ECN

Energy research Centre of the Netherlands

Estimation of Variability and Predictability of Large-scale Wind Energy in The Netherlands

M. Gibescu (Delft University of Technology)

A.J. Brand, ECN, Petten

W.L. Kling (Delft University of Technology)

This report has been published in Wind Energy,
Volume 12, Issue 3 (April 2009), pp. 241-260

ECN-W--09-011

WIND ENERGY
Wind Energ. (2008)
Published online in Wiley Interscience
(www.interscience.wiley.com) DOI: 10.1002/we.291

Estimation of Variability and Predictability of Large-scale Wind Energy in The Netherlands

Madeleine Gibescu*, Department of Electrical Power Engineering, Delft University of Technology, The Netherlands
Arno J. Brand, Wind Energy Unit, Energy Research Centre of The Netherlands
Wil L. Kling, Department of Electrical Power Engineering, Delft University of Technology, Delft, The Netherlands

Key words:
wind energy;
energy markets;
power system balancing;
programme-responsible parties;
wind fluctuations;
wind predictability

This paper presents a data-driven approach for estimating the degree of variability and predictability associated with large-scale wind energy production for a planned integration in a given geographical area, with an application to The Netherlands. A new method is presented for generating realistic time series of aggregated wind power realizations and forecasts. To this end, simultaneous wind speed time series—both actual and predicted—at planned wind farm locations are needed, but not always available. A 1-year data set of 10-min averaged wind speeds measured at several weather stations is used. The measurements are first transformed from sensor height to hub height, then spatially interpolated using multivariate normal theory, and finally averaged over the market resolution time interval. Day-ahead wind speed forecast time series are created from the atmospheric model HiRLAM (High Resolution Limited Area Model). Actual and forecasted wind speeds are passed through multi-turbine power curves and summed up to create time series of actual and forecasted wind power. Two insights are derived from the developed data set: the degree of long-term variability and the degree of predictability when Dutch wind energy production is aggregated at the national or at the market participant level. For a 7.8 GW installed wind power scenario, at the system level, the imbalance energy requirements due to wind variations across 15-min intervals are ±14% of the total installed capacity, while the imbalance due to forecast errors vary between 53% for down- and 56% for up-regulation. When aggregating at the market participant level, the balancing energy requirements are 2–3% higher. Copyright © 2008 John Wiley & Sons, Ltd.

Received 29 May 2007; Revised 23 May 2008; Accepted 21 July 2008

1. Introduction

As a consequence of the efforts to curb global warming, the penetration of emission-free renewables is increasing in modern power systems. Among renewable energy sources, wind energy holds the highest promise for large-scale integration, given the abundance of the resource and the relatively mature technology. The European Wind Energy Association expects wind to achieve about 20% share in the European Union (EU) energy pro-

* Correspondence to: M. Gibescu, Department of Electrical Power Engineering, Delft University of Technology, PO Box 5031, Delft, 2600GA, The Netherlands.
E-mail: m.gibescu@tudelft.nl
Contract/grant sponsor: Dutch Ministry of Economic Affairs.

duction by 2020. In The Netherlands—the case study for this work—permits for 6000–8000 MW of offshore wind power are under consideration, in addition to the existing ca 1.75 GW onshore. However, the variability and limited predictability of wind energy are expected to impact the required amounts of balancing (reserve and regulation) energy needed in the Dutch power system.

This work presents an estimate of the requirements for balancing energy at system and market participant level due to the integration of these large amounts of wind energy in The Netherlands. These requirements are calculated with a new method for generating realistic time series of aggregated wind energy production and forecast for system integration studies. A study on wind integration in The Netherlands, which focuses on the capabilities of conventional units to balance wind power, including up-/down-regulation and reserve provision, was performed by the Delft University of Technology in cooperation with the Dutch system operator TenneT and reported in Ummels *et al.*[1] (the study did not consider the need for wind energy curtailment due to transmission network constraints). This paper goes into a detailed presentation of how the wind-related input data for the above study was developed. As a conservative assumption, the imbalance due to wind forecasting errors is estimated based on day-ahead predictions only, which means that the potential for more accurate hour-ahead predictions as applicable to intra-day markets is not considered.

The organization of this paper is as follows. First, Section 2 gives a short review of the existing methods for generating wind energy time series for system integration studies. Subsequently, the new method to generate wind energy time series is introduced. This includes, in Section 3, the three transformations applied to measured wind speed data:

- from sensor height to hub height, using a slightly stable wind speed profile (with location-dependent average stability lengths) and the standard deviation (SD) as an estimate for the friction velocity[2];
- from measured to planned locations, by a spatio-temporal interpolation method that first estimates average wind patterns and then expresses the covariance between locations as a function of their distance (extending the work of Gibescu *et al.*[3]); and
- from 10- to 15-min averages, which is the market resolution interval for the Dutch balancing market.

Furthermore, in Section 4, the creation of wind speed forecast time series and in Section 5 the creation of wind power time series—both realized and forecasted—are presented. These are based on smoothed, park-aggregated power curves. Finally, in Section 6, the variability and degree of predictability of wind energy for the 2020 growth scenario in The Netherlands is quantified, and the resulting balancing requirements at system and market participant levels are derived based on a statistical analysis of the data.

2. Review of Existing Methods to Generate Wind Power Time Series

System integration studies require time series of measured and forecasted wind energy, with the wind power time series originating from wind speed time series. Methods using Weibull probability distributions of wind speed are not sufficient for such studies, as these can only give an idea of the expected energy production at a particular geographical location over long periods of time, such as total yearly output. Taking the spatial correlation of time-averaged wind speeds between multiple sites into account is required because wind farms will be concentrated in areas with favourable wind conditions; hence, their outputs may be strongly correlated. These cross-correlations are essential when assessing system-wide properties, such as the variability and predictability in large-scale wind production, which in turn affect the system requirements for reserve and regulation energy and, hence, the volumes cleared in the imbalance markets.

The methods introduced in existing system integration studies[4–8] fall short of using a principled approach for the creation of *both* measured and forecasted wind power time series for hypothetical wind energy growth scenarios.

These studies identify three methods to generate time series of actual wind data, namely:

- by using actually measured wind speed time series;
- by using synthesized wind time series data[5]; and
- by using a combination of measured and synthesized wind speed time series.[4,6,8]

DOI: 10.1002/we

In order to correctly account for the random nature—in particular, the spatial and temporal correlations—of wind in the area under consideration, measured wind speed time series must be used. Our method does make assumptions regarding the Markov property and the exponential decay of covariance with distance, but otherwise derives the relevant statistical properties of the interpolated series from the available data. In addition, in order to accurately model the balancing market in The Netherlands, 15-min or lower averaged wind speed data are used.

Regarding wind forecast data, there exist two methods to generate it, namely, by using real wind forecasts[7] and by using synthesized wind forecasts.[8,9] In order to correctly account for the limitations inherent in a particular forecasting method and for the degree of uncertainty in forecasted wind speed, real, rather than synthesized, wind forecasts must be used. Unlike the alternative, this method does not require assumptions on the distribution, correlation and increase with the look-ahead time of wind speed forecasting errors.

In summary, in this work, the measured data originate from actually measured wind speeds, and the forecasted data originate from real wind speed forecasts. Some preliminary results were presented in Gibescu and Brand.[10]

3. Wind Speed Measurements

3.1 Original Data Set

For the investigation of balancing requirements due to large-scale wind power, wind power production was modelled using weather data and multi-turbine power curves. First, wind speed data was obtained from the Royal Dutch Meteorological Institute (KNMI). The data set contains 10-min wind speed averages with a resolution of 0.1 m s^{-1} for 16 locations in The Netherlands and its coastal waters (six onshore, four coastal and six offshore; see Figure 1) measured between 31 May 2004 and 1 June 2005. In addition, associated standard deviations (SDs) are available for the onshore locations and can be estimated for the offshore locations. These will be useful for the height transformation performed in Section 3.2. The chosen time series reflects the spatial distribution of present and future installed wind power in The Netherlands. This historical wind speed data are employed to estimate the average daily patterns and to express the covariance between speeds at locations as

Figure 1. Onshore, coastal and offshore wind speed measurement sites

DOI: 10.1002/we

a function of distance between them. Then, for a new set of locations where wind farms are planned, wind time series for the 1-year study period are generated such that the spatial correlation between the sites at every point in time is taken into account.

