PREDICTION OF WIND POWER USING TIME-VARYING COEFFICIENT-FUNCTIONS

T.S. Nielsen * **H. Aa. Nielsen** * **H. Madsen** *

* *Informatics and Mathematical Modelling,*
The Technical University of Denmark,
DK-2800 Lyngby, Denmark

Abstract: A method for adaptive and recursive estimation in a class of non-linear autoregressive models with external input is proposed. The model class considered is conditionally parametric ARX-models (CPARX-models), which is conventional ARX-models in which the parameters are replaced by smooth, but otherwise unknown, functions of a low-dimensional input process. These coefficient-functions are estimated adaptively and recursively without specifying a global parametric form, i.e. the method allows for on-line tracking of the coefficient-functions. The usefulness of the method is illustrated using prediction of power production from wind farms as an example. A CPARX model for predicting the power production is suggested and the coefficient-functions are estimated using the proposed method. 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 previously identified parametric models.

Keywords: Adaptive algorithms; Recursive algorithms; Estimation algorithms, Nonlinear models; Time-varying systems; Nonparametric regression; Forecasts; Windmills.

1. INTRODUCTION

The conditional parametric ARX-model (CPARX-model) is a non-linear model formulated as a linear ARX-model in which the parameters are replaced by smooth, but otherwise unknown, functions of one or more explanatory variables. These functions are called coefficient-functions. In (Nielsen, Nielsen & Madsen 1997) this class of models is used in relation to district heating systems to model the non-linear dynamic response of network temperature on supply temperature and flow at the plant. A particular feature of district heating systems is, that the response on supply temperature depends on the flow. This is modelled by describing the relation between temperatures by an ARX-model in which the coefficients depend on the flow.

For on-line applications it is advantageous to allow the function estimates to be modified as data become available. Furthermore, because the system may change slowly over time, observations should be down-weighted as they become older. For this reason a time-adaptive and recursive estimation method is proposed. Essentially, the estimates at each time step are the solution to a set of weighted least squares regressions and therefore the estimates are unique under quite general conditions. For this reason the proposed method provides a simple way to perform adaptive and recursive estimation in a class of non-linear models. The method is a combination of the recursive least squares with exponential forgetting (Ljung & Söderström 1983) and locally weighted polynomial regression (Cleveland & Devlin 1988). In the paper *adaptive estimation* is used to denote, that old observations are down-weighted, i.e. in the sense of *adaptive in time*.

The method is illustrated using prediction of power production from wind farms as an example. Conditional parametric models are used to describe the relationship between observed power production and meteorological forecasts of wind speed and wind direction – the power curve – as well as the wind direction dependency in the dynamic behavior of a wind farm. These relationships are difficult to parametrize explicitly, but can, as it will be shown, readily be captured by conditional parametric models.

The time-adaptivity of the estimation is an important property in this application of the method as the total system consisting of wind farm, surroundings and numerical weather prediction (NWP) model will be subject to changes over time. This is caused by effects such as aging of the wind turbines, changes in the surrounding vegetation and maybe most importantly due to changes in the NWP models used by the weather service.

The proposed models are implemented in an on-line application for wind power prediction - the Wind Power Prediction Tool (WPPT) - which is used operationally by several of the Danish electrical utilities. WPPT has previously used more traditional (linear) parametric models for power prediction but it will be shown that conditional parametric models implies a significant improvement of the prediction performance compared to more traditional parametric models. The two models for short-term prediction 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 (Marti, Nielsen, Madsen, Navarro & Barquero 2001), where the performance of the new power prediction model is evaluated for various wind forecasts.

2. MODEL AND ESTIMATION METHOD

When using a conditional parametric model to model the response y_s the explanatory variables are split in two groups. One group of variables $\mathbf{x}_s$ enter globally through coefficients depending on the other group of variables $\mathbf{u}_s$, i.e.

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

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, and $s = 1, \ldots, N$ are observation numbers.

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

$$\hat{\theta}(\mathbf{u}) = \underset{\theta_u}{\operatorname{argmin}} \sum_{s=1}^{N} w_u(\mathbf{u}_s)(y_s - \mathbf{x}_s^T \theta_u)^2. \tag{2}$$

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

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

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.

$$w_s(\mathbf{u}) = W\left(\frac{||\mathbf{u}_s - \mathbf{u}||}{h(\mathbf{u})}\right). \tag{4}$$

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

$$w_s(\mathbf{u}) = \prod_j W\left(\frac{|u_{j,s} - u_j|}{h(\mathbf{u})}\right), \tag{5}$$

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. (Cleveland & Devlin 1988). 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 (Cleveland & Devlin 1988).

