2018 European Control Conference (ECC)
June 12-15, 2018. Limassol, Cyprus

Wind Farm Power Forecasting for Less Than an Hour Using Multi Dimensional Models

Torben Knudsen and Thomas Bak and Tom Nørgaard Jensen

Abstract— **This research focuses on prediction of wind farm power for horizons of 0-10 minutes and not more than one hour using statistical methods. These short term predictions are relevant for both transmission system operators, wind farm operators and traders. Previous research indicates that for short time horizons the persistence method performs as well as more complex methods. However, these results are based on accumulated power for an entire wind farm. The contribution in this paper is to develop multi-dimensional linear methods based on measurements of power or wind speed from individual wind turbine in a wind farm. These multi-dimensional methods are compared with the persistence method using real 1 minute average data from the Sheringham Shoal wind farm with 88 turbines. The results show that the use of measurements from individual turbines reduce the prediction errors 5-10% and also improves the prediction error variance estimate compared to the persistence method. We also present convincing examples showing that the predictions follow the wind farm power over a window of an hour.**

ABBREVIATIONS AND NOTATION

E( )	Expectation	MARX	Multi-dimensional auto regressive with exogenous input
NID( )	Normal independend distributed		
N( )	Normal distributed	PM	Persistence Method
V( )	Variance	Residual	One step prediction error
AR	Auto regressive	WF	Wind farm
EWS	Effective wind speed	WFP	Wind farm power
MAR	Multi-dimensional auto regressive	WS	Wind speed
		WT	Wind turbine

I. INTRODUCTION

In wind energy, power production is driven by the wind speed (WS) which is stochastic and therefore not exactly known in the future. On the other hand, WS is not described by white noise so it can be predicted to some extent. Different applications are associated with different relevant prediction horizons.

For the electricity grid the farm power production over the next hour is important [1]–[3]. The Danish transmission system operator points to 10 minutes as the most relevant prediction horizon [4] and the uncertainty on the forecast is as important as the point forecast itself [3]. For trading in the balancing market, the inter-hour forecast also has some value.

Models from meteorology, used for weather forecast, are based on the laws of physics and useful for forecast from one to approximately 10 days ahead [5]. For shorter horizons, e.g. less than one hour, these models become less useful and for very short horizons in the range of minutes, statistical models using available measurements are superior [6]. The persistence method (PM) [2], [7], where the most recent measurement is used as the prediction, is close to optimal. The PM provides point predictions, but a good method for calculating the prediction uncertainty is still needed.

Section of Automation and Control, Aalborg University, Fredrik Bajers Vej 7C2-212 DK-9220 Aalborg, Denmark tk@es.aau.dk

The majority of work in this field has focused on prediction of hourly average total wind farm power (WFP) with a prediction horizon of one day. Some of the methods only forecast a representative farm WS [8] and leaves the WFP forecasting to others. The probabilistic models used have often been complicated as e.g. artificial neural network [8] and the measurements used have been total WFP.

The contributions here are to explore other ideas:

- Predict *available* WFP with focus on horizons of 0-10 minutes and not more than one hour.
- Use measured or estimated WS for *each and all* wind turbines (WT).
- Use simple *multi-dimensional auto regressive* (MAR) models.
- Benchmark the performance against the PM using *real wind farm* (WF) data.

The hypothesis is that the assumed correlation between wind at one WT and its upwind neighbouring WTs can be exploited in the models to provide not only a good point prediction but also an uncertainty prediction.

The remaining of this paper starts by developing the methods used. Then the data from the real full scale wind farm Sheringham Shoal is presented. Then results are provided and discussed and finally, a conclusion is drawn.

Notice that some information are intentionally left out due to confidentiality e.g. in Fig. 3-4.

II. METHODS AND MODELS

Before discussing mathematical models and methods some general choices must be made. First of all, we chose to produce not only a WFP point forecast but also estimate the associated uncertainty. Moreover, what is really useful is not the *actual* produced WFP but the *available*. The difference between the two arises if some WTs are derated. One way to pursue the available WFP is to use *effective wind speed* (EWS) estimation and wind farm wake models [9], [10]. The measurements necessary for this are not available in this work and can therefore not be used. Instead a measured WS is used to replace EWS. Notice that the term EWS is still used where appropriate to keep the distinction.

Another important choice is between two different forecasting approaches:

1) Model and forecast WS and then use the WT power curve to transform from forecasted WS to forecasted WFP.
2) Directly model and forecast the power.

The one minute average WS is not a stationary linear Gaussian process. However, it can be approximated by one with parameters varying due to weather changes. Therefore, WS should be well modelled by basic statistical time series methods, which is a strength of approach 1). The weakness of 1) is that the WT power curve will add to the uncertainty. Approach 2) avoids the uncertainty added by using the power curve. However, the downside is that WT power is very different from a linear Gaussian process: Around cut-in WS the distribution has a long right tail, in medium wind the skewness is close to zero, around rated wind the distribution has a long left tail and for higher WS the power is close to constant at rated power.

The third choice is whether to work with individual WT measurements or accumulated WF measurements.

Potentially, any combination of these choices could make a good method. The combinations chosen here are presented below.

A. *Persistence based on WF power including uncertainty*

PM is in the literature only presented as a point forecast. The PM forecast corresponds to an optimal forecast if the signal in question is assumed to be a stochastic Wiener process. With this assumption the corresponding uncertainty is developed below. From a statistical point of view this is not new but still a small contribution to the wind energy forecast area.

