ZEPHYR – THE PREDICTION MODELS

T.S. Nielsen, H. Madsen, H. Aa. Nielsen
Informatics and Mathematical Modelling,
The Technical University of Denmark,
DK-2800 Lyngby, Denmark
tsn@imm.dtu.dk
L. Landberg, G. Giebel
Risø National Laboratory
DK-4000 Roskilde, Denmark

Abstract This paper briefly describes new models and methods for predicting the wind power output from wind farms. The system is being developed in a project which has the research organization Risø and the department of Informatics and Mathematical Modelling (IMM) as the modelling team and all the Danish utilities as partners and users.
The new models are evaluated for five wind farms in Denmark as well as one wind farm in Spain. It is shown that the predictions based on conditional parametric models are superior to the predictions obtained by state-of-the-art parametric models.
Keywords: Forecasting Methods, Models (Mathematical), Adaptive Estimation, Statistics.

1 INTRODUCTION

Historically there has been two models used in Denmark to predict the power production from wind farms – the Prediktor model developed at Risø and the Wind Power Prediction Tool (WPPT) developed at IMM. Both models use weather predictions from numerical weather prediction (NWP) models as input. The way this input is used is different for the two models: *Prediktor* uses mainly physical relations to transform the predicted wind into predicted power whereas *WPPT* uses statistical models to describe the relationship between observed power production and the weather predictions. The WPPT models are all adaptive to changes in the layout of the park and to changes in the roughness of the surroundings.

This presentation describes briefly a new modelling system for wind power prediction – Zephyr – which is under development in a co-operation between Risø and IMM and all the Danish electrical utilities. The main goal of the Zephyr project is to merge Prediktor and WPPT to obtain synergy between the physical and the statistical approach.

A central part of this system is statistical models for short-term predictions of the wind power production in wind farms. Recent research has demonstrated that conditional parametric models implies a significant improvement of the prediction performance compared to more traditional parametric models. Two models for short-term predictions are outlined and the performances are compared for six different wind farms - five in Denmark and one from the Zaragoza region in Spain (La Muela). The wind farm at La Muela is investigated further in [1], where the performance of the new power prediction model described later in this paper is evaluated for various wind forecasts.

A Client/Server architecture for an implementation of the wind power prediction has been constructed as described in [2]. This implementation is build upon JAVA which enables an user interface based on internet browsers.

2 THE ESTIMATION METHOD

A method for on-line estimation is used in both WPPT and Zephyr. The non-linear dynamic relationship between wind power and the explanatory variables has been estimated using local regression and conditional parametric models, for which the estimation problem has close resemblance to that known from ordinary linear models.

The models are of the form

$$y_s = \mathbf{x}_s^T \theta(\mathbf{u_s}) + \mathbf{e_s};\ \ \mathbf{s} = \mathbf{1}, \ldots, \mathbf{N}, \tag{1}$$

where the response y_s is a stochastic variable, $\mathbf{u}_s$ and $\mathbf{x}_s$ are explanatory variables, e_s is i.i.d. $N(0, \sigma^2)$, $\theta(\cdot)$ is a vector of unknown but smooth functions with values in $\mathbf{R}^p$, and $s = 1, \ldots, N$ are observation numbers.

2.1 Off-line estimation

Estimation in the model (1) aims at estimating the functions $\theta(\cdot)$ within the space spanned by the observations of $\mathbf{u}_s$; $s = 1, \ldots, N$. The functions are only estimated for distinct values of the argument $\mathbf{u}$. Below $\mathbf{u}$ denotes one single of these fitting points and $\hat{\theta}(\mathbf{u})$ denotes the estimates of the coefficient-functions, when the functions are evaluated at $\mathbf{u}$.