To suppress heteroscedasticity in the wind speed data set—that is, increasing variance with the mean—we apply a variance-stabilizing transformation (Section 9.2 in Brockwell and Davis[11]), and from here onwards, we will work with wind speed's natural logarithm. This will simplify the regression model presented later in this section by allowing for a single, mean-independent variance of the interpolated logarithm wind speeds at different locations.

3.2 Height Transformation

The measured wind speeds generally refer to a sensor height, which may differ per location. To convert to a chosen (common) hub height, one can use the method provided in Walker and Jenkins[12] or in Brand.[2] They both use the logarithmic vertical wind speed profile to estimate wind speeds at heights other than the sensor height. Whereas in Walker and Jenkins[12] the local roughness length is used, which is difficult to determine for onshore locations, in Brand[2] this need is eliminated. Instead, two location-dependent parameters are used: friction velocity u_* and the *average* Monin–Obukhov (or stability) length L_{esti}. The sample wind speed SD provides an estimate of the friction velocity and originates from the measured wind speed data set. If this SD is not available *and* the location is offshore, the friction velocity can still be estimated directly from the vertical wind speed profile. Regarding the second parameter, estimating the actual stability length by its average is warranted by the high-resolution atmospheric model (HiRLAM) that calculates a positive average heat flux over the North Sea and over The Netherlands, implying that the average vertical wind speed profile *is* stable.[13]

Given $\mu(z_s)$ and $\sigma(z_s)$—the 10-min average wind speed and associated SD at sensor height z_s—then the wind speed and SD at hub height z_h can be calculated from:

$$\hat{\mu}(z_h) = \mu(z_s) + \sigma(z_s)\left(\ln\left(\frac{z_h}{z_s}\right) + 5\frac{z_h - z_s}{L_{\text{esti}}}\right) \tag{1}$$

and

$$\hat{\sigma}(z_h) = \sigma(z_s) \tag{2}$$

where L_{esti} is the location-dependent average stability length.

When only $\mu(z_s)$ is available, and provided that the location is offshore, we estimate the wind speed and SD at hub height by

$$\hat{\mu}(z_h) = \mu(z_s) + 2.5u_*\left(\ln\left(\frac{z_h}{z_s}\right) + 5\frac{z_h - z_s}{L_{\text{esti}}}\right) \tag{3}$$

and

$$\hat{\sigma}(z_h) = 2.5u_* \tag{4}$$

where u_* is computed according to the equation

$$\mu(z_s) - 2.5u_*\left(\ln\left(\frac{z_s g}{Au_*^2}\right) + 5\frac{z_s}{L_{\text{esti}}}\right) = 0 \tag{5}$$

and $g = 9.81$ m s^{-2} is the gravitational acceleration and $A = 0.011$ is Charnock's constant. Equations (3) to (5) use the factor 2.5, which is $1/\kappa$, where $\kappa = 0.4$ is the von Karman constant.

DOI: 10.1002/we

3.3 Wind Speed Spatial and Temporal Interpolation

Wind speeds are measured and recorded routinely at weather stations. This measurement data will be used to generate time series of wind speeds at other locations, such as (planned) wind farms. This section describes how wind speeds at locations of interest can be sampled conditionally on the wind speeds at measurement locations. We use a multivariate Gaussian model, where we make certain assumptions on the spatial and temporal covariance structure. The original method was first introduced by the authors,[3] and is extended here to include temporal correlations.

Let $W(x, t)$ be the log wind speed at a location x in Cartesian coordinates and time $t = (d, k)$ defined by the day of the year d and the time of day k. The reason for taking the logarithm is twofold. As already mentioned, there is a pronounced heteroscedasticity in the raw wind speeds, which is stabilized by the log transformation. The second reason is even more compelling: it is well known that the Weibull provides the best fit for the distribution of raw wind speeds. Unfortunately, the Weibull distribution has some practical drawbacks. For instance, there does not seem to be a unique and natural extension of the univariate Weibull to the multivariate case. The log-normal distribution has a superficial resemblance to the Weibull, but it does not have any of its drawbacks. Indeed, upon taking logarithms, we are in the (multivariate) normal case, which allows us to make extensive use of conditioning. Let us briefly review the mathematical statistics background, which may be found in Brockwell and Davis.[11]

Consider a random vector $\mathbf{X}$, which is distributed according to the multivariate normal distribution with mean vector μ and covariance matrix $\mathbf{\Sigma}$. Suppose we partition $\mathbf{X}$ into two sub-vectors, where $^{(1)}$ corresponds to the sampled data and $^{(2)}$ to the observed data set:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}$$

Correspondingly, we partition the mean vector and covariance matrix into

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \text{ and } \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$$

Now, if $\det(\Sigma_{22}) > 0$, then the conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)}$ is again multivariate normal. The conditional mean is

$$\boldsymbol{\mu}^{(1)} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\left(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)}\right) \tag{6}$$

and the conditional covariance matrix is

$$\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} \tag{7}$$

Now, let us return to the log wind speeds $W(x, t)$ at location x and time $t = (d, k)$. We propose the following model:

$$W(x, t) = \mu(x, k) + \varepsilon(x, t) \tag{8}$$

where μ is a deterministic function representing the daily wind pattern by location and ε is a zero-mean random process representing the variations around the mean. Note that we are assuming that μ depends on $t = (d, k)$ only through the time of day k. In other words, our model does not include any seasonal effects. Time series analysis tools can be used to find any other trends and periodic components that may be present in the data. For our 1-year data set, the seasonal component was checked and found that its inclusion did not substantially improve the accuracy of the interpolation.

In Figure 2 we plot the average daily wind pattern for each of the 16 measured locations. The lower curves were found to correspond to locations on land and the higher curves to those out at sea. Hence, the plot suggests that the model should have a daily effect that varies smoothly with geographical location. Measurement sites on land display a typical unimodal pattern with a maximum around midday, while sites offshore have a much flatter daily profile, with a higher overall average. The coastal locations fall somewhere in between.

DOI: 10.1002/we

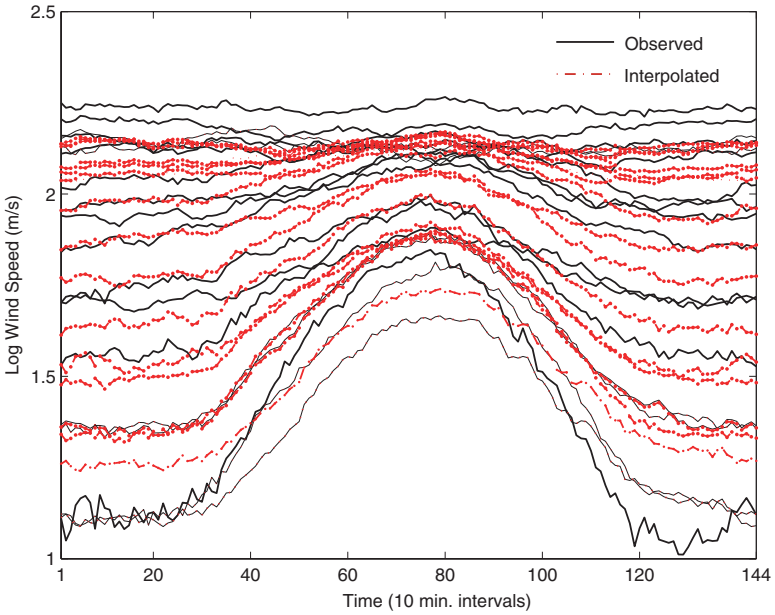

Figure 2. Daily wind speed pattern for measured and interpolated sites

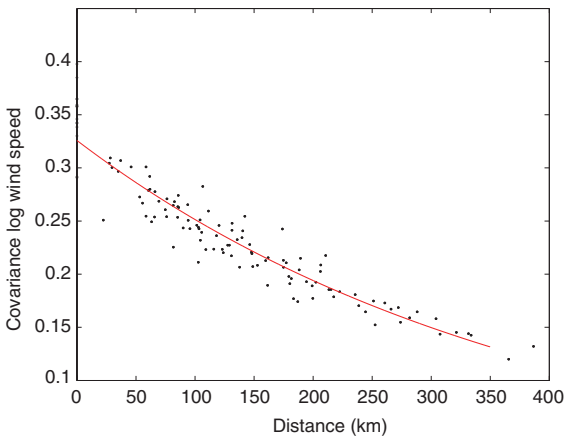

Figure 3. Wind speed covariance versus distance for 16 measurement sites in the Netherlands