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 $\mathbf{p}_2(\mathbf{u}) = [1 \; u_1 \; u_2 \; u_1^2 \; u_1 u_2 \; u_2^2]^T$. Furthermore, let $\mathbf{x}_s = [x_{1s} \ldots x_{ps}]^T$. With

$$\mathbf{z}_s^T = \left[x_{1,s}\mathbf{p}_{d(1)}^T(\mathbf{u}_s) \ldots x_{p,s}\mathbf{p}_{d(p)}^T(\mathbf{u}_s)\right] \tag{6}$$

and

$$\hat{\phi}^T(\mathbf{u}) = [\hat{\phi}_1^T(\mathbf{u}) \ldots \hat{\phi}_j^T(\mathbf{u}) \ldots \hat{\phi}_p^T(\mathbf{u})], \tag{7}$$

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

$$y_s = \mathbf{z}_s^T \phi(u) + e_s; \;\; s = 1, \ldots, N, \tag{8}$$

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

$$\hat{\theta}_j(\mathbf{u}) = \mathbf{p}_{d(j)}^T(\mathbf{u})\, \hat{\phi}_j(\mathbf{u}); \;\; j = 1, \ldots p. \tag{9}$$

This method is identical to the method described in (Cleveland & Devlin 1988) when $\mathbf{x}_j = 1$ for all j with the exception that in (Cleveland & Devlin 1988) 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.

Interpolation is used for approximating the estimates of the coefficient-functions for other values of the arguments than the fitting points. This interpolation should only have marginal effect on the estimates. Therefore, it sets requirements on the number and placement of the fitting points. If a nearest neighbour bandwidth is used it is reasonable to select the fitting points according to the density of the data as it is done when using *k-d* trees (Chambers & Hastie 1991, Section 8.4.2). However, in this paper the approach is to select the fitting points on an equidistant grid and ensure that several fitting points are within the (smallest) bandwidth so that linear interpolation can be applied safely.

3. ADAPTIVE ESTIMATION

As pointed out in the previous section local polynomial estimation can be viewed as local constant estimation in a model derived from the original model. This observation forms the basis of the method suggested. For simplicity the adaptive estimation method is described as a generalization of exponential forgetting. However, the more general forgetting methods described by Ljung & Söderström (1983) could also serve as a basis.

Using exponential forgetting and assuming observations at time $s = 1, \ldots, t$ are available, the adaptive least squares estimate of the parameters ϕ relating the explanatory variables $\mathbf{z}_s$ to the response y_s using the linear model $y_s = \mathbf{z}_s^T \phi + e_s$ is found as

$$\hat{\phi}_t = \operatorname*{argmin}_{\phi} \sum_{s=1}^{t} \lambda^{t-s} (y_s - \mathbf{z}_s^T \phi)^2, \tag{10}$$

where $0 < \lambda < 1$ is called the forgetting factor, see also (Ljung & Söderström 1983). The estimate can be seen as a local constant approximation in the direction of time. This suggests that the estimator may also be defined locally with respect to some other explanatory variables $\mathbf{u}_t$. If the estimates are defined locally to a fitting point $\mathbf{u}$, the adaptive estimate corresponding to this point can be expressed as

$$\hat{\phi}_t(\mathbf{u}) = \operatorname*{argmin}_{\phi_u} \sum_{s=1}^{t} \lambda^{t-s} w_u(\mathbf{u}_s)(y_s - \mathbf{z}_s^T \phi_u)^2, \tag{11}$$

Following (Nielsen, Nielsen, Joensen, Madsen & Holst 2000) the solution to (11) can be found recursively as

$$\begin{aligned} \hat{\phi}_t(\mathbf{u}) = {} & \hat{\phi}_{t-1}(\mathbf{u}) + \\ & w_u(\mathbf{u}_t)\mathbf{R}_{u,t}^{-1}\mathbf{z}_t \left[y_t - \mathbf{z}_t^T \hat{\phi}_{t-1}(\mathbf{u}) \right]. \end{aligned} \tag{12}$$

where

$$\mathbf{R}_{u,t} = \lambda \mathbf{R}_{u,t-1} + w_u(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T \tag{13}$$

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 (13) 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 (13) has to be modified to ensure that the past is weighted down only when new information become available, i.e.

