PSO 2007 (FU 5766)
Improved Wind Power Prediction

Spatio-temporal modelling of short-term wind power prediction errors

Ewelina Kotwa
Henrik Aalborg Nielsen
Henrik Madsen
Julija Vlasova

October 30, 2007

Contents

1 Introduction

Forecasts of wind power generation are more and more frequently used in various management tasks related to integration of wind power generation in power systems. The quality of the forecast is very important, and a reliable estimate of the uncertainty of the forecast is known to be essential. Today the forecasts of wind power generation are provided without a proper consideration to the spatio-temporal dependencies observed in the wind power generation field. State-of-art prediction systems, like *Wind Power Prediction Tool* (WPPT [1]), typically provide forecast for a single wind farm. Predictions for a larger region with wind farms are then obtained by an upscale of the individual farm predictions. This means that the spatio-temporal relations are not adequately considered.

The aim of this work is to investigate weather it is possible to improve the wind power forecasting system WPPT, developed in Denmark, by examining the spatio-temporal correlation of the prediction errors. The paper is organized in the following way: Section 2 presents the data set used in the work. Then in Section 3 a pre-modelling analysis of the data structure is provided. The results of the correlation study are presented and discussed. Further, the modelling step is performed: three types of the models for WPPT error analysis are discussed: ARX, Threshold and Conditional Parametric Models. This takes place in Section 4. The description of each model consists of 4 parts *Modeling*, *Estimation*, *Application* and *Results*. The first two are theoretical: *Modeling* is a general description of the model; *Estimation* deals with how the parameters can be estimated. The last two parts show respectively how the model was applied to the data and the results obtained. Section 5 describes validation methods used in this study for checking the adequacy of the performance of the fitted models. The paper concludes in Section 6 with a small discussion on the general results and possible future work.

[1] For a detailed description of WPPT see www.enfor.dk

2 Data

The data selected for this work comes from 24 wind farms owned by Energinet.dk (previously ELSAM) where Wind Power Prediction Tool (WPPT) is used to make forecasts of the power production based on information from the Danish Meteorological Institute (DMI) HIRLAM model. The power production is recorded hourly matching the temporal resolution of the HIRLAM predictions. A new version of WPPT was installed in the fall of 2003 and as several parameters are estimated adaptively, some time is needed for the model to burn in. Therefore, it was decided to disregard data from before 01-01-2004. For this project data from the first seven months of 2004 is used.

The HIRLAM data is delivered in a 40 by 42 grid covering Denmark and surroundings in particular a large part of the North Sea. Every six hours a new 48 hour forecast with hourly steps is calculated. It takes 2-3 hours to calculate the forecast so in practice a 45-46 hour forecast is produced. Wind prediction in different levels of the HIRLAM model are available, but in this work we use wind at height 10 and 3000 meters a.g.l. (above ground level) only. The motivation for this choice is that the 10m wind minds local landscape characteristics while the 3000m is considered as a level where the wind is ”undisturbed”. However, after correlation analysis of the data was performed (see Section 3 for details) the 10 meters a.g.l. wind showed better performance and was chosen for the modelling step.

The WPPT model contains two parts: a static part, describing the power curve including dependence on wind speed and direction, and a dynamical part, which includes an autoregressive part. While calculating the prediction errors, it was chosen to use the output from the combined performance of the two parts. The errors were created as the difference between the power predictions and productions normalized by the installed wind power.

2.1 Managing the data

In order to reduce the influence of local behavior and to concentrate on global phenomena it was decided to group the data according to the location and correlation in errors in $lag\,0$. The following groups were selected:

The location of the mentioned farms and groups is given in the Figure 1. It is noticed, that only data from the 22 farms is used. The remaining two farms were excluded at this point since the correlation study showed that they can not be reasonably pooled into groups together with the other farms.

The group errors were calculated as an average of the errors within the groups. In the same

Group No	farms included
1	wab, war, wbs, wrb, whoe, wtj
2	wbi, wfj, wve, wvm
3	wgv, woek, wnv, wnr, wkm, whh
4	wto, wrv, wtu, who
5	wdr, wdg

Table 1: Wind farm groups

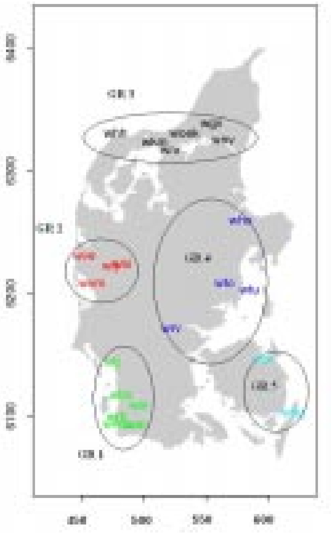

Figure 1: Selected groups of wind farms

way the group wind speed was calculated. For the directions in groups the geometrical approach was chosen. Wind direction at each of the farms was presented as a vector (wind speed normalized). The resultant vector (corresponding angle) was taken to represent the wind direction in groups.

3 Identification of the data structure

Before proceeding with the model building procedure, the structure of the data must be identified. This section presents some of the standard tools used for capturing the dependecies among the data (for more details see [11]). The following notation regarding the tools is introduced:

ACF - Auto-Correlation Function
PACF - Partial Auto-Correlation Function
CCF - Cross-Correlation Function
PCCF - Partial Cross- Correlation Function

The main questions at this point are:

1. Is there a significant linear dependency within and between the groups?
2. Do the variables *wind direction* and *wind speed* influence the strength of this dependency?

In the following we will answer these questions.

3.1 Dependency within the groups

Firstly we investigate the influence that the previous values of each time series have on its current state. Hence, we apply the ACF and PACF. Assigning the time series of interest as X_t (which is assumed to be stationary), the formula for the ACF in lag k is given below:

$$ACF(k) = Cor[X_t, X_{t-k}] = \frac{E[(X_t - \mu)(X_{t-k} - \mu)]}{\sigma^2} \tag{1}$$

where μ is the mean of the time series X_t and σ is its standard deviation. The coefficient takes values in the interval $[-1, 1]$. Obviously for $k = 0$ it is equal to unity. The problem with ACF is that since the estimates of auto-correlations are correlated within themselves, the pattern at the lower lags can be propagated on the larger lags. In order to remove the influence of the lower lags, we use PACF. This is given by the following formula:

$$PACF(k) = Cor[X_t, X_{t-k}|X_{t-1}, ..., X_{t-k+1}] = Cor(R_{X_t}, R_{X_{t-k}}) \tag{2}$$

which shows the "real" dependency between X_t and X_{t-k} excluding the influence of all the observations in between. It can be treated as correlation of R_{X_t} and $R_{X_{t-k}}$, which are residuals after regressing respectively X_t and X_{t-k} on $X_{t-1}, ..., X_{t-k+1}$.