We can estimate the mean log wind speed $\mu(x, k)$ at all measurement locations by the daily averages shown in Figure 2. To obtain estimates for the locations of interest within the convex hull formed by the measurement sites, linear spatial interpolation was used. For locations outside that hull, nearest neighbour interpolation was used. The results are shown as dotted lines in Figure 2.

A model for the random part $\varepsilon(x, t)$ is now required. As explained above, a zero-mean, multivariate normal distribution is assumed for the log wind speeds minus the daily pattern. We must pay special attention to the covariance structure of this process because it is certainly not realistic to assume that the variations of wind speed about the daily mean are independent across space or time. This is especially true for the relatively small region where the Dutch wind farms are operated.

In Figure 3, we plot the sample covariance between the log wind speeds at all pairs of (measurement) locations versus the distance between them. Looking at the plot, and taking into account that the covariance should vanish at very large distances, it seems reasonable to assume an exponential decay with distance. This automatically implies that the resulting covariance matrix is positive definite. We propose

DOI: 10.1002/we

$$\mathrm{Cov}(\varepsilon(x_i,t),\varepsilon(x_j,t)) = \alpha_0 \mathrm{e}^{-\beta\|x_i - x_j\|} \tag{9}$$

where $\|x_i - x_j\|$ is the Euclidean distance between the two locations. To be able to sample wind speed time series, we also need to take the temporal dependence into account. Similar to equation (9), we propose

$$\mathrm{Cov}(\varepsilon(x_i,t),\varepsilon(x_j,t-1)) = \alpha_1 \mathrm{e}^{-\beta\|x_i - x_j\|} \tag{10}$$

We can jointly estimate the parameters α_0, α_1 and β by a least squares fit. The fit for α_0 and β is shown in Figure 3, where $\hat{\alpha} = 0.32$ and $1/\hat{\beta} = 392.36$ (km). The latter term is also known as the *characteristic distance* or *decay parameter*. If we translate the parameters of this decay fit from logarithmic to pure wind speeds, and look at correlation coefficients (i.e. covariance normalized by the product of the two SDs) between location pairs, we obtain a value of 610 km for the characteristic distance. This value is in line with the 723 km reported in Chapter 6 of Giebel,[4] which is based on measurements from 60 locations spread throughout the EU, and the 500 km reported in Landberg *et al.*[14] and Holttinen,[6] using Danish only and Scandinavian data, respectively. This suggests that these values are generic.

Finally, the Markov property for the sampled time series is assumed: conditionally on $W(x, t-1)$, $W(x, t)$ does not depend on $W(x, t-2)$, $W(x, t-3)$, . . . etc. Consequently, we do not need to specify the covariance between $W(x_i, t)$ and $W(x_j, s)$ when $|s - t| > 1$.

We should point out that the equations (9) and (10) do not depend on the time. Thus, any daily or seasonal changes in the covariance structure are ignored. We did try to identify such effects, but found that they were not very large, and not particularly systematic; hence, they would not have a substantial effect on the time series that the method ultimately generates.

Now, everything is in place to describe the proposed sampling scheme. At each stage, we sample a collection of normal random variables, conditional on some other normal random variables. The mean and covariance structure of all random variables involved is fully described, and therefore we can use the general theory from equations (6) and (7), where subset (1) denotes the unobserved wind speeds at time t, and subset (2) denotes both observed speeds at times t and $t-1$, and unobserved, but already interpolated values at time $t-1$.

1. Conditionally on $\{W(x, 1)$: x observed$\}$ sample $\{W(x, 1)$: x unobserved$\}$
2. Conditionally on $\{W(x, 1), W(x, 2)$: x observed$\}$ and $\{W(x, 1)$: x sampled at step 1$\}$, sample $\{W(x, 2)$: x unobserved$\}$

   . . .

t. Conditionally on $\{W(x, \mathrm{t}-1), W(x, t)$: x observed$\}$ and $\{W(x, \mathrm{t}-1)$: x sampled at step $t-1\}$, sample $\{W(x, t)$: x unobserved$\}$

Once the log wind speeds for the locations of interest are sampled, we simply exponentiate to obtain the wind speeds themselves. Of course, the time series produced in this way will reflect the assumptions we have made. However, this does *not* mean that they will look like samples from the multivariate log-normal distribution. Our method provides nothing more than linear interpolations of the measured time series, and so their Weibull character will be preserved to a great extent. To evaluate the effectiveness of our method, we use cross-validation, leaving one measurement location out of the data set and using the remaining $n-1$ locations to 're-create' it.

First, we want to verify that our method preserves the marginal Weibull parameters. As an illustration, Figure 4 shows the histogram of the original data for a coastal location in The Netherlands (IJmuiden) together with a Weibull fit of the original and interpolated data. As expected, some smoothing has occurred in the interpolated data due to the weighted averaging, but not much.

Another interesting point to verify is whether or not our method reproduces the (auto)covariance structure of the original data. Figure 5 shows the lag-one autocorrelations for the original and cross-validated data, with the straight line indicating a perfect match. Even though some over- and underestimation of the autocovariances can be observed from Figure 5, there does not seem to be any structural bias.

DOI: 10.1002/we

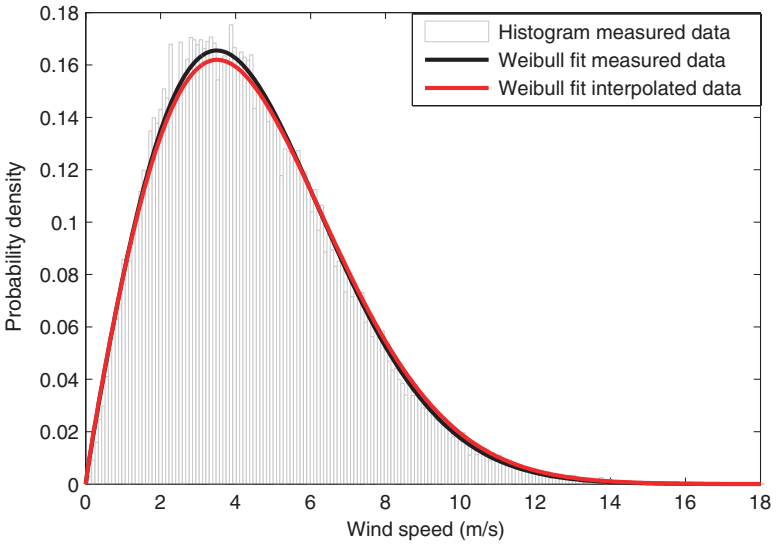

Figure 4. Wind speed histogram and fit to Weibull at IJmuiden

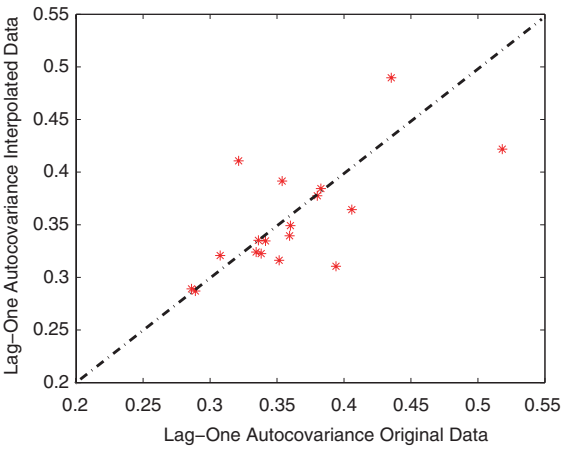

Figure 5. Lag-one autocovariance, original versus interpolated data

In terms of the limitations of this method, it should be kept in mind that the interpolation weights are determined by the assumption of the exponential decay of the covariance with distance. Thus, if this decay does not hold, then the covariance structure of the generated series will not be correct. Naturally, the estimated time series are only as good as the input data allows. For instance, under complex terrain, measured data at closer distances would be required to correctly track local changes in wind behaviour.

