
Sónia Liléo¹, Erik Berge², Ove Undheim², Rickard Klinkert¹ and Rolv E. Bredesen²

¹Kjeller Vindteknikk (KVT), Drottninggatan 71C, 111 36 Stockholm, Sweden
²Kjeller Vindteknikk (KVT), Gunnar Randers vei 24, Postboks 122, NO-2027 Kjeller, Norway

E-mail: sonia.lileo@vindteknikk.com, Tel:+46 737529574

Abstract

This paper presents some of the main results obtained in a comprehensive study performed on the long-term correction of wind measurements. Wind measurements are typically performed during a relatively short time period. Long-term data considered representative of the site's long-term wind conditions is therefore needed in order to estimate the long-term wind conditions at the site. This study shows that reanalysis mesoscale datasets with fine spatial resolution do not necessarily result in larger hourly correlation coefficients of the relationship to measured wind speed, at least in terrain with low complexity. MERRA reanalysis data (0.5 x 0.7 degree lat, lon) is seen to result, for the majority of the analyzed cases, in larger hourly correlation coefficients than other long-term datasets with finer spatial resolutions. This study also shows that the total uncertainty in the estimated long-term corrected wind speed is in average about 2 % varying normally up to about 4 % provided the following conditions: one year of site wind measurements with a data coverage after quality control filtering larger than 85 % is used; a reference long-term dataset with fine spatial resolution and with a hourly correlation coefficient of the relationship to the site wind measurements larger than 80 % is chosen; one of the commonly used long-term correction methods is applied; and finally, a reference period with a length of 15 to 20 years is used. The use of shorter or longer reference periods is seen to result in larger uncertainties. Using two years of site wind measurements has shown to slightly reduce the total uncertainty to about 1.8 %. The use of complete years of measurements is preferable, since an uneven representation of different months appears to increase the uncertainty. These values also include the uncertainty associated with the assumption of the past wind climate being a predictor of the future wind conditions. Note that possible changes in the future wind climate are here assumed to be of the same magnitude as possible changes occurred in the past wind climate.

1. Introduction

The description of the temporal variability of the wind conditions is essential when assessing the wind conditions at sites potentially suitable to wind power development. Wind measurements are typically performed during a relatively short time period (~1-3 years), that is commonly not representative of the long-term wind conditions at the site. Long-term correction of the wind measurements is therefore needed in order to estimate the expected long-term wind climate that best represents the site.

The main steps involved in the long-term correction of wind measurements are the following: performance of wind measurements at the site; choice of a long-term reference dataset; choice of a long-term correction method; choice of a past period to be used as representative of the future wind variations; and finally, estimate of the long-term corrected wind conditions and of the associated uncertainty. A research project co-financed partly by the Swedish research program Vindforsk III and partly by Kjeller Vindteknikk (KVT) has been conducted aiming to provide a collection of state-of-the-art knowledge on the long-term correction methodology. The results of this research project are presented in Liléo et al. (2013) [1]. This paper intends to present a summary of the main results obtained in this project. We refer to the full report for further details and discussions.

Proceedings of the EWEA conference 2013
2. Using reanalysis data to describe the local wind climate in terrain with low complexity

There are several types of long-term reference datasets. These may be reanalysis datasets or wind measurements from weather stations and from satellites. Only reanalysis datasets have been considered in this study. Results on the use of satellite wind data as long-term reference data may be found in Hasager et al. (2006) [2] and Harstveit et al. (2012) [3].

Reanalysis datasets may be categorized into two different groups: reanalysis global datasets and reanalysis mesoscale datasets. Reanalysis global datasets are produced using a constant data assimilation system that ingests worldwide observational data spanning a large time period back in time. The observational data used have a rather wide range of sources and correspond typically to different observation times and different spatial resolutions. When ingested by an assimilation system, the observational data are used as input to a Numerical Weather Prediction model (often referred to as a General Circulation Model, GCM) in order to create a description of the state of the atmosphere at uniformly spaced time instants and on an uniform horizontal grid covering the entire globe. The spatial resolution of these datasets are though typically rather coarse (coarser than a grid of 30-50 km side). Mesoscale numerical models may be used to downscale reanalysis global data into finer horizontal grids (typically 1 to 10 km side) and with finer temporal resolution (typically 1 hour). The resultant downscaled datasets are designated as mesoscale reanalysis datasets, or even virtual met masts or virtual time series.