Plots 2 and 3 show the ACF and PACF for Group 5.

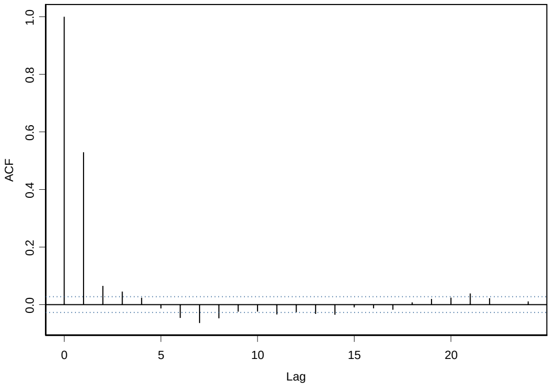

Figure 2: ACF for Group 5

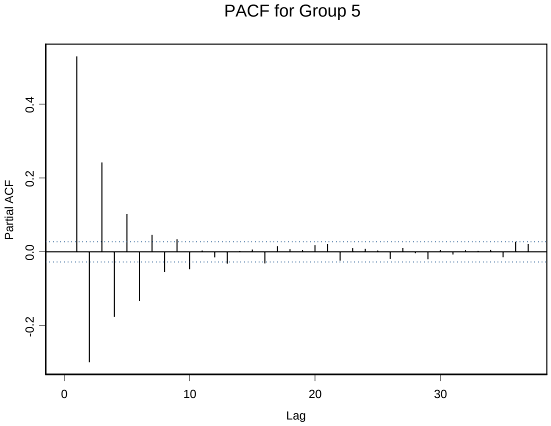

Figure 3: PACF for Group 5

3.2 Dependency between the groups

Once the auto-correlation of the group errors is identified, we proceed with investigation of the cross-dependencies among the data. The common practice is to check if the errors at different farm groups are cross-correlated. Especially the information about the dependency in the time lag would be of the great importance, because of the potentials for improving the future model forecasting ability. If such pattern is discovered, we can speak about the errors propagation between the groups.

Simple Correlation- and Partial Correlation Coefficients will be used here to identify the association between the variables. The formulas are given below respectively:

$$CCF(k) = Cor[Y_t, X_{t-k}] = \frac{E[(Y_t - \mu_Y)(X_{t-k} - \mu_X)]}{\sigma_Y \sigma_X} \tag{3}$$

$$PCCF(k) = Cor[Y_t, X_{t-k}|X_t, X_{t-1}, ..., X_{t-k+1}] = Cor(R_{Y_t}, R_{X_{t-k}}) \tag{4}$$

Here μ_X and μ_Y are the means of X_t and Y_t respectively, while σ_X and σ_Y - standard deviations of the relevant time series. Analogically to PACF, PCCF can be presented as the correlation between residuals R_{Y_t} and $R_{X_{t-k}}$ which are obtained after regressing Y_t and X_{t-k} on $X_t, X_{t-1}, ..., X_{t-k+1}$.

After examining the auto- and cross-correlation for all the groups, it was noted that Group 5 seems to be the most promising one. The reason is most likely that Group 5 is located upwind from the other groups when the wind direction is Western (which is dominant for that part of Denmark). The results for the Group 5 are shown in (Table 2). Note that this is not the correlation matrix but rather the table containing the correlation coefficients between Group 5 at time t and and the remaining Group values in the past up to the time t. Note that the last raw contains Auto-correlation values.

		lag					
	Group	0	1	2	3	4	5
Cross-Correlation	1	0.1755	0.2818	**0.3072**	0.2166	0.1191	0.0696
	2	0.1915	0.1866	0.1689	0.1631	0.1382	0.0795
	3	0.1578	0.1481	0.1140	0.0814	0.0743	0.0597
	4	0.2893	**0.3198**	0.2601	0.1391	0.0503	0.0183
Auto-correlation	5	1.0000	0.5267	0.0589	0.0403	0.0194	-0.0162

Table 2: Cross- and Auto-Correlation for Farm Group 5

Similar table was constructed for the Partial Correlation results (Table 3).

		lag				
	Group	1	2	3	4	5
Cross-Correlation	1	0.2791	0.1147	-0.0105	-0.0605	-0.0594
	2	0.1885	0.0439	0.0137	-0.0062	-0.0324
	3	0.1424	0.0103	-0.0163	-0.0119	-0.0138
	4	0.3219	0.0326	-0.0713	-0.0802	-0.0581
Auto-Correlation	5	0.5294	-0.1935	-0.0855	-0.0584	-0.0539

Table 3: Partial Cross- and Auto-Correlation for Farm Group 5

From the Tables 2-3 it can be inferred that the biggest influence on Group 5 is from Groups 1 and 4. The corresponding plots of the most significant dependencies are presented in figures below.

It is noted that for the Groups 5 and 1 the highest correlation is observed in lag 2, whereas for the Groups 5 and 4 the peak is reached in lag 1. The reason is most likely the geographical layout, especially the distance between the groups.

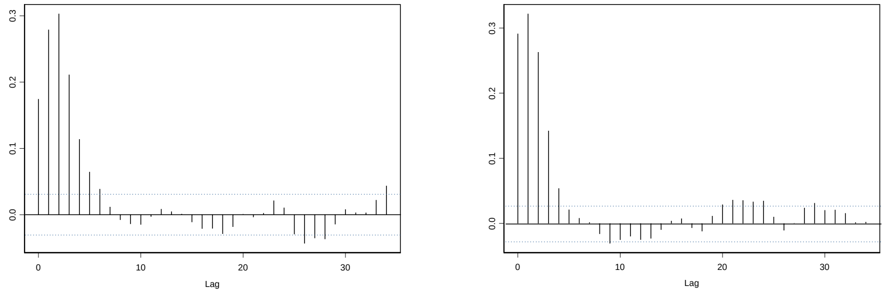

Figure 4: CCF for the Groups 5 and 1 (left), and Groups 5 and 4 (right)

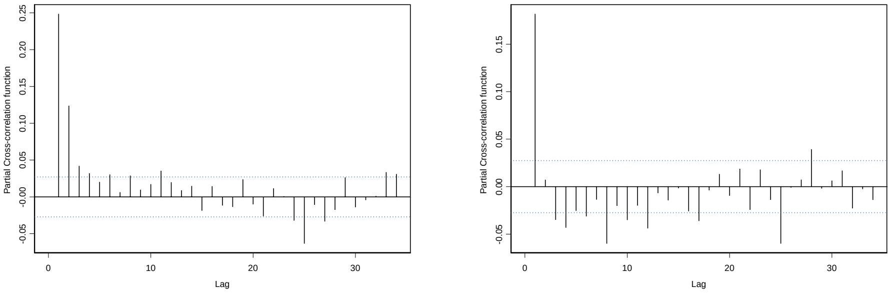

Figure 5: PCCF for the Groups 5 and 1 (left), and Groups 5 and 4 (right)

3.3 Dependency on the wind direction

The next step is to check how the wind direction influences the dependency. Again Group 5 is chosen and other groups are decided to play the role of explanatory variables. However, similar results could be obtained trying to explain the errors of other groups.