3.4 Averaging over the Market Resolution Interval

The change from 10- to 15-min averages was required by the design of the Dutch balancing market and can be easily accomplished as shown below.

Given that $\mu(k)$, $\mu(k+1)$, $\mu(k+2)$, . . . are the consecutive 10-min wind speed averages, we compute $m(k)$, $m(\mathrm{k}+1)$, . . . , which are the consecutive 15-min wind speed averages:

DOI: 10.1002/we

$$m(k) = \frac{2}{3}\mu\left(\frac{3(k-1)}{2}+1\right)+\frac{1}{3}\mu\left(\frac{3(k-1)}{2}+2\right)$$

$$m(k+1) = \frac{1}{3}\mu\left(\frac{3(k-1)}{2}+2\right)+\frac{2}{3}\mu\left(\frac{3(k-1)}{2}+3\right)$$

4. Wind Speed Forecasts

The 15-min average wind speed forecast time series are generated for a subset of locations where measurements are also available. These day-ahead forecasts originate from the wind power forecasting method AVDE,[15] a physical forecasting method with an output statistics module. In an operational sense, AVDE is a post-processor to the high-resolution atmospheric model HiRLAM or any weather prediction model that delivers the required input data (two horizontal wind speed components, temperature and pressure in two vertical levels on a horizontal grid covering the sites to be considered) in the required format (GRIB). If wind speed and/or wind power realizations are available, the output statistics module of the AVDE can be used in order to compensate for systematic errors in the forecasts. The forecasts are meant to guide wind producers in a day-ahead market, and are completed at 12:00 the previous day, thus carrying an increasing delay of 12–36 h. Using a method similar to the one used for the spatial interpolation of wind speed measurements, appropriately correlated forecast error time series are then generated for the planned wind farm locations. Since we do not analyse the variability of wind forecast errors over *successive* time intervals, we assume that, conditional on the forecast errors at the observed locations, the forecast errors at the computed locations at time t are independent of the errors experienced at time $t-1$.

In Figure 6, we present the geographical locations of the seven wind speed forecast sites (labelled), together with the projected offshore wind farm locations for the year 2020 (black stars) and, in shades of grey, the current density of onshore wind energy capacity by province in The Netherlands.

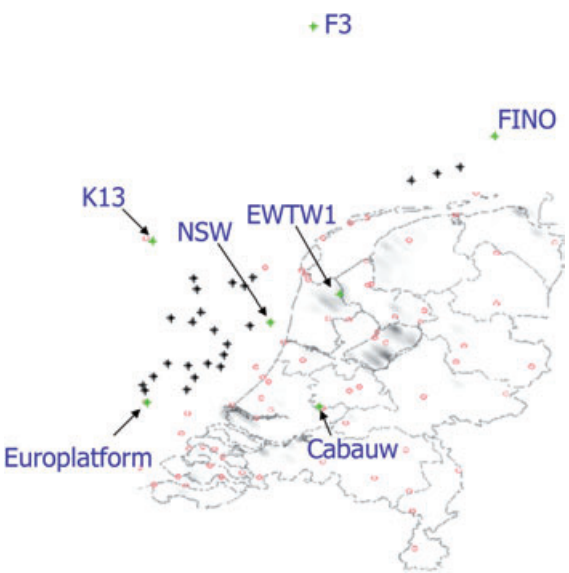

Figure 6. Wind speed forecast sites (labelled), onshore (shaded grey) and offshore (black stars) wind farm sites for the Basic 2020 scenario

DOI: 10.1002/we

Söder[9] addresses the problem of generating future time series of forecasting errors when only past observations are available. This is quite different from our problem of generating forecasting errors at certain locations, given those at other locations for the same moment in time. Similar to the interpolation model for wind speeds, we model the forecast error as the sum between a deterministic term, derived from the average daily pattern (Figure 7), and a random term, which obeys a covariance matrix derived from the exponential fit presented in Figure 8. Note that the logarithmic transformation was not necessary in the case of interpolating forecast errors, as the variance of the error does not significantly increase with its mean. In order to correctly take into account the changes in the covariance structure due to the look-ahead time, $24 \times 4 = 96$ separate exponential decay curves were fitted as shown in Figure 8.

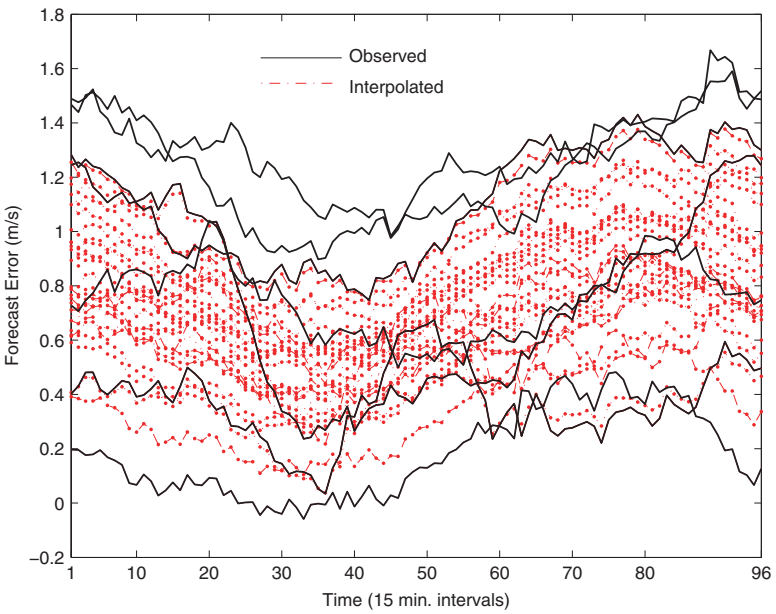

Figure 7. Daily wind speed forecast error pattern for measured and interpolated sites

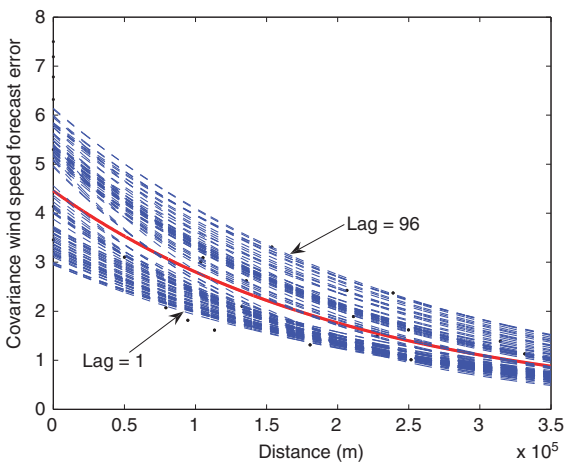

Figure 8. Wind speed forecast error covariance versus distance for various forecast horizons

DOI: 10.1002/we

5. Conversion to Wind Power Output

5.1 Multi-Turbine Power Curves

For each of the projected locations, a wind farm aggregated model has been created using regionally averaged power curves, which depend on the area of the farm and the SD of the wind speed distribution at the given farm's location. As the name suggests, regional averaging provides the average power of a set of wind turbines placed in an area where the wind climate is known, assuming the turbines do not affect each other. The multi-turbine curve is created by applying a Gaussian filter to a single-turbine power curve, and is not to be confused with a wind farm power curve, which brings the wind shadow of turbines into account.

