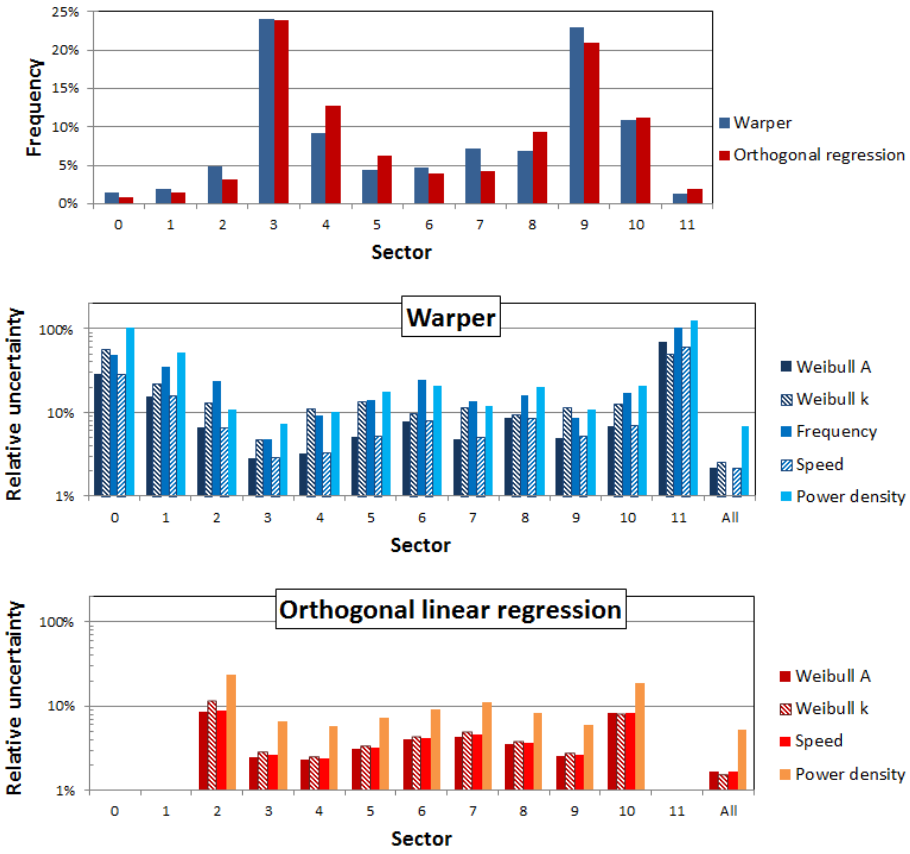


Figure 27: a) Frequency, b) relative jackknife uncertainty estimates by diffusion-based transformation and c) relative uncertainty by linear orthogonal regression using data from WM05 Napier during Jan-Mar 2011. See table 1 for definition of sector centreline directions.

methods have unusually large uncertainty. This may indicate that our failure criteria are not strict enough.

sector-wise estimates

Figure 27 shows an example of sector-wise uncertainties of Weibull parameters, mean wind speed and mean wind power density. For the diffusion-based transformation we include the uncertainty of the estimated frequency of occurrence in each sector. The data are calculated by the same data as for figure 26, except only shown for 20 jackknife blocks and for the methods of diffusion-based transformation (numbers listed in table 1) and orthogonal linear regression. We see that estimates for individual sectors are higher than

Table 1: Sector-wise relative uncertainty estimates by diffusion-based transformation using data from WM05 Napier during Jan-Mar 2011.

Sec	θ	fr	σ_A/A	σ_k/k	σ_{fr}/fr	σ_U/U	σ_P/P
0	0°	1.6%	29%	57%	49%	28%	102%
1	30°	1.9%	16%	22%	35%	16%	52%
2	60°	4.9%	6.5%	13%	24%	6.5%	11%
3	90°	24.0%	2.8%	4.7%	4.8%	2.9%	7.3%
4	120°	9.2%	3.2%	11.2%	9.2%	3.3%	10%
5	150°	4.5%	5.1%	14%	14%	5.1%	18%
6	180°	4.8%	7.9%	9.6%	24%	8.0%	21%
7	210°	7.2%	4.8%	11%	14%	5.0%	12%
8	240°	7.0%	8.7%	9.4%	16%	8.4%	20%
9	270°	22.9%	5.0%	11%	8.7%	5.1%	11%
10	300°	10.9%	6.8%	12%	17%	6.9%	21%
11	330°	1.3%	69%	50%	103%	61%	125%
All		100%	2.1%	2.5%	-	2.1%	6.9%

Table 2: Assessment of sector-wise linear regression fits $\hat{u}_p = a + b\,u_r$ using orthogonal linear regression and data from WM05 Napier during Jan-Mar 2011.