In order to examine whether the wind direction has the influence on the dependency between the groups, we divide the data according to the wind direction near Group 5 at time moment t. The division is performed by constructing four intervals: (0-90], (90-180], (180-270], (270-360] which match the cardinal directions. Then Cross-Correlation between Group 5 and the remaining groups was calculated for each interval. Below, only the most significant results for Groups 5 and 1 as well as Groups 5 and 4 are presented (Table 4 and Table 5, respectively).

It is easy to observe that in direction $(180-270]$ (S-E) and $(270-360]$ (N-E) the dependency is stronger than in the remaining cases and stronger than the overall cross correlation coefficient for relevant groups. The reason is most likely that those directions correspond to the geographical layout of the groups. Group 5 is located upwind from group 1 when the wind is South-East, and upwind form group 4 when the wind is North-East. The lags where the highest correlations are observed match the distance between

	Regime			
Lag	(0-90]	(90-180]	(180-270]	(270-360]
0	0.0457	0.1472	0.2240	0.1580
1	0.0499	0.2856	0.3597	0.2361
2	0.0672	0.3103	**0.4213**	0.2219
3	0.0358	0.1810	0.3218	0.1542
4	-0.0166	0.0985	0.2193	0.0519
5	0.0115	0.1130	0.1347	-0.0099

Table 4: *Directional* correlation for Groups 5 and 1

	Regime			
Lag	(0-90]	(90-180]	(180-270]	(270-360]
0	0.1390	0.3200	0.2615	0.3460
1	0.2212	0.2691	0.2570	**0.4514**
2	0.1788	0.2049	0.2075	0.3762
3	0.1288	0.1555	0.0978	0.1831
4	0.1014	0.0965	0.0158	0.0485
5	0.0252	0.0735	0.0102	-0.0157

Table 5: *Directional* correlation for Groups 5 and 4

the groups: the larger the distance, the higher the lag. These results will be essential in the further parts of this paper.

3.4 Dependency on the wind speed

Another potential explanatory variable to be examined is the forecasted *wind speed*. We will first check whether the wind speed influences the current correlation coefficients in the similar fashion as in Section 3.3. Namely, the data is grouped into five intervals according to the wind speed [m/s] near Group 5 at time t as shown: 1 - [0,4), 2 - [4,6), 3 - [6,8), 4 - [8,10), 5 - [10,25)

Again, the cross-correlation coefficients were calculated for each case separately. Figure 6 shows how the correlation coefficients vary among the lags on different wind speed levels. The data considered comes form direction [180-270) only. From the picture one can conclude that there is a certain tendency: the higher the wind speed is, the bigger dependency can be observed for lags $1-5$. For lags $5-7$ higher dependency is observed for lower wind speed levels. However this analysis has some restrictions. One should keep in mind that the number of observations is not the same among the intervals. Each of them contains $200-400$ data points which makes the results difficult to compare.

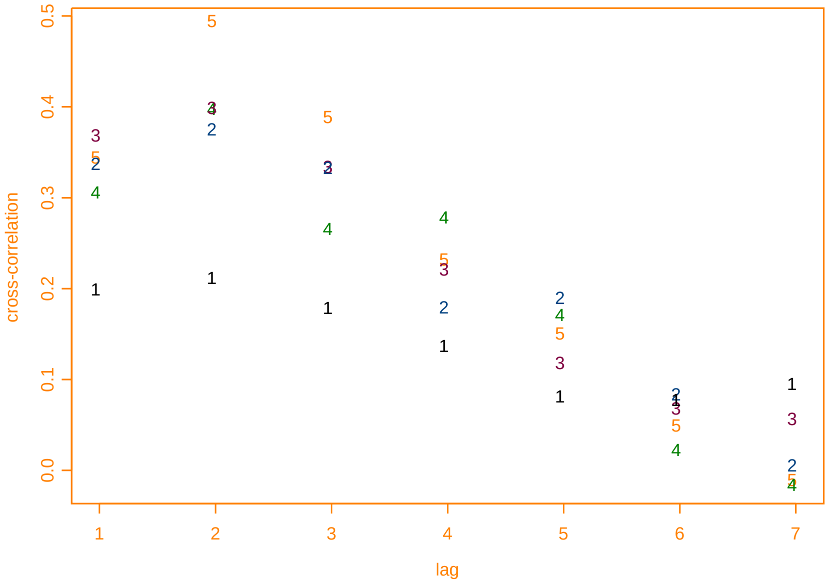

Figure 6: Cross-correlation for Groups 5 and 1 in direction (180,270] for different speed levels ("1"-"5")

3.5 Results

From the analysis carried out above one can clearly infer the dependency of the wind power in group 5 on four factors. Firstly, it is the influence of its own previous values, secondly, the dependency on the errors of remaining groups, and, finally, on wind direction and speed. Though the auto-correlation seems to play the major role (value = 0.5267), after dividing the process according to the wind direction, we can observe fairly big influence of Group 1 (0,4213 in direction $(180-270]$ at $lag = 2$) and Group 4 (0,4514 in direction $(270-360]$ at $lag = 1$) plus largest influence for higher wind speeds.

4 Models

After identifying the structure of the data, the modeling part is to be performed. This takes place in the following section which is the core of this report. All the considered models aim at improving the one hour error predictions. However, it is assumed that analogous methodology could be applied for longer term predictions. Since the highest Cross-Correlation coefficient was discovered between errors in Groups 5 and 4 (see Section 3), it was decided to use the errors of Group 5 at time t as a *dependent variable* and assign it as Y_t. Errors of Groups 1-4 will be denoted as $X_{1,t}, ..., X_{4,t}$ and called the *explanatory variables.* This notation is used throughout this report. Since the models aim at one obtaining one hour predictions, all the explanatory variables must be available at least one hour prior to Y_t

This section is divided into four subsections, each corresponding to the particular model type. As a general rule, every subsection begins with two theoretical parts: Modelling and Estimation, followed by Application and Results in which models are used in practice, and finally results presented, compared and discussed.

