Models and Methods for Predicting Wind Power

February 1996

Henrik Madsen (Ed.)

ELSAM, Sep, IMM

Contents

Preface

This book aims to provide a unified treatment of how to establish on-line and short term (say up to 36 hours) predictions of wind power. Wind power models and methods for obtaining predictions for a single wind mill farm as well as predictions for a large region are considered. Furthermore, it is demonstrated how to use meteorological forecasts as input to the models in order to improve the predictions on longer horizons.

Models describing relations between the wind speed, wind direction and the wind power are described. The book focuses on the statistical models; but also practical methods for handling data and tools for implementing the models are considered.

Most of the models, methods and tools are developed by IMM[1], ELSAM[2] and Sep[3] during the Joule II project *Wind Power Prediction Tool in Control Dispatch Centres.* During this project a system for on-line wind power predictions has been developed and implemented at the ELSAM load dispatch centre.

Several people have contributed significantly by writing sections of the book: Henrik Madsen, Ken Sejling, Torben S. Nielsen and Henrik Aa. Nielsen (IMM); Uffe S. Jensen, Henning Parbo and Bjarne Korshøj (ELSAM); Nic Halberg, Eppie Pelgrum and René Beune (Sep).

H. Madsen,
February 1996

[1] Inst. Math. Modelling, Tech. Univ. of Denmark
[2] The Jutland–Funen Power Pool
[3] Dutch Electricity Generating Board

Summary

This book describes operational methods and models for short term (say up to 36 hours) prediction of wind speed and wind power. Most of the models are constructed for on-line applications where an updated prediction is needed rather often (say every 30 minutes). A major part of the models are self-learning, which means that they are easily ported to a new site.

The models are developed during the Joule project *Wind Power Prediction Tool in Control Dispatch Centres* and as a special follow up study of the project *Integration of Wind Power in the Danish Generation System* (ELSAM, 1989) and the Joule project *European Wind Power Integration Study* (ELSAM, 1992). During these projects a software system for on-line wind power predictions has been developed. The software is running on a computer at ELSAM[4] and gives predictions of the total wind power in the ELSAM area 0.5 to 12 hours (alternatively 0.5 to 36 hours) ahead in half hourly steps. The predictions are given with an estimate of the variance of the prediction error. The software is running under both UNIX and VMS and a graphical user interface is implemented under both operating systems. Furthermore, an off-line version of the software is installed both at Sep in Holland and at ELSAM. This software runs under UNIX and DOS.

The on-line WPPT system has been used in the ELSAM dispatch centre since July 1994. In periods it has been rather difficult to obtain an acceptable quality of the on-line measurements in the seven wind mill parks. The economic value of the WPPT system for load dispatch cannot be calculated at this moment. The operators conclude, however, that the economic benefit will be positive both in normal operation situations and in failure situations. Furthermore, the operators conclude that the development of the WPPT system should be continued and that meteorological forecasts should be integrated into the next version of WPPT.

For the region controlled by ELSAM, measurements at seven reference wind farms form the basis for an upscaling of predictions for the individual wind farms to on-line predictions of the total wind power in the region. The measurements are collected every 5 minutes and transferred to the ELSAM dispatch centre. The installed capacity of the seven wind farms is 39 MW, which corresponds to about 10% of the total installed wind power capacity in the ELSAM area.

Different methods for sampling are investigated. It is concluded that fast sampling followed

[4] ELSAM is responsible for production and transmission of electrical power in Jutland and Funen, Denmark

by simple averaging is appropriate. In order to handle errors in the data, methods for robust estimation and for detection of outliers (bad measurements) have been developed. If measurements are classified as erroneous this is automatically indicated in the implemented system.

The problem of estimating the total wind power in a region, like the ELSAM area, based on the predictions of wind power in some wind farms is referred to as the upscaling problem. Since the relation between the wind speed and the wind power is specific for a given brand and type of mill, and since the types of private mills in general differ from the larger wind mills in the seven wind farms, the upscaling must somehow depend on the wind speed. In the report this is investigated using measurements from nearby private wind mills and wind mills owned by companies under ELSAM. For the moment a very simple upscaling using a constant is used in the implementation at ELSAM.