2.1. Reanalysis global datasets

The main properties of different reanalysis global datasets developed during the last decades are summarized in Liléo et al. (2013) [1]. A description of each of the datasets and references to further information are also presented in that report. Moreover, a comparison of the temporal and spatial characteristics of the reanalyses R1, JRA-25, ERA-Interim, MERRA and CFSR has been performed and is discussed in Liléo et al. (2013) [1].

2.2. Reanalysis mesoscale datasets

There is a large number of different reanalysis mesoscale datasets available in the market. However, none of them are publicly available. For this reason, only the mesoscale datasets produced at KVT have been used in this study. These datasets are named WRF FNL and WRF ERA-Interim and are produced by KVT using the Weather Research and Forecast (WRF) mesoscale model (Skamarock et al. 2008 [5]) driven by FNL and by ERA-Interim data, respectively. WRF FNL data is available for the time period 2000 to the present with a temporal resolution of 1 hour and a horizontal resolution of 4 km x 4 km. WRF ERA-Interim data is available for the period 1992 to the present with 1 hour temporal resolution and 6 km x 6 km horizontal resolution. A more detailed description of these datasets may be found in Liléo et al. (2013) [1].

2.3. Strength of the relationship to wind measurements performed in terrain with low complexity

A database composed by data recorded at 24 met masts placed in sites potentially suitable for wind power development and at 18 masts belonging to meteorological stations, has been used in this study. These masts are located in Norway, Denmark and Sweden, in terrain with rather low complexity. The precise location of the masts is not presented in this report for confidentiality reasons. Data from the meteorological stations were retrieved through NCDC's Land-based Data webpage [6] for the period 2002 to 2009. Wind speed and direction data from each of the masts included in the database have been inspected. Erroneous data and data influenced by the formation of ice on the sensors have been removed. The measurements are from 10, 50, 80 and 100 m height and the measurement period varies between 1 and 8 years.

The strength of the relationship between the wind speed data measured at each of the masts included in our database and the wind speed data from the nearest located reanalysis grid point, has been measured by means of the Pearson correlation coefficient, R. Concurrent data
at the highest possible temporal resolution have been used, i.e., no averaging is involved. The term concurrent data is here used to designate data with identical time stamps.

Surface and geostrophic wind speed data from ERA-Interim, MERRA and CFSR/CFSv2, as well as surface wind data from WRF FNL and WRF ERA-Interim have been included in this analysis. The median values of the obtained correlation coefficients are plotted with blue bars in Figure 1. Each bar corresponds to a reanalysis dataset. The median values are also explicitly displayed in the figure. The lower and the upper whiskers mark the minimum and the maximum values of the obtained correlation coefficients, respectively. The lower and the upper edges of the white boxes mark the first and the third quartiles, respectively. Note that the quartiles divide the samples into four equally sized parts, and that the second quartile is the same as the median (shown with the blue bars and the displayed values). Within each white box is located half of the samples. Larger boxes represent a larger dispersion of the results. Such a box-and-whisker plot showing the minimum, the three quartiles and the maximum of the results is considered to adequately represent the distribution of the results, since the correlation coefficient has most likely a non-normal distribution (Gorsuch and Lehmann, 2010 [7]).

Figure 1. Box-and-whisker plot of the correlation coefficient (R) of the relationship between wind speed measurements and wind speed data from the nearest located reanalysis grid point. Data from the surface and from the geostrophic levels of different reanalyses are used. The spatial resolution of each dataset is given in the figure.

The results presented in Figure 1 show that the relationship between measured wind speed data and reanalysis geostrophic wind speed data (850 hPa level) is weaker than the relationship with reanalysis wind speed data from the surface level. This conclusion was expected since the weather patterns in the atmosphere are shifted in time with height and because the strength of the relationship given by the hourly correlation coefficient is related to simultaneity (i.e. simultaneous occurrence in time).

It is also seen that the relationship between measured and MERRA wind speed data is, for the majority of the analyzed cases, stronger than the relationship with the remaining reanalysis datasets. The larger correlation coefficients obtained for MERRA as compared to WRF FNL and WRF ERA-Interim suggests that a finer spatial resolution of the long-term reference data may not necessarily result in a larger hourly correlation coefficient, at least in terrain with low complexity. The correlation coefficient calculated based on hourly values is first and foremost a
measure of how well the short-term fluctuations in the reference and in the measured wind speed data agree in phase. The correlation coefficient does not measure how well the mean wind speed level of the reference data agrees with that of the measured wind. Modeled datasets with fine spatial resolution, such as WRF FNL and WRF ERA-Interim, may capture some properties of the local wind climate, such as the mean wind speed level, more accurately than datasets with coarser spatial resolution (e.g. MERRA). This property is of high relevance for wind resource mapping for example, but not as relevant in the long-term correction of wind measurements.