As a starting point, the Linear Regression, i.e. AR and ARX models [11], is considered. The Threshold Models governed by the external signal extend the topic by letting the coefficients vary among some selected regimes. In the last part Conditional Parametric Models are considered.

4.1 Linear Models

During the study related to the error analysis different linear models were fitted to data: univariate, auto-regressive AR and ARX (see [2; 11; 12] for more details). The latter model includes the influence of bothauto-regressivee part and external input, It turns out that the ARX models result in a better fit. Thus the univariate AR model is disregarded and only the ARX is presented in this report.

4.1.1 Modeling

The structure of the ARX model used in this work is

$$Y_t = \beta_0 + \sum_{l=1}^{p} \beta_l Y_{t-l} + \sum_{i=1}^{n} \sum_{j=1}^{k_i} \beta_{i,j} X_{i,t-j} + \epsilon_t \tag{5}$$

where the dependent variable Y_t is explained by its p previous values in the auto-regressive part, and in addition by n external input variables, each up to lag k_i. All the coefficients are put into a vector $\boldsymbol{\beta} = [\beta_0 ... \beta_p, \beta_{1,1} ... \beta_{n,k}]$, and $\{\epsilon_t\}$ is a noise sequence with the zero mean and constant variance.

4.1.2 Estimation

The estimation is performed using Least Squares (LS). The main idea behind this method is to minimize the residual sum of squares (RSS), which means finding an estimate $\hat{\boldsymbol{\beta}}$ of a real value $\boldsymbol{\beta}$, for which expression

$$RSS \equiv \sum_t \left(Y_t - \left(\beta_0 + \sum_l \beta_l Y_{t-l} + \sum_i \sum_j \beta_{i,j} X_{i,t-j} \right) \right)^2 \tag{6}$$

has the smallest value.

4.1.3 Application

The first step in our application is to select the explanatory variables. The obvious choice is to select the input variables which has the highest correlation coefficients with the dependent variable. After consulting with the results of CCF and PCCF together with the geographical location of the groups, it was decided to use X_1 and X_4 as the explanatory variables. In order to decide on the lags of the predictors the *one-in-one-out* method was used: starting from the bigger model (as suggested by ACF and PACF analysis together with Akaike Information Criterion [12]), the number of lags is then gradually decreased and the results compared. As a simple rule we decided to disregard a variable if the decrease of R-squared value is smaller than 0.005. Furthermore, we eliminate the variables in case of large *p-value* which means that according to the *t-statistics* the value of the corresponding coefficient does not significantly differ from zero. Finally we arrived at the following model:

$$Y_t = \beta_0 + \sum_{i=1}^{7} \beta_i Y_{t-i} + \sum_{j=1}^{3} \beta_{1,j} X_{1,t-j} + \sum_{l=1}^{2} \beta_{4,l} X_{4,t-l} + \epsilon_t \tag{7}$$

4.1.4 Results

Here, results of final models are presented in comparison with some other simple linear models fitted to data. The structures of the models can be seen in Table 4.1.4. The models are compared using two criteria: R-squared (R^2) and Root Mean Squared Error (RMSE) (see [12; 14] for more details).

It is clearly seen that the largest contribution toward explaining the variability of errors in group 5 is the Auto-regressive part of the model. However, 13 lags (as indicated by the Akaike Criterion) seem not to be the best choice. After decreasing the number of lags to 10 the R-squared and RMSE remain almost the same. Finally, we decide to use order 7 since by further increasing it, does only imply a small increase in R-squared value 0.005. By adding the cross-components to the model we obtain a further improvement of 0.05. In the end, according to R-squared value, it is possible to explain almost 48% of variation of Y with the final model.

The fact that the largest contribution comes from the auto-regressive part of the model,

Model No	No of Lags (X_1, X_4, Y)	R squared	RMSE
1	3, 0, 0	0.1189	0.0761
2	3, 2, 0	0.1744	0.0736
3	0, 0, 7	0.4213	0.0618
4	0, 0, 10	0.4271	0.0613
5	0, 0, 13	0.4272	0.0613
6	3, 2, 7	0.4794	0.0585

Table 6: Results of Linear Models (ARX)

indicates, that WPPT itself could possibly be improved for single area predictions by buildingauto-regressivee models within single areas.

4.2 Threshold Models

In this section we introduce the idea of what we call *directional correlation* and *directional regression*. The main purpose is to discover and capture the dependency between wind direction and propagation of prediction errors among the groups. The idea is supported by the intuitive, physical knowledge. We claim that if the wind direction is compatible with the direction of the vector, having its beginning in the group A and ending in B, then the error dependency in those groups should be higher than in case of different directions. As before we will start with a theoretical section and proceed with the applications and results.

4.2.1 Modeling

Threshold models extend the idea of linear models by letting coefficients vary among regimes. Regimes are defined by the threshold values, which are the upper bounds of the intervals in which the given 'sub-model' is active. The switch of the regimes in threshold models can be governed by previous values of the dependent variable, external signals or unobservable stochastic processes. In this paper we consider the second type of the models.

Define intervals $R_1 \cup ... \cup R_k = \Re$ such that $R_i \cap R_j = \emptyset, i \neq j$. Each interval is given by $R_i = (r_{i-1}, r_i]$. The values $r_0, ..., r_k$ are called thresholds (see e.g [2; 12; 16]). In general the threshold values are to be estimated from the data, but in this report we consider the case when those values are known in advance. The motivation for such an assumption is that we analyze the case when the regimes are governed by the wind directions. Then by applying physical knowledge and intuition it is not difficult to form the regimes. The general form of the models examined further is

$$Y_t = \beta_0^{(J_t)} + \sum_{l \in L_y^{(J_t)}} \beta_l^{(J_t)} Y_{t-l} + \sum_{i \in G^{(J_t)}} \sum_{j \in L_{x_i}^{(J_t)}} \beta_{i,j}^{(J_t)} X_{i,t-j} \tag{8}$$

where

$$J_t = \begin{cases} 1 & U_t \in R_1 \\ 2 & U_t \in R_2 \\ \vdots & \\ k & U_t \in R_k \end{cases}$$

where U_t is an external signal which determines the regime switch, t is the time index, Y_t is the output variable, $\boldsymbol{X}_t$ are input variables, $\{\epsilon_t\}$ is zero mean white noise, L_y and L_{x_i} are sets of non-negative integers defining the auto-regressive and input lags in the model, G is a set of positive integers defining which input variables to include into the model, J_t indicates the regime and, finally, $\beta_{j,i}$ are coefficients to be estimated.

