PREDICTION OF REGIONAL WIND POWER

T.S. Nielsen[1], H. Madsen[1], H. Aa. Nielsen[1], L. Landberg[2], G. Giebel[2]
[1]Informatics and Mathematical Modelling,
The Technical University of Denmark,
DK-2800 Lyngby, Denmark
Tel: +45 4525 3428, Fax: +45 4588 2673, E-mail: tsn@imm.dtu.dk
[2]Department of Wind Energy
Risø National Laboratory
DK-4000 Roskilde, Denmark

ABSTRACT This paper presents a new concept for predicting the total wind power production in a larger region based on a combination of on-line measurements of power production from selected wind farms, power measurements for all wind turbines in the area and numerical weather predictions of wind speed and wind direction. The models are implemented in the Zephyr/WPPT system – an on-line software system for calculating short-term predictions of wind power currently being developed by IMM and Risø in coorporation with Elsam, Eltra, Elkraft and SEAS – the major electrical utilities with respect to wind power in Denmark.
Zephyr/WPPT employs statistical models to describe the relationship between power production and the numerical weather predictions. The statistical models belong to the class of conditional parametric models – a model class particular useful for estimating non-linear relationships on-line. The estimation is furthermore made adaptively in order to allow for slow changes in the system e.g. caused by the annual variations of the climate.
Keywords: Forecasting Methods, Models (Mathematical), Adaptive Estimation, Statistics.

1 INTRODUCTION

The amount of wind power installed is increasing rapidly during these years and today wind power constitutes a substantial part of the total installed power production capacity in some regions.

It is clear, however, that in order to incorporate a substantial wind power production efficiently and economically into the existing production system reliable short-term predictions of the available wind power are a necessity.

This paper presents a new concept for predicting the total wind power production in a larger region based on a combination of on-line measurements of power production from selected wind farms, power measurements for all wind turbines in the area and numerical weather predictions of wind speed and wind direction. The models are implemented in the Zephyr/WPPT wind power prediction system [1] – an on-line software system for calculating short-term predictions of wind power currently being developed by IMM and Risø in coorporation with Elsam, Eltra, Elkraft and SEAS – the major electrical utilities with respect to wind power in Denmark.

If necessary the total region is broken into a number of sub-areas. The predictions for the total region are then calculated using a two branch approach as illustrated in figure 1.

- In the first model branch predictions of wind power are calculated for a number of wind farm using on-line measurements of power production as well as numerical weather predictions as input. The prediction of the total power production in the area is calculated by upscaling the sum of the predictions for the individual wind farms.
- The second model branch predicts the area power production explicitly by using a model linking off-line measurements of area power production to the numerical weather predictions.

For both model branches the power prediction for the total region is calculated as a sum of the predictions for the sub-areas. The final prediction of the wind power production for the total region is then calculated as a weighted average of the predictions from the two model branches.

A central part of this system is statistical models for short-term predictions of the wind power production in wind farms or areas. Recent research has demonstrated that conditional parametric models implies a significant improvement of the prediction performance compared to more traditional parametric models [2].

The conditional parametric is a non-linear model formulated as a linear 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. 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 applied. 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 described 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 [3] and locally weighted polynomial regression [4]. 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 time-adaptivity of the estimation is an important property in this application of the method as the total system consisting of wind farm or area, 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 as well as changes in the population of wind turbines in the wind farm or area.

2 MODEL OVERVIEW

The Zephyr/WPPT modelling system described in the following calculates predictions of the available wind power from wind turbines in a region. For a larger region this is done by separating the region into a number of sub-areas. Wind power predictions are then calculated for each sub-area and hereafter summarized to get a prediction for the total region.