A further discussion is conducted in Liléo et al. (2013) [1] on the use of the correlation coefficient, based on hourly and monthly data, as a measure of the data's representativeness. By representativeness it's meant how well the long-term data represents the long-term wind conditions at a given site. The authors conclude that neither the hourly nor the monthly correlation coefficients are an ideal measure of the long-term data's representativeness. The definition of a more appropriate measure of representativeness should be further investigated.

3. Long-term correction methods

Different methodologies have been developed to long-term correct wind measurements. A common property of all of them is their main purpose: to describe the long-term wind climate at a specific site, based on short-term measurements performed at the site (short-term site measurements), and long-term data available from a representative location (long-term reference data). The main differences between the different methodologies consist on how the comparison between site and reference data is performed and on how the sought long-term wind climate at the site is calculated.

Table 1 presents commonly used long-term correction methods (LTC methods) grouped according to the methodology used in the comparison of the site and reference data. The different LTC methods may be grouped into two main categories: regression LTC methods and non-regression LTC methods. A short description of each of these methods and further references are presented in Liléo et al. (2013) [1].

<table>
<thead>
<tr>
<th>Category</th>
<th>Method's name</th>
<th>Developer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression LTC methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least squares regression</td>
<td>Regression MCP</td>
<td>EMD, WindPRO</td>
</tr>
<tr>
<td></td>
<td>Least Squares method</td>
<td>GL-GH, WindFarmer</td>
</tr>
<tr>
<td>Principal component regression</td>
<td>PCA method</td>
<td>GL-GH, WindFarmer</td>
</tr>
<tr>
<td>Quantile regression</td>
<td>U&amp;N method</td>
<td>KVT, internal use</td>
</tr>
<tr>
<td>Non-regression LTC methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear scaling methods</td>
<td>Weibull Scale MCP</td>
<td>EMD, WindPRO</td>
</tr>
<tr>
<td></td>
<td>Wind Index MCP</td>
<td>EMD, WindPRO</td>
</tr>
<tr>
<td></td>
<td>T&amp;N method</td>
<td>KVT, internal use</td>
</tr>
<tr>
<td></td>
<td>KH method</td>
<td>KVT, internal use</td>
</tr>
<tr>
<td>Probabilistic LTC methods</td>
<td>Matrix Method MCP</td>
<td>EMD, WindPRO</td>
</tr>
</tbody>
</table>

Table 1. Categorization of commonly used long-term correction methods.
3.1. Uncertainty associated with different LTC methods

A set composed by 16 different masts have been selected from the database presented in Section 2.3, to analyze the performance of the different LTC methods presented in Table 1. The selected masts fulfill the following criteria: more than 24 months of data; data coverage larger than 85%; hourly correlation coefficient (R) between site and reference data larger than 80%.

The main goal of this analysis is to estimate the error in the long-term corrected mean wind speed by testing the results from different LTC methods against a known result, i.e., performing a self-prediction test. The first year of measurements has been considered as the short-term period for each case, while the long-term period was defined as the maximum number (N) of complete years with data available within the measurement period for each of the selected masts. N ranges between 2 and 8 years for the selected masts. The reference dataset has been chosen as the data, among ERA-Interim, MERRA, CFSR/CFSv2, WRF FNL and WRF ERA-Interim, whose relationship to the site long-term data shows the highest correlation coefficient (calculated on a hourly basis). The percentage difference between the estimated and the measured long-term mean wind speeds has been defined as the prediction error. Figure 2 shows the absolute prediction error averaged over the 16 analyzed cases for each of the analyzed LTC methods. The error bars show one standard deviation from the mean value. The term "tr orig" means that the corresponding regression function was forced through the origin.

The results show that the absolute error obtained in the estimate of the long-term mean wind speed is in average about 1.5 to 2% for the tested LTC methods. The maximum standard deviation observed is about 1.8%, meaning that for about 68% of the analyzed masts the prediction error is at maximum about 4%. It is important to point out that the ranking shown in Figure 2 is not statistically significant, meaning by this that the ranking is susceptible to the set of site and reference data used in the analysis. Particularly, the correlation coefficient between the site and reference data used in the analysis is expected to affect the performance of the different methods (Liéo et al. 2013 [1]). The performance of a similar analysis based on cases
characterized by a low correlation coefficient would be of interest in order to evaluate the performance of the different models for such cases.

