WIND DATA SUiTE Report Series

WIND DATA SUiTE Report No. 2017-10

Dr. Helmut Frey:

Multiple synoptic scale correlate predict (MSSCP).

Citation:

Frey, H. (2017): Multiple synoptic scale correlate predict (MSSCP). WIND DATA SUiTE, Germany, Report No. 2017-10, 4 pp.
https://www.WindDataSuite.de/wds/publics/WDS-Report201710-DrHF-MSSCP.pdf

Published at DEWEK 2017.

MULTIPLE SYNOPTIC SCALE CORRELATE PREDICT (MSSCP)

Dr. Helmut Frey
WIND DATA SUiTE, Böttcherstraße 8, D-21365 Adendorf, Germany, info@WindDataSuite.de

Summary

A new Measure Correlate Predict (MCP) method called MSSCP (Multiple Synoptic Scale Correlate Predict) is presented. MSSCP considers the wind field variability on the synoptic scale and thus incorporates physical processes which are essential for a more accurate statistical long-term extrapolation. It is shown that MSSCP reduces the long-term extrapolation errors in the wind power densities by approximately fifty percent on average.

1. Introduction

1.1 Background
Every wind energy project is based on an investigation of the local wind conditions in order to get an estimate of the long-term wind speed distribution at the location of the projected wind park (called location "WP" in the following). For this purpose, a Measure Correlate Predict (MCP) is applied which extrapolates a long-term wind distribution from a certain reference location (called location "Ref" in the following) onto the location WP, where a short-term wind measurement has been conducted. Several MCP methods exist already [1] using different extrapolation methods, from linear and non-linear regression models over variance methods, Weibull fit methods, matrix methods, and many more, up to methods using neural network techniques and CFD modelling.
In the following, emphasis is laid on MCP methods based on regression models.

1.2 The general problem
Since every MCP method is based on a statistical approach, the results of the long-term extrapolations are more or less erroneous. Most of the MCP methods reproduce the wind speed frequency distributions with overestimated Weibull shape parameter values [2] and with errors in the wind power densities (WPD) of 20 % and higher, though the errors in the predicted mean wind speeds may be smaller by far.
Beneath the linear regression model with vertical distances, which is the proven least-square fit for normally distributed independent variables (as for example wind speed data), several other algorithms, as for example linear regression with perpendicular (orthogonal) distances, or any polynomial regression, or any other algorithms, or certain combinations of them all, are suggested by different MCP modules. Some modules even offer a choice which algorithm to use for a certain situation.
However, the physical laws are the same anywhere and anytime. These physical laws describe the temporal evolution of the wind fields at locations WP and Ref. There is no evidence that the correlations between the physical processes at two locations could be approximated in different situations best by different statistical models. May be, a certain MCP method in general yields better results than an other MCP method, but there is no evidence that interchanging the MCP methods situational could reflect situational effects of physical processes.
In order to improve the results of statistical extrapolations in general, the errors arising from the statistical approach have to be reduced reliably and independently of time ranges and locations. This can be achieved by putting more information about the underlying physical processes into the system of the MCP.
In the following, emphasis is laid on MCP methods based on the linear regression model with vertical distances.

2. Theoretical background of MSSCP

2.1 The frequency domain of time series
In order to improve the results of an extrapolation, the set of regressions has to resolve those physical processes whose variances contribute to a great extend to the total variance of the time series.
These processes can be detected by converting a time series into its Fourier Transform. The total variance V_{ts} of the wind speed time series is given by

$$(1)\quad V_{ts} = \frac{1}{N}\sum_{i=1}^{N}(v_i - \bar{v})^2$$

where N is the length of the time series, v_i are the wind speed values of the time series, and $\bar{v}$ is the arithmetic mean of the wind speeds. The variance $V_k(p_k)$ of a certain partial wave with the period p_k can be determined from the amplitude spectrum of the wind speed Fourier Transform, calculated with a Fast Fourier Transform (FFT) algorithm, and is then given by