Sec	θ	a	b	σ_a	σ_b	R^2	N_{obs}	N_{eqv}	OK
0	0°	0	1	0	0	0	3	***3***	-
1	30°	2.363	0.8233	0.859	0.344	***0.470***	9	7	-
2	60°	1.337	1.0092	0.259	0.0709	0.634	37	14	√
3	90°	1.345	1.0003	0.0198	0.002	0.903	751	65	√
4	120°	0.368	1.0004	0.0192	0.0023	0.915	504	61	√
5	150°	-0.111	0.9998	0.026	0.0037	0.896	275	41	√
6	180°	0.071	0.9983	0.094	0.0162	0.839	75	22	√
7	210°	1.049	1.0048	0.116	0.0212	0.895	52	14	√
8	240°	0.055	1.0017	0.0741	0.0105	0.844	186	24	√
9	270°	0.264	1.0015	0.0726	0.0094	0.89	175	22	√
10	300°	1.181	1.0008	0.288	0.0328	0.929	41	10	√
11	330°	0	1	0	0	0	4	***3***	-
All							2112	106	95.6%

for all wind directions and quite high for sectors 0, 1 and 11, which have low frequency of occurrence. These are also the sectors where MCP by orthogonal linear regression failed due to lack of data, see table 2. Despite the 90 day sample period, we simply have too few independent observations $N_{\text{eqv}} < 5$ or too low coefficients of determination $R^2 < 0.5$ for reliable MCP predictions in these sectors. For these sectors, representing 4.4% of the long-term data, the reference winds speeds are used with no correction, i.e. $a = 0, b = 1$.

duration and season

A systematic study of jackknife uncertainty for all our data would involve repeated calculations for multiple jackknife samples, multiple jackknife block sizes, multiple durations of observations periods and multiple met mast sites. Such a study would be very time consuming for the diffusion-based mapping method, but it is feasible for the MCP methods. Figure 28 presents relative uncertainties of long-term corrected wind power density by MCP with orthogonal linear regression and jackknife uncertainty estimates. The displayed statistics are average values of multiple uncertainty estimates, with the number of jackknife blocks varying from 5 to 20 but only including estimates with MCP failure rates less than 10%. Some curves in figure 28 have gaps where the MCP failure rate was too high for all jackknife block sizes. This is mainly a problem for samples with a duration of 30 days and also seem to vary between sites, see table 3.