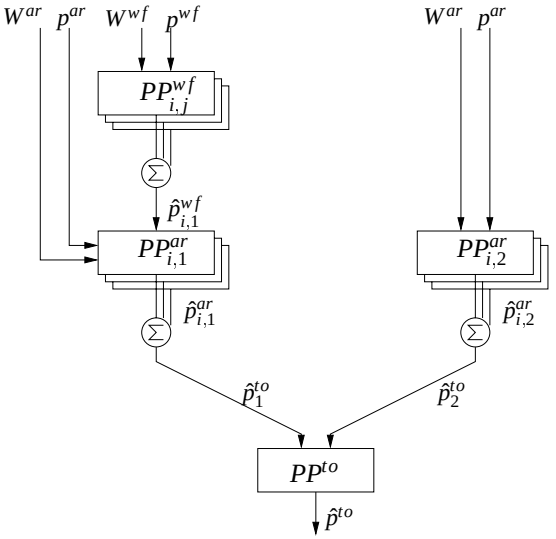

Figure 1: Overview of the model structure in Zephyr/WPPT. Two different predictions are calculated for the wind power production in a region: In the left model branch the wind farm models, $PP^{wf}_{i,j}$, are used to calculate power predictions for the reference wind farms in sub-area i. The predictions for the reference wind farms in sub-area i are summarized to $\hat{p}^{wf}_{i,1}$, which hereafter is upscaled by the upscaling model $PP^{ar}_{i,1}$ to a power prediction, $\hat{p}^{ar}_{i,1}$, for all wind turbines in the sub-area. The predictions for the sub-areas are then summarized to get the power prediction of the left model branch for the total region, $\hat{p}^{to}_1$. In the right model branch power predictions of the power production in sub-area i, $\hat{p}^{ar}_{i,2}$, are calculated directly by the area model $PP^{ar}_{i,2}$. The predictions for the sub-areas are then summarized to get the power prediction of the right model branch for the total region, $\hat{p}^{to}_2$. The final power prediction for the region, $\hat{p}^{to}$, is calculated by model $\hat{p}^{to}$ as a weighted average of the predictions from the two model branches.

The predictions are calculated using on-line production data from a number of wind farms in the area (reference wind farms), off-line production data for the remaining wind turbines in the area and numerical weather predictions of wind speed and wind direction covering the area. The predictions covers a horizon corresponding to the prediction horizon of the numerical weather predictions hours – typical from 0 to approximately 48 hours ahead in time. The time resolution of the predictions can be chosen freely but a reasonable choice for the longer prediction horizons is to use the same time resolution as the numerical weather predictions.

The predictions for the total region are calculated using a two branch approach as illustrated in figure 1.

- In the left model branch predictions of wind power are calculated for a number of reference wind farm using on-line measurements of power production as well as numerical weather predictions as input (see section 3.1). The predictions from the reference wind farms in a sub-area are summarized and hereafter upscaled to get the prediction of power production of all wind turbines in the sub-area (see section 3.2). This model branch takes advantage of the auto-correlation which is present in the power production for prediction horizons less than approximately 12 hours.

- The right model branch predicts the power production in a sub-area explicitly by using a model linking off-line measurements of total power production in the sub-area to the numerical weather predictions (see section 3.3). This model branch takes advantage of the smooth properties of the total production as well as the fact that the numerical weather models perform well in predicting the weather patterns but less well in predicting the local weather at a particular wind farm.

For both model branches the power prediction for the total region is calculated as a sum of the predictions for the sub-areas. The final prediction of the wind power production for the total region is then calculated as a weighted average of the predictions from the model two branches (see section 3.4).

3 PREDICTION MODELS

Conditional parametric models are used to describe the relationship between observed power production in wind farms or areas 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 is shown in [2], readily be captured by conditional parametric models.