The best models and methods without the meteorological forecasts are implemented in both an on-line and an off-line system. For the on-line system a graphical user interface is developed. Predictions from the on-line system are available in the ELSAM dispatch centre in Skærbæk. The off-line version is installed both at ELSAM and at Sep in Holland. The main purpose of the off-line version is to provide a possibility for an investigation of the prediction performance for different prediction horizons using historical data.

A major part of the research in the project has concentrated on finding the best mathematical model for on-line predictions of the wind power. Both linear and non-linear models have been formulated. Among the non-linear models neural networks have been considered. The main investigation is based on data from about 3 months (July to September 1993) of half hourly data of the wind speed and wind power from one of the wind farms. A preliminary study has shown how meteorological forecasts can be used as input to the model to improve the long term forecasts.

In the following a summary of the most important models is given.

- **Models for Predictions of Wind Speed.**

  Both linear and non-linear models have been considered. The most promising non-linear models are the bilinear model and the smooth threshold autoregressive model. The non-linear models are, however, only slightly better than the linear AR(1) model.

- **Models Relating the Wind Power Production to the Wind Speed.**

  Several types of functional relations have been considered. A comparison of the prediction performance shows only a minor difference between some of the functional relations; but the Compertz curve has been used in the implementations.

  Models relating the power to the wind speed and the wind direction show a clear and significant dependence on both variables. Using a Compertz curve where the parameters show a trigonometric wind direction dependence a variance reduction on 24.5% and 42.6% was found for the two wind farms considered compared to models with fixed parameters. Due to different physical layout of the farms it is quite natural to expect a difference in the variance reduction between parks.

In cases where the wind direction is missing it has been shown that an indirect method can be used where the residuals from the Compertz curve are described by an ARMA(1,1) model. This is due to the persistence in the variations of the wind direction. Using this indirect method the further improvement by having measurements of the wind direction is minimal.

The curves are used e.g. for estimating the actual number of wind turbines in operation.

- **Models for Multi-Step Prediction of Wind Power Production**

  Both models which take into account the relations to wind speed and models which consider the wind power production alone have be considered. Models for multi-step prediction, 0.5 hours to 36 hours ahead, are considered.

  A rich class of models containing for instance auto-regression, moving averages, seasonal auto-regression and deterministic diurnal variation has been considered, and an optimal model is found for each prediction horizon. Also non-linear terms enter some of the models. Models found as one-step prediction models and dedicated k-step prediction models are found to give almost the same results.

  A non-linear dependence on wind speed is significant for prediction horizons less than or equal to 6 hours. For prediction horizons larger than 8 hours wind speed does not enter the optimal model for prediction of wind power production. A deterministic diurnal variation is seen in a summer period, but not in a (short) winter period. Compared with the naive forecast, which corresponds to predicting the future value as the most recently observed, the improvement is between 2% and 23% depending on the prediction horizon.

  A square root transformation of the power and the wind speed are found to give nearly symmetrical confidence intervals. Hence, this transformation is used in calculating confidence intervals of the prediction error.

  The models are estimated using adaptive recursive least squares estimation with an exponential forgetting. Based on the analysis a candidate for a best overall adaptive k-step prediction model is suggested.

- **Model for Using Meteorological Forecasts**

  Most of the models are prepared for using meteorological forecasts as input. It is obvious that meteorological forecast will improve the predictions of wind power, especially on horizons greater than, say, 12 hours.

  A method for using meteorological forecasts which are available, say every 12 hour, is suggested. Compared with the naive forecast the improvement is up to 36.6% depending on the prediction horizon.

- **Neural Network Models for Prediction of Wind Power Production**

  It is investigated whether neural networks, as a non-linear predictor, can be used for predicting the wind power production. The structure of the feed-forward neural network used is found using the Bayes Information Criterion. Only prediction horizons up to 3 hours have been considered. It is found that the neural network is inferior in prediction performance compared to both the naive predictor and the most promising overall adaptive k-step prediction model found.