One solution to the estimation problem is to replace $\theta(\mathbf{u}_s)$ in (1) with a constant vector θ_u and fit the resulting model locally to $\mathbf{u}$, using weighted least squares. Below two similar methods of allocating weights to the observations are described. For both methods the weight function $W : \mathbf{R}_0 \to \mathbf{R}_0$ is a nowhere increasing function. In this paper the tri-cube weight function

$$W(u) = \begin{cases} (1-u^3)^3, & u \in [0;1] \\ 0, & u \in [1;\infty[\end{cases} \tag{2}$$

is used. Hence, $W : \mathbf{R}_0 \to [0,1]$

In the case of a spherical kernel the weight on observation s is determined by the Euclidean distance $||\mathbf{u}_s - \mathbf{u}||$

between $\mathbf{u}_s$ and $\mathbf{u}$, i.e.

$$w_s(\mathbf{u}) = W\left(\frac{||\mathbf{u}_s - \mathbf{u}||}{h(\mathbf{u})}\right). \tag{3}$$

A product kernel is characterized by distances being calculated for one dimension at a time, i.e.

$$w_s(\mathbf{u}) = \prod_j W\left(\frac{|u_{j,s} - u_j|}{h(\mathbf{u})}\right), \tag{4}$$

where the multiplication is over the dimensions of $\mathbf{u}$. The scalar $h(\mathbf{u}) > 0$ is called the bandwidth. If $h(\mathbf{u})$ is constant for all values of $\mathbf{u}$ it is denoted a fixed bandwidth. If $h(\mathbf{u})$ is chosen so that a certain fraction (α) of the observations fulfill $||\mathbf{u}_s - \mathbf{u}|| \leq h(\mathbf{u})$ it is denoted a nearest neighbor bandwidth. If $\mathbf{u}$ has the dimension two or larger, scaling of the individual elements of $\mathbf{u}_s$ before applying the method should be considered, see e.g. [3]. Rotating the coordinate system in which $\mathbf{u}_s$ is measured may also be relevant. In this study the models have been estimated using a product kernel with a fixed bandwidth.

If the bandwidth $h(\mathbf{u})$ is sufficiently small the approximation of $\theta(\cdot)$ as a constant vector near $\mathbf{u}$ is good. This implies that a relatively low number of observations is used to estimate $\theta(\mathbf{u})$, resulting in a noisy estimate or large bias if the bandwidth is increased. See also the comments on kernel estimates in [3].

It is, however, well known that locally to $\mathbf{u}$ the elements of $\theta(\cdot)$ may be approximated by polynomials, and in many cases these will be good approximations for larger bandwidths than those corresponding to local constants. Let us describe how local polynomial approximations are used in a local least squares setting. Let $\theta_j(\cdot)$ be the j'th element of $\theta(\cdot)$ and let $\mathbf{p}_d(\mathbf{u})$ be a column vector of terms in a d-order polynomial evaluated at $\mathbf{u}$, if for instance $\mathbf{u} = [u_1\ u_2]^T$ then $\mathbf{p}_2(\mathbf{u}) = [1\ \ u_1\ \ u_2\ \ u_1^2\ \ u_1u_2\ \ u_2^2]^T$. Furthermore, let $\mathbf{x}_s = [x_{1s} \ldots x_{ps}]^T$. With

$$\mathbf{z}_s^T = \left[x_{1s}\mathbf{p}_{d(1)}^T(\mathbf{u}_s) \ldots x_{js}\mathbf{p}_{d(j)}^T(\mathbf{u}_s) \ldots x_{ps}\mathbf{p}_{d(p)}^T(\mathbf{u}_s)\right] \tag{5}$$

and

$$\hat{\phi}^T(\mathbf{u}) = [\hat{\phi}_1^T(\mathbf{u}) \ldots \hat{\phi}_j^T(\mathbf{u}) \ldots \hat{\phi}_p^T(\mathbf{u})], \tag{6}$$

where $\hat{\phi}_j(\mathbf{u})$ is a column vector of local constant estimates at $\mathbf{u}$ corresponding to $x_{js}\mathbf{p}_{d(j)}(\mathbf{u}_s)$, estimation is handled as described above, but fitting the linear model

$$y_s = \mathbf{z}_s^T \phi_u + e_s;\ \ s = 1, \ldots, N, \tag{7}$$

locally to $\mathbf{u}$. Hereafter the elements of $\theta(\mathbf{u})$ is estimated by

$$\hat{\theta}_j(\mathbf{u}) = \mathbf{p}_{d(j)}^T(\mathbf{u})\,\hat{\phi}_j(\mathbf{u});\ \ j = 1, \ldots p. \tag{8}$$