4.2.2 Estimation

Since the thresholds are known, the estimation problem is solved by fitting different linear models to the data in each of the regimes. The technique used for estimation is again Least Squares, as described in the previous section.

4.2.3 Application

Let us assign the sequence of the forecasted wind direction at the Group 5 as U_t. The time series Y_t is divided according to U_t values into 4 groups. The structure of the regimes is shown below:

$$J_t = \begin{cases} 1 & for \quad U_t \in (0, 90] \text{ (North-East wind)} \\ 2 & for \quad U_t \in (90, 180] \text{ (South-East wind)} \\ 3 & for \quad U_t \in (180, 270] \text{ (South-West wind)} \\ 4 & for \quad U_t \in (270, 360] \text{ (North-West wind)} \end{cases}$$

The decision of fixing the threshold values was dictated by easiness in interpreting the influence of wind direction which in this case is compatible with the geographical cardinal directions. Furthermore, other divisions were checked and the improvement in model fit was considered insignificant or none.

As a result we obtain a four-regime model determined by the external signal (forecasted wind direction for time t). Linear models are fitted for each of the regimes. The optimal number of lags is decided separately for each regime as it was done in the previous section. Table 7 shows the structure of the final threshold model.

J_t	Lags in Group 1	Group 2	Group 3	Group 4	AR
1	-	-	4th	1st	10
2	1st	-	-	1st	5
3	1st, 2nd, 4th	-	-	1st	6
4	1st and 3rd	-	-	1st	6

Table 7: Threshold model structure

The choice of the variables seems to be reasonable if the position of the farm groups is taken into account (see Figure 1), e.g. for the direction (270,360] which corresponds to the northeast wind, influence of Groups 1 and 4 is seen to be most significant. Maximum lags taken for Groups 1 and 4 conform with the *directional distance* from Group 5 in this regime. By the *directional distance* in this case we consider a projection of the distance between the corresponding groups on the axis following the middle wind direction of the current regime (=315 for Regime 4).

4.2.4 Results

The performance of the Threshold model is shown in Table 8.

J_t	R^2	RMSE
1	0.3840	0.05763
2	0.4614	0.04352
3	0.4932	0.06832
4	**0.5490**	0.05579
overall model	**0.4991**	0.05742

Table 8: Threshold Model results

The overall R^2 of the Threshold model is 0.4991 and the $RMSE$ is 0.057. For comparison, for the best ARX model we obtain 0.4797 and 0.059, respectively. It is thus concluded that dividing the data according to directions is quite successful. For the overall model we obtain more than 2% improvement in R^2 compared to the ARX. Considering directions from regimes 3 and 4, the value of R-squared exceeds 0.50. The Root Mean Square Error Criterion also shows the advantage of the Threshold Models. It is due to the fact that quite a lot of wind data is available in those directions and location of Group 1 and Group 4 is relevant. On the contrary, considerably less observations are gathered in the remaining regimes and, what is more important, there is no groups of wind farms situated in these directions from the Group 5. That is the motivation for in the further work to focus on the data from directions (180, 270] and (270, 360], only.

4.3 Conditional Parametric Models

Now we will try to improve the directional model observed in the previous section by including wind speed information into the model building procedure. The idea is that the time delay might depend on the wind speed. For this purpose a new class of non-linear models, namely the conditional parametric models are to be considered.

4.3.1 Modeling

Conditional parametric models are a class of models in which the coefficients are allowed to vary as smooth functions of other variables (for broader description see [1; 4]). The model has the form

$$Y_i = \boldsymbol{z}_i^T \boldsymbol{\theta}(\boldsymbol{x}_i) + e_i; \quad i = 1, \ldots, N, \tag{9}$$

where Y_i is a measure of the response, $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are the explanatory variables, e_i are independent normal variables with $E(e) = 0$ and $Var(e) = \sigma^2$. $\boldsymbol{\theta}(\cdot)$ is a vector of unknown smooth functions, which we are aiming at estimating using non-parametric methods. The observation number is indicated by $i = 1, ..., N$. Here we will consider the case when $\boldsymbol{\theta}(\cdot)$ is approximated by local polynomials.

4.3.2 Estimation

Let $\boldsymbol{\theta}_j(\cdot)$ denote the j'th element of $\boldsymbol{\theta}(\cdot)$ and $\boldsymbol{P}_d(\boldsymbol{x})$ a vector containing terms of d-order polynomials at $\boldsymbol{x}$, e.g. $\boldsymbol{P}_2(\boldsymbol{x}) = [1 \; x_1 \; x_2 \; x_1^2 \; x_1 x_2 \; x_2^2]^T$ when $\boldsymbol{x} = [x_1 \; x_2]^T$. The estimation is performed by fitting the model

$$Y_i = \boldsymbol{u}_i^T \boldsymbol{\phi}(\boldsymbol{x}) + e_i; \quad i = 1, \ldots, N, \tag{10}$$

locally to $\mathbf{x}$. Where

$$\boldsymbol{u}_i^T = \left[z_{1i} \boldsymbol{P}_{d(1)}^T(\boldsymbol{x}_i) \ldots z_{ji} \boldsymbol{P}_{d(j)}^T(\boldsymbol{x}_i) \ldots z_{pi} \boldsymbol{P}_{d(p)}^T(\boldsymbol{x}_i) \right] \tag{11}$$

and

$$\hat{\boldsymbol{\phi}}^T(\boldsymbol{x}) = [\hat{\boldsymbol{\phi}}_1^T(\boldsymbol{x}) \ldots \hat{\boldsymbol{\phi}}_j^T(\boldsymbol{x}) \ldots \hat{\boldsymbol{\phi}}_p^T(\boldsymbol{x})], \tag{12}$$

$\boldsymbol{z}_i = [z_{1i} ... z_{pi}]^T$ and $\hat{\boldsymbol{\phi}}_j(\boldsymbol{x})$ is a vector of local constant estimates evaluated at $\boldsymbol{x}$ corresponding to the relevant elements of $\boldsymbol{u}_i^T$. Finally, we can write our estimate as:

$$\hat{\theta}_j(\boldsymbol{x}) = \boldsymbol{P}_{d(j)}^T(\boldsymbol{x}) \, \hat{\boldsymbol{\phi}}_j(\boldsymbol{x}); \quad j = 1, \ldots p. \tag{13}$$

At the beginning we assumed that the residuals e_i are normally distributed with mean 0 and variance σ^2. Note that while estimating $\hat{\theta}_j(\boldsymbol{x})$ we use observations only in the neighborhood of $\boldsymbol{x}$ so we can allow heteroschadasity of residuals among the whole data set. Variance should however be locally constant within each neighborhood of $\boldsymbol{x}_i$. The LFLM (Local Fitting of Linear Model) software by Henrik Aalborg Nielsen from Danish Technical University [13] was used to estimate the parameters.