Chapter 1

Introduction

Recently several European countries have offered incentives for wind power, including Denmark, Germany, Italy, the Netherlands, Spain and the United Kingdom. In Denmark, for instance, the politicians have announced goals to install further wind power plants such that a total installed capacity of 1200 MW is reached in the year 2000. Table 1.1 shows estimates[1] of wind power capacity for operational and projected installations in Europe.

The steady increase of wind energy being connected to the electrical grids not only in Europe, but all over the world, makes it still more important to have forecasting methods and tools for the prediction of wind power production in the day-to-day planning of operations and also in long term planning studies.

This book contains a description of models and tools for short term (say up to 36 hours with a time step of 0.5 hour) prediction of the wind power. Methods for both local predictions, e.g. in a wind farm, and in a region, as e.g. the ELSAM[2] supply area, are described.

For load dispatching it is very important that the reliability of the predictions is known. Therefore special attention is paid to the problem of finding reasonable confidence intervals for the predictions.

Models for both wind speed and wind power are suggested. Most of the models are well suited for on-line predictions for use in the short term planning. Also a simple method for using meteorological forecasts as input to a model is suggested. This is, of course, important for obtaining a reasonable prediction on horisons greater than, say, 24 hours.

Special emphasis is laid upon models that can be used as a simple and operational description of the correlations and variations. They need to be simple since the operators need updated forecasts rather frequently, say every 30 minutes. The models are identified by using statistical methods for dynamical systems and time series analysis. Both linear and non-linear models have been formulated. Most of the models are adaptive or self-learning, which implies that the models and the tools rather easily can be moved from system to

[1] Source: EPRI, IEA, AWEA and J.K. Vesterdal, ELSAM.

[2] The ELSAM supply area consists of a major part of Denmark, namely Jutland and Funen.

Amounts in MW		
Year	1994	2000
Denmark	540	1200
Germany	643	1000
The Netherlands	153	500
Ukraine	4	500
United Kingdom	145	300
Sweden	38	300
Italy	22	300
Spain	73	150
Greece	35	150
Ireland	7	150
Portugal	9	50
Belgium	7	50
Finland	6	20
Norway	4	5
Totals	1686	4975

Table 1.1: Estimates of installed wind power capacity for operational and projected installations in Europe

system.

An implementation of a method for on-line data collection and transmission is also described. Software tools for both on-line and off-line short term prediction of wind power are developed and described briefly.

It is believed that the models and methods presented are widely useful for short term predictions of wind power. The book is, however, also a summary report for the Joule project *Wind Power Prediction Tool in Central Dispatch Centres.* A more detailed description of this project can be found in (IMM, ELSAM, & SEP, 1995). The aim of the project has been to develop, implement and test a forecasting model for the prediction of wind power production to be used in the day-to-day planning of operations. In the ELSAM supply area the installed power is around 400 MW of electric power, while the peak load is approximately 3500 MW.

1.1 Outline

The book addresses the following main topics:

- Description of measurements

- Frequency analysis of wind behaviour
- Models for predicting wind speed
- Models for predicting wind power
- Models for the relations between wind power and wind speed
- Models and methods for upscaling the predictions of some wind farms to a prediction of the total power production of a region
- On-line and off-line software tools
- Hardware
- Practical experience

The most important data, and the wind and power measurements used in developing the models, are described in Chapter 2. Other chapters use specific data, these are described in the corresponding chapters.

In Chapter 3 the sampling frequency is discussed and, using spectral analysis, different methods for sampling are investigated.

Hereafter follows the main part of the book, where a substantial No. of different models are investigated. This is reported in Chapters 4 and 5.

Methods for upscaling predictions from few farms to a prediction for a total region are discussed in Chapter 6.

In Chapter 7 both on-line and off-line software tools are described. This software is developed during the Joule project *Wind Power Prediction Tools in Central Dispatch Centres.* Also the hardware selected for this project is described.

Chapter 8 describes the practical experience so far with the hardware and software. Also the value for load dispatching is briefly discussed.