3.3. Dependence of the prediction error on the measurement period length

The database used in the previous section has also been used to investigate on the influence of the measurement period length on the error of the estimated long-term wind speed. The short-term period for each mast has been varied from 2 months to the maximum possible number of months within the long-term period. Steps of 2 months have been used. The long-term period is, as in Section 3.2., defined as the number of complete years (N) with data available within the measurement period for each of the selected masts. The prediction error (percentage difference between estimated and measured long-term mean wind speeds) has then been calculated for each of the considered cases. Figure 3 shows the mean absolute prediction error obtained using the in-house developed LTC methods KH, T&N and U&N.

![Figure 3](image)

Figure 3. Mean absolute prediction error as a function of the measurement period length. The results are based on data from 16 different masts with a hourly correlation coefficient between site and reference data larger than 80 %. The LTC methods used are the in-house developed KH, T&N and U&N methods.

The mean absolute prediction error is seen to strongly decrease with the increase of the measurement period length from 2 to 12 months. A prediction error of about 1.5 to 2 % is obtained using 1 year of site measurements with high data coverage, independently on the LTC method applied and provided a hourly correlation coefficient between measurements and reference data greater than 80 % (Figures 2 and 3). Increasing the length of the measurement period from 1 to 2 years is seen to reduce the prediction error to slightly over 1 % when using the KH and the T&N methods (Figure 3). Note also the increase in the prediction error when slightly more than 1 and than 2 years of measurements are used. This might be related to a bias introduced by the unequal representation of each month. The use of complete years of measurements is recommended. The accuracy of the U&N method appears to become lower than the accuracy of the KH and T&N methods for longer measurement periods. The reason for the lower performance of the U&N method (quantile regression method) for longer measurement periods should be further investigated.
4. Assuming the past as a predictor of the future wind conditions

Wind speed data from the 20CRv2 reanalysis surface level (Liléo et al. 2012 [1] and references therein) for the period 1920 to 2010 is used to estimate the uncertainty associated with the assumption that the future wind climate will vary in a similar way as in the past. Furthermore, this data is also used to analyze how the uncertainty in the long-term corrected wind depends on the choice of the reference period length. The 20CRv2 reanalysis dataset has been used in this study since it covers a long time period back in time. Figure 4 shows the annual wind index averaged over the region considered in this study (Norway, Denmark, Sweden, Finland and the Baltic countries), calculated based on the wind speed data from the 20CRv2 dataset. The annual wind index is here defined as the ratio between the annual mean wind speed and the wind speed averaged over the period 1920 to 2010. A year with a wind index larger/smaller than 100 % corresponds to a year with mean wind speed larger/smaller than normal. Figure 5 also shows the 5-year moving average of the annual wind index in red and the 5-year moving average of the North Atlantic Oscillation (NAO) index for the winter months (December, January, February and March) in dark blue. Wind speed data prior to 1920 is not considered in this analysis since the visual inspection of the data showed somewhat suspicious data prior to 1920 for some of the 20CRv2 grid points.

There is a clear relationship between the variations in the wind index calculated based on 20CRv2 data and in the NAO winter index, particularly during the period 1935 to 2010 (this issue is further discussed in Liléo et al. (2012) [1]). Note though that the NAO index has no defined periodicity, being impossible to predict how it will vary in the future. Figure 4 also shows that the period 1989 to 1995 was characterized by unusual high annual mean wind speeds associated with a large positive peak in the NAO index.

![Wind index averaged over the entire region](image)

**Figure 4.** The blue bars show the annual wind index averaged over the entire region considered in this study and based on wind speed data from the 20CRv2 reanalysis. The wind index relates the annual mean wind speed with the wind speed averaged over the period 1920 - 2010. The red curve shows the 5-year moving average of the annual wind index and the blue curve the 5-year moving average of the NAO index for the winter months (December, January, February and March).

4.1. Dependence of the prediction error on the reference period length
The following analysis concerns the evaluation of the error in the predicted future mean wind speed as a function of the length of the chosen reference period. The methodology used is the following: the length of the future period is set to 20 years (typical lifetime of a wind farm) and the period is allowed to move within the interval 1952 to 2010. The reference period is always chosen as adjacent (precedent) to the future period and is composed by consecutive years. The length of the reference period is allowed to vary between 1 and 30 years within the interval 1951 to 1989. Note that the reference period length could vary up to 39 years, but since the number of possible periods longer than 30 years is rather low, a maximum period length of 30 years is defined. The percentage difference between the mean wind speeds of the reference and future periods is calculated for each case and is designated as the prediction error. The average and the standard deviation of the absolute value of the prediction error, calculated over all the cases corresponding to a given reference period length, are then computed. This process was repeated using wind speed data from each of the 20CRv2 grid points located in the analyzed region. The relationship between the absolute prediction error and the reference period length obtained for each grid point was then averaged over all the grid points. The resultant mean curve is plotted in Figure 5. The dashed lines show one standard deviation below and above the mean curve.