Let x denote the signal in question, here WFP. If $\hat{x}(t+\tau|t)$ is the optimal predictor for $x(t+\tau)$ given measurements of x until and including time t, then the PM is by definition

$$\hat{x}(t+\tau|t) = x(t) \quad \forall\, \tau \geq 0 \,. \tag{1}$$

From a time series point of view this is consistent with using a Wiener process for x

$$x(t) \in W(q) \,, \tag{2}$$

where $W(q)$ denotes a Wiener process in continuous time with *incremental* variance q. By definition this means that x is Gaussian distributed with independent increments whose variance is proportional to time lag τ. From this, the predictor and its prediction error variances follows from (3) which holds for any $\tau \geq 0$.

$$x(t+\tau) - x(t) \in \mathrm{N}(0, q\tau) \Rightarrow \tag{3a}$$

$$x(t+\tau)|x(t) \in \mathrm{N}(x(t), q\tau) \tag{3b}$$

$$\hat{x}(t+\tau|t) \triangleq \mathrm{E}\left(x(t+\tau)|x(t)\right) = x(t) \Rightarrow \tag{3c}$$

$$\tilde{x}(t+\tau|t) \triangleq x(t+\tau) - \hat{x}(t+\tau|t) \in \mathrm{N}(0, q\tau) \Rightarrow \tag{3d}$$

$$\mathrm{E}(\tilde{x}(t+\tau|t)^2) = \mathrm{V}(\tilde{x}(t+\tau|t)) = q\tau \tag{3e}$$

The only parameter in this model is q. As it is not known it has to be estimated. Assuming the measured data are $x(t_i), i = 0, \ldots, N$ this is done by (4).

$$\hat{q}(N) = \frac{1}{N} \sum_{i=1}^{N} \frac{\tilde{x}(t_i|t_{i-1})^2}{t_i - t_{i-1}} \tag{4}$$

It is well known that the variance on WFP is not the same all the time. Consequently, an adaptive estimator is needed. The expression (4) is an *offline* estimator which alternatively can be written as a *online/recursive* estimator as follows

$$\begin{aligned} \hat{q}(N) &= \frac{1}{N}\left(\sum_{i=1}^{N-1} \frac{\tilde{x}(t_i|t_{i-1})^2}{t_i - t_{i-1}} + \frac{\tilde{x}(t_N|t_{N-1})^2}{t_N - t_{N-1}}\right) \\ &= \frac{N-1}{N}\hat{q}(N-1) + \frac{1}{N}\frac{\tilde{x}(t_N|t_{N-1})^2}{t_N - t_{N-1}} \\ &= \lambda(N)\hat{q}(N-1) + (1-\lambda(N))\frac{\tilde{x}(t_N|t_{N-1})^2}{t_N - t_{N-1}} \,, \\ \lambda(N) &= \frac{N-1}{N} = 1 - \frac{1}{N} \,. \end{aligned} \tag{5}$$

The weights on the previous estimate and the new information $\lambda(N)$ and $1 - \lambda(N)$ add to 1 but the former tends to one. If λ is fixed, the *adaptive forgetting factor* estimator (6) is obtained.

$$\begin{aligned} \hat{q}(N) &= \lambda\hat{q}(N-1) + (1-\lambda)\frac{\tilde{x}(t_N|t_{N-1})^2}{t_N - t_{N-1}} \,, \\ & 0 < \lambda < 1 \end{aligned} \tag{6}$$

This results in *exponential forgetting* where λ must be close to one as the effective averaging window is $1/1-\lambda$ [11].

The resulting forecast can be given by the point forecast plus/minus 2 standard deviations for the prediction error given by

$$\hat{x}(t+\tau|t) = x(t) \,, \; \hat{\sigma}_{\tilde{x}(t+\tau|t)} = \sqrt{\hat{q}(t)\tau} \,. \tag{7}$$

B. *Models using EWS at individual WT*

As discussed in Section I, the principal idea here is to use upwind EWS to predict down wind EWS. If significant derating is used, these EWSs should, ideally, be transformed into corresponding full load EWSs, using a wake model, before using the WT power curve to transform them into WFP. This is however not done here because the necessary WT parameters are confidential.

1) Model: Below a model including all EWSs in the WF is outlined in (8). Assume now that the time unit is the constant sampling time. The idea is that EWS at all WTs at time t to some degree is explained by EWS at all WTs at time $t-1$. The exogenous input term u is not necessary here but is included for later use. The model has several names, a *state space* model, a vector ARX model or a MARX model. A MARX model is a MAR model with exogenous input.

$$\begin{aligned} x(t) &= ax(t-1) + bu(t-1) + e(t-1) \,, \\ & e(t) \in \mathrm{NID}(\underline{0}, \Sigma) \,, \end{aligned} \tag{8a}$$

$$x(t) = \begin{bmatrix} w_1(t) \\ \vdots \\ w_{n_t}(t) \end{bmatrix}, \; u(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_{n_u}(t) \end{bmatrix}, \quad e(t) = \begin{bmatrix} e_1(t) \\ \vdots \\ e_{n_t}(t) \end{bmatrix}, \tag{8b}$$

where w_i is EWS for WT i. The parameters are $a, \Sigma \in \mathbb{R}^{n_t \times n_t}$ and $b \in \mathbb{R}^{n_t \times n_u}$, n_t is the number of WTs and n_u are the number of inputs, here one or none.

2) Parameter estimation: Estimation of constant parameters can be formulated using matrix notation in the well known *General Linear Model* (GLM) fashion [12] as shown in (9), where measurements of x and u are available from sample $t_1 - 1$ to t_2.

$$x(t) = ax(t-1) + bu(t-1) + e(t-1) \Leftrightarrow \tag{9a}$$

$$x(t)^{\mathsf{T}} = \begin{bmatrix} x(t-1)^{\mathsf{T}} & u(t-1)^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} a^{\mathsf{T}} \\ b^{\mathsf{T}} \end{bmatrix} + e(t-1)^{\mathsf{T}} \Leftrightarrow \tag{9b}$$

$$Y = X\Theta + E , \tag{9c}$$

$$Y \triangleq \begin{bmatrix} x(t_1)^{\mathsf{T}} \\ \vdots \\ x(t_2)^{\mathsf{T}} \end{bmatrix}, \; X \triangleq \begin{bmatrix} x(t_1-1)^{\mathsf{T}} & u(t_1-1)^{\mathsf{T}} \\ \vdots & \\ x(t_2-1)^{\mathsf{T}} & u(t_2-1)^{\mathsf{T}} \end{bmatrix}, \quad E \triangleq \begin{bmatrix} e(t_1-1)^{\mathsf{T}} \\ \vdots \\ e(t_2-1)^{\mathsf{T}} \end{bmatrix}, \; \Theta \triangleq \begin{bmatrix} a^{\mathsf{T}} \\ b^{\mathsf{T}} \end{bmatrix} \tag{9d}$$

The above model formulation leads to the below parameter and covariance estimates

$$\widehat{\Theta} \triangleq \arg\min_{\Theta} |Y - X\Theta|_F^2 \Rightarrow \widehat{\Theta} = (X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}Y \tag{9e}$$

$$\widehat{\Sigma} = \frac{1}{t_2 - (t_1 - 1)} \mathcal{E}^{\mathsf{T}}\mathcal{E} , \; \mathcal{E} = Y - X\widehat{\Theta} . \tag{9f}$$

In (9e) the input u is assumed sufficiently exciting to make $X^{\mathsf{T}}X$ invertible and F is the Frobenius norm.