Chapter 2

On-line Wind and Power Measurements

2.1 The Seven Wind Farms

The on-line wind power prediction tool is developed and implemented on the basis of an individual power prediction at each of the seven wind farms which have been selected to give a good representation of the total wind power in the ELSAM area. Table 2.1 lists the installed power and the number of wind mills for each of the seven farms. At Dræby

Wind farm	Number of Mills	Power Capacity [kW]	
Ryå	23	2580	20×99 + 3×200
Nørrekær Enge	78	17280	36×130 + 42×300
Hollandsbjerg	32	4500	30×130 + 2×300
Vedersø Kær	27	6075	27×225
Torrild	16	2700	15×150 + 1×450
Dræby Fedsodde	11/12	2420/2640	11/12×220
Brøns	8	3200	8×400

Table 2.1: Farm statistics.

Fedsodde one additional mill has been installed on October 26, 1993 so the installed power is

increased from 2420 to 2640 kW. Figure 2.1 shows the location of the seven reference farms. Predictions of the total wind power in the Elsam area is obtained by a so-called upscaling of the predictions for the individual farms.

2.2 The On-line Measurements

It was decided to measure *wind speed* and total *power production* in each of the seven reference farms. The wind direction is not measured, although it is known to have a significant impact on the power production. The importance of using the measurements of wind direction for predictions is described in Chapter 4. This chapter also suggests a method for an indirect description of the dependency of the wind direction. Using this method the drawback of not having the wind direction available is shown to be minor.

The measurements for the on-line wind power prediction tool are sampled at five-minute intervals, and the main model study is based on five minutely data collected from the beginning of August until the beginning of December 1993. The collection of data has, however, been going on since May 1993 and the quality of the data has increased during the period.

Some examples of the measurements are shown on figures later on in this chapter. The figures show the measured wind speed and power during September 1993 for Nørrekær Enge, Vedersø Kær and Torrild. In these figures each measurement is shown as a dot. The range of the ordinate axes are chosen so all measurements are shown on the figures (also obviously erroneous measurements). By comparing the figures it is clearly seen that the wind speed in East Jutland (Torrild) is less that the wind speed in West and North Jutland (Vedersø Kær and Nørrekær Enge, respectively). A complete set of figures of the data used is found in (IMM et al., 1995).

Furthermore, the figures demonstrate that it is possible, in general, to obtain data of a reasonable quality. However, the constant wind speed during short periods for Nørrekær Enge and during a longer period period for Torrild indicates errors in the measuring system.

2.3 Examples of Errors in the Data

It is, of course, impossible to measure wind speed and power, and to transfer the data to the central dispatch center, without any error. This section describes briefly some different kinds of errors which were seen in the data.

- For some of the wind farms (Ryå, Nørrekær Enge, Hollandsbjerg, Torrild and Dræby Fedsodde) it is observed that the measurements remain on a constant level for longer periods (An example is seen on Figure 2.2). These observations are clearly erroneous as it is not at all likely that the wind speed nor the power are completely constant. Only in situations where the wind speed is close to 0 m/s will the power be completely