4.3.3 Application

Now we will try to improve the Threshold model fitted in the previous section by assuming that its coefficients are not constants but smooth functions of a forecasted wind speed level at time t. A model to fit looks like:

$$Y_t = \sum_{k=1}^{4} \left(\sum_{j=set.k} \theta_j^k(s_t) X_{k,t-j} \right) + \sum_{j=set.5} \theta_j^5(s_t) Y_{t-j} + \epsilon_t \tag{14}$$

Here, k indicates a group number and set.k is a set of non-negative integers showing which lags of Group k to include into the model. Similar to the way it was done in the previous section we define 4 models depending on the wind direction forecasted for Group 5 at time t. For each of the models we take corresponding lags to the ones from the Threshold Model (see Table 7).

Variable s_t indicates wind speed level at time t. Firstly, it was decided to take wind speed near Group 5 at time t as a representative of s_t. Of course, in this case the information about wind speed levels near other groups at different time moments was lost and not involved into the model. Having it in mind it was decided that s_t should be a summary of wind speed information near all the farms included into the model. For making such summary weighted linear regression was chosen. The weights were selected according to the correspondent coefficients of simple linear regression of Y_t on the errors from the other groups included into the model. Then, for instance, the forecasted wind speed near group 1 at time $(t-2)$ has the same weight for the summary of wind speed as the linear regression coefficient near $X_{1,t-2}$.

4.3.4 Results

Table 9 shows the corresponding results for the Conditional Parametric Model and Threshold Model from Section 4.2.

Conditional Parametric Model			Threshold Model		
Regime	R^2	RMS	Regime	R^2	RMS
1	0.485	0.053	1	0.384	0.075
2	0.483	0.043	2	0.461	0.044
3	0.556	0.064	3	0.493	0.068
4	0.577	0.054	4	0.549	0.056

Table 9: Conditional Parametric Model Results

The improvement in R^2 and RMS values seems significant, especially for directions $[0, 90)$ and $[180, 270)$. Those results look very promising. Simple diagnostics for Conditional

Parametric Model in regimes 3 and 4 are shown in Figure 7. One can see that the model does not describe the data well in the tails. After checking the correlation of residuals, a bootstrapping technique was applied to check the level of uncertainty associated with the estimates of coefficients. The results are shown below in Figures 8-9. It can be observed that when the wind speed is very high or very low, the uncertainty level is much higher. This is related to an insufficient amount of data available for those intervals.

For the coefficients describing the dependency on the wind speed level, the influence of AR- part decreases when the wind speed level grows. The influence of the other groups, on the contrary, increases. The change in coefficient values is most significant for the groups which are located upwind from group 5.

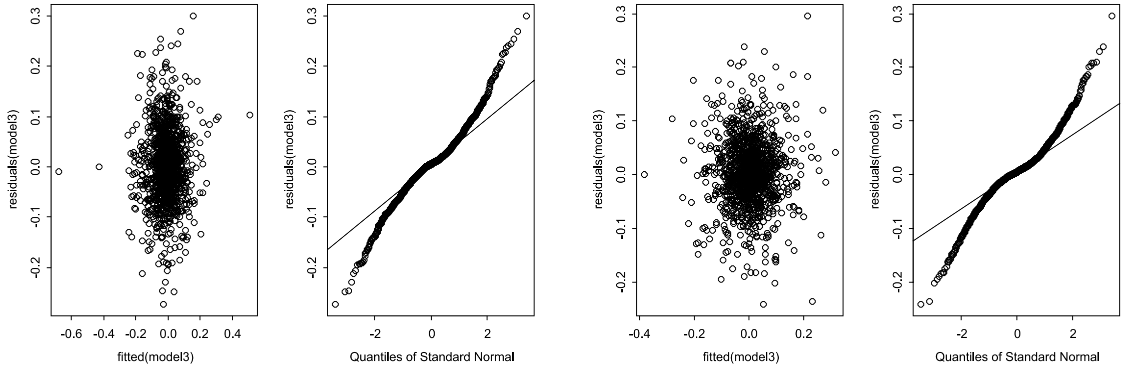

Figure 7: Simple diagnostics for the fit corresponding to Conditional Parametric Model in regimes 3 (left) and 4 (right)

5 Cross-Validation

All the R^2 and RMSE estimates of the fitted models shown previously in this study were obtained for the same data set as was used in estimation steps. We were comparing and giving preferences to one or another model based on those characteristics without taking cross-validation results into the account. However, to draw a final inference, cross-validation test will be performed.

We start comparing the adequacy of the two models which showed the best performance in the previous sections of this study - Threshold model and conditional Parametric model. 3-fold cross correlation test is applied (see [2; 12]). The data in each regime is divided into 3 equal subsets. Two of the constructed subsets are used for parameter estimation and the third subset is used for checking the model performance.The results are presented in Table 5.