$$(2)\quad V_k(p_k) = \frac{1}{p_k}A_k^2\int_{t=0}^{p_k}\sin^2(\omega_k t)\,dt = \frac{1}{2}A_k^2 \quad , \quad k = 1,\dots,K$$

where k is the number of the partial wave, A_k its amplitude, ω_k its angular frequency, $p_k = N/k$ its period, and where the frequency domain $K = N/2$ covers the entire wave spectrum of the FFT, and where it holds that

$$(3)\quad \sum_{k=1}^{K} V_k = V_{ts}$$

if the length of the time series is adjusted to a power of 2 number, $N = 2^m$, due to the FFT algorithm. The declaring variance $V_d(p)$ of a certain period p with a bandwidth $w; 0 < w < 1$ is then given by

$$(4)\quad V_d(p) = 100\frac{\sum_{p'=p(1-w)}^{p'=p(1+w)} V(p')}{V_{ts}} \quad [\%]$$

In order to illustrate a distribution of declaring variances, a Fast Fourier Transform (FFT) algorithm has been applied to the time series data of the hourly values of the wind speeds of the DWD (Deutscher Wetterdienst) weather station Schwerin (station number 04625) within the 21-years time range from 1995/01/01 to 2015/12/31 (called the "Schwerin data" in the following). Fig.1 shows the declaring variances for the Schwerin wind speed data with a bandwidth of 10 % ($w = 0.1$). The 1-day wave period (diurnal wave) declares 7.1 % of the total variance of the wind speed time series, the 1-year wave period (seasonal wave) 3.5 %, and all the wave periods within the period band from 2 to 10 days together declare 39.6 %.

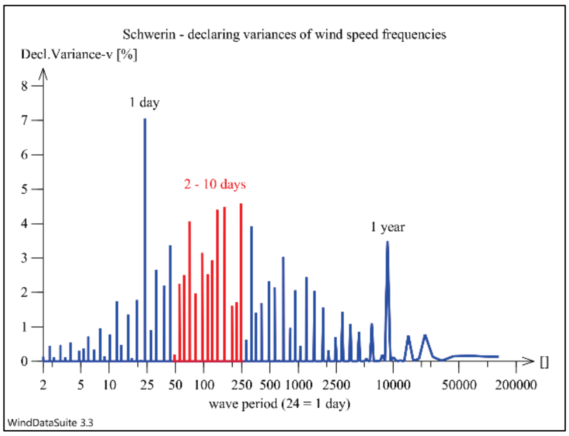

Fig.1: Declaring variances of the wind speed spectrum

The strongest signal is the diurnal wave which can be considered as a planetary wave caused by solar irradiation locally modulated by Earth's rotation. The diurnal wave is a nearly stationary wave when considering the locally limited area containing the locations of the wind park and the reference stations. Thus, for this diurnal signal strong correlations are expected, independent of location and time range.
Also a strong signal is the seasonal wave caused by Earth's planetary orbit. The seasonal wave can be considered as a stationary wave within the locally limited area. Thus, for this seasonal signal also strong correlations are expected, independent of location and time range.
All processes with periods below one day can be considered partly as seasonal modulations of the diurnal wave and partly as local processes with high frequencies where turbulence acts at the highest frequencies and disperses waves into stochastic decay.
The processes with periods between one day and one year are a superposition of physical processes which can be described by propagating waves and their interaction with each other and with the air density field, dependent of time range and location.
Especially the processes within the period band from 2 to 10 days are determined by highly interacting barotropic and baroclinic processes and describe the synoptic scale with its phenomena like frontogenesis, frontolysis, eddy genesis and decay, and the passing of the low- and high-pressure areas. These waves cannot be considered as stationary at all. So, on the one hand the synoptic scale contributes to a great extend to the total time series variance, and on the other hand the correlations between two locations will be quite weak due to the dispersion of the transient signals if not appropriately resolved.