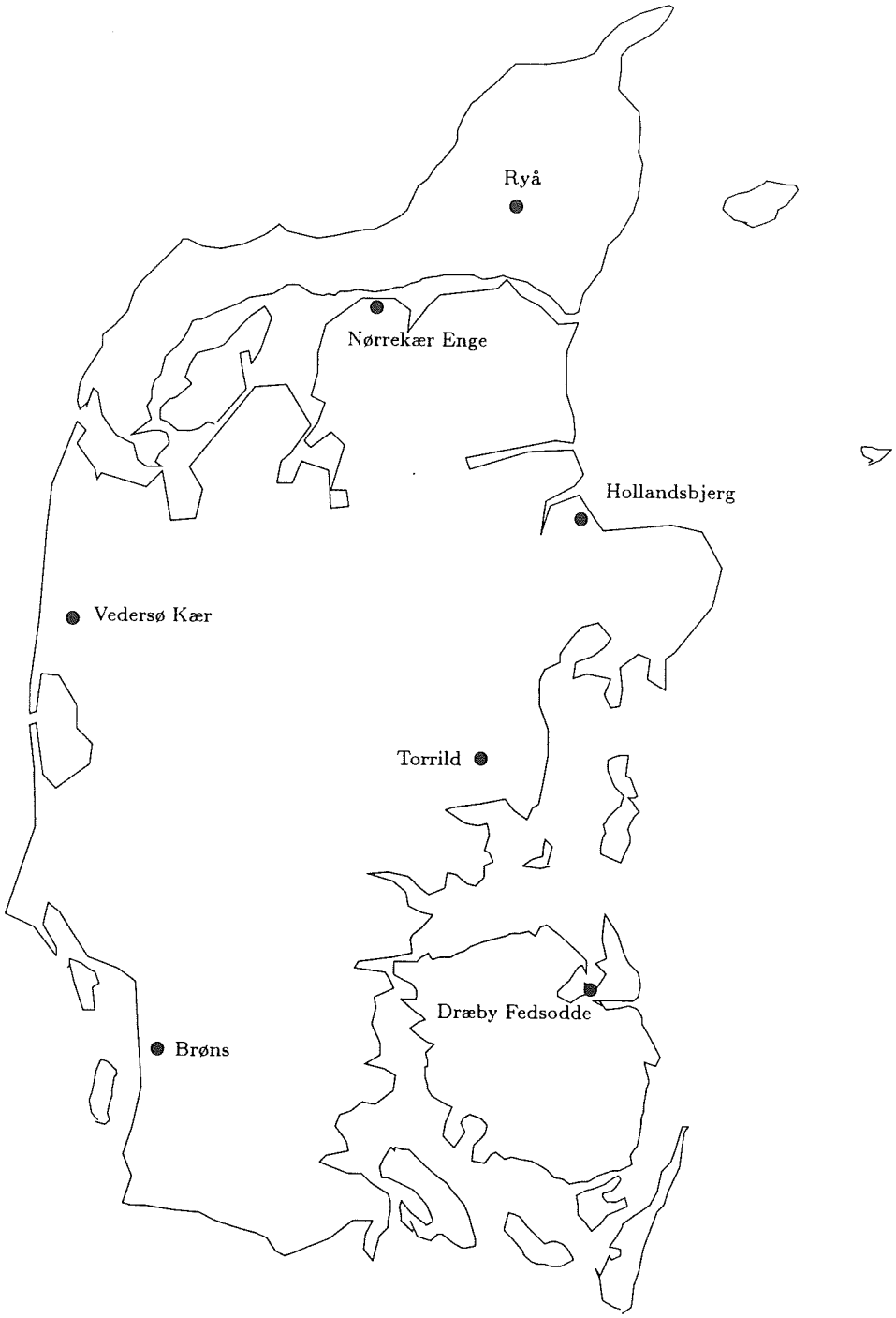