3.1 The wind farm model ($PP^{wf}_{i,j}$)

The wind farm model uses wind direction dependent power curves in the transformation of forecasted wind speed and wind direction to power. The model for the jth wind farm in the ith sub-area is given as

$$\begin{aligned}\hat{p}^{pc}_{i,j}(t+k) &= f(w^{wf}_{i,j}(t+k), \theta^{wf}_{i,j}(t+k), k) \\ \hat{p}^{wf}_{i,j}(t+k) &= a(\theta^{wf}_{i,j}(t+k), k)\, p^{wf}_{i,j}(t) + \\ & \quad b(\theta^{wf}_{i,j}(t+k), k)\, \hat{p}^{pc}_{i,j}(t+k) + h^{24}(\theta^{wf}_{i,j}(t+k), k) \quad (1)\end{aligned}$$

where $p^{wf}_{i,j}(t)$ is the observed power at time t, $w^{wf}_{i,j}(t+k)$ and $\theta^{wf}_{i,j}(t+k)$ are local forecasts of wind speed and wind direction, respectively, and f, a, b, and h^{24} are smooth time-varying functions to be estimated. The difference between observed and forecasted diurnal variation of wind speed is contain in the h^{24} term.

The wind farm model takes advantage of the autocorrelation which is present in the power production for prediction horizons less than approximately 12 hours.

In [2] the performance of the proposed model is evaluated 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 [5] and [6], where the performance of the wind farm model is evaluated for various wind forecasts.

3.2 The upscaling model ($PP^{ar}_{i,1}$)

The predicted power production in sub-area i is calculated by multiplying the summarized power predictions for the wind farms in the sub-area by a upscaling function, which depends on area forecasts of wind speed and wind direction. The model is given as

$$\hat{p}^{ar}_{i,1}(t+k) = b(w^{ar}_i(t+k), \theta^{ar}_i(t+k), k) \sum_j \hat{p}^{wf}_{i,j}(t+k) \quad (2)$$

where $w^{ar}_i(t+k)$ and $\theta^{ar}_i(t+k)$ are area forecasts of wind speed and wind direction, respectively, and b is a smooth time-varying function to be estimated.

3.3 The area model ($PP^{ar}_{i,2}$)

The area model transforms area forecasts of wind speed and wind direction to power in a way similar to the wind farm power curve model by explicitly linking weather forecasts for the area to off-line observations of the power production in the area. For sub-area i the model is given as

$$\hat{p}^{ar}_{i,2}(t+k) = f(w^{ar}_i(t+k), \theta^{ar}_i(t+k), k). \quad (3)$$

where f is a smooth time-varying function to be estimated.

This model takes advantage of the smooth properties of summarized power productions and the fact that the numerical weather models perform well in predicting the weather patterns but less well in predicting the local weather at a particular wind farm.

3.4 The total model (PP^{to})

The prediction of the total power production in the region is calculated using the total predictions from the two model branches in figure 1. The prediction is calculated as a prediction horizon dependent weighted average of the power predictions for the two model branches using Root Mean Square (RMS) as weighting criterion. The model is given as

$$\hat{p}^{to}_{t+k} = b_1(k)\hat{p}^{ar}_1(t+k) + b_2(k)\hat{p}^{ar}_2(t+k) \quad (4)$$

where $\hat{p}^{ar}_1(t+k)$ and $\hat{p}^{ar}_1(t+k)$ are the power predictions for model branch 1 and 2, respectively, and b_1 and b_2 are smooth time-varying functions to be estimated.

The predictions from the two model branches are closely correlated especially for the longer prediction horizons. Thus a regularized estimation procedure must be used to ensure stable estimates of the b_1 and b_2 functions. Here Ridge Regression [7] have been used.

4 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.

$$y_s = \mathbf{x}_s^T \theta(\mathbf{u}_s) + e_s; \;\; s = 1, \ldots, N, \quad (5)$$

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 (5) 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 (5) with a constant vector θ_u and fit the resulting model locally to $\mathbf{u}$, using weighted least squares

$$\hat{\theta}(\mathbf{u}) = \underset{\theta_u}{\operatorname{argmin}} \sum_{s=1}^{N} w_u(\mathbf{u}_s)(y_s - \mathbf{x}_s^T \theta_u)^2. \quad (6)$$

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

$$W(u) = \begin{cases} (1-u^3)^3, & u \in [0;1] \\ 0, & u \in [1;\infty[\end{cases} \quad (7)$$

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.

$$w_s(\mathbf{u}) = W\left(\frac{||\mathbf{u}_s - \mathbf{u}||}{h(\mathbf{u})}\right). \quad (8)$$

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

$$w_s(\mathbf{u}) = \prod_j W\left(\frac{|u_{j,s} - u_j|}{h(\mathbf{u})}\right), \quad (9)$$

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. [4]. 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 [4].