Although inspired by and having the same effect as the Gaussian filter in the multi-turbine approach of Norgard and Holttinen,[16] the SD in the new filter correctly originates from the local wind climate alone. Unlike the Norgard–Holttinen method, our filter does not require estimating the turbulence intensity, which incidentally is a measure of variation in a 10-min period in a given location rather than a measure of variation in the same 10-min period at different locations. Nor does our method apply a moving block average to the wind speed time series with the time slot arbitrarily based on the local average wind speed.

An example of a multi-turbine power curve constructed for an offshore wind park of installed power 405 MW, at a location where the SD of the wind speed is 4.6 m s^{-1} is shown in Figure 9.

The width σ_F of the Gaussian filter is given by an estimate for the SD that describes the regional variation of wind speeds at different locations in the same wind climate (Appendix A):

$$\sigma_F = \sigma\sqrt{\frac{1}{2}\left(1 - e^{-d_{\text{ave}}/D_{\text{decay}}}\right)} \tag{11}$$

where σ is the SD of the wind speed distribution, d_{ave} is the average distance between the locations and D_{decay} is the characteristic distance of the decay of correlation (as estimated in Section 3). If the individual locations are not known, as is the case in this study, an estimate for d_{ave} is (see Appendix B):

$$d_{\text{ave}} = \frac{2}{3}\sqrt{\frac{A}{\pi}\left(1 + \frac{2}{\sqrt{M}}\right)} \tag{12}$$

where A is the area of the region and M is the number of locations in that area. In this study, the area relates to a province for the onshore wind power and to an individual wind farm for the offshore wind power. The area of an individual farm is approximated by the area of a rectangle whose sides depend on the number of turbines, the rotor diameter and the spacing between turbines.

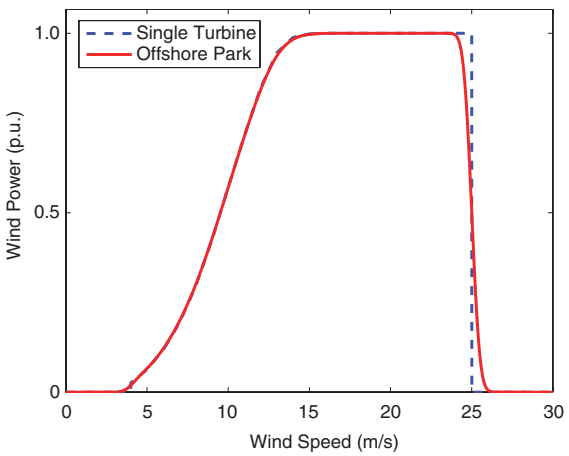

Figure 9. Example of an aggregated wind speed to power conversion curve

DOI: 10.1002/we

The method to determine the regional variation of wind speeds at different locations in the same wind climate was verified by using the measured data introduced in Section 3. The method to determine the multi-turbine power curve for a given area is still in need of verification data.

5.2 Wind Energy Growth Scenarios in The Netherlands

Our method uses offshore wind growth scenarios, consistent with the renewable policy goals in The Netherlands over the next 15 years. Based on these rough estimates, on the current onshore wind farm placement, and the pending applications for environmental permits for offshore wind farms, the most likely locations and installed capacities are chosen for the years 2010, 2015 and 2020. Three offshore wind energy growth scenarios were created for this project: Low, Basic and Advanced. These are summarized in Table I, rounded to multiples of 10 MW. Only one scenario was created for the growth of onshore wind installed capacity.

5.3 Levels of Aggregation

The power of the individual wind farms is first aggregated at the system level. This gives a good initial estimate for the degrees of variability and predictability that come with large-scale wind energy; however, it ignores the real situation where wind power is integrated by several sub-levels, as owned and operated by the individual market parties. Note that in The Netherlands, wind power is fully integrated in the day-ahead and imbalance market structures. Market participants known as programme responsible parties (PRPs), governing a portfolio consisting of both renewable and conventional energy resources, will submit to the Transmission System Operator (TSO) balanced schedules for energy delivered to and absorbed from the system during a 15-min interval, or programme time unit (PTU). This arrangement provides some insulation from the full exposure to imbalance charges for the wind producer, as conventional units in the PRP's portfolio may act to correct energy programme deviations due to wind variability and unpredictability. To that effect, we define seven PRPs, each owning a unique combination of installed powers and geographical spread of onshore and offshore wind farms, as described in Table II. For reasons of confidentiality, we gave these parties fictitious names; however, the installed powers are consistent with the current and planned developments of actual businesses in The Netherlands.

Table I. Scenarios for offshore and onshore wind capacity (MW)

	Year		
	2010	2015	2020
Low Offshore	720	2010	3800
Basic Offshore	1180	3110	6030
Advanced Offshore	1520	4110	8000
Onshore	1750	1800	1800

Table II. PRPs, Basic 2020 growth scenario

PRP	Offshore (MW)	Onshore (MW)	Total (MW)
Anton	881	840	1721
Berta	1792	593	2385
Cesar	800	0	800
Dora	2520	140	2660
Emil	40	0	40
Friedrich	0	92	92
Gustav	0	135	135
System	6033	1800	7833

DOI: 10.1002/we

6. Results

6.1 Overview

Two insights are derived from the developed data set, namely, the degree of variability and the degree of predictability to be expected in the nationwide and producer-level wind energy production, for a given scenario, in the course of the year under study. These two attributes are statistically analysed in terms of mean, SD and confidence intervals (CIs). The 99.7% CI derived from the data (and not from analytical approximations to the empirical distributions) was chosen, as this would mean that reported values may be exceeded in no more than approximately 1 day per year. Subsequently, requirements for the amount of positive and negative imbalance energy that needs to be dispatched by the TSO, or by each of the producers, to cover deficit or excess of production due to wind variability and wind forecasting errors are estimated.

6.2 Interpolation Results

An example of the first-week speed profile for a number of measured and estimated sites is shown in Figure 10. The top subplot shows a week's worth of measured wind speeds for two locations selected from the measurement set, which are the *most* and the *least* correlated with the two computed time series shown in the bottom subplot. This picture illustrates the idea of simultaneous variation among highly correlated sites.

Since converting wind speed to power involves a non-linear transformation of the data through the power curve introduced in Section 5.1, the following simulation steps have to be taken to obtain the expected wind power output per time point per location:

- sample from the distribution of wind speeds using equations (6) and (7), mean, and covariance derived from the measurement data;
- sample from the distribution of wind speed forecast errors using equations (6) and (7), mean, and covariance derived from the forecast data;
- pass the sampled wind speed through the park-aggregated power curve for a given location;
- add the sampled wind speed to the sampled forecast error;

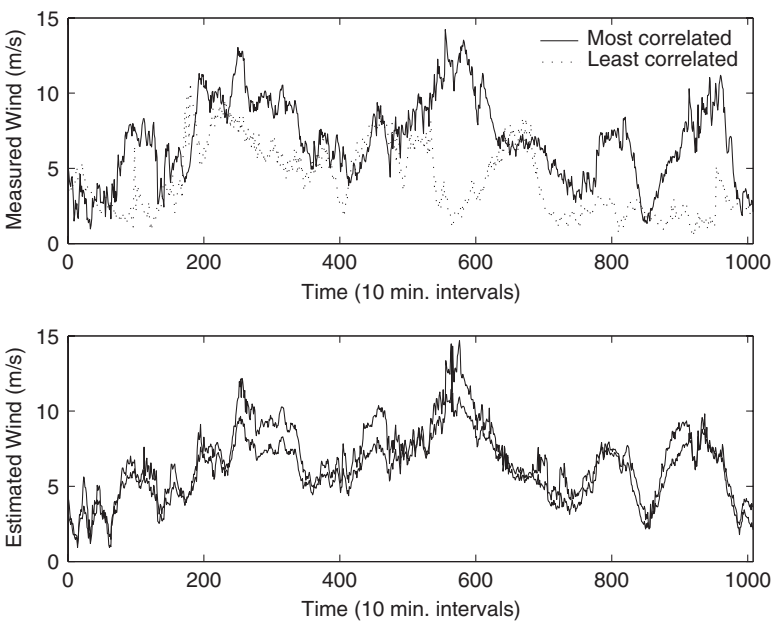

Figure 10. Example measured and estimated wind speed profiles for 1 week

DOI: 10.1002/we

- pass the resulting wind speed forecast through the park-aggregated power curve; and
- the average over the sample set of the wind power and, respectively, of the wind power forecast from this simulation is the best estimate.