Notice that (9) can be interpreted as a linear regression with vector output (9b) and then each column in $\widehat{\Theta}$ (9e) could be produced by scalar output regression for one output component at a time. This means that the statistics for the estimator can be derived as below where a Matlab like notation is used.

$$\widehat{\Theta}(:,j) = (X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}Y(:,j) , \quad \widehat{\Theta}(:,j) = \arg\min_{\Theta(:,j)} |Y(:,j) - X\Theta(:,j)|^2 \Rightarrow \tag{10a}$$

$$\mathrm{E}\left(\widehat{\Theta}(:,j)\right) = \Theta(:,j) , \quad \mathrm{Cov}\left(\widehat{\Theta}(:,j)\right) = (X^{\mathsf{T}}X)^{-1}\Sigma(j,j) \tag{10b}$$

3) Recursive estimation: WSs can be modelled with simple models like MARX but there is no reason to believe that the parameters are constant. On the contrary, dominating frequencies as well as variance will generally increase with the average WS but they also depend on e.g. atmospheric stability. Obviously, the correlation between neighbouring WTs will also depend on wind direction. Therefore, it is also relevant to have a recursive and adaptive estimator. Using the exponential forgetting version [11, Sect. 11.2] the below can be derived.

$$\phi(t) = \begin{bmatrix} x(t-1) \\ u(t-1) \end{bmatrix} , \tag{11a}$$

$$R(t) = \lambda R(t-1) + \phi(t)\phi(t)^{\mathsf{T}} , \tag{11b}$$

$$\hat{x}(t) = \phi(t)^{\mathsf{T}}\widehat{\Theta}(t-1) , \tag{11c}$$

$$\epsilon(t) = x(t) - \hat{x}(t) , \tag{11d}$$

$$\widehat{\Theta}(t) = \widehat{\Theta}(t-1) + R(t)^{-1}\phi(t)\epsilon(t)^{\mathsf{T}} , \tag{11e}$$

$$\widehat{\Sigma}(t) = \lambda\widehat{\Sigma}(t-1) + (1-\lambda)\epsilon(t)\epsilon(t)^{\mathsf{T}} . \tag{11f}$$

The recursive parameter estimation can be initiated by first collecting a batch of data ending at sampling time t_s and then calculating the off-line estimate corresponding to (9) (but including exponential forgetting) and then start from $R(t_s), \widehat{\Theta}(t_s), \widehat{\Sigma}(t_s)$. This batch startup will be used here. An alternative is using preset starting values [11].

The choice of λ is based on the fact that the recursive estimator (11) fulfils

$$\widehat{\Theta}(t) = \arg\min_{\Theta} \left| \sum_{s=t_s}^{t} \lambda^{t-s}\epsilon(s)\epsilon(s)^{\mathsf{T}} \right|^2 . \tag{12}$$

If the *forgetting time constant* τ_λ is defined by $\lambda^{\tau_\lambda/T_s} = e^{-1}$ where T_s is the sampling time, then τ_λ is the time it takes the weight in (12) to decrease to e^{-1} and the corresponding forgetting factor will be $\lambda = e^{-T_s/\tau_\lambda}$. The time window τ_λ should be as large as possible to reduce uncertainty but not larger than the time interval where the parameters are assumed to be fairly constant to reduce bias. For WS, τ_λ could be 3 hours corresponding to 3×60=180 samples assuming a one minute sampling time.

Another issue is over parametrisation. The number of samples should be larger, preferably much larger, than the number of parameters. In the MARX model (8) there are $n_t = 88$ parameters in a single row of a, where each row corresponds to the e.g. 180 samples of each component in x. The choice of $\tau_\lambda = 180$ is a compromise between a smaller time window to be able to follow faster changes in the wind dynamic and a larger window to be able to estimate the many parameters.

C. Multistep prediction in MARX models

For a MARX model, multistep prediction can be formulated as a recursion. Assume the parameters are known either as a function of time or they are assumed to remain constant at the most recent value from the recursive parameter estimator. For $\tau \geq 0$ the prediction and corresponding uncertainty are given by (13b) and (13d) starting with the initialization (13e).

$$x(t+\tau+1) = ax(t+\tau) + bu(t+\tau) + e(t+\tau) \Rightarrow \tag{13a}$$

$$\hat{x}(t+\tau+1|t) = a\hat{x}(t+\tau|t) + bu(t+\tau) \Rightarrow \tag{13b}$$

$$\tilde{x}(t+\tau+1|t) = a\tilde{x}(t+\tau|t) + e(t+\tau) \Rightarrow \tag{13c}$$

$$\mathrm{Cov}\left(\tilde{x}(t+\tau+1|t)\right) = a\,\mathrm{Cov}\left(\tilde{x}(t+\tau|t)\right) a^{\mathrm{T}} + \Sigma\,, \tag{13d}$$

$$\hat{x}(t|t) = x(t)\ ,\ \mathrm{Cov}\left(\tilde{x}(t|t)\right) = \underline{\underline{0}}\,. \tag{13e}$$