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 $\mathbf{x}_s = [x_{1s}\dots x_{ps}]^T$. With

$$\mathbf{z}_s^T = \left[x_{1,s}\mathbf{p}_{d(1)}^T(\mathbf{u}_s)\dots x_{p,s}\mathbf{p}_{d(p)}^T(\mathbf{u}_s)\right] \tag{10}$$

and

$$\hat{\phi}^T(\mathbf{u}) = [\hat{\phi}_1^T(\mathbf{u})\dots\hat{\phi}_j^T(\mathbf{u})\dots\hat{\phi}_p^T(\mathbf{u})], \tag{11}$$

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

$$y_s = \mathbf{z}_s^T\phi(u) + e_s;\ \ s = 1,\dots,N, \tag{12}$$

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

$$\hat{\theta}_j(\mathbf{u}) = \mathbf{p}_{d(j)}^T(\mathbf{u})\,\hat{\phi}_j(\mathbf{u});\ \ j = 1,\dots p. \tag{13}$$

This method is identical to the method described in [4] when $\mathbf{x}_j = 1$ for all j with the exception that in [4] 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 [8, 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.

5 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 [3] could also serve as a basis.

Using exponential forgetting and assuming observations at time $s = 1,\dots,t$ 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

$$\hat{\phi}_t = \underset{\phi}{\operatorname{argmin}} \sum_{s=1}^{t} \lambda^{t-s}(y_s - \mathbf{z}_s^T\phi)^2, \tag{14}$$

where $0 < \lambda < 1$ is called the forgetting factor, see also [3]. 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

$$\hat{\phi}_t(\mathbf{u}) = \underset{\phi_u}{\operatorname{argmin}} \sum_{s=1}^{t} \lambda^{t-s} w_u(\mathbf{u}_s)(y_s - \mathbf{z}_s^T\phi_u)^2, \tag{15}$$

Following [9] the solution to (15) can be found recursively as

$$\begin{aligned}\hat{\phi}_t(\mathbf{u}) &= \hat{\phi}_{t-1}(\mathbf{u}) + \\ &\quad w_u(\mathbf{u}_t)\mathbf{R}_{u,t}^{-1}\mathbf{z}_t\left[y_t - \mathbf{z}_t^T\hat{\phi}_{t-1}(\mathbf{u})\right].\end{aligned} \tag{16}$$

where

$$\mathbf{R}_{u,t} = \lambda\mathbf{R}_{u,t-1} + w_u(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T \tag{17}$$

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

$$\begin{aligned}\mathbf{R}_{u,t} &= \lambda v(w_u(\mathbf{u}_t);\lambda)\mathbf{R}_{u,t-1} \\ &\quad + w_u(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T,\end{aligned} \tag{18}$$

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 (18) becomes

$$\begin{aligned}\mathbf{R}_{u,t} &= (1-(1-\lambda)w_u(\mathbf{u}_t))\mathbf{R}_{u,t-1} + \\ &\quad w_u(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T.\end{aligned} \tag{19}$$

It is resonable to denote

$$\lambda_{eff}^u(t) = 1-(1-\lambda)w_u(\mathbf{u}_t) \tag{20}$$

the *effective forgetting factor* for point $\mathbf{u}$ at time t. For a further discussion of adaptive estimation of conditional parametric models see [10].

5.1 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 (5).