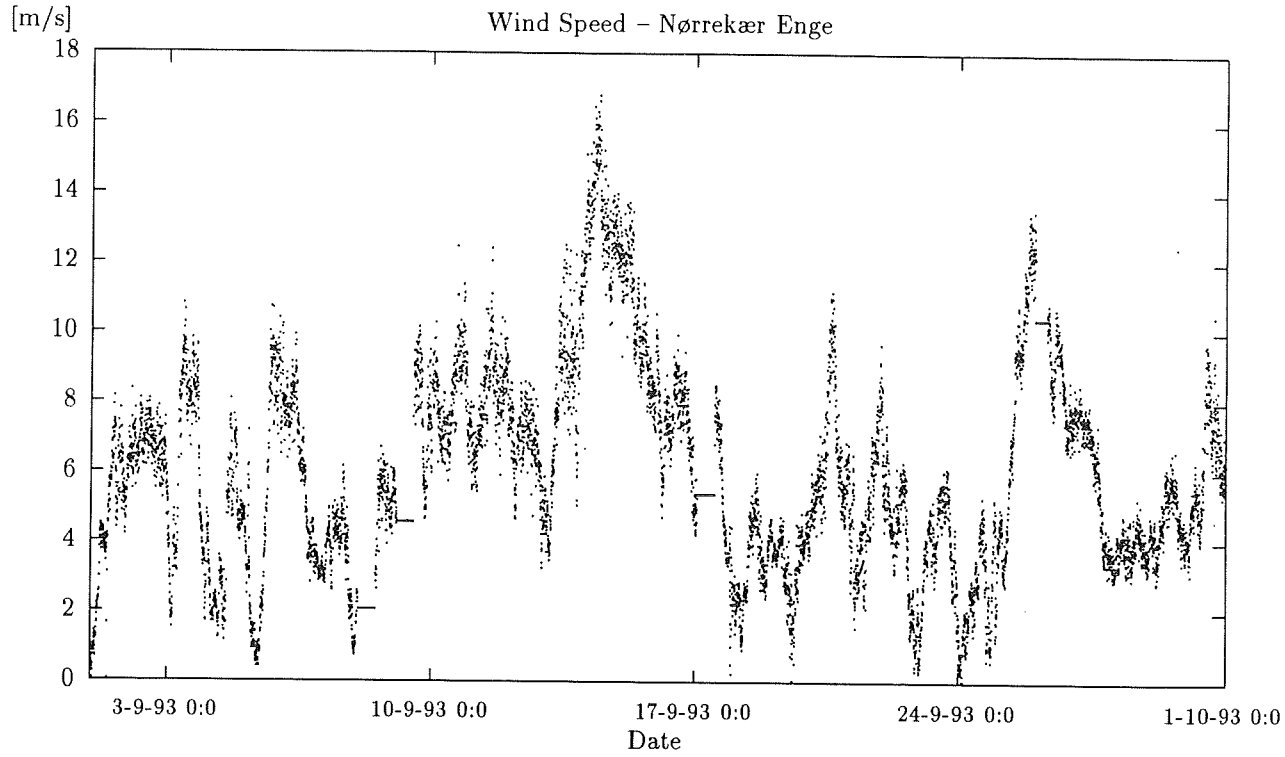
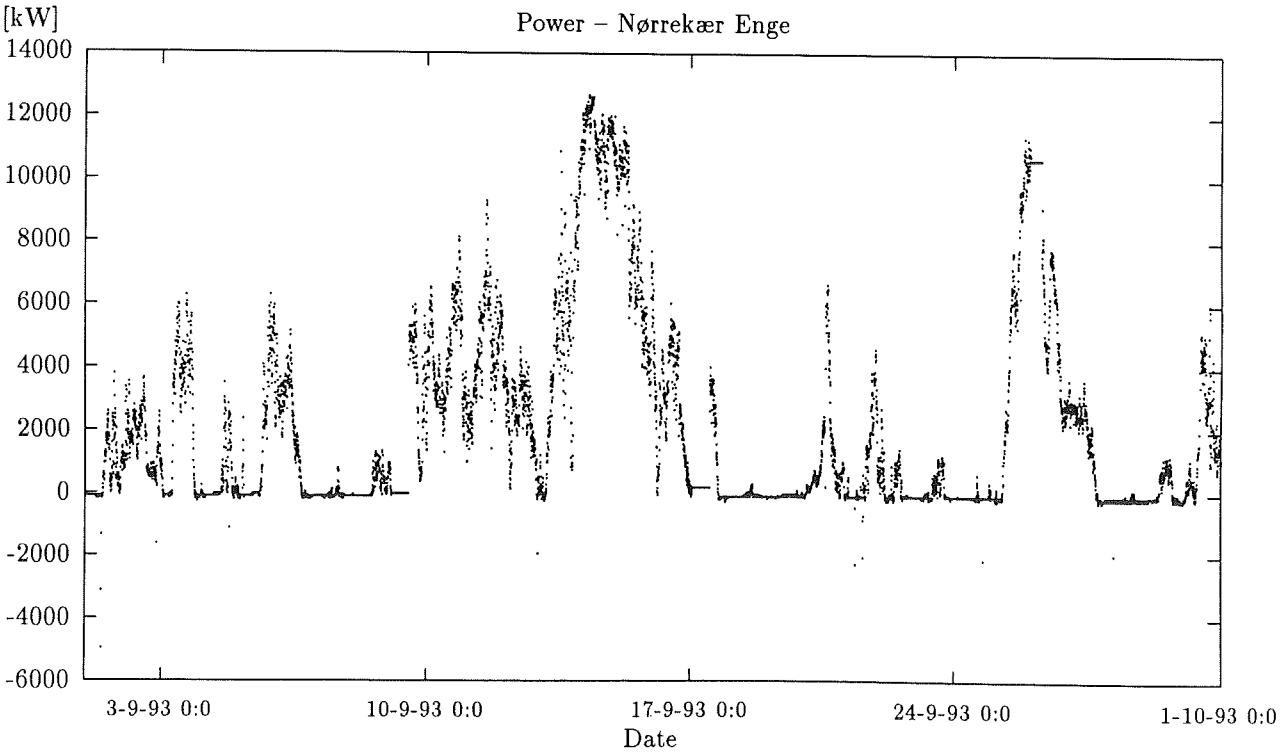