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

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 (14) becomes

$$\begin{aligned} \mathbf{R}_{u,t} = {} & (1 - (1-\lambda)w_u(\mathbf{u}_t))\mathbf{R}_{u,t-1} + \\ & w_u(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T. \end{aligned} \tag{15}$$

It is resonable to denote

$$\lambda_{eff}^u(t) = 1 - (1-\lambda)w_u(\mathbf{u}_t) \tag{16}$$

the *effective forgetting factor* for point $\mathbf{u}$ at time t. For a further discussion of adaptive estimation of conditional parametric models see (Joensen, Madsen, Nielsen & Nielsen 1999).

3.1 *Summary of the method*

To clarify the method the actual algorithm is briefly described in this section. It is assumed that at each time step t measurements of the output y_t and the two sets of inputs $\mathbf{x}_t$ and $\mathbf{u}_t$ are received. The aim is to obtain adaptive estimates of the coefficient-functions in the non-linear model (1).

Besides λ in (13), prior to the application of the algorithm a number of fitting points $\mathbf{u}^{(i)};\ i=1,\ldots,n_{fp}$ in which the coefficient-functions are to be estimated has to be selected. Furthermore the bandwidth associated with each of the fitting points $h^{(i)};\ i=1,\ldots,n_{fp}$ and the degrees of the approximating polynomials $d(j);\ j=1,\ldots,p$ have to be selected for each of the p coefficient-functions. For simplicity the degree of the approximating polynomial for a particular coefficient-function will be fixed across fitting points. Finally, initial estimates of the coefficient-functions in the model corresponding to local constant estimates, i.e. $\hat{\phi}_0(\mathbf{u}^{(i)})$, must be chosen. Also, the matrices $\mathbf{R}_{u^{(i)},0}$ must be chosen. One possibility is $\mathrm{diag}(\varepsilon,\ldots,\varepsilon)$, where ε is a small positive number.

In the following description of the algorithm it will be assumed that $\mathbf{R}_{u^{(i)},t}$ is non-singular for all fitting points. In practice we would just stop updating the estimates if the matrix become singular. Under the assumption mentioned the algorithm can be described as:

For each time step t: Loop over the fitting points $\mathbf{u}^{(i)};\ i=1,\ldots,n_{fp}$ and for each fitting point:

- Construct the explanatory variables corresponding to local constant estimates using (6): $\mathbf{z}_t^T=[x_{1,t}\mathbf{p}_{d(1)}^T(\mathbf{u}_t)\ldots x_{p,t}\mathbf{p}_{d(p)}^T(\mathbf{u}_t)]$.
- Calculate the weight using e.g. (4) and (3): $w_{u^{(i)}}(\mathbf{u}_t)=(1-(||\mathbf{u}_t-\mathbf{u}^{(i)}||/h^{(i)})^3)^3$, if $||\mathbf{u}_t-\mathbf{u}^{(i)}||<h^{(i)}$ and zero otherwise.
- Find the effective forgetting factor using (16): $\lambda_{eff}^{(i)}(t)=1-(1-\lambda)w_{u^{(i)}}(\mathbf{u}_t)$.
- Update $\mathbf{R}_{u^{(i)},t-1}$ using (15): $\mathbf{R}_{u^{(i)},t}=\lambda_{eff}^{(i)}(t)\mathbf{R}_{u^{(i)},t-1}+w_{u^{(i)}}(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T$.
- Update $\hat{\phi}_{t-1}(\mathbf{u}^{(i)})$ using (12): $\hat{\phi}_t(\mathbf{u}^{(i)})=\hat{\phi}_{t-1}(\mathbf{u}^{(i)})$ $+w_{u^{(i)}}(\mathbf{u}_t)\mathbf{R}_{u^{(i)},t}^{-1}\mathbf{z}_t\left[y_t-\mathbf{z}_t^T\hat{\phi}_{t-1}(\mathbf{u}^{(i)})\right]$.
- Calculate the updated local polynomial estimates of the coefficient-functions using (9): $\hat{\theta}_{j,t}(\mathbf{u}^{(i)})=\mathbf{p}_{d(j)}^T(\mathbf{u}^{(i)})\,\hat{\phi}_{j,t}(\mathbf{u}^{(i)});\ \ j=1,\ldots p$

The algorithm could also be implemented using the matrix inversion lemma as in (Ljung & Söderström 1983).

4. WIND POWER PREDICTION MODELS

The development of the Wind Power Prediction Tool (WPPT) began in 1992 and the first test version was installed at a Danish power utility in 1995. WPPT went into operational use in 1998 and has since then been used operationally by must of the Danish power utilities. During WPPT's life time several studies have been carried out to improve the performance of the power prediction models. Much effort have been dedicated to make best possible use of the available meteorological forecasts e.g. by introducing wind direction dependency in the power curve and employing additional explanatory variables besides forecasted wind speed and wind direction. This section first gives 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 4) is outlined.