If the predicted x in (13) is the vector of all WS as in section II-B then a transformation to predicted WFP p_f is given by

$$p_f(t) = \sum_{i=1}^{n_t} p_t(x_i(t))\ , \tag{14}$$

where p_t is the power curve (function) for a single WT. If the power curve is assumed known, the point estimate and standard deviation for WFP can be approximated by

$$\hat{p}_f(t+\tau|t) = f(\hat{x}(t+\tau|t))\ , \tag{15a}$$

$$\hat{\sigma}_{\tilde{p}_f(t+\tau|t)} = \sqrt{(\nabla f)\,\mathrm{Cov}\left(\tilde{x}(t+\tau|t)\right)\nabla f^{\mathrm{T}}}\ , \tag{15b}$$

where

$$f(x(t)) \triangleq p_f(t)\ ,\ \nabla f \triangleq \left.\frac{\partial f(x)}{\partial x^{\mathrm{T}}}\right|_{x=\hat{x}(t+\tau|t)}\ . \tag{15c}$$

A good alternative to the linearization approximation (15) is to use the *unscented transform* [13]–[15] to calculate both the conditional mean and variance for the predictor. This can formally be expressed as

$$[\hat{p}_f(t+\tau|t), \hat{\sigma}^2_{\tilde{p}_f(t+\tau|t)}] = \mathrm{U}\left(f, \hat{x}(t+\tau|t), \mathrm{Cov}\left(\tilde{x}(t+\tau|t)\right)\right)\ , \tag{16}$$

where U denotes the unscented transform, which needs the three inputs: function, mean and covariance of the input. Compared to the linearization method the unscented transform gives similar results for mildly nonlinear transformations but for strongly nonlinear transformations it is superior. For this reason, it is used here.

If the predicted x holds the power from each WT, p_t in (14) would be the identity and f is simply the sum of its elements. The above can still be used but now the mean and covariance in (15) and (16) will be exact since f is now linear. The same is true if x is the WFP.

D. Power curve estimation

The single wind turbine power curve has not been made available for this research. Therefore, the WT power curve relating WS to power must be estimated. Here nacelle WS must be used as this is the only available WS. For the power curve different parametrisations can be used. The power curve shall be able to capture the basics, which are: from zero WS to cut-in WS the power is zero, from cut in WS to rated WS the power increases nonlinearly from zero towards rated power where it stays as the wind increases towards cut-out WS. The choice here is the below nonlinear regression model (17) with only two parameters b_1, b_2.

$$P = P_n(1 - e^{-b_1(w-b_3)^{b_2}})\ , \quad P_n = 3.6MW\ ,\ b_3 = 0 \tag{17}$$

The parameter b_3 is difficult to estimate and can be set to zero with little influence on the fit. The parameter P_n can normally be fixed to the rated power for modern wind turbines as they are well controlled in high wind.

III. Data from a full-scale wind farm

As the focus here is prediction horizon in minutes, the standard 10 minutes average SCADA data are not sufficient. Statoil has been helpful to provide data from the Sheringham Shoal wind farm.

A. Sheringham Shoal wind farm

Sheringham Shoal is owned by Statkraft, Statoil and Green Investment Bank through the joint venture company Scira Offshore Energy Limited and is situated north east of London as show in Fig. 1. The WF consists of 88 Siemens SWT-3.6-107 (3.6 MW 107 m diameter rotor) which is a modern state of the art variable speed pitch controlled WT. The distances to the nearest neighbours varies from 6-7 rotor diameters. There are two offshore substations with a meteorological station on one of the substations. The WTs are bottom fixed with monopiles.

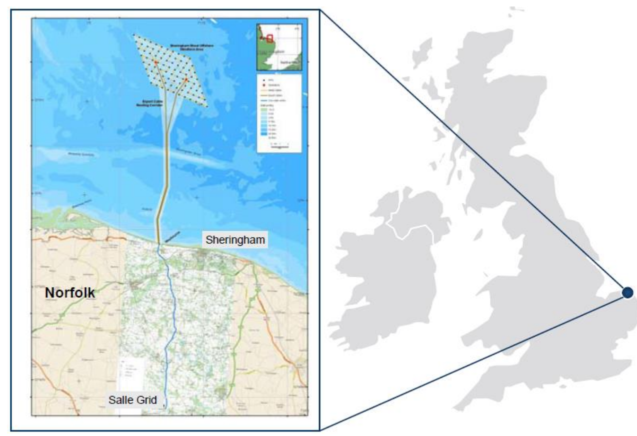

Fig. 1. Geographical placement of Sheringham Shoal WF.

B. Selection of data

As already mentioned, it would be preferable to estimate the effective wind speed. However, this is not possible with the available data. The measurement used here is the WS measured at the nacelle and the generated electrical power. Therefore, the data should be with as little derating as possible. The measurements sampling time is approximately 1 second and they are therefore down sampled to a one minute sampling time.

The following criteria was used to choose a period from available data: Wind speed both below and above rated, not too much derating, and availability of data for a high number of turbines. The date 2014-01-27 was selected based on this. Notice that the avoidance of derating is due to the available data not a limitation in the methods based on EWS.

The farm geometry with the WTs included in the data is seen in Fig. 2.

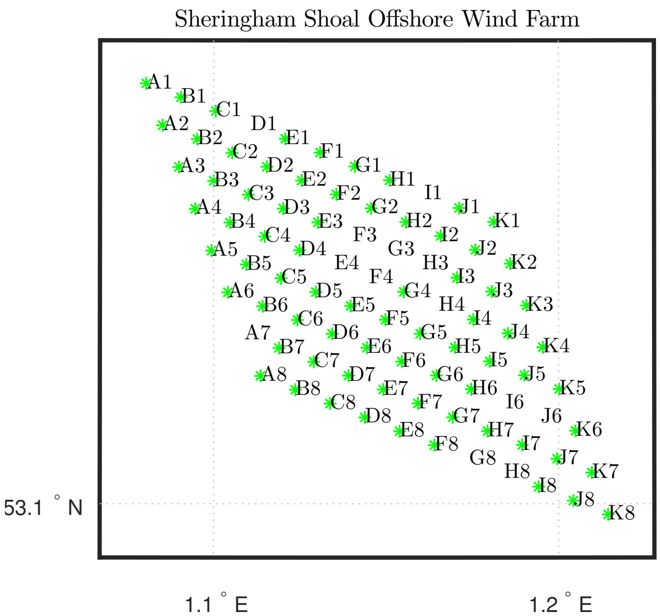

Fig. 2. Farm geometry with the WTs included in the data for 2014-01-27 shown in green.