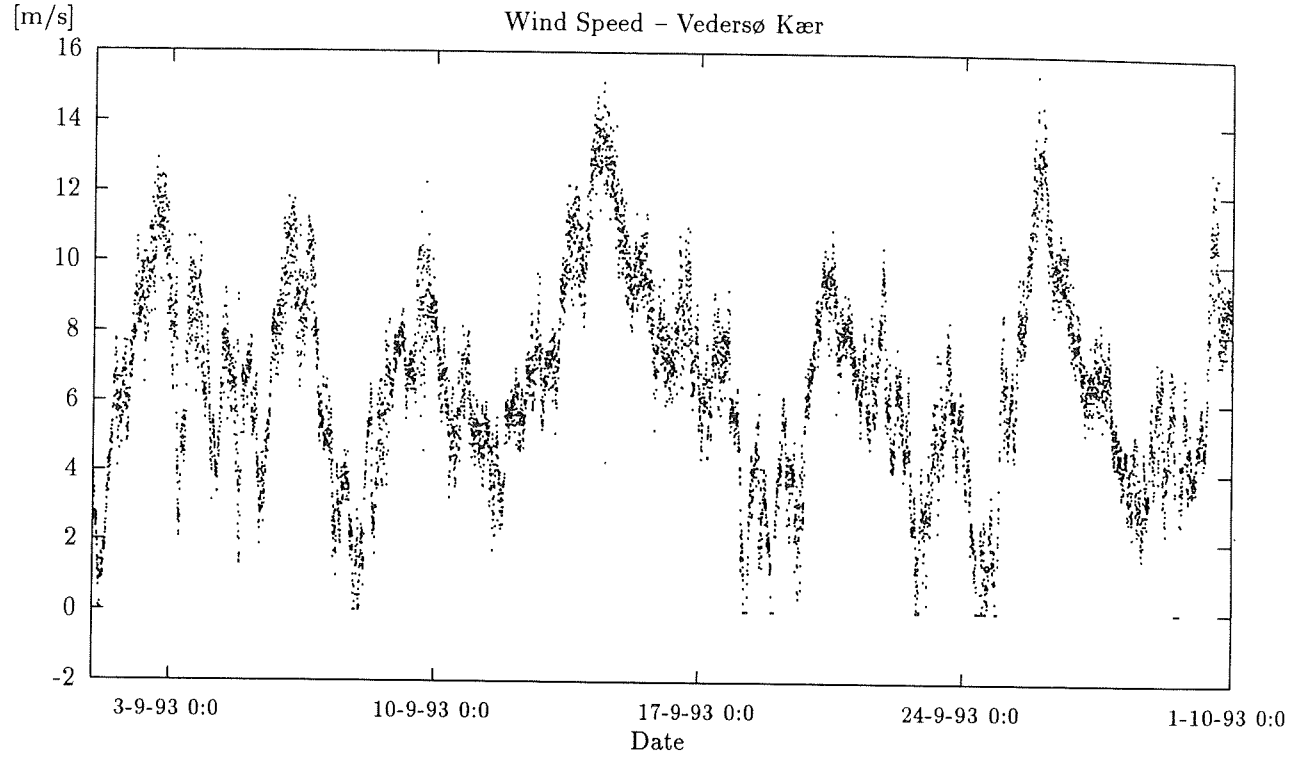