As an example, the first-week time series of produced and forecasted wind powers, aggregated at the system level, are shown in Figure 11 for the Basic 2020 scenario.

6.3 Estimation of Wind Variability

Given the locations and installed powers for future wind farms, the estimation method of the previous sections can be used in combination with the aggregated power curve presented in Section 5.1 to compute the system-wide average wind power generated per 15-min time interval for the duration of an entire year. The total installed wind power chosen for the simulations correspond to the sum of the onshore growth and the 'Basic' growth offshore scenario for year 2020. By differentiating the wind power production vector, we obtain an estimate of the variability of aggregated production across 15-min time intervals and above.

This quantity and its sign are of special interest, as simultaneous load and wind variations are to be balanced by the remaining conventional generation units via the up- or down-ramping of their outputs.

Table III presents the 99.7% CIs and the extreme values (smallest and largest) of the 15-min, 30-min, 1-h and 6-h variations at the system level by the year 2020 for the Basic scenario (7830 MW). The sorted positive and negative variations in wind production over various time ranges are also shown in Figure 12 for the same scenario. Based on the 99.7% CI, the system-wide variations across 15-min intervals are in the range of ±14% of the installed power for this scenario.

Table IV shows the statistics of the 15-min variations for each of the seven PRPs individually. These variations are in the range of ±12–26% of the power installed by the PRP, depending on the geographical spread of its owned locations. The collective requirement for balancing 15-min variations becomes approximately ±16% of the system's installed power, which is 2% more than the requirement at the system level.

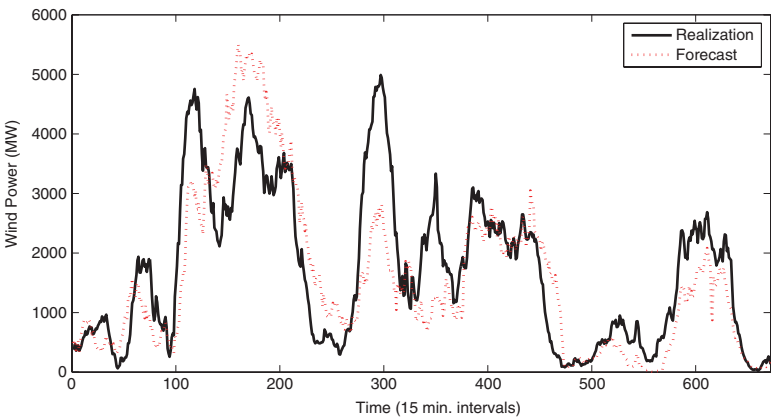

Figure 11. Produced and forecasted wind power time series for 1 week, 7830 MW installed wind power

Table III. Statistics of wind variability, Basic 2020 scenario

Time range	Minimum (MW)	Maximum (MW)	99.7% CI (MW)
15 min	−2411	2883	−1090.8 to 1054.2
30 min	−2411	2883	−1252.9 to 1309.6
1 h	−3133	3634	−1968.0 to 1846.0
6 h	−7211	6790	−5157.8 to 5105.4

DOI: 10.1002/we

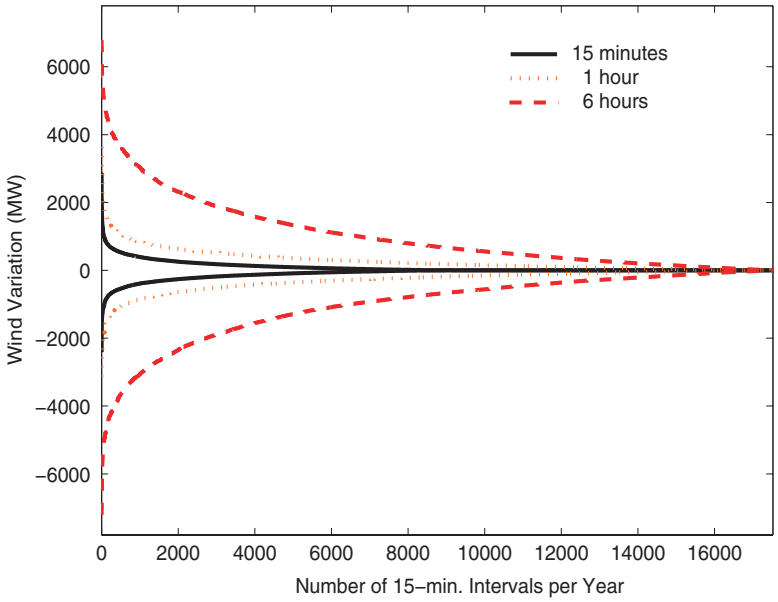

Figure 12. Variations in 7830 MW aggregated wind power

Table IV. Statistics of 15-min variability, PRP level

PRP	Minimum (MW)	Maximum (MW)	99.7% CI (MW)
Anton	−367	481	−201 to 197
Berta	−755	825	−323 to 327
Cesar	−482	448	−191 to 188
Dora	−1264	1420	−461 to 484
Emil	−36	38	−9 to 9
Friedrich	−48	42	−22 to 24
Gustav	−74	78	−26 to 27
Total PRPs	−3026	3332	−1233 to 1256
System	−2411	2883	−1091 to 1054

6.4 Estimation of Wind Predictability

System reserve is allocated among online generators to account for equipment outages and uncertainties in load and wind forecast errors. It is quite obvious that the higher the forecast uncertainty, the larger the amount of reserve needed to achieve the same reliability level. In this subsection, we perform a statistical analysis of the imbalance error aggregated over the entire wind production of The Netherlands. In Figure 13, we present the normalized histogram for the system-aggregated forecast error corresponding to the Basic 2020 scenario, together with the fit to a double-exponential probability density function, which was found to be a more accurate analytical representation of the data than the normal distribution.

Table V presents the 99.7% CI plus the average, SD, minimum and maximum of the imbalance (equal to the wind power forecast error) by 2020. Statistics for the positive and negative imbalance are presented in the second and third rows of Table V, respectively. Based on the 99.7% CI, in the scenario with 7830 MW of wind power by 2020, the positive (up-regulation or reserve) balancing energy requirement is about 56%, and the negative (down-regulation) requirement is about 53% of the installed capacity.

Table VI shows the predictability at the PRP level. Balancing energy requirements for an individual PRP are in the range 45–82% of its installed capacity for up-regulation or reserve, and 46–72% for down-regulation.