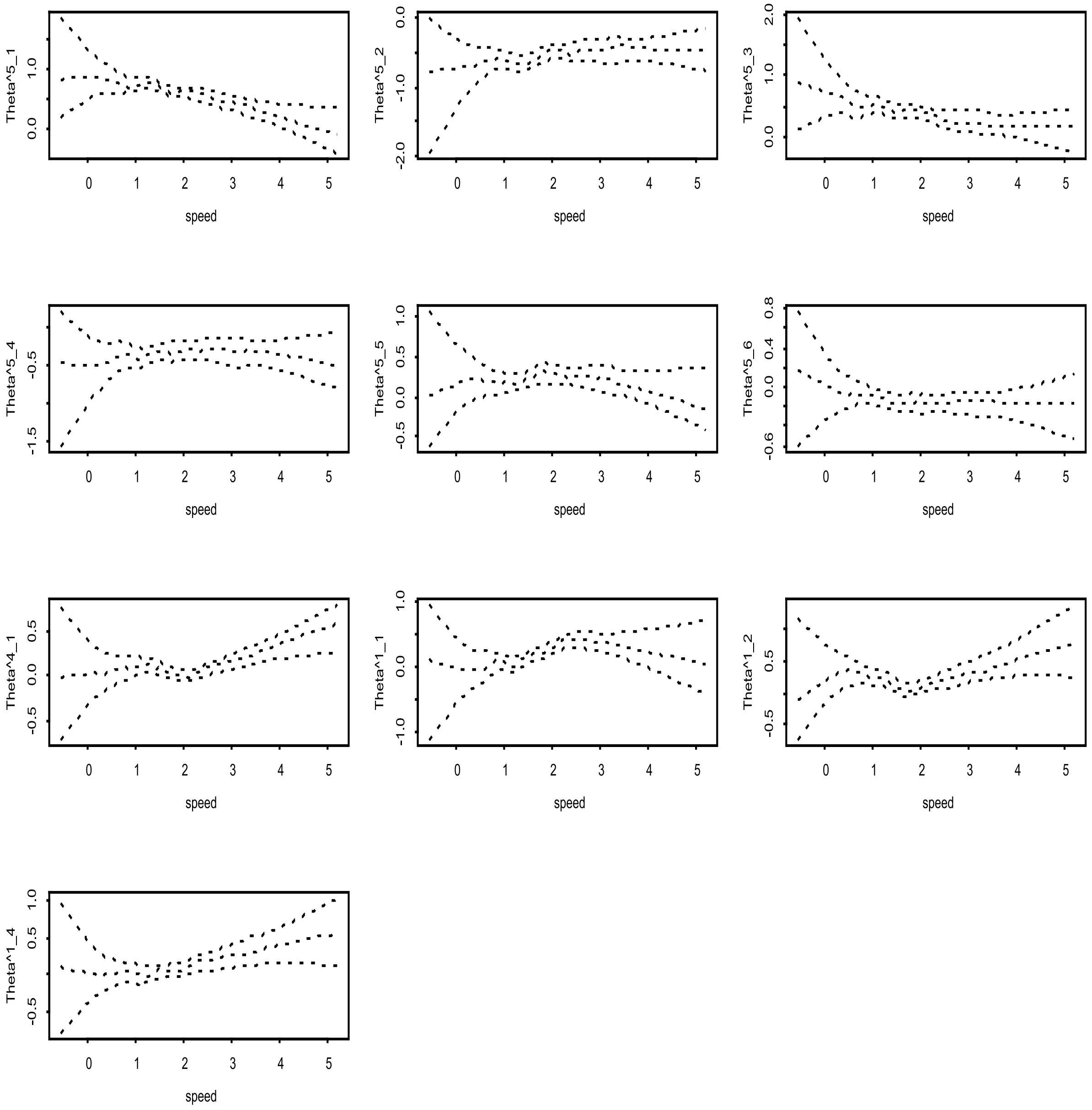

Figure 8: Obtained coefficients and 95% standard normal intervals based on 200 bootstrap replicates of Conditional Parametric Model in regime 3. Here Theta k j stands for θ_j^k and speed for s

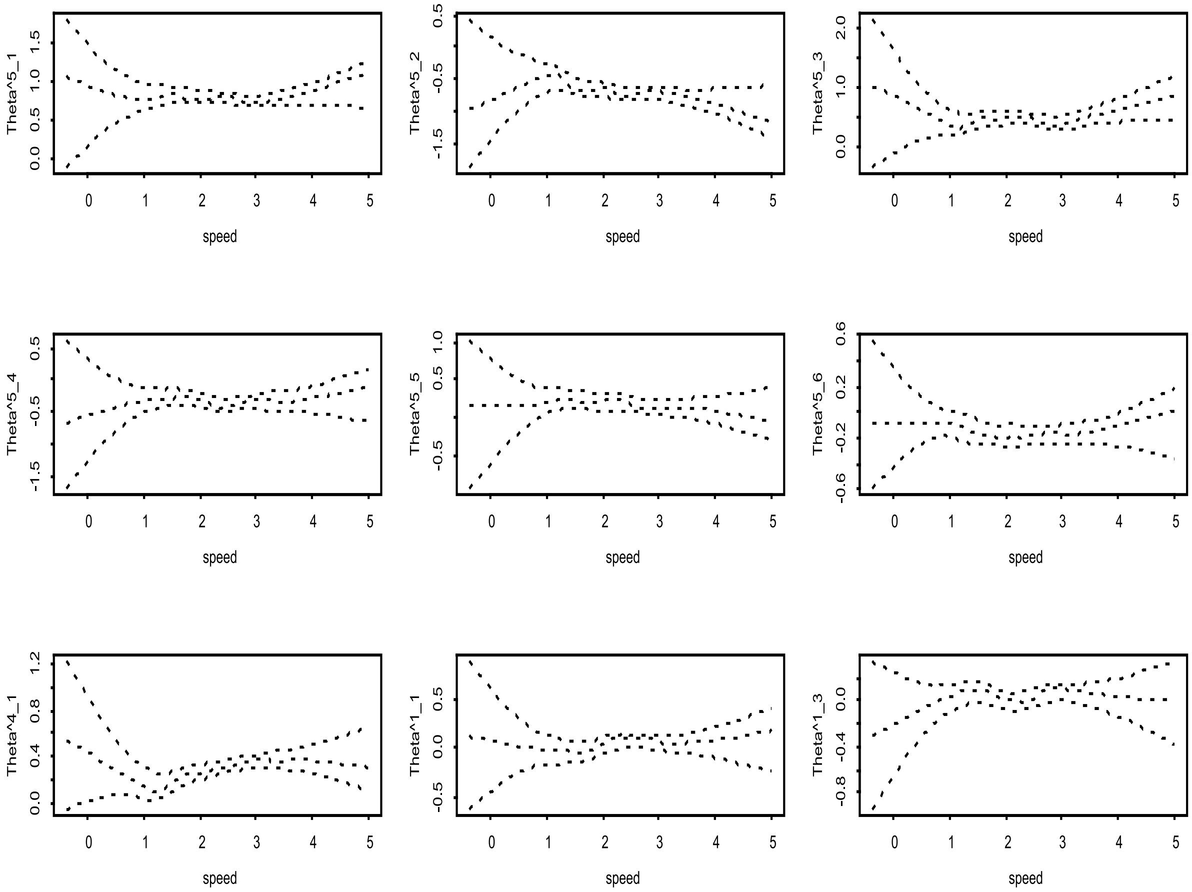

Figure 9: Obtained coefficients and 95% standard normal intervals based on 200 bootstrap replicates of Conditional Parametric Model in regime 4. Here Theta k j stands for θ_j^k and speed for s