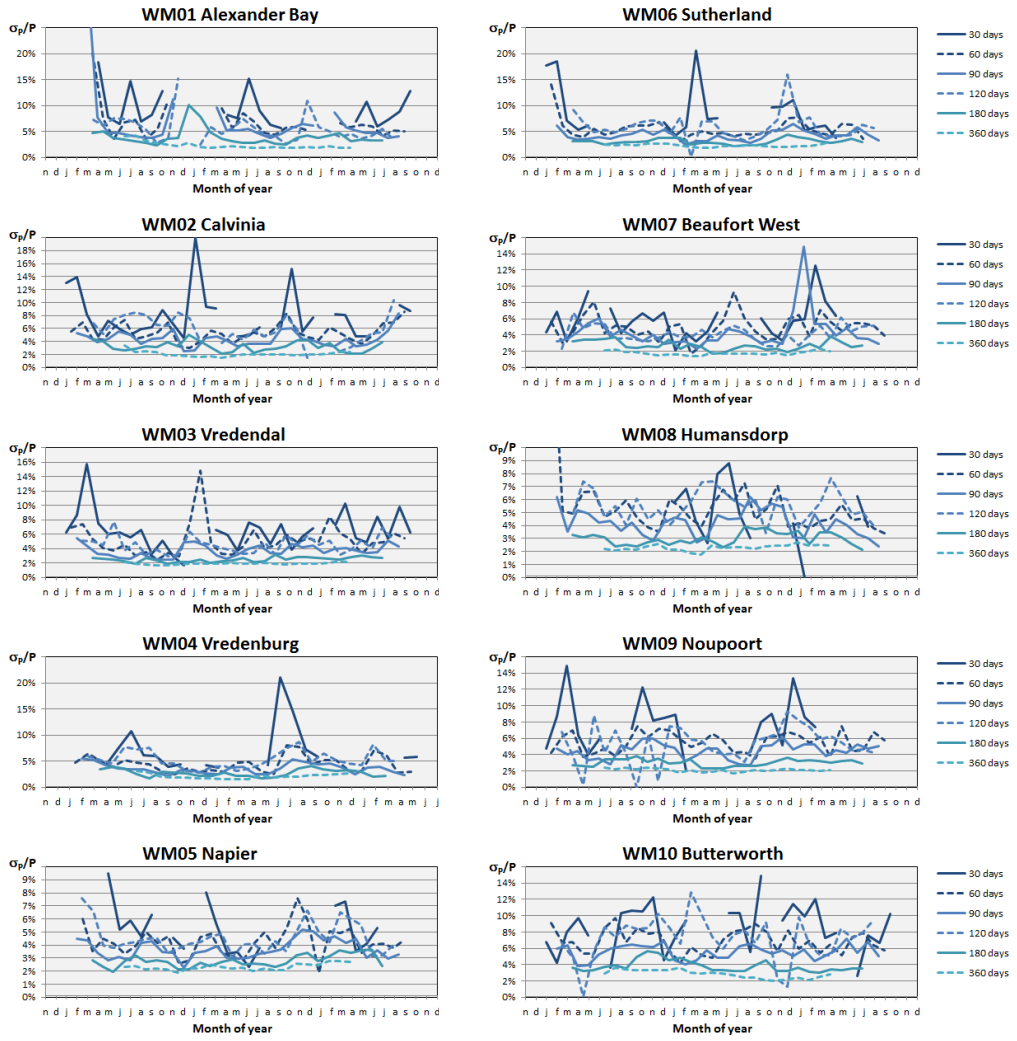


Figure 28: Relative uncertainty of the wind power density as estimated by MCP with orthogonal linear regression using jackknife resampling.

Table 3: Number of periods with no uncertainty estimates in figure 28.

	Mast ID number									
Duration	01	02	03	04	05	06	07	08	09	10
30 days	14	5	3	12	12	13	7	17	13	5
60 days	8				1	1				
90 days	4									
120 days	2									

8 Conclusions

new diffusion-based transformation

This report presents a new method for long-term correction of observed wind data. The method is based on transformation from wind speed and direction to 2D Gaussian variables. Inspired by the cartogram method of Gastner & Newman (2004), we define the transformation by an analogy with linear diffusion and calculate efficiently by fast Fourier transformation and integration with adaptive time steps. Such transformations are constructed for observed wind data and concurrent reference wind data and later applied for

transformation of a long time series of reference data to Gaussian domain and back to a long time series representative for the site of observation. Depending on the application we may or may not add a random component before backward transformation for realistic statistical decorrelation between reference data and local wind conditions.

time series simulation

The traditional input for wind resource assessment is the observed wind climate, i.e. joined statistics of local wind speed and direction. However, some applications need realistic time series of wind conditions, and for this purpose it is appropriate to include statistical decorrelation. This conditional simulation is done in frequency domain since the correlation of the two data sets depends on the time scale of the variation. It is possible to predict the time-dependent conditional mean wind and error band, and our stochastic simulations hit this target with realistic deviation.

data for the study

The new long-term correction method is tested by data from ten reference sites used for the Wind Atlas of South Africa and having different wind climates. Both direct transformation and conditional simulation works well when calibrated by a reasonable long period of concurrent data. Conditional simulation is less successful for short periods of concurrent data indicating problems with spectral analysis of very short time series. Another problem, also relevant for direct mapping, is that the long-term reference time series may contain wind conditions which never occurred during the period of concurrent data and therefore are impossible to correct by empirical transformations.

MCP methods

For comparison we implemented a selection of traditional MCP methods based on sector-wise linear regression. For short periods of concurrent data these methods are more robust than the new diffusion-based method, since they are able to extrapolate from low to high wind speeds. Thus we apply a contingency strategy and predict by MCP when the new method fails.

failure rate

Even MCP methods may fail when too few independent observations are available in certain wind sectors. Based on our test data we conclude that all methods, both MCP and the new one, have high failure rates for short measurement campaigns like 30 days and acceptable failure rates when half a year of observation are available. Unfortunately, it is not easy to predict exactly how long a measurement campaign needs to be, and the impression is that this depends on the local wind climate and season. The best strategy for practical field work is probably to make regular failure rate assessments of long-term corrections based on the accumulating database, and continue the campaign until a sufficient range of wind conditions are sampled. This is similar to the strategy for wind turbine power-curve measurements.