Besides λ in (17), prior to the application of the algorithm a number of fitting points $\mathbf{u}^{(i)};\ i = 1,\dots,n_{fp}$ in which the coefficient-functions are to be estimated has to be selected. Furthermore the bandwidth associated with each of the fitting points $h^{(i)};\ i = 1,\dots,n_{fp}$ and the degrees of the approximating polynomials $d(j);\ j = 1,\dots,p$ 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 (10):
  $\mathbf{z}_t^T = [x_{1,t}\mathbf{p}_{d(1)}^T(\mathbf{u}_t)\ldots x_{p,t}\mathbf{p}_{d(p)}^T(\mathbf{u}_t)]$.
- Calculate the weight using e.g. (8) and (7):
  $w_{u^{(i)}}(\mathbf{u}_t) = (1-(||\mathbf{u}_t-\mathbf{u}^{(i)}||/h^{(i)})^3)^3$, if
  $||\mathbf{u}_t-\mathbf{u}^{(i)}|| < h^{(i)}$ and zero otherwise.
- Find the effective forgetting factor using (20):
  $\lambda_{eff}^{(i)}(t) = 1-(1-\lambda)w_{u^{(i)}}(\mathbf{u}_t)$.
- Update $\mathbf{R}_{u^{(i)},t-1}$ using (19):
  $\mathbf{R}_{u^{(i)},t} = \lambda_{eff}^{(i)}(t)\mathbf{R}_{u^{(i)},t-1} + w_{u^{(i)}}(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T$.
- Update $\hat{\phi}_{t-1}(\mathbf{u}^{(i)})$ using (16):
  $\hat{\phi}_t(\mathbf{u}^{(i)}) = \hat{\phi}_{t-1}(\mathbf{u}^{(i)})$
  $+w_{u^{(i)}}(\mathbf{u}_t)\mathbf{R}_{u^{(i)},t}^{-1}\mathbf{z}_t\left[y_t-\mathbf{z}_t^T\hat{\phi}_{t-1}(\mathbf{u}^{(i)})\right]$.
- Calculate the updated local polynomial estimates of the coefficient-functions using (13):
  $\hat{\theta}_{j,t}(\mathbf{u}^{(i)}) = \mathbf{p}_{d(j)}^T(\mathbf{u}^{(i)})\,\hat{\phi}_{j,t}(\mathbf{u}^{(i)})$; $j = 1,\ldots p$

The algorithm could also be implemented using the matrix inversion lemma as in [3].

6 SUMMARY

In this paper a new method for short-term prediction of wind power in a region is proposed. The predictions are calculated using a combination of on-line measurements of power production from selected wind farms, off-line power measurements for all wind turbines in the region and numerical weather predictions of wind speed and wind direction as input.

The prediction models are formulated as conditionally parametric models, which can be described as conventional linear models 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. Methods for on-line estimation of parameters in such models are outlined in the paper.

References

[1] G. Giebel, L. Landberg, T. S. Nielsen, and H. Madsen. Zephyr – the next generation prediction system. In *Proceedings of the 2002 Global Windpower Conference*, Paris, France, 2002.

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

[3] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, MA, 1983.

[4] 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.

[5] 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.

[6] I. Marti, T. S. Nielsen, H. Madsen, A. Roldan, and S. Perez. Improving prediction models in complex terrain. In *Proceedings of the 2002 Global Windpower Conference*, Paris, France, 2002.

[7] A. E. Hoerl and R. W. Kennard. Ridge regression: biased estimation of nonorthogonal problems. *Technometrics*, 8:27–51, 1970.

[8] John M. Chambers and Trevor J. Hastie, editors. *Statistical Models in S*. Wadsworth, Belmont, CA, 1991.

[9] 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.

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

[11] T. W. Anderson, K. T. Fang, and I. Olkin, editors. *Multivariate Analysis and Its Applications*, chapter Coplots, Nonparametric Regression, and conditionally Parametric Fits, pages 21–36. Institute of Mathematical Statistics, Hayward, 1994.