2.2 Resolving the synoptic scale processes
The u and v components of the horizontal wind speed, in x- and y-direction in the Cartesian co-ordinate system, respectively, can be calculated directly from the wind speeds and the wind directions of the wind measurement. The time derivatives result from applying a forward-time differencing scheme such that

$$(5.x)\quad \frac{\partial u}{\partial t}(t) = \frac{u(t) - u(t-\Delta t)}{\Delta t} \quad , \; t = 0, \dots, N-1$$

$$(5.y)\quad \frac{\partial v}{\partial t}(t) = \frac{v(t) - v(t-\Delta t)}{\Delta t} \quad , \; t = 0, \dots, N-1$$

where t is the time step of the time series, Δt is the sampling rate, and N is the length of the time series. The boundary condition at $t = 0$ is $u, v(-\Delta t) = u, v(0)$.
The dynamics of the horizontal wind field is described comprehensively by the Navier-Stokes equations which are given by

$$(6.x)\quad \frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} - w\frac{\partial u}{\partial z} + fv - \frac{1}{\rho}\frac{\partial p}{\partial x} - F_x$$

$$(6.y)\quad \frac{\partial v}{\partial t} = -u\frac{\partial v}{\partial x} - v\frac{\partial v}{\partial y} - w\frac{\partial v}{\partial z} - fu - \frac{1}{\rho}\frac{\partial p}{\partial y} - F_y$$

where u and v are the horizontal wind velocity components, w is the vertical wind velocity component in z-direction, f is the Coriolis parameter, ρ is the air density, p is the air pressure, and F are friction terms (turbulence Reynolds stresses tensor).
$\frac{\partial u}{\partial t}$ and $\frac{\partial v}{\partial t}$ in equations (6) describe the total local temporal change of the wind field due to all physical processes, and that is exactly what the wind measurement has recorded when using equations (5). Indeed, equations (6) contain the synoptic scale processes, but all other high-frequency processes are also contained.
So, in order to resolve the synoptic scale processes, the horizontal wind velocity components u and v gained from the wind measurement have to be digitally low-pass filtered such that all signals with a wave period equal to and less than one day will be extinguished totally.
This digitally filtering is equivalent to neglecting all physical processes in equations (6) which yield the high-frequency wave signals, such that the processes according to the quasi-geostrophic equations will be represented now better by far by the remaining lower frequency signals.
Since an accurate signal separation is essential, the response function of an appropriate digital low-pass filter has to prove an almost extinguished Gibbs oscillation, a clean 50%-energy transmittance at the specified cut-off period, and a 0%-energy transmittance at the 1-day wave period already. The time derivatives $\partial u'/\partial t$ and $\partial v'/\partial t$ for the synoptic scale are then given by

$$(7.x)\quad \frac{\partial u'}{\partial t}(t) = \frac{u'(t) - u'(t-\Delta t)}{\Delta t} \quad , \; t = 0, \dots, N-1$$

$$(7.y)\quad \frac{\partial v'}{\partial t}(t) = \frac{v'(t) - v'(t-\Delta t)}{\Delta t} \quad , \; t = 0, \dots, N-1$$

where u' and v' are the digitally low-pass filtered horizontal wind speed components. In order to illustrate the effect of the digital low-pass filter, Fig.2 depicts a temporal clipping of the u and v components and of the digitally low-pass filtered components u' and v' of the Schwerin data.