The WPPT2 model (from 1999 – see (Nielsen, Madsen, Nielsen & Tøfting 1999)) was identified on basis of data from the same five Danish wind farms as is used in this paper. The model utilizes local power measurements from the wind farm as well as forecasts of wind speed from the national weather service. That is the relationship power production and forecasted wind speed is independent of forecasted wind direction. The model is given as

$$\begin{aligned}p_{t+k} = {} & a_1 p_t + a_2 p_{t-1} + b_1^m w_{t+k|t}^m + b_2^m (w_{t+k|t}^m)^2 + \\ & \sum_{i=1}^{2}[c_i^c \cos\frac{2i\pi h_{t+k}^{24}}{24} + c_i^s \sin\frac{2i\pi h_{t+k}^{24}}{24}] + \\ & m + e_{t+k} \end{aligned} \tag{17}$$

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 up to 39 hours are updated every hour.

The new WPPT models (WPPT4) uses conditional parametric estimates of wind direction dependent power curves in the transformation of forecasted wind speed and wind direction to power. The model is given as

$$p_{t+k}^{pc} = f(w_{t+k|t}^m, \theta_{t+k|t}^m, k) + e_{t+k} \tag{18}$$

$$\begin{aligned}p_{t+k}^{pp} = {} & a(\theta_{t+k|t}^m,k)p_t + b(\theta_{t+k|t}^m,k)p_{t+k}^{pc} + \\ & c^c(\theta_{t+k|t}^m,k)\cos\frac{2\pi h_{t+k}^{24}}{24} + \\ & c^s(\theta_{t+k|t}^m,k)\sin\frac{2\pi h_{t+k}^{24}}{24} + e_{t+k}\end{aligned} \tag{19}$$

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 depends on the calculation time of the last wind forecast received. At present the wind forecast from the Danish Meteorological Institute (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.

5. THE PREDICTION PERFORMANCE

The performance of WPPT2 and WPPT4 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 onward 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 (17km 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 onward.

Figure 1 summaries the prediction performance obtained for the WPPT2 and WPPT4 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 WPPT4 model gives a clear improvement compared to the WPPT2 model and for no wind farms does WPPT4 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 WPPT4 model is found for La Muela. The reason for this, at first glance

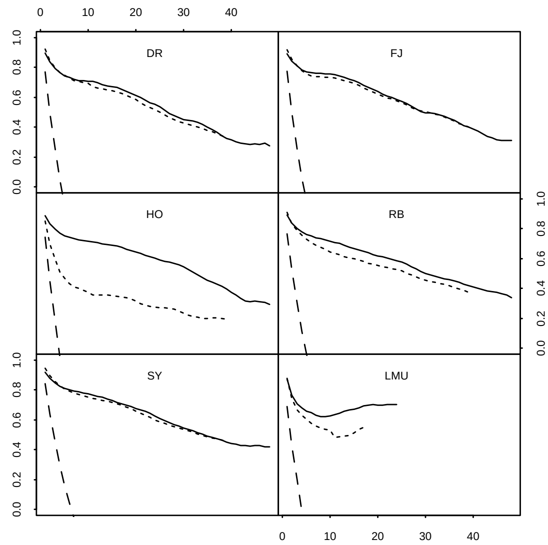

Fig. 1. Degree of explanation for WPPT4 (full line), WPPT2 (dotted line) and the naive predictor (dashed line) as a function of prediction horizon [hours]. From top left to bottom right the results are for the wind farms at Dræby (DR), Fjaldene (FJ), Hollandsbjerg (HO), Rejsby (RB), Sydthy (SY) and La Muela (LMU), Spain.

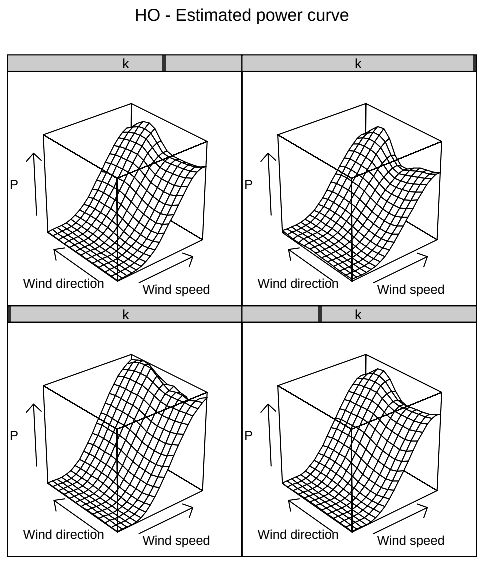

Fig. 2. The estimated power curve for Hollandsbjerg. From bottom left to top right the power curves correspond to prediction horizons of 0 hours (the analysis), 12 hours, 24 hours and 36 hours.

unexpected result, can be found in (Marti et al. 2001), which shows that it is clearly advantages to use the forecasts of the 10 meter winds as input to the WPPT4 models instead of the forecasts of the model level winds.

For the two wind farms at Hollandsbjerg and La Muela the score of the WPPT4 models is much better than the WPPT2 models. This can be explained by a very

LMU - Estimated power curve

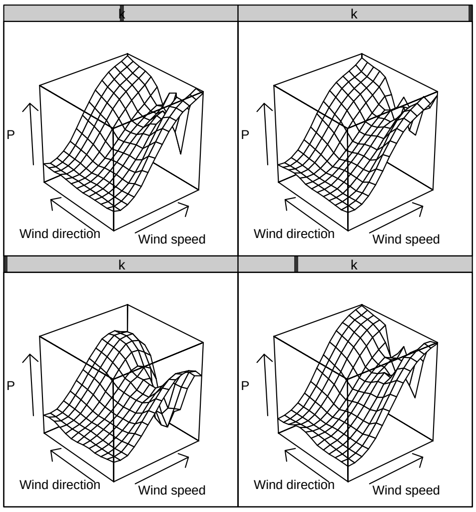

Fig. 3. The estimated power curve for La Muela. From bottom left to top right the power curves correspond to prediction horizons of 0 hours (the analysis), 6 hours, 12 hours and 24 hours.

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 WPPT4.

From Figure 1 it is seen that at La Muela the performance of the WPPT4 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.

6. SUMMARY

In this paper methods for adaptive and recursive estimation in a class of non-linear autoregressive models with external input are proposed. The model class considered is conditionally parametric models, which is a conventional linear model in which the parameters are replaced by smooth, but otherwise unknown, functions of a low-dimensional input process. These functions are estimated adaptively and recursively without specifying a global parametric form.

The methods can be seen as generalizations or combinations of recursive least squares with exponential forgetting (Ljung & Söderström 1983), local polynomial regression (Cleveland & Devlin 1988), and conditional parametric fits (Anderson, Fang & Olkin 1994). Hence, the methods constitutes an extension to the notion of local polynomial estimation.

The method is illustrated using power prediction for wind farms as an example. Both parametric and conditional parametric models are considered. 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.

7. REFERENCES

Anderson, T. W., Fang, K. T. & Olkin, I., eds (1994), *Multivariate Analysis and Its Applications*, Institute of Mathematical Statistics, Hayward, chapter Coplots, Nonparametric Regression, and conditionally Parametric Fits, pp. 21–36.

Chambers, J. M. & Hastie, T. J., eds (1991), *Statistical Models in S*, Wadsworth, Belmont, CA.

Cleveland, W. S. & Devlin, S. J. (1988), 'Locally weighted regression: An approach to regression analysis by local fitting', *Journal of the American Statistical Association* **83**, 596–610.

Joensen, A., Madsen, H., Nielsen, H. A. & Nielsen, T. S. (1999), 'Tracking time-varying parameters using local regression', *Automatica* **36**, 1199–1204.

Ljung, L. & Söderström, T. (1983), *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA.

Marti, I., Nielsen, T. S., Madsen, H., Navarro, J. & Barquero, C. G. (2001), Prediction models in complex terrain, *in* 'Proceedings of the European Wind Energy Conference', Copenhagen, Denmark.

Nielsen, H. A., Nielsen, T. S., Joensen, A., Madsen, H. & Holst, J. (2000), 'Tracking time-varying coefficient functions', *International Journal of Adaptive Control and Signal Processing* **14**, 813–828.

Nielsen, H. A., Nielsen, T. S. & Madsen, H. (1997), Arx-models with parameter variations estimated by local fitting, *in* 'Proceedings of 11th IFAC Symposium on System Identification', Vol. 2, Kitakyushu, Japan, pp. 475–480.

Nielsen, T. S., Madsen, H., Nielsen, H. A. & Tøfting, J. (1999), Using meteorological forecasts in online predictions of wind power, Technical report, Eltra, Fredericia, Denmark.