bias and RMS error

The power-density bias and root-mean-square error of long-term corrected time series were examined for all sites, for a range of methods, and for variable duration of the concurrent data. Based on predictions without model failures, the wind power density was found to be on the order of 10%, 5% and 2% for 30, 90 and 360 days, respectively. All methods converged to the ideal case of little bias and low RMS error for increasing duration, except the two MCP methods using simple linear regression or speed-up factors, which continued to have negative bias even with one year of data. The RMS error seemed to have a seasonal dependence, but this often differed between sites and was therefore hard to generalize. The new method had problems for short measurement periods, in particular when using conditional simulation rather than direct transformation, but it worked fine when calibrated by one year of data. The best methods for short measurement campaigns seem to be MCP by orthogonal linear regression or the variance ratio method.

jackknife resampling

For practical field work we need uncertainty estimates of long-term corrections using all available data. This is possible with a variant of the jackknife resampling method using blocks of the time series. Uncertainty estimates depend on the number of blocks and may fail due to lack of data, so it is recommended to average estimates for multiple block sizes. Jackknife resampling can produce estimates of statistical measures like Weibull parameters, mean wind speed, mean power density and similar. Uncertainty estimates for individual sectors are higher than uncertainties for all wind directions and may fail for sectors with too few independent observations.

Acknowledgements

This work started in the DTU Wind Energy *Mesoscale generalization and data* internal development project of 2017. Thanks to my colleagues Andrea N. Hahmann, Jake Badger, Niels G. Mortensen and Brian O. Hansen for useful discussions and to Niels G. Mortensen and Patrick Volker for communicating the data. These data were originally produced by the *Wind Atlas for South Africa phase I* project sponsored by the *South African National Energy Development Institute*.

The work continued in 2018 as part of the *RECAST - Reduced Assessment Time* project lead by Rozenn Wagner (DTU) and sponsored by the *Danish Innovation Fund, contract no. 7046-00021B*. Thanks to all project partners, especially the participants of work package one: Morten L. Thøgersen (EMD), Lee Cameron (RES), Juan Pablo M. Leon and Yavor V. Hristov (Vestas), Rozenn Wagner, Andrea N. Hahmann, Elin Svensson, Andreas Bechmann and Bjarke T. Olesen (DTU).

References

Abramowitz, M. & Stegun, I. A. (1964). *Handbook of Mathematical Functions*, US National Bureau of Standards.

Carta, J. A., Velázquez, S. & Cabrera, P. (2013). A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site, *Renew. Sust. Energ. Rev.*, **27**, 362–400.

Frigo, M. & Johnson, S. G. (2005). The Design and Implementation of FFTW3, *Proc. IEEE*, **93**, 216–231.

Gastner, M. T. & Newman, M. E. J. (2004). Diffusion-based method for producing density-equalizing maps, *Proc. Natl. Acad. Sci.*, **101**, 7499–504.

Hahmann, A., Lennard, C., Badger, J., Vincent, C., Kelly, M., Volker, P., Argent, B. & Refslund, J. (2015). *Mesoscale modeling for the Wind Atlas of South Africa (WASA) project*, DTU Wind Energy, Denmark.

Hahmann, A. N., Vincent, C. L., Pea, A., Lange, J. & Hasager, C. B. (2015). Wind climate estimation using WRF model output: method and model sensitivities over the sea, *International Journal of Climatology*, **35**(12), 3422–3439.

Mortensen, N., Hansen, J., Kelly, M., Prinsloo, E., Mabille, E. & Szewczuk, S. (2014). Wind Atlas for South Africa (WASA) Station and Site Description Report, *DTU Wind Energy E-0071*, DTU Wind Energy, Denmark.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1992). *Numerical Recipes in C, second edition*, Cambridge University Press.

Rogers, A. L., Rogers, J. W. & Manwell, J. F. (2005). Comparison of the performance of four measure-correlate-predict algorithms, *JWEIA*, **93**, 243–264.

Teisier, G. (1948). *Biometrics*, **4**, 14.

Worthing, A. G. & Geffner, J. (1946). *Treatment of experimental data*, John Wiley and Sons, New York.

York, D. (1966). Least-square fitting of a straight line, *Can. J. Phys.*, **44**, 1079–1086.