All 1440 one minute samples are shown for all WT in Fig. 3. These data cover both partial and full load as desired. Variation in both time and space (between WTs) can be seen for both WS and wind direction. This is expected for WS because of wakes from the upwind WTs. For the wind direction it is less expected. Large deratings for a few WTs from full to half power are also seen in the beginning of the day.

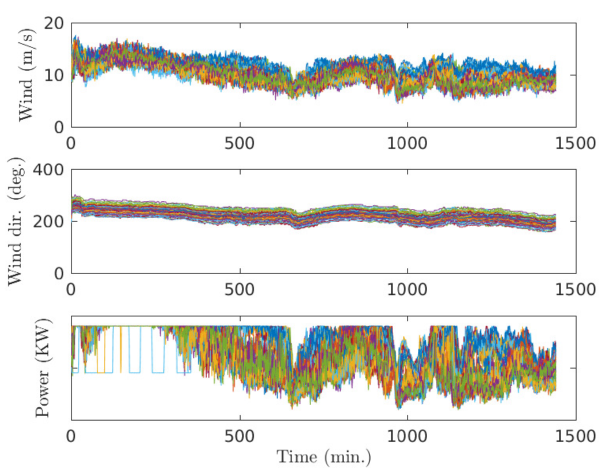

Fig. 3. Measurements from 2014-01-27 for the whole day. One minute sampling time. Tick labels for power are missing due to confidentiality.

IV. Results

First the power curve estimation is presented. The PM results based on WFP is then briefly given. The MAR model using individual WT WS is presented in more detail including model tests. The rest of the section presents and compares the time series methods against the PM.

A. Power curve estimation

As all turbines are of the same type it makes sense to estimate one common power curve for all. The results of this is seen in Fig. 4 and the RMS error is: 0.269 MW or 7.5% of rated power. The power curve used is given by (17)

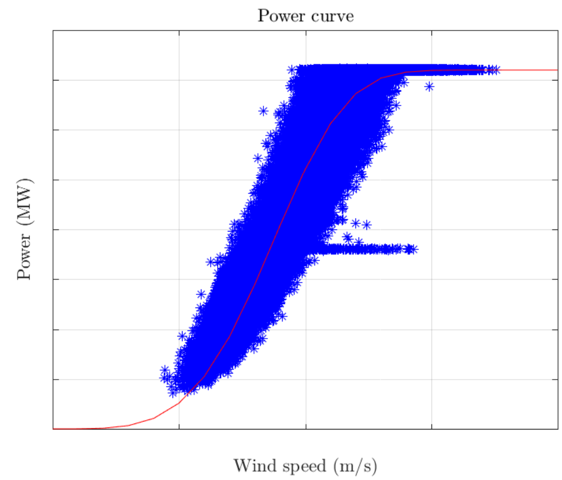

Fig. 4. Power curve based on all WTs (2014-01-27). Notice the derating to half power. Blue is simultaneous measurements of WS and WT power, red is the estimated power curve. Tick labels are missing due to confidentiality.

with the estimated parameters $\hat{b}_1, \hat{b}_2$ inserted.

B. Persistence based on WFP

The results from using the PM are shown in Fig. 5. The example is for a prediction horizon of 5 samples (minutes). The top figure shows the measurements (blue) and the predictions (red) which are just measurements delayed 5 samples. Subplot two shows the estimated incremental covariance in the Wiener process. Notice that the covariance is very small during the time where all WT are in full load. Subplot three shows the prediction error with 95% confidence limits i.e. $\pm$ 2 standard deviations. This fit is acceptable even though a test for variance one for the normalized error is rejected with p-value $\sim$ 0. The bottom plot is the auto correlation for the 5 step prediction error. Ideally this should only be non zero for lags less than 5 [16, sect. 5.7]. Clearly this is not perfect either.

C. MARX models based on all WT WSs

For the MARX model there are some initial choices of forgetting factor and tracking of average WS before results regarding the prediction performance can be obtained.

1) Forgetting factor, average tracking and stability: When using the MARX model (8) where the state x is WS for all WTs, it is necessary to choose the forgetting factor λ (11) and whether to use the input parameter b in (8). If $u \triangleq 0$ and b is not used then if a has eigenvalues inside the unit circle, the stationary mean of x is zero. This is not in accordance with the WS mean value being larger than zero. On the other hand, if the dynamics are slow, with eigenvalues close to the unit circle, the conditional mean i.e. the prediction $\hat{x}(t + \tau|t) =$

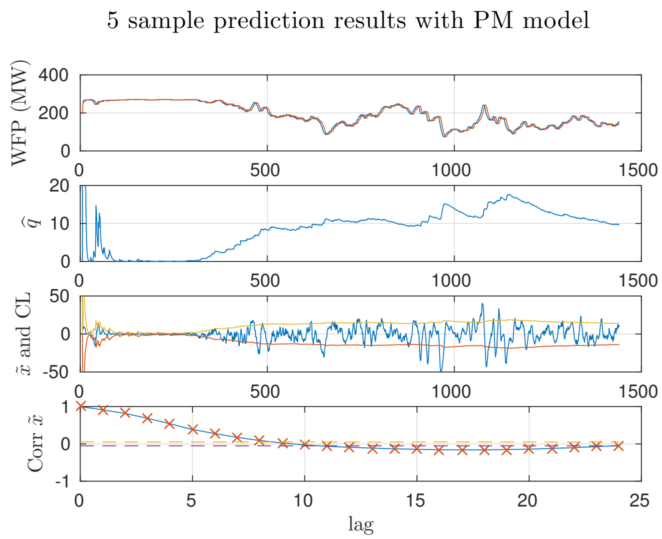

Fig. 5. PM used for WFP 5 minutes prediction (2014-01-27). The x-axis is time in minutes. The top subplot shows the measurements (blue) and the predictions (red). Subplot two shows the estimated incremental covariance in the Wiener process. Subplot three shows the prediction error with 95% confidence limits. The bottom plot is the auto correlation for the 5 step prediction error shown with both x and a line and also including 95% confidence limits.