This method is identical to the method described in [3] when $\mathbf{x}_j = 1$ for all j with the exception that in [3] the elements of $\mathbf{u}_s$ used in $\mathbf{p}_d(\mathbf{u}_s)$ are centered around $\mathbf{u}$ and hence $\mathbf{p}_d(\mathbf{u}_s)$ must be recalculated for each value of $\mathbf{u}$ considered.

By combining the method with an exponential forgetting, such that the influence of older observations is downgraded, the estimation becomes adaptive such that variations in time of the system can be taken into account. The combined method can be seen as a generalization of exponential forgetting. Using matrix notation it can be shown (see [4]) that the off-line adaptive solution locally to some fixed point $\mathbf{u}$ can be expressed as

$$\hat{\phi}_t(\mathbf{u}) = \left(\mathbf{Z}_t^T \Lambda_t \mathbf{W}_{ut} \mathbf{Z}_t\right)^{-1} \mathbf{Z}_t^T \Lambda_t \mathbf{W}_{ut} \mathbf{y}_t, \tag{9}$$

where $\mathbf{W}_{ut} = \mathrm{diag}(w_u(\mathbf{u}_1), \ldots, w_u(\mathbf{u}_t))$ is a diagonal weight matrix in which the weights depend on the observations $\mathbf{u}_s$; $s = 1, \ldots, t$, and $\Lambda_t = \mathrm{diag}(\lambda^{t-1}, \lambda^{t-2}, \ldots, \lambda, 1)$ controls the weight on older observations, and thus introduces an adaptive estimation.

2.2 On-line estimation

It can be shown ([4]) that (9) is the solution to the following weighted least squares problem

$$\mathrm{Min}_\phi \sum_{s=1}^{t} \lambda^{t-s} w_u(\mathbf{u}_s)(y_s - \mathbf{z}_s^T \phi)^2. \tag{10}$$

The off-line (non-recursive) solution is shown in (9), but in the on-line implementation in the prediction tools WPPT and Zephyr the estimate is written down recursively.

Following [4] the solution can be found recursively as

$$\begin{aligned}\hat{\phi}_t(\mathbf{u}) &= \hat{\phi}_{t-1}(\mathbf{u}) + \\ &\quad w_u(\mathbf{u}_t)\mathbf{R}_{u,t}^{-1}\mathbf{z}_t\left[y_t - \mathbf{z}_t^T\hat{\phi}_{t-1}(\mathbf{u})\right].\end{aligned} \tag{11}$$

where

$$\mathbf{R}_{u,t} = \lambda \mathbf{R}_{u,t-1} + w_u(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T \tag{12}$$

It is observed that existing numerical procedures for recursive least squares estimation can be applied by replacing $\mathbf{z}_t$ and y_t with $\mathbf{z}_t\sqrt{w_u(\mathbf{u}_t)}$ and $y_t\sqrt{w_u(\mathbf{u}_t)}$, respectively.

When $\mathbf{u}_t$ is far from $\mathbf{u}$ it is clear from (12) that $\mathbf{R}_{u,t} \approx \lambda \mathbf{R}_{u,t-1}$. This may result in abruptly changing estimates if $\mathbf{u}$ is not visited regularly. This is considered a serious practical problem and consequently (12) has to be modified to ensure that the past is weighted down only when new information become available, i.e.