A Bivariate Weibull distribution

Consider a four-dimensional Gaussian distribution of the variables x_1, y_1, x_2, y_2 modelling orthogonal components of vectors at two sites. The correlation between variables at the two sites is set to r, but there is no correlation among components of different direction. The scale of all parameters is set to unity and the joint probability distribution becomes

$$p(x_1, y_1, x_2, y_2) = \frac{1}{4\pi^2(1-r^2)} \exp\left[-\frac{x_1^2 + x_2^2 - 2r\, x_1 x_2 + y_1^2 + y_2^2 - 2r\, y_1 y_2}{2(1-r^2)}\right] \quad \text{(A.37)}$$

We transform this to polar coordinates, $s_i = \sqrt{x_i^2 + y_i^2}$ and $\theta_i = \mathrm{atan2}(x_i, y_i)$, and integrate over directional variables to achieve a joint probability for the radial ones.

$$\begin{aligned} p(s_1, s_2) = & \int_0^{2\pi}\int_0^{2\pi} \frac{s_1 s_2}{4\pi^2(1-r^2)} \exp\left[-\frac{s_1^2 + s_2^2 - 2r s_1 s_2 \cos(\theta_1 - \theta_2)}{2(1-r^2)}\right] d\theta_1 d\theta_2 \\ & = \frac{s_1 s_2}{1-r^2} \exp\left[-\frac{s_1^2 + s_2^2}{2(1-r^2)}\right] \mathrm{I}_0\left[\frac{r\, s_1 s_2}{1-r^2}\right] \end{aligned} \quad \text{(A.38)}$$

The last factor in this expression, $\mathrm{I}_0(x)$, is a modified Bessel function of the first kind and zero order (Abramowitz & Stegun 1964). The marginal distribution of each variable is a Raleigh distribution, which has the cumulated probability $F_i(s_i) = 1 - \exp(-s_i^2/2)$. We can transform this into the more general Weibull distribution by

$$u_i = A_i s_i^{2/k_i} 2^{-1/k_i} \quad \text{(A.39)}$$

We then find the joint probability model we were looking for

$$\begin{aligned} p(u_1, u_2) &= p(s_1, s_2) \frac{ds_1}{du_1}\frac{ds_2}{du_2} \\ &= \frac{k_1 k_2\, \eta_1^{k_1-1}\eta_2^{k_2-1}}{A_1 A_2 (1-r^2)} \exp\left[-\frac{\eta_1^{k_1} + \eta_2^{k_2}}{1-r^2}\right] \mathrm{I}_0\left[\frac{2r\, \eta_1^{k_1/2}\eta_2^{k_2/2}}{1-r^2}\right] \end{aligned} \quad \text{(A.40)}$$

where we use scaled wind speeds $\eta_i = u_i/A_i$ for brevity. Moments of this distribution are found by insertion of equation A.39 followed by integration.

$$\begin{aligned} \langle u_1^{n_1} u_2^{n_2} \rangle &= A_1^{n_1} A_2^{n_2} \int_0^\infty \int_0^\infty \frac{s_1^{2n_1/k_1} s_2^{2n_2/k_2}}{2^{n_1/k_1 + n_2/k_2}} p(s_1, s_2) ds_1 ds_2 \\ &= A_1^{n_1} A_2^{n_2}\, \Gamma\left(1 + \tfrac{n_1}{k_1}\right) \Gamma\left(1 + \tfrac{n_2}{k_2}\right) F\left(-\tfrac{n_1}{k_1}, -\tfrac{n_2}{k_2}, 1; r^2\right) \end{aligned} \quad \text{(A.41)}$$

Here $F(a, b, c; z)$ is the hypergeometric function (Abramowitz & Stegun 1964). The expression is ill-defined in the limit of maximum correlation $r = 1$, but here the Raleigh distributed variables are equal $s_1 = s_2$ and the wind speeds are related by $(u_1/A_1)^{k_1} = (u_2/A_2)^{k_2}$. We therefore deduce

$$\begin{aligned} \langle u_1^{n_1} u_2^{n_2} \rangle |_{r=1} &= \int_0^\infty u_1^{n_1} \left(A_2 \left(u_1/A_1\right)^{k_1/k_2}\right)^{n_2} p(u_1) du_1 \\ &= A_1^{n_1} A_2^{n_2} \Gamma\left(1 + \frac{n_1}{k_1} + \frac{n_2}{k_2}\right) \end{aligned} \quad \text{(A.42)}$$

From equations A.41 and A.42 we obtain an expression for the wind-speed correlation.

$$\rho = \begin{cases} \dfrac{\left[F(-k_1^{-1},-k_2^{-1},1;r^2)-1\right]\Gamma(1+k_1^{-1})\Gamma(1+k_2^{-1})}{\sqrt{\left(\Gamma(1+2k_1^{-1})-\Gamma^2(1+k_1^{-1})\right)\left(\Gamma(1+2k_2^{-1})-\Gamma^2(1+k_2^{-1})\right)}} & \text{for } r^2 < 1 \\ \dfrac{\Gamma(1+k_1^{-1}+k_2^{-1})-\Gamma(1+k_1^{-1})\Gamma(1+k_2^{-1})}{\sqrt{\left(\Gamma(1+2k_1^{-1})-\Gamma^2(1+k_1^{-1})\right)\left(\Gamma(1+2k_2^{-1})-\Gamma^2(1+k_2^{-1})\right)}} & \text{for } r^2 = 1 \end{cases} \tag{A.43}$$

B Additional plots

This appendix displays mean wind uncertainty estimates, quite similar to the figures inside the report presenting mean power density uncertainty estimates.