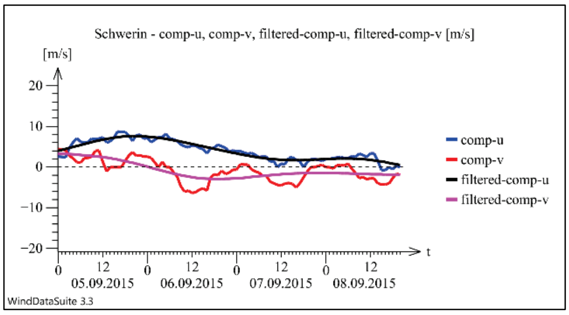

Fig.2: u and v and the digitally filtered u' and v'

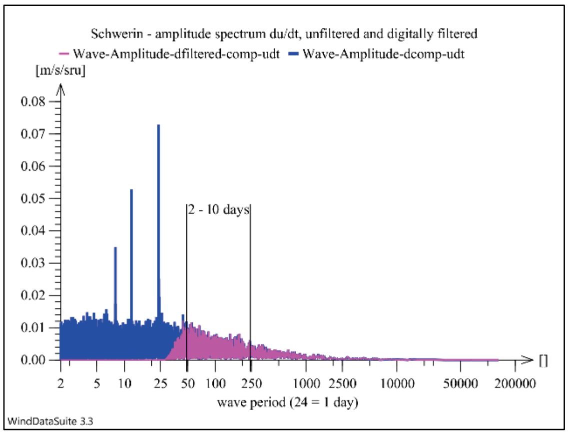

Fig.3: Amplitude spectrum of $\partial u/\partial t$ and $\partial u'/\partial t$

Fig.3 depicts the FFT amplitude spectrum of the $\partial u/\partial t$ values and of the digitally low-pass filtered values $\partial u'/\partial t$ of the Schwerin data. It can clearly be seen that in $\partial u'/\partial t$ all signals with a period below 25 hours have vanished, that the signals with a period above 48 hours equal the signals of the unfiltered $\partial u/\partial t$ values, and that especially the periods from 48 (2 days) to 240 (10 days) contain now most of the energy.

In order to resolve the synoptic scale signals for the set of regressions, so-called "synoptic scale clusters" with certain interval bounds will be searched for the values of $\partial u'/\partial t\,(t)$ and $\partial v'/\partial t\,(t)$ such that they span a 2-dimensional $M\ x\ M$ matrix where M may be any odd number greater than zero. The number of synoptic scale clusters N_{SSC} is then $N_{SSC} = M^2$. This is illustrated here for $M = 5, N_{SSC} = 25$ in Fig.4 where $U = \partial u'/\partial t\,(t)(t)$ and $V = \partial v'/\partial t$.

This $M\ x\ M$ matrix thus resolves the synoptic scale signals not only with respect to their magnitude but also with respect to their propagation direction which is essential for capturing the passing of low- and high-pressure areas.

The interval bounds $\varepsilon_{yk}, \varepsilon_{xk}\ , k = 1, \ldots, M-1$, with $\varepsilon_{yk}, \varepsilon_{xk} < 0\ \forall\ k < M/2$ and $\varepsilon_{yk}, \varepsilon_{xk} > 0\ \forall\ k > M/2$, of the synoptic scale cluster intervals will be determined preferably by quantiles such that the number of values in each synoptic scale cluster is approximately of the same size.