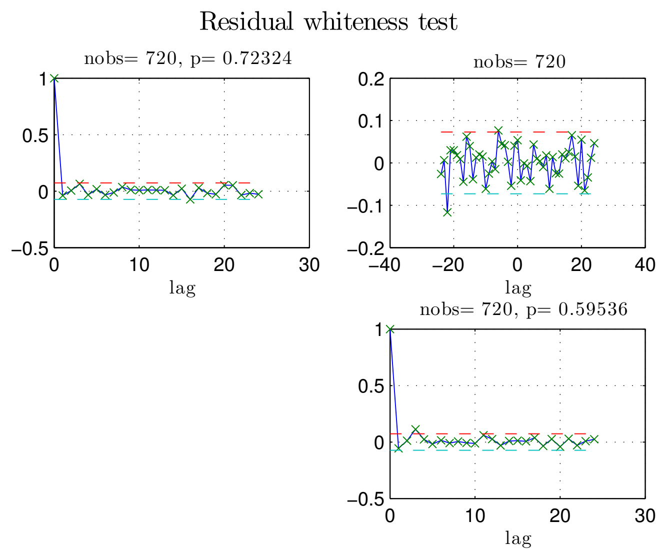

Fig. 6. Test for residual whiteness in the MAR model including WT F6 and E7 which are "wind direction" neighbours. Auto (upper left and lower right subplot) and cross (upper right) correlations with confidence limits are shown. The p parameters above the subplots are the Portmanteau test p-value using 24 lags and nobs is the number of observations used for the calculations.

$a^\tau x(t)$ will only slowly decay to zero with τ. If $u \triangleq 1$ then the parameter $b \in \mathbb{R}^{n_t}$ can be used to estimate a mean value. When the focus is on short term forecasting it is not obvious which approach is best. The 5 min forecast RMS error was slightly smaller using $u = 0$ compared to $u = 1$. For simplicity, $u = 0$ is therefore chosen here so a MAR model is used.

The forgetting time constant τ_λ has been set to 3 hours, corresponding to $\lambda = e^{-T_s/\tau_\lambda} = 0.9945$, by using manual tuning to minimize the above mentioned 5 min forecast RMS error. Stability of the $\widehat{a}$ matrix is another issue to consider as the standard recursive least squares (RLS) method does not guaranty stability. In general, the eigenvalues for $\widehat{a}$ are larger when b is not used and they are occasionally outside the unit circle. The maximal absolute eigenvalues are 1.0005 and 1.0012 for $u = 0$ and $u = 1$ respectively but there are far more instances of instability for $u = 0$. In summary, the maximal absolute eigenvalues are so small for both models that it will not be a problem for the short prediction horizon in focus here.

2) Model test: If the model structure is correct the *residuals* is uncorrelated i.e. white noise [11, Sect. 16.6]. For the models used here the one step prediction error is the residual [11, Sect. 16.6]. Also, it can be of interest to check the residual distribution e.g. if it is close to Gaussian. These two residual tests are therefore performed. In general, the residuals seems to be white even though the Portmanteau test [16] often shows significance but the correlations are numerically small. An example test of two turbines in the wind direction of each other is shown in Fig. 6. The departure from Gaussianity is also in general modest as the example in Fig. 7 indicates.

3) Forecasting performance: An example of 5 minutes forecasting using the MAR model is shown in Fig. 8 which is similar to Fig. 5 for the PM. Notice that in Fig. 8 only the last 715 samples are included to avoid all initial convergence effects. Notice that the PM has no initial effects. The parameter estimates are not shown because there are n_t^2 parameters in $\hat{a}$ alone. Otherwise, the difference to the PM is that the auto correlation is closer to insignificant from sample 5 onward as it should for a correct model.

Apart from the RMS forecasting errors it is also important how the methods capture fast changes. Fig. 9 shows a good example where the WT WS measurements until time 18:00 are used to predict WFP from time 18:00 and two hours forewords using multistep prediction (13). As seen the MAR WTWS method forecast takes the right direction and the WFP stays more or less inside the confidence limits during the first hour. Notice that the PM would just give a constant forecast starting from the last measurement at time 18:00.

D. Comparing forecasting performance

Fig. 10 compares the performance of the 5 minutes forecasting statistics for the different methods shown in table I. Notice that the first three methods in table I is based on produced power either for the whole farm (PM and AR-WFP) or from all WT (MAR-WTP). Only the last method (MAR-WTWS) uses the wind speed and must therefore use the estimated power curve to predict the WFP. The uncertainty of the power curve (see Fig. 4) then adds to the total prediction uncertainty only for the last method. To access the uncertainty contribution from the power curve, artificial WT powers has been calculated by transforming the WT WS though the estimated power curve. The yellow graphs in subplot 2 and 4 are results where the data are

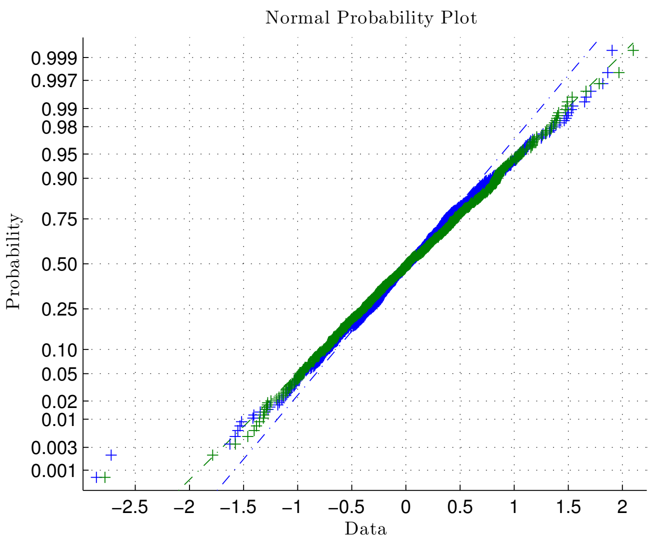

Fig. 7. Test for Gaussian distributions of residuals in the MAR model including WT F6 (blue) and E7 (green) which are "wind direction" neighbours. Observations are marked with + and the dashed lines are the corresponding Gaussian distributions.