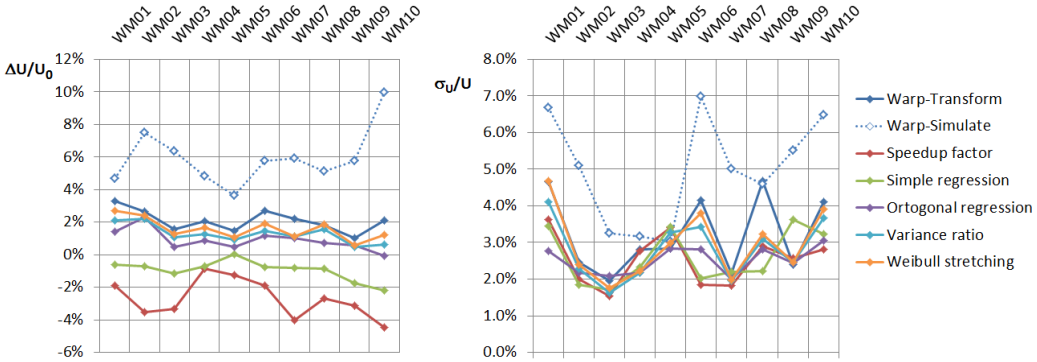


Figure 29: Bias and standard deviation of predicted mean wind speed for correlation period of 360 days, similar to figure 24.