Figure 5. Mean absolute prediction error averaged over all the 20CRv2 grid points located within the region considered in this study. The dashed lines show one standard deviation below and above the mean value.

Figure 5 shows that the mean absolute prediction error decreases significantly with the increase of the reference period length from 1 to 12-15 years. For longer reference periods the mean prediction error is about 1.5 % and is seen to normally vary between 0.8 and 2.3 %. The results suggest that a reference period of 15 to 20 years may be the most adequate. Longer reference periods are associated with a slightly larger standard deviation.

5. Inter-annual variability of the wind speed

Figure 6 shows the standard deviation of the annual mean wind speed based on MERRA data and on WRF ERA-Interim data for the period 1992-2011. The results show that the inter-annual variability of the wind speed is site specific varying between about 3 and 7 % in the analyzed region. The inter-annual variability of the wind speed should therefore be calculated specifically for each site.
Figure 6. Inter-annual variability of the wind speed based on the MERRA dataset (left panel) and on the WRF ERA-Interim dataset (right panel).

6. Conclusions in terms of guidelines on uncertainty reduction and on expected values for the uncertainty interval

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>Expected unc. in wind speed (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean value</td>
</tr>
<tr>
<td>Choice of the long-term reference data, of the LTC method and of the measurement period length.</td>
<td></td>
</tr>
<tr>
<td>1 year measurements</td>
<td>1.5</td>
</tr>
<tr>
<td>2 years measurements</td>
<td>1.0</td>
</tr>
<tr>
<td>3-4 years measurements</td>
<td>0.7</td>
</tr>
<tr>
<td>4-6 years measurements</td>
<td>0.5</td>
</tr>
<tr>
<td>Past used as a predictor of the future wind conditions assuming a reference period of 15-20 years.</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Total uncertainty</strong></td>
<td></td>
</tr>
<tr>
<td>1 year measurements</td>
<td>2.1</td>
</tr>
<tr>
<td>2 years measurements</td>
<td>1.8</td>
</tr>
<tr>
<td>3-4 years measurements</td>
<td>1.7</td>
</tr>
<tr>
<td>4-6 years measurements</td>
<td>1.6</td>
</tr>
</tbody>
</table>
The table above shows the expected mean value of the uncertainty in wind speed resultant from the long-term correction of wind measurements. The mean value plus one standard deviation is also shown. The combined contribution of the uncertainty sources "Choice of the long-term reference data", "Choice of the long-term correction method" and "Measurement period length" is presented as the first post. The second post is related to the uncertainty resultant from the assumption of the past being a predictor of the future wind conditions.

The results obtained in this study indicate an average uncertainty in the long-term corrected wind speed when using one year of site wind measurements of about 2 %, provided the following conditions: the data coverage of the measurements after quality control filtering is larger than 85 %; a reference dataset with fine spatial resolution and corresponding to a hourly correlation coefficient of the relationship between measured and reference data larger than 80 % is used; a reference period length of 15 to 20 years is chosen. These conditions are considered to favor the reduction of the uncertainty in the long-term corrected wind speed. The increase of the length of the measured time series from 1 to 2 years is seen to result in a small decrease of the total uncertainty from about 2 % to 1.8 %. Note that these results are based on data from sites in terrain with rather low complexity and that possible changes in the future wind climate are here assumed to be of the same magnitude as possible changes occurred in the past wind climate.

Besides the evaluation of the uncertainty associated with the long-term correction of wind measurements (Table 2), this study has also presented results on the uncertainty associated with inter-annual variability of the wind speed (Figure 6). The results have shown that the inter-annual variability of the wind speed is rather site specific varying between 3 and 7 % in the analyzed region and should therefore be calculated specifically for each site.

Acknowledgements

The authors would like to acknowledge ECMWF and NCEP for the development of the ERA-Interim and CFSR reanalysis projects, respectively, and the Global Modeling and Assimilation Office and the Goddard Earth Sciences Data and Information Services Center for the dissemination of MERRA data. The Twentieth Century Reanalysis Project (20CR) received support by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment program, Office of Biological and Environmental Research and by the National Oceanic and Atmospheric Administration Climate Program Office. This project was co-financed by the Swedish research program Vindforsk III.

Bibliography