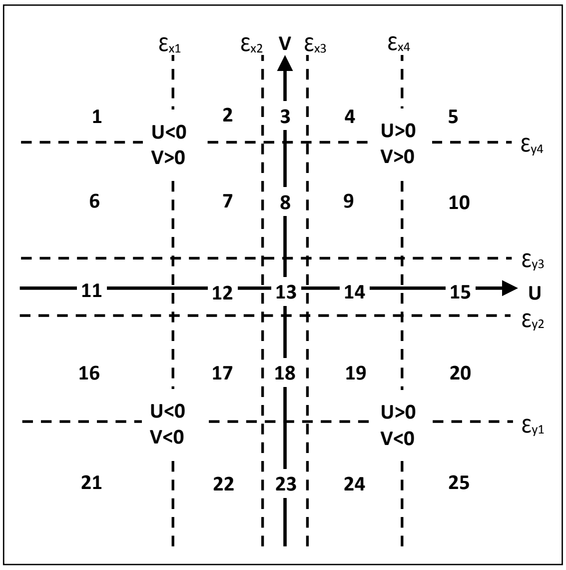

Fig.4: The 5 x 5 synoptic scale cluster matrix

For the MSSCP method, the synoptic scale clusters are needed only for the data of the reference stations (Ref locations), but are needed for the long-term time series as well as for the short-term time series which must have a time range common with the short-term time series data of the wind measurement conducted at the location WP. The synoptic scale clusters will be searched in such a way, that the interval bounds for the long-term time series are identical to those of the short-term time series and that each of the synoptic scale clusters is populated in the long-term time series as well as in the short-term time series. Thus, finally, it is possible to calculate separate regressions for each of the synoptic scale clusters and to use them adequately for the long-term extrapolation.

3. The MSSCP method

MSSCP calculates a set of linear regressions for the wind speeds. The extrapolations will be sampled in the 6-dimensional matrix $A^6 = \{H \times R \times C \times SRef \times VC \times SWP\}$ where $H_i, i = 1, \ldots, I$ are the selected measurement heights of the wind measurement at the WP location, $R_j, j = 1, \ldots, J$ are the reference stations to be used, $C_k, k = 1, \ldots, K$ are the synoptic scale clusters for each of the reference stations, $SRef_l, l = 1, \ldots, L$ are the wind direction sectors of the reference station data in each of the synoptic scale clusters, and $VC_m, m = 1, \ldots, M$ are the wind speed classes and $SWP_n, n = 1, \ldots, N$ are the wind direction sectors of the wind park data in each of the reference station wind direction sectors. For each element in the 4-dimensional matrix $A^4 = \{H \times R \times C \times SRef\}$, the optimal phase shift between the short-term wind speed data of the wind park and the reference station will be determined separately. For each element in the 5-dimensional matrix $A^5 = \{H \times R \times C \times SRef \times VC\}$, the regression parameters will be calculated separately. Thus, the number N_{LR} of linear regressions to be calculated (the size of the set) is given by $N_{LR} = I \times J \times K \times L \times M$. For each of the N_{LR} linear regressions only those wind speeds and wind directions of the

reference station data and of the wind park data will be used, which are fulfilling the respective conditions of the respective matrix element M_{ijklmn}.
If MSSCP is calculated with multiple reference station ($J > 1$), then the final combined extrapolation will be a weighted combination of the reference station specific results, where the distinct weighting factors are a combination of weights with respect to the distance from the wind park and/or the wind direction sector specific coefficients of determination and/or the wind direction sector specific location of the reference station relative to the wind park location.

4. MSSCP validation

4.1 Hindcasts and used data
Several long-term time series data of DWD (Deutscher Wetterdienst) weather stations (WSt) and long-term reanalysis data of respective nearby located MERRA-2 (Modern Era Retrospective-analysis for Research and Analysis - Version 2) points have been sampled into the groups (Merra2-286-302, WSt Bremen, WST Brake), (Merra2-277-307, WST München, WSt Ingolstadt), (Merra2-277-304, WSt Stötten), and (Merra2-287-306, WSt Schwerin, WST Boltenhagen). All data have a temporal resolution of 1 hour.
Long-term extrapolation hindcasts have been calculated such that within each group each of the stations served as the short-term "wind measurement" by extracting a time range of 1 year of the data while the other stations in the group served as the long-term references by extracting a time range of at least 20 years of the data. The long-term extrapolations (for one measurement height) have been calculated with synoptic scale clusters by using the matrix $A^{5+} = \{R \times C \times SRef \times VC \times SWP\}$ (see chapter 3) and without synoptic scale clusters by using the matrix $A^{4-} = \{R \times SRef \times VC \times SWP\}$. Thus, those error reductions can be determined by the hindcasts which result solely from incorporating the synoptic scale clusters into MSSCP.