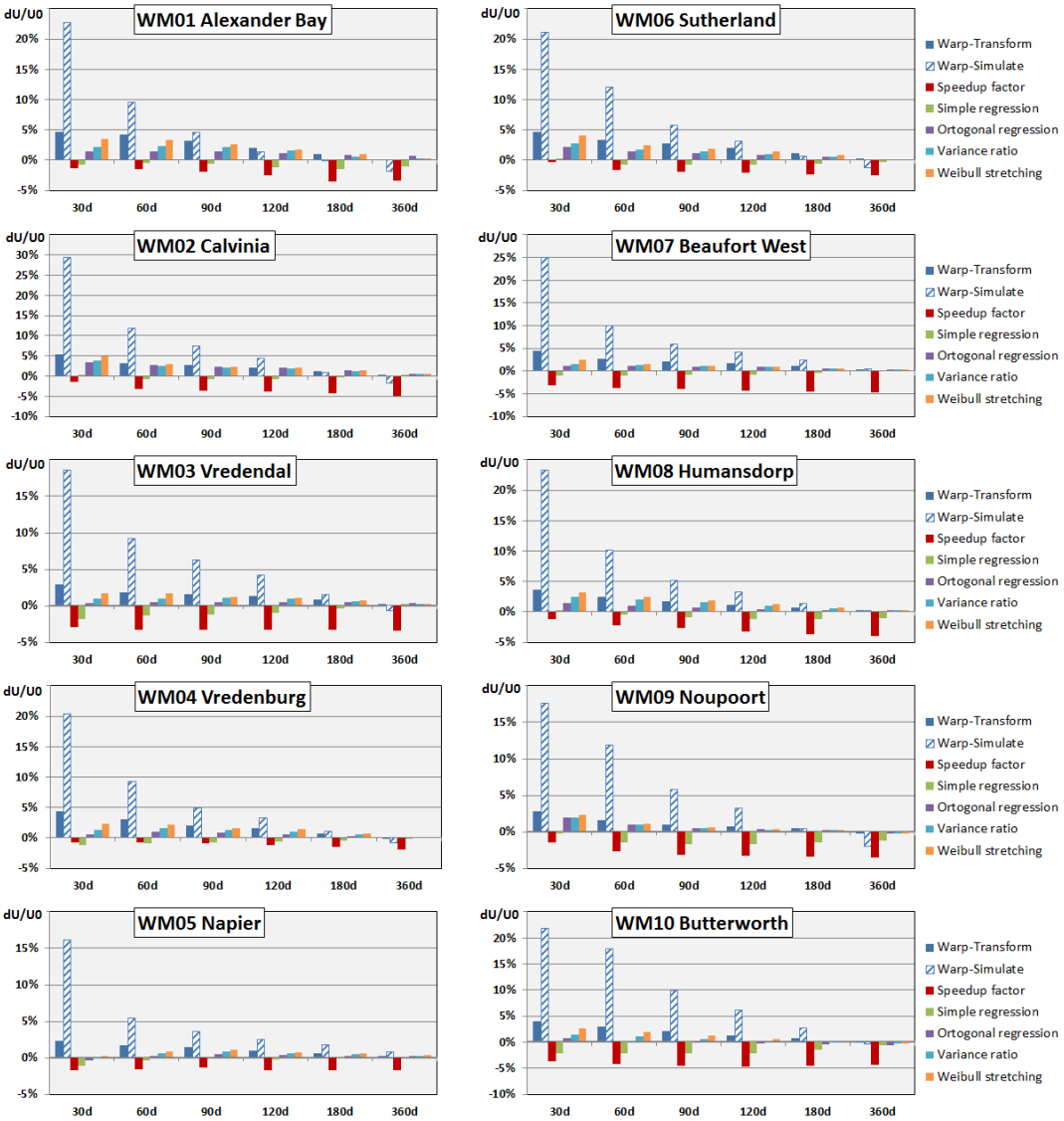


Figure 30: Bias of predicted mean wind speed for different correlation models as a function of the length of the correlation period, similar to figure 19.

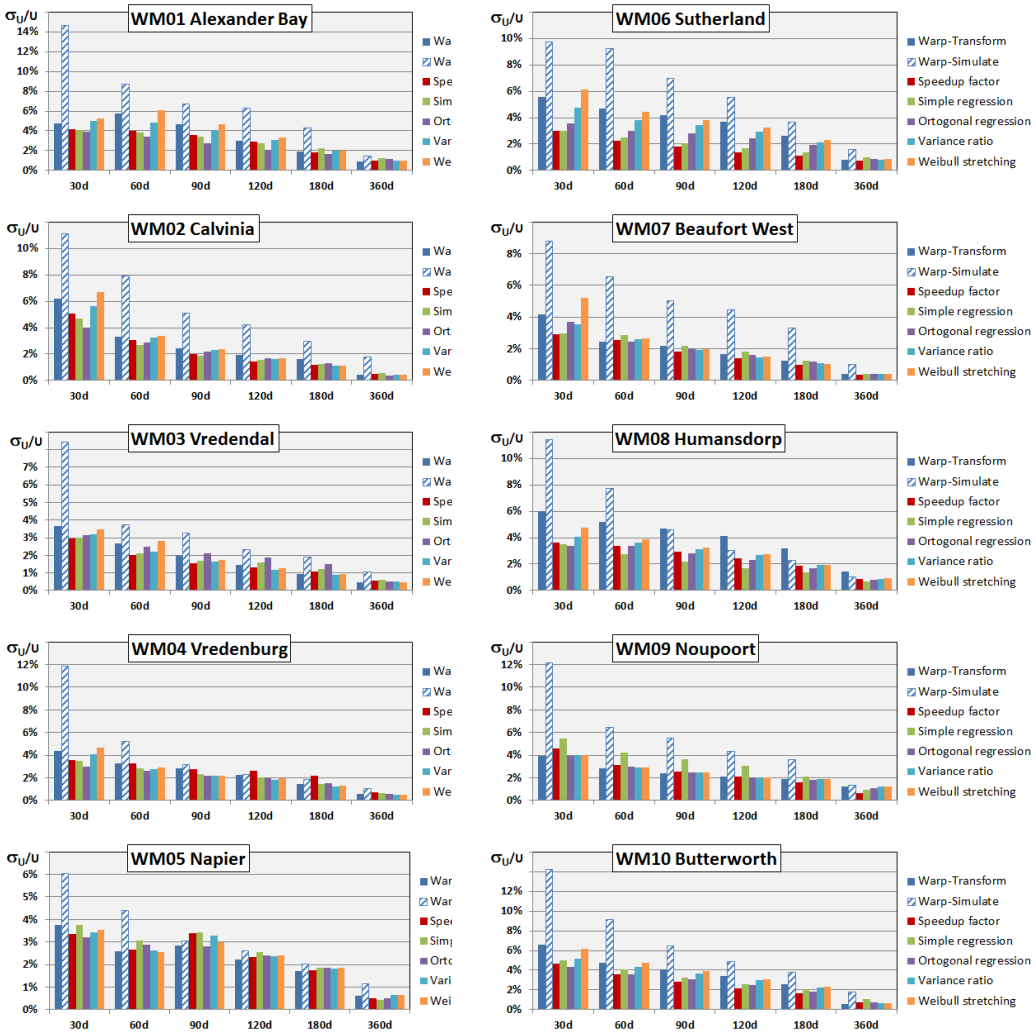


Figure 31: Standard deviation of predicted mean wind power density for different correlation models as a function of the length of the correlation period, similar to figure 20.