Model	subset1		subset2		subset3	
regime	R^2	RMSE	R^2	RMSE	R^2	RMSE
Conditional Parametric, regime 3	0.4785	0.0589	0.4040	0.0778	0.4921	0.0803
Threshold, regime 3	0.4811	0.0593	0.4778	0.0720	0.4749	0.0796
Conditional Parametric, regime 4	0.5253	0.0558	0.5159	0.0581	0.5700	0.0689
Threshold, regime 4	0.5341	0.0543	0.5292	0.0556	0.5529	0.0660

Table 10: 3-fold cross validation results for Conditional Parametric and Threshold models in regimes 3-4

Cross-correlation results show that the threshold model describes the data more adequately than the Conditional Parametric one . R^2 and RMSE values are stable for the Threshold model and do not decrease in comparison with the results obtained for the whole data set (see Section 4.2). On the contrary, the results of Conditional Parametric model show more abrupt jumps within the constructed subsets and R^2 declines after performing cross-validation test. Another remark is, even though there was significant difference between the R^2 and RMSE values of those models obtained for the entire data set (Table 9), the performance on a 'new' data was fairly similar. All these arguments indicate that the Threshold model describes the data most adequately and efficient among all the models examined in this study.

However, another important remark is the amount of data we used in this study. As it was already mentioned at the beginning of the report, the data comes from the 7 month period. Which means, that when we make a 3-fold cross-validation tests, each of the constructed subsets consists of the data from 2.33 months period only. Taking into the account, that the data has to be divided according to wind direction and different wind speed levels have to be considered, it is easy to conclude that the 7 month period is not sufficient to draw a final inference on the results of the cross-validation test. Therefore, based on the amount of data we had for this study, it is impossible to finally conclude which of the constructed models described the data the best: Conditional Parametric or Threshold model.

6 Conclusions and Future Work

In this work new models and methods for improving on-line short-term predictions of wind power were derived and examined. The study was focused on the improvement of the one-hour wind power predictions. However, the methodology used in the analysis, could be applied for a longer-term predictions in the similar manner. however this calls for data for a larger geographical area.

The results of the work show a great potential in improving the WPPT by modelling the spatio-temporal correlation of the errors. The captured error propagation appeared to be dependent on the forecasted weather situation (mainly wind direction) and geographical position of the wind farms. Methods applied in the project captured a non-linear behavior of error dependency. The best results were obtained while fitting threshold models with regime switching according to the forecasted wind direction. The model adequately explains more than 47% of the error variation for one-hour predictions. During the study more complex models were fitted to data. Varying coefficient models were fitted in order to capture the dependency on the forecasted wind speed level as well. However, the further diagnosis of this approach showed that the improvements it gives in comparison with the threshold model are doubtful having in mind the results of the cross-validation tests. A final inference in this case could not be drawn due to insufficient amount of data (only data from 7 month period was available)

The promising results obtained in the study revealed broad perspectives for the future work. First step could be trying to apply the derived methods for the longer-term predictions. In this case the information from a larger geographical region should be taken into account. Data from e.g. all Denmark or oil platforms in the North Sea might lead to huge improvements on a longer time horizon. Since a dependency on the weather situation was noticed, it would be of a high interest to include forecast information obtained from different sources. Possibly, it could improve the quality of the predictions, reducing the uncertainty level associated with the weather forecasts.

Several attempts taken in this study aiming to include wind speed level into the model building procedure did not give reliable improvements. However, investigation held in the identification step indicated that some dependency presents. It could be another push up for the future work to investigate it further. One of the possible approach to take could be Markov Regime Switching Model which assumes that the analyzed time series depends not only on the known variables, but also on some hidden processes. The motivation for this can be the fact that many weather related phenomena are complex and sometimes difficult to capture. Once the Markov approach evidences the dependency on such a hidden process, it would be a motivation for further examination and attempts to determine this process. Possibly, wind speed level could somehow be related to this unobserved process since, as already mentioned before, Identification step showed some dependency on it.

Acknowledgments

The research has been financially supported by the Danish utilities fund (FU-5766). The financial support is hereby greatly acknowledged. The data has been provided by Energi E2 A/S (now DONG Energy) in Denmark, and the meteorological forecasts have been provided by the Danish Meteorological Institute (DMI). The authors wish to thank the suppliers of the data and meteorological forecasts.

References

[1] Chambers J.M., Hastie T.J. eds (1991) *Statistical models in S*, Wadsworth, Belmont, CA.

[2] Chatfield C., *The Analysis of Time Series, an introduction*, fifth edition

[3] Cleveland W. S., Loader C., *Smoothing by Local Regression: Principles and Methods*

[4] Cleveland W. S., Devlin, S.J.,(1988) *Locally Weighted Regression: An approach to regression analysis by locall fitting*, Journal of American Statistical Assocination 83, 596-610.

[5] Eiliers P. H. C., Marx B.D. (2004) *Splines, Knots and Penalties*,

[6] Gregor Giebel *The State-Of-The-Art in Short-Term Prediction of Wind Power. A literature Overview*. Project ANEMOS, cont. nr ENK5-CT-2003-00665

[7] Hamilton James D. *Time Series Analysis*, (1994) Princeton University Press, Princeton.

[8] Hastie, T. J. and Tibshirani, R.J. *Varying-coefficient models*.

[9] Hastie, T. J. and Tibshirani, R.J. (1990) *Generalized Additive Models*, Chapman and Hall, London/ New York.

[10] Kotwa E., Vlasova J. Master thesis on *Spatio-temporal modeling of short-term wind power prediction errors*, Sweden, Lunds universitet (2007:E20).

[11] Madsen H. (2007) *Time Series Analysis*, Chapman Hall.

[12] Madsen H., Holst J. (2000), *Modelling Non-Linear and non-Stationary Time Series*, IMM.

[13] Nielsen H.A. (1997) *An S-PLUS/ R library for locally weighted fitting of linear models*

[14] Perce D. A., "R^2 Measures for Time Series", *Journal of the American Statistical Association*, June 1979, Vol. 74, 901-909.

[15] Pinson P., Madsen H. eds, (2007) *Regime-switching modelling of the fluctuations of offshore wind generation*,

[16] Tong H. (1990), *Non-Linear Time Series -A Dynamical System Approach*, Oxford University Press.