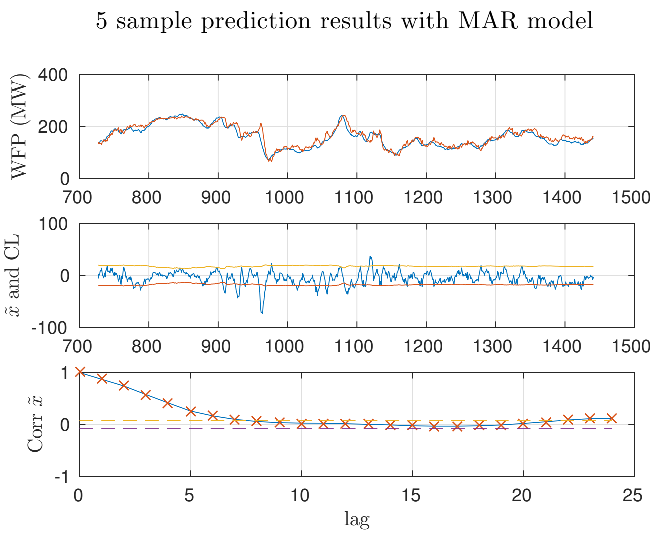

Fig. 8. MAR method used with individual WS inputs to do 5 min predictions of measured WFP (2014-01-27). This figure is similar to Fig. 5 except no parameter estimates are shown.

with artificial WT power and the blue graphs in subplot 1 and 3 are with the original WT power data. The first two subplots show the RMS forecasting error for WFP in MW where the baseline is the results for PM. The two last subplots are the RMS on the forecasting error divided by the estimated standard deviation for this error. If this estimated standard deviation is correct the above RMS will be one on the average.

The first subplot (blue bars) shows that using the auto regressive (AR) model for WFP gives no improvement. The estimate $\hat{a}$ for the AR model is actually very close to 1, which means that this model corresponds to the persistence model. Including all WT powers in a multi-dimensional model reduces the RMS by a 6% and also improves the

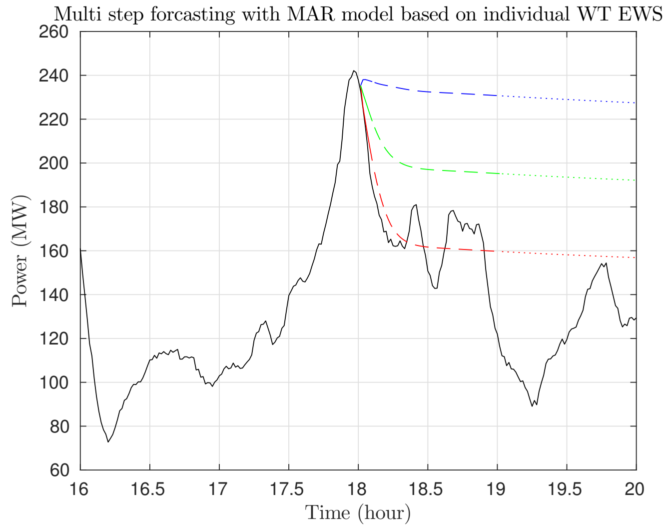

Fig. 9. Multistep prediction using the MAR method with individual WT WS inputs to forecast measured WFP from time 18:00 (2014-01-27). The blue and red curves are upper and lower 95% confidence limits respectively and the green curve is the forecast.

Abbrevation	Desription	Signal modelled	Power curve used
PM	Persistence method	WFP	No
AR-WFP	One-dimensional AR	WFP	No
MAR-WTP	Multi-dimensional AR	All WT powers	No
MAR-WTWS	Multi-dimensional AR	All WT WS	Yes

TABLE I

MODELS USED FOR PREDICTION PERFORMANCE TEST.

uncertainty as seen in subplot three. Using the MAR model for WT WS gives a 7% larger RMS compared to PM but a slightly better uncertainty estimate. When the effect of the power curve uncertainty is removed, the yellow bars shows similar results except for the MAR-WTWS, which is now superior with a reduction of RMS at 11%.

V. CONCLUSION

This research has focused on prediction of wind farm power for time horizons of 0-10 minutes and not more than one hour. This is useful for both transmission system operators, wind farm operators and traders. The idea explored in this work was to use measurements from all turbines to forecast the available farm power. To accomplish this, multi-dimensional auto regressive (MAR) models have been used. The parameter estimate for these models amount to a least squares problem with a closed form solution. The hypothesis is that the assumed correlation between wind at one WT and its upwind neighbour WTs can be exploited in such forecasting models.

Data from real wind farms have limitations. The disturbed wind speed measurements from each turbine are used to measure wind speed at turbine locations. Furthermore, the power curve of the wind turbines is not publicly available and had to be estimated.