DOI: 10.1002/we

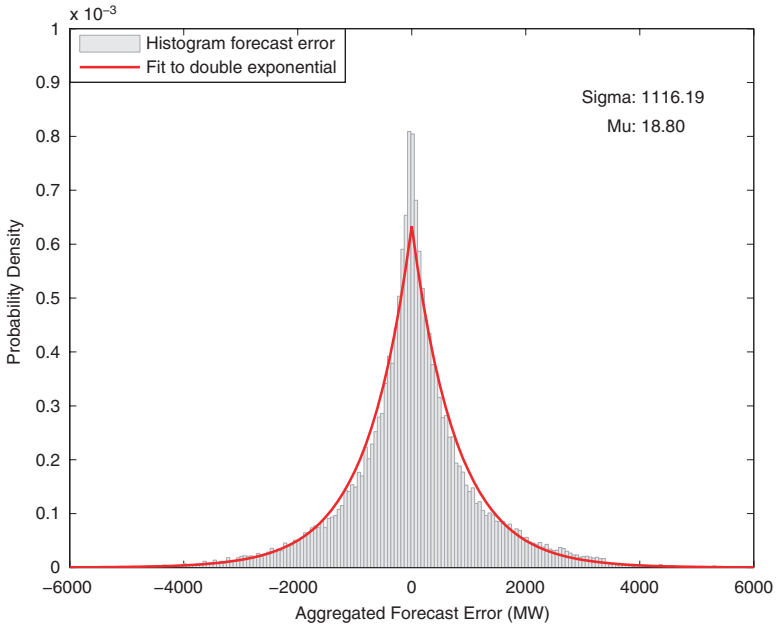

Figure 13. Aggregated forecast error histogram and probability density function for 7830 MW installed wind power

Table V. Statistics of wind predictability (MW), Basic 2020 scenario

Imbalance	Minimum	Maximum	99.7% CI	Mean	SD
Total	−5366	5692	−4112.9 to 4370.6	18.8	1116.2
Positive	0	5692	1.2 to 4765.2	789.8	821.1
Negative	−5366	0	−4471.8 to −1.0	−754.7	790.5

Table VI. Statistics of predictability, PRP level

PRP	Minimum (MW)	Maximum (MW)	99.7% CI (MW)
Anton	−1057	1156	−787 to 773
Berta	−1604	1621	−1239 to 1242
Cesar	−696	798	−575 to 660
Dora	−1941	2169	−1577 to 1726
Emil	−35	39	−26 to 26
Friedrich	−78	67	−56 to 45
Gustav	−116	109	−85 to 74
Total PRPs	−5527	5959	−4345 to 4546
System	−5366	5692	−4113 to 4371

The collective requirement is about 58% of the total capacity for up-regulation or reserve, and about 56% for down-regulation. These collective requirements are larger by up to 3% than if the balancing actions were taken at the system level.

In addition, we perform a statistical analysis of the MW forecast error data by look-ahead time, which in this case varies between 48 and 144 PTUs (i.e. 15-min intervals), or 12–36 h. The results are shown in Figure 14, expressed in percentage with respect to the installed power. As expected, the performance of the forecast

DOI: 10.1002/we

degrades slightly with look-ahead time. The best values obtained are for the 12-h-ahead forecast, where the SD is 12.7%, and the 99.7% CI is [–50% to +48%]. The SD for the 36-h-ahead forecast goes up to 17.2%.

Table VII presents the overall forecast error measures: the root mean square error (RMSE) and the mean absolute error (MAE). The values are presented both in MW and in percentage with respect to the average power and with respect to the nominal power. The percent RMSE value of 14.2% for the system level is smaller compared with the 17–19% for the single wind farm level, and the percent MAE (9.8%) is also smaller compared with the 12–14%, both reported in Madsen *et al.*[17] in percent of installed power for lead times between 12 and 36 h. The values calculated in percentage with respect to the average power (equal to 3434.5 MW for the 7800 MW installed capacity scenario) are understandably higher. Note that in the presence of an intra-day market, the aggregated forecast errors could drop to about half of the day-ahead values, as also simulated in Ummels *et al.*[1]

7. Conclusion

A statistical interpolation method to generate time series of system- and participant-aggregated wind power production and forecast values has been presented in this paper. The method takes into account the spatial and temporal correlations among multiple sites, as derived from the measurement and forecast data. In addition, a method for deriving park-aggregated power curves with smooth cut-in and cut-out, that takes into account the local wind climate, was introduced.

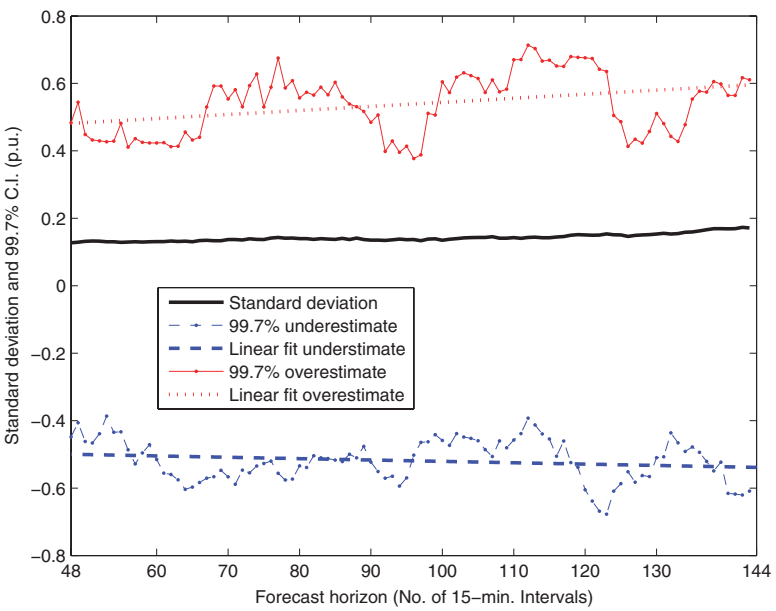

Figure 14. SD and 99.7% CI of the aggregated forecast error by forecast horizon

Table VII. Overall predictability statistics, 7830 MW wind power

	MW	$P_{average}$ (%)	$P_{installed}$ (%)
RMSE	1116.2	32.8	14.2
MAE	772.3	22.7	9.8

DOI: 10.1002/we

In the Dutch scenario with 7.8 GW wind power for the year 2020, at the system level the imbalance energy requirements due to wind variations across 15-min intervals are ±14%, while the imbalance energy requirements due to forecast errors vary between 53% down-regulation and 56% up-regulation or reserve. If 15-min variations and forecast errors are compensated individually by market participants, the collective balancing energy requirements are larger: ±16% for the 15-min wind variations, and between 56% down-regulation to 58% up-regulation or reserve for the forecast errors. Thus, a difference of 2–3% exists between the two aggregation levels.

Such quantitative information is relevant for regulators and system operators when making decisions about adjusting the market design to accommodate high penetrations of wind energy in the power system. Similar conclusions were reached in precedent studies such as Giebel,[4] Doherty and O'Malley,[5] and Holttinen,[6] and Norgard *et al*.,[8] and more recently, the European Wind Integration Study.[18] In particular, we show that assigning balancing responsibility to market parties, as is currently done in The Netherlands, introduces some inefficiencies when compared with taking care of wind energy-related imbalances at the national system operator level. In future studies, the reserve requirements due to the integration of wind energy should be estimated in conjunction with other sources of uncertainty, such as load forecasting errors and forced outages of conventional power plants.

Acknowledgements

This work was funded by the Dutch Ministry of Economic Affairs partly under the EOS-LT project RegelDuurzaam (Control and Reserve Power in Sustainable Electricity Supply Systems) and partly under the BSIK programme We@Sea, project SIBH, *System Integration and Balance Control for Large-Scale Offshore Wind*, We@Sea/BSIK 2004-010. The measured wind data originate from KNMI, ECN and DEWI. The HiRLAM data originate from KNMI. The authors would like to thank the reviewers who prompted us to substantially improve our method. In addition, we thank all our project partners from Ecofys, TenneT and KEMA for their valuable comments.