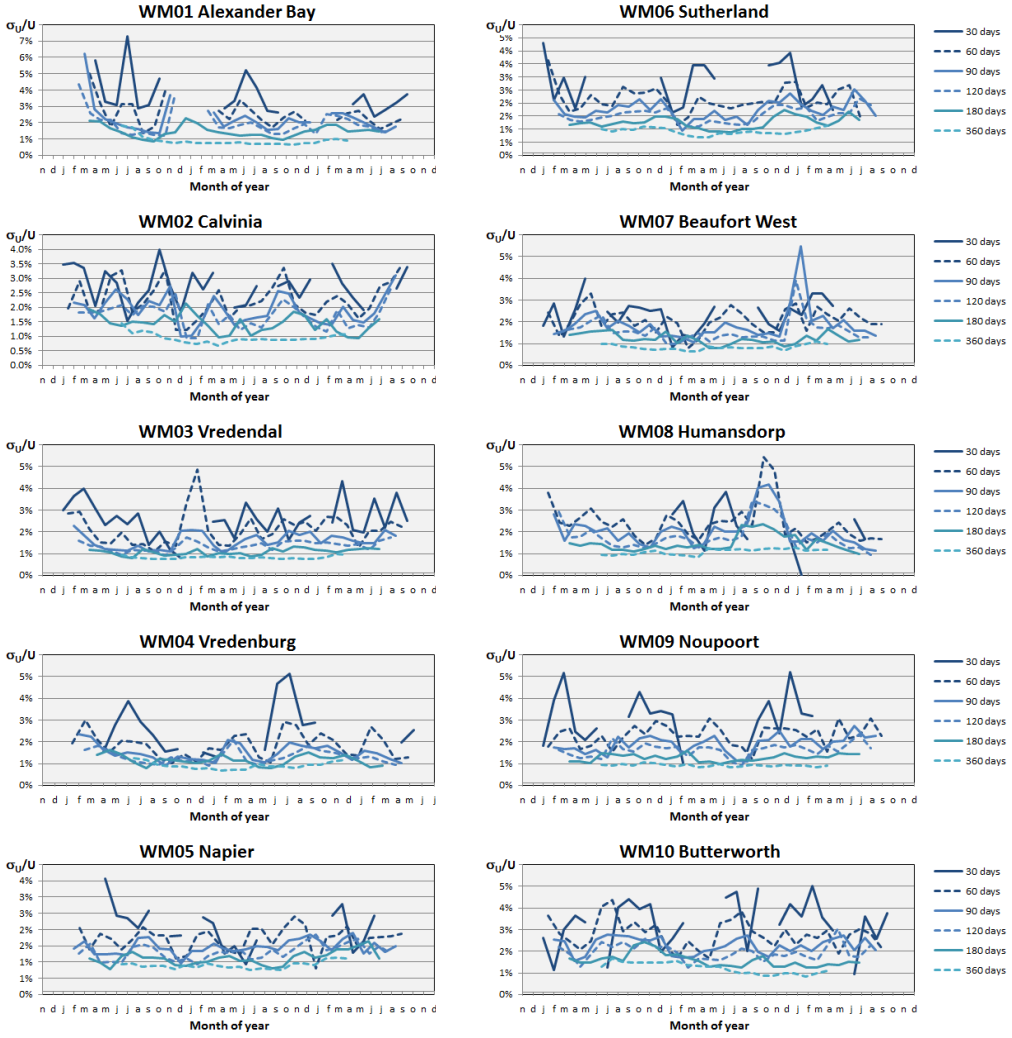


Figure 32: Seasonal variation of the error of the predicted mean wind power density for diffusion-based mapping with correlation periods of variable duration at ten stations, similar to figure 22.

DTU Wind Energy is a department of the Technical University of Denmark with a unique integration of research, education, innovation and public/private sector consulting in the field of wind energy. Our activities develop new opportunities and technology for the global and Danish exploitation of wind energy. Research focuses on key technical-scientific fields, which are central for the development, innovation and use of wind energy and provides the basis for advanced education at the education.

We have more than 240 staff members of which approximately 60 are PhD students. Research is conducted within nine research programmes organized into three main topics: Wind energy systems, Wind turbine technology and Basics for wind energy.

Technical University of Denmark
Department of Wind Energy
Frederiksborgvej 399
Building 118
4000 Roskilde
Denmark
Telephone 46 77 50 85

info@vindenergi.dtu.dk
www.vindenergi.dtu.dk