$$\begin{aligned}\mathbf{R}_{u,t} &= \lambda v(w_u(\mathbf{u}_t); \lambda)\mathbf{R}_{u,t-1} \\ &\quad + w_u(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T,\end{aligned} \tag{13}$$

where $v(\cdot\,;\lambda)$ is a nowhere increasing function on $[0;1]$ fulfilling $v(0;\lambda) = 1/\lambda$ and $v(1;\lambda) = 1$. Note that this requires that the weights span the interval ranging from zero to one. Here only the linear function

$$v(w;\lambda) = 1/\lambda - (1/\lambda - 1)w,$$

is considered. Thus (13) becomes

$$\begin{aligned}\mathbf{R}_{u,t} &= (1 - (1-\lambda)w_u(\mathbf{u}_t))\mathbf{R}_{u,t-1} + \\ &\quad w_u(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T.\end{aligned} \tag{14}$$

It is obvious to denote $1 - (1-\lambda)w_u(\mathbf{u}_t)$ the effective forgetting factor for point $\mathbf{u}$ at time t; $\lambda_{eff}^u(t)$. For a further discussion of the adaptive estimation see [5].

3 WIND POWER PREDICTION MODELS

This section gives first an overview of the model used in the version of WPPT which is operational in Denmark today (WPPT version 2). Later on the new model (WPPT version 3), which also will be the implemented in Zephyr is outlined.

3.1 The old parametric prediction model

The WPPT2 model (from 1999 – see [6]) utilizes local power measurements from the wind farm as well as forecasts of wind speed from the national weather service. The model is given as

$$
\begin{aligned}
p_{t+k} \quad = \quad & a_1 p_t + a_2 p_{t-1} + b_1^m w_{t+k|t}^m + b_2^m (w_{t+k|t}^m)^2 + \\
& \sum_{i=1}^{2} [c_i^c \cos \frac{2i\pi h_{t+k}^{24}}{24} + c_i^s \sin \frac{2i\pi h_{t+k}^{24}}{24}] + \\
& m + e_{t+k} \qquad (15)
\end{aligned}
$$

where p_t is the observed power at time t, $w_{t+k|t}^m$ is the forecasted wind speed at $t+k$ given at time t, h_{t+k}^{24} is time of day at time $t+k$, e_{t+k} is a noise term, and a_1, a_2, b_1^m, b_2^m, c_1^c, c_1^s and m are the time-varying model parameters which are estimated adaptively.

Predictions of the wind power with an prediction horizon from 1 hour upto 39 hours are updated every hour.

3.2 The new conditional parametric prediction model

The new semi-parametric WPPT models (WPPT3) uses semi-parametric estimates of wind direction dependant power curves in the transformation of forecasted wind speed and wind direction to power. The model is given as

$$
p_{t+k}^{pc} = f(w_{t+k|t}^m, \theta_{t+k|t}^m, k) + e_{t+k} \qquad (16)
$$

$$
\begin{aligned}
p_{t+k}^{pp} \quad = \quad & a(\theta_{t+k|t}^m, k) p_t + b(\theta_{t+k|t}^m, k) p_{t+k}^{pc} + \\
& c^c(\theta_{t+k|t}^m, k) \cos \frac{2\pi h_{t+k}^{24}}{24} + \\
& c^s(\theta_{t+k|t}^m, k) \sin \frac{2\pi h_{t+k}^{24}}{24} + e_{t+k} \qquad (17)
\end{aligned}
$$

where p_{t+k}^{pc} is the predicted power production from the power curve model, p_{t+k}^{pc} is the final power prediction where also autoregressive and diurnal effects are included, $\theta_{t+k|t}^m$ is the forecasted wind direction and f, a, b, c^c and c^s are smooth time-varying functions to be estimated as described previously.

Power curve predictions, p^{pc}, with an prediction horizon from 1 hour to 48 hours are updated every six hours whenever a new wind forecast becomes available. The final power prediction, p^{pp}, are updated every hour, but here the maximum prediction horizon dependings on the calculation time of the last wind forecast received. At present the wind forecast from DMI is available two hours after the calculations are initiate, which means that the maximum prediction horizon for the final power prediction model varies between 46 hours and 40 hours.

4 THE PREDICTION PERFORMANCE

The performance of WPPT2 and WPPT3 has been compared for five wind farms in Denmark sited at Dræby, Fjaldene, Hollandsbjerg, Rejsby and Sydthy and for a wind farm in Spain sited at La Muela in the Zaragoza region.

For the five Danish wind farms the data set consists of hourly values of observed power production as well as forecasted wind speed and wind direction from the lowest model level (level 31) of the Danish HIRLAM DKV model (17km grid size) with a prediction horizon from 1 hour to 48 hours in steps of 1 hour. The data set covers almost an entire year from 1997-05-26 01:00 to 1998-05-18 00:00. In order to exclude effects of model initialization from the results only the data from 1998-01-19 00:00 and onwards has been used in the model evaluation.

The Spanish data set consists of hourly values of observed power production for five of the wind turbines from the wind farm at La Muela and forecasted values of the 10 meter wind speed and wind direction from the Spanish HIRLAM model (0.2°grid size) with a prediction horizon from 1 hour to 24 hours in steps of 1 hour. The data set covers the period from 2000-01-31 12:00 to 2000-08-16 18:00 and again only data from the last part of the period is used in the model evaluation – here from 2000-06-16 05:00 and onwards.

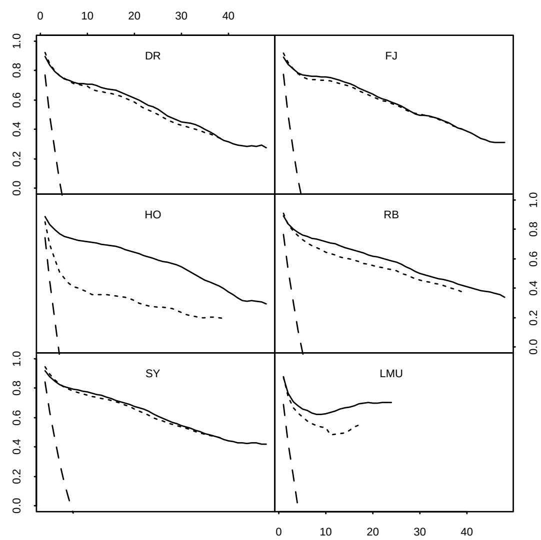

Figure 1: *Degree of explanation for WPPT3 (full line), WPPT2 (dotted line) and the naive predictor (dashed line) as a function of prediction horizon [hours]. From top left to bottom right the results are for the wind farms at Dræby (DR), Fjaldene (FJ), Hollandsbjerg (HO), Rejsby (RB), Sydthy (SY) and La Muela (LMU), Spain.*

Figure 1 summaries the prediction performance obtained for the WPPT2 and WPPT3 models as well as the naive (what you see is what you get) predictor. Degree of explanation (r^2), which describes how large a part of the variability of the observed value is explained by the prediction, is used as a performance measure. r^2 should be a number between 0 and 1 where 0 is the score obtained by the mean value predictor and 1 is the score of the perfect model, i.e. all variability of the observed value is explained by the model.

From the figure it is seen that for most of the wind farms the WPPT3 model, which also will be used in Zephyr, gives a clear improvement compared to the WPPT2 model and for no wind farms does WPPT3 perform worse that WPPT2. r^2 range from approximately 0.9 for a prediction horizon of 1 hour down to 0.45 to 0.50 for a prediction horizon of 36 hours depending on the wind farm.

The Spanish wind farm at La Muela are situated in semi-complex terrain as opposed to the Danish wind farms which all are situated in rather flat terrain. Never the less the best performance of the WPPT3 model is found for La Muela. The reason for this, at first glance unexpected result, can be found in [1], which shows that it is clearly advantages to use the forecasts of the 10 meter winds as input to the

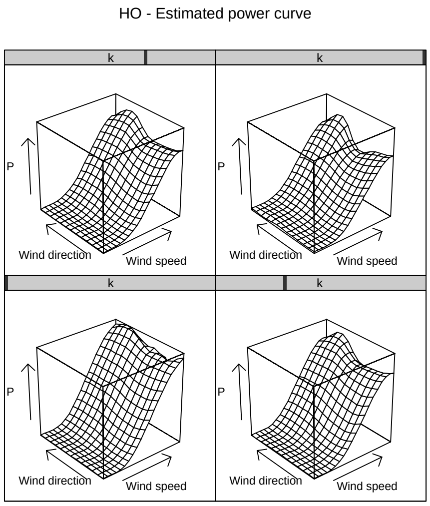

Figure 2: *The estimated power curve for Hollandbjerg. From bottom left to top right the power curves correspond to prediction horizons of 0 hours (the analysis), 12 hours, 24 hours and 36 hours.*

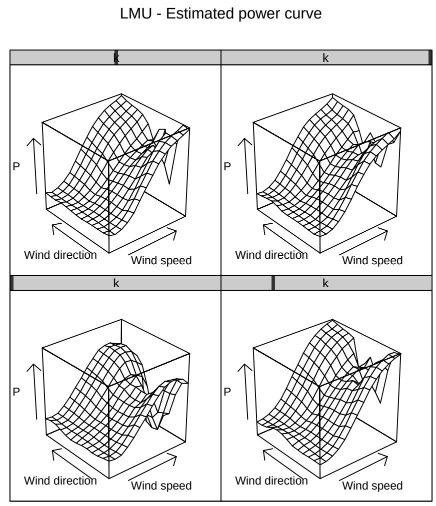

Figure 3: *The estimated power curve for La Muela. From bottom left to top right the power curves correspond to prediction horizons of 0 hours (the analysis), 6 hours, 12 hours and 24 hours.*

WPPT3 models instead of the forecasts of the model level winds.

For the two wind farms at Hollandbjerg and La Muela the score of the WPPT3 models is much better than the WPPT2 models. This can be explained by a very pronounced wind direction dependency in the estimated power curve for these two wind farms – see Figure 2 and 3, which only can be handled by the more advanced power curve model in WPPT3.

From Figure 1 it is seen that at La Muela the performance of the WPPT3 models gets better as the prediction horizon increases. Some of the improvement can be attributed to a slightly increasing performance of the wind forecasts as the prediction horizon increases, and some can be attributed to a strong diurnal variation in the wind speed (and power production) at La Muela. The model structure in the power prediction model is probably sub-optimal for a site with a strong diurnal variation and a model where p_t has been replaced with a weighted power prediction $p^w_{t+k} = w(k)p_t + (1-w(k))p_{t+k-24}$ is likely to be better suited for such sites.

5 SUMMARY

Models for short-term prediction of the wind power in wind farms are proposed and discussed. Both parametric and conditional parametric models are considered. Methods for on-line estimation of parameters in such models are briefly outlined.

The model parameters are estimated adaptively using the Recursive Least Squares algorithm with exponential forgetting in order to accommodate slow changes in the system.

The predictions based on conditional parametric models are shown to be superior to the predictions obtained by state-of-the-art parametric models. The degree of explanation varies from 0.90 for a one-hour prediction horizon to 0.45 to 0.50 for a 36 hour prediction horizon.

References

[1] I. Marti, T. S. Nielsen, H. Madsen, J. Navarro, and C. G. Barquero. Prediction models in complex terrain. In *Proceedings of the European Wind Energy Conference*, Copenhagen, Denmark, 2001.

[2] G. Giebel, L. Landberg, T. S. Nielsen, and H. Madsen. Zephyr – the next generation prediction system. In *Proceedings of the European Wind Energy Conference*, Copenhagen, Denmark, 2001.

[3] William S. Cleveland and Susan J. Devlin. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83:596–610, 1988.

[4] H. Aa. Nielsen, T. S. Nielsen, A. Joensen, H. Madsen, and J. Holst. Tracking time-varying coefficient functions. *International Journal of Adaptive Control and Signal Processing*, 14:813–828, 2000.

[5] A. Joensen, H. Madsen, H. Aa. Nielsen, and T. S. Nielsen. Tracking time-varying parameters using local regression. *Automatica*, 36:1199–1204, 1999.

[6] T. S. Nielsen, H. Madsen, H. Aa. Nielsen, and J. Tøfting. Using meteorological forecasts in on-line predictions of wind power. Technical report, Eltra, Fredericia, Denmark, 1999.