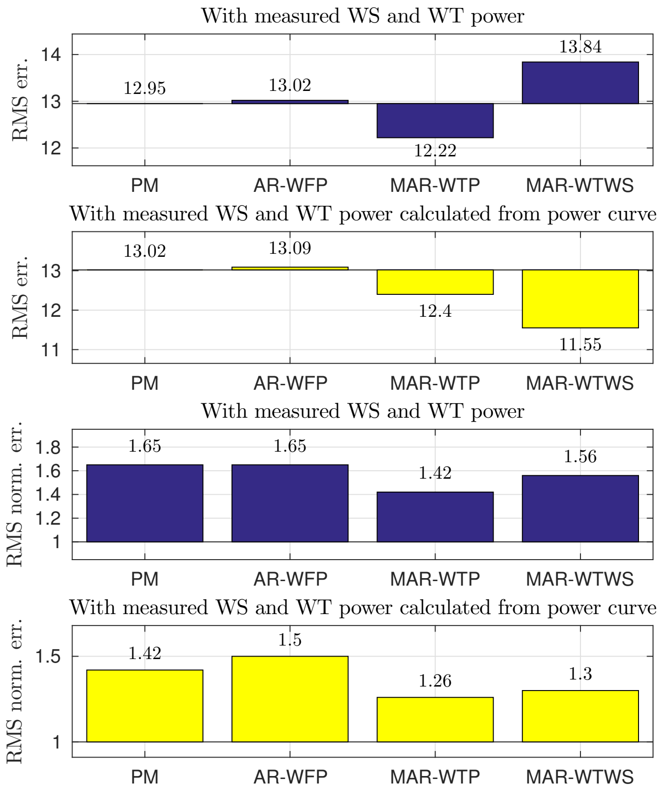

Fig. 10. Prediction performance statistics for WFP 5 minutes prediction using the methods in table I (2014-01-27). The first two subplots are RMS for the prediction error. The last two subplots are RMS for the prediction error divided by the estimated standard deviation which then ideally should be one. The blue bars are based on measured power data. The yellow bars are based on "artificial" wind turbine power (see text for details).

When avoiding wind speed and directly using measured power a modest improvement is shown as a result of using multi-dimensional models compared to one-dimensional including persistence.

The methods for modelling and forecasting wind speed, which taken through the power curve to get power forecasts, have been tested with two data sets. Based on measured wind farm power the best result for one particular day is that the MAR models perform slightly worse compared to the persistence method regarding RMS prediction error but slightly better regarding uncertainty. Based on artificial wind turbine power calculated using the estimated power curve, the result is a reduction of RMS prediction error of 11% for the multi-dimensional AR model compared to persistence.

The overall conclusion is, that in comparison to persistence there is a gain using multi-dimensional AR models. This is probably due to the correlation between turbines in the farm. The MAR model based on wind speeds with adaptive parameter estimation is the most convincing as the residual whiteness test are generally passed.

Further work in this direction should assess the effect of using the concept of effective wind speed in place of measured wind speed at the nacelle. Further results with other days and/or longer time series, especially the performance during front passages or other significant weather event will be interesting.

VI. Acknowledgment

This work has been (partially) funded by Norwegian Centre for Offshore Wind Energy (NORCOWE) under grant 193821/S60 from Research Council of Norway (RCN). NORCOWE is a consortium with partners from industry and science, hosted by Christian Michelsen Research.

Statkraft, Statoil and Green Investment Bank are acknowledged for providing the data.

References

[1] G. Giebel and G. Kariniotakis, "Best practice in short-term forecasting - a users guide," Risø National Laboratory for Sustainable Energy, DTU, Denmark, Tech. Rep., 2009.

[2] W. Glassley, J. Kleissl, C. C. van Dam, H. Shiu, J. Huang, G. Braun, and R. Holland, "California renewable energy forecasting, resource data and mapping," California Institute for Energy and Environment, Reports 500-99-013, 2010, appendix B Wind Energy Forecasting: A Review of State-of-the-Art and Recommendations for Better Forecasts.

[3] L. Landberg, "Taking the guesswork out of wind power forecasting," in *EWEA2012, Monday 16 - Thursday 19 April 2012, Copenhagen, Denmark*, EWEA. EWEA, 2012, slide presentation.

[4] P. Pinson, "Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions," *Journal of the Royal Statistical Society, Appl. Statist.*, vol. 61, pp. 555–576, 2012.

[5] Y. Shaojin, P. Yongqing, and W. Jianzhong, "Determination of kolmogorov entropy of chaotic attractor included in one-dimensional time series of meteorological data," *Advances in Atmospheric Sciences*, vol. 8, no. 2, pp. 243–250, 1991.

[6] G. Giebel, R. Brownsword, G. Kariniotakis, , M. Denhard, and C. Draxl, "The state of the art in short-term prediction of wind power - a literature overview, 2nd edition," Risø DTU, Wind Energy Division, Frederiksborgvej 399, 4000 Roskilde, Denmark, Tech. Rep., 2011, dELIVERABLE REPORT - ANEMOS.plus and SafeWind.

[7] G. Giebel and G. Kariniotakis, "Best practice in short-term forecasting - a users guide," in *Conference proceedings (online)*, EWEA. European Wind Energy Association (EWEA), 2007.

[8] A. Sapronova, C. Meissner, and M. Mana, *Improving an Accuracy of ANN-Based Mesoscale-Microscale Coupling Model by Data Categorization: With Application to Wind Forecast for Offshore and Complex Terrain Onshore Wind Farms*. Cham: Springer International Publishing, 2014, pp. 61–66. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-13290-7_5

[9] M. Soltani, T. Knudsen, M. Svenstrup, R. Wisniewski, P. Brath, R. Ortega, and K. Johnson, "Estimation of rotor effective wind speed: A comparison," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 4, pp. 1155–1167, 2013.

[10] T. Knudsen, T. Bak, and M. Svenstrup, "Survey of wind farm control - power and fatigue optimization," *Wind Energy*, vol. 18, no. 8, pp. 1333–1351, August 2015, published online 9 May 2014 in Wiley Online Library (onlinelibrary.wiley.com).

[11] L. Ljung, *System Identification, Theory for the User*, 2nd ed., ser. Prentice-Hall information and system sciences series. Prentice-Hall, 1999.

[12] H. Madsen and P. Thyregod, *Introduction to General and Generalized Linear Models*. Chapman & Hall, 2011, iSBN: 978-1-4200-9155-7.

[13] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proceedings of the IEEE 1995 American Control Conference*, vol. 3, Jun 1995, pp. 1628–1632.

[14] T. Knudsen, "Wind turbine state estimation," Automation and Control, Department of Electronic Systems, Aalborg University, Tech. Rep., 2014.

[15] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, 2006.

[16] H. Madsen, *Time Series Analysis*. Chapman & Hall, 2008, iSBN: 978-1-4200-5967-0.