4.2 Results
The relative errors in the wind power densities (WPD) resulting from the extrapolations are depicted in Fig.5 and are determined by

$$(8) \quad WPD = \frac{1}{2}\rho v^3$$

$$(9) \quad e_h^- = 100\frac{|WPD_h^- - WPD_h^m|}{WPD_h^m} \quad [\%] \quad , \quad h = 1, \dots, H$$

$$e_h^+ = 100\frac{|WPD_h^+ - WPD_h^m|}{WPD_h^m} \quad [\%] \quad , \quad h = 1, \dots, H$$

where ρ is the air density, v is the horizontal wind speed, e_h^- denotes the relative error resulting from hindcast run h without synoptic scale clusters, e_h^+ denotes the relative error resulting from hindcast run h with synoptic scale clusters, WPD_h^m is the mean WPD of the real measured long-term data, WPD_h^- is the mean WPD resulting from the long-term extrapolation without synoptic scale clusters, WPD_h^+ is the mean WPD resulting from the long-term extrapolation with synoptic scale clusters, and $h = 1, \dots, H$ is the number of the distinct hindcast run.

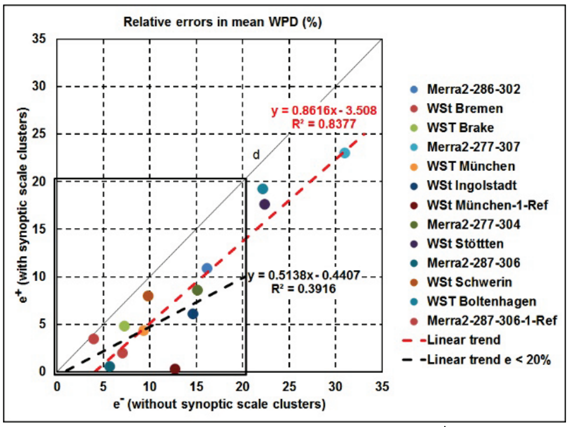

Fig.5: The relative errors e_r^- and e_r^+

Fig.5 shows that in all runs the WPD error has been reduced when extrapolating with synoptic scale clusters (all points are located below the diagonal d). When considering all runs, then a correlation seems to exist (coefficient of determination $r^2 = 0.83$), but this may be just an accident due to the three runs yielding relative errors greater than 20 %. These three runs should be considered as outliers since in general the correlations between the used data are obviously too bad. When considering only the runs yielding errors less than 20 % (the points within the black rectangle in Fig.5), then the correlation is quite weak (coefficient of determination $r^2 = 0.39$), thus indicating that most probably any systematic relation between the error reduction and the error itself does not exist. The mean error reduction is given by

$$(10) \quad \overline{red_e} = \frac{1}{H}\sum_{h=1}^{H} red_{e,h} = \frac{100}{H}\sum_{h=1}^{H}\frac{e_h^- - e_h^+}{e_h^-} \quad [\%]$$

When considering all runs then $\overline{red_e} = 45\,\%\ (min = 14\,\%, max = 98\,\%)$, when considering only the runs yielding errors less than 20 % then $\overline{red_e} = 53\,\%\ (min = 16\,\%, max = 98\,\%)$.

5. Conclusions

By incorporating so called "synoptic scale clusters" into the regression analysis, MSSCP improves the results of long-term extrapolations to a great extent.

Several extrapolation hindcasts have been performed yielding an error reduction in the wind power densities of approximately fifty percent on average when applying the synoptic scale clusters technique.

6. References

[1] Ritter Sheppard, C. J. (2009): Analysis of the Measure-Correlate-Predict Methodology for ´Wind Resource Assessment. Master Thesis at the Faculty of Environmental Resources Engineering at Humboldt State University, 2009, 118 pp.
[2] Rogers, A. L., Rogers, J. W., and Manwell, J. F. (2005): Comparison of the Performance of Four Measure-Correlate-Predict Algorithms. Journal of Wind Engineering and Industrial Aerodynamics, 93, 243-264, 2005.