Appendix A: Regional Variation of Wind Speeds

Consider the matrix w_{ij} with observations of the wind speed at different moments ($i = 1, 2, 3, \ldots, N$) and in different locations ($j = 1, 2, 3, \ldots, M$) in otherwise the same wind climate, and give an estimate for the variation of the wind speed over the different locations at the same moment. First, group the observations per location. This yields the local averages μ_j and the local SDs σ_j of the wind speed observations, which, since all locations are in the same wind climate, are equal to the average μ and the SD σ of the wind speed distribution:

$$\mu = \mu_j \equiv \frac{1}{N}\sum_{i=1}^{N} w_{ij} \tag{13}$$

and

$$\sigma^2 = \sigma_j^2 \equiv \frac{1}{N}\sum_{i=1}^{N} (w_{ij} - \mu_j)^2 \tag{14}$$

Next, group the observations per moment, which gives the averages μ_i and the SDs σ_i of the wind speed observations at the moment i:

$$\mu_i \equiv \frac{1}{M}\sum_{j=1}^{M} w_{ij} \tag{15}$$

and

$$\sigma_i^2 \equiv \frac{1}{M}\sum_{j=1}^{M} (w_{ij} - \mu_i)^2 = \frac{1}{M}\sum_{j=1}^{M} w_{ij}^2 - \mu_i^2 \tag{16}$$

DOI: 10.1002/we

The μ_i (measures for the average of the wind speed in the M locations at the same moment) as well as the σ^2_i (measures for the variance between the wind speeds in the M locations at the same moment) constitute a time series with N elements. The average over μ_i is the regional average μ_R of the wind speed. It is straightforward to show that μ_R is equal to the average μ of the wind speed distribution. The average of σ^2_i is the expected value of the variance between the wind speeds in the M locations at the same moment and is to be interpreted as the regional variance σ^2_R of the wind speed. By summing over the N values of σ^2_i, it follows that:

$$\sigma_R^2 = \frac{1}{M}\left[(M-1) - \frac{2}{M}\sum_{j=1}^{M-1}\sum_{k=j+1}^{M} r_{jk}\right]\sigma^2 \tag{17}$$

where r_{jk} is the correlation between the wind speed in the locations j and k at the same moment. By employing the exponential correlation[4,7,13]

$$r_{jk} \equiv \mathrm{e}^{-\frac{d_{jk}}{D_{\mathrm{decay}}}} \tag{18}$$

where d_{jk} is the distance between the locations j and k, and the average distance d_{ave} between the locations

$$d_{\mathrm{ave}} \equiv \frac{1}{M(M-1)}\sum_{j=1}^{M-1}\sum_{k=j+1}^{M} d_{jk} \tag{19}$$

(recall $d_{jj} = 0$ and $d_{kj} = d_{jk}$), it can be shown that

$$\sum_{j=1}^{M-1}\sum_{k=j+1}^{M} r_{jk} \approx \frac{1}{2}M(M-1)\mathrm{e}^{-\frac{d_{\mathrm{ave}}}{D_{\mathrm{decay}}}} \tag{20}$$

It follows that the regional variance of the wind speed in the M locations is estimated by

$$\sigma_R^2 \approx \left(1 - \frac{1}{M}\right)\left[1 - \mathrm{e}^{-\frac{d_{\mathrm{ave}}}{D_{\mathrm{decay}}}}\right]\sigma^2 \tag{21}$$

Since the number of locations ranges from 2 to ∞, the limits of the regional variance are

$$\frac{1}{2}\left[1 - e^{-\frac{d_{\mathrm{ave}}}{D_{\mathrm{decay}}}}\right]\sigma^2 \le \sigma_R^2 \le \left[1 - e^{-\frac{d_{\mathrm{ave}}}{D_{\mathrm{decay}}}}\right]\sigma^2 \tag{22}$$

Appendix B: Average Distance between Locations

Consider a region of area A with M locations, without any knowledge on the shape of the area or the position of the locations, and estimate the average distance between the locations. First, introduce a length scale L that characterizes the area:

$$L = \sqrt{\frac{A}{\pi}} \tag{23}$$

On the basis of L, an estimate for the largest distance $d_{\max}$ between two locations is

$$d_{\max} = 2L = 2\sqrt{\frac{A}{\pi}} \tag{24}$$

Next, introduce the area per location:

$$A_{\mathrm{loc}} = \frac{A}{M} \tag{25}$$

DOI: 10.1002/we

and the corresponding characteristic length scale:

$$L_{\text{loc}} = \sqrt{\frac{A}{\pi M}} \tag{26}$$

which subsequently gives an estimate for the smallest distance d_{min} between two locations:

$$d_{\text{min}} = 2L_{\text{loc}} = 2\sqrt{\frac{A}{\pi M}} \tag{27}$$

Together, d_{max} and d_{min} provide an estimate for the average distance between two locations:

$$d_{\text{ave}} = \frac{d_{\text{max}} + 2d_{\text{min}}}{3} = \frac{2}{3}\sqrt{\frac{A}{\pi}\left(1 + \frac{2}{\sqrt{M}}\right)} \tag{28}$$

in the case that only the area of the region and the number of locations are known.

References

1. Ummels BC, Gibescu M, Pelgrum E, Kling WL, Brand AJ. Impacts of wind power on thermal generation unit commitment and dispatch. *IEEE Transactions on Energy Conversion* 2007; **22**: 44–52.
2. Brand AJ. Observed and predicted wind speed time series in The Netherlands and the North Sea. *Energie Centrum Nederland (ECN) Technical Report* No ECN-C–06-007, 2006.
3. Gibescu M, Ummels BC, Kling WL. Statistical wind speed interpolation for simulating aggregated wind energy production under system studies. *9th International Conference on Probabilistic Methods Applied to Power Systems*, Stockholm, Sweden, 2006; 1–7.
4. Giebel G. On the benefits of distributed generation of wind energy in Europe. PhD Dissertation, Carl von Ossietzky University, Oldenburg, Germany, 2000.
5. Doherty R, O'Malley M. A new approach to quantify reserve demand in systems with significant installed wind capacity. *IEEE Transactions on Power Systems* 2005; **20**: 587–595.
6. Holttinen H. Hourly wind power variations in the Nordic countries. *Wind Energy* 2005; **8**: 173–195.
7. Lange M. Focken U. *Physical Approach to Short-term Wind Power Prediction*. Springer-Verlag: Berlin, 2006.
8. Norgard P, Giebel G, Holttinen H, Söder L, Petterteig A. Fluctuations and predictability of wind and hydro power. *Risoe National Laboratory Technical Report* No Riso-R-1443(EN), 2004.
9. Söder L. Simulation of wind speed forecast errors for operation planning of multiarea power systems. *International Conference on Probabilistic Methods Applied to Power Systems*, Ames, Iowa, 2004; 723–728.
10. Gibescu M, Brand AJ. Estimation of system balancing requirements due to the integration of large-scale wind energy. *6th International Workshop on Large-scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms*, Delft, The Netherlands, 2006; 1–8.
11. Brockwell PJ, Davis RA. *Time Series: Theory and Methods* (2nd edn). Springer: New York, 1991.
12. Walker JF, Jenkins N. *Wind Energy Technology*. Wiley: New York, 1997.
13. Brand AJ, Hegberg T. Offshore wind atlas. *Energie Centrum Nederland (ECN) Technical Report* No ECN-CX–04-136, 2004.
14. Landberg L, Hansen MA, Vesterager K, Bergstrom W. Implementing wind forecasting at a utility. *Risoe National Laboratory Technical Report* No Riso-R-929(EN), 1997.
15. Brand AJ, Kok JK. Aanbodvoorspeller Duurzame Energie Deel 2:Korte-termijn prognose van windvermogen. *Energie Centrum Nederland (ECN) Technical Report* No ECN-C–03-049, 2003.
16. Norgard P, Holttinen H. A multi-turbine power curve approach. *Nordic Wind Power Conference*, Göteborg, Sweden, 2004; 1–5.
17. Madsen H, Pinson P, Kariniotakis G, Nielsen HA, Nielsen TS. Standardising the performance evaluation of short-term wind power prediction models. *Wind Engineering* 2005; **29**: 475–489.
18. ETSO. European Wind Integration Study (EWIS) towards a successful integration of wind power into European electricity grids. *European Transmission System Operators, Final Report* [Online] Available: www.etso-net.org. (Accessed September 2008).

DOI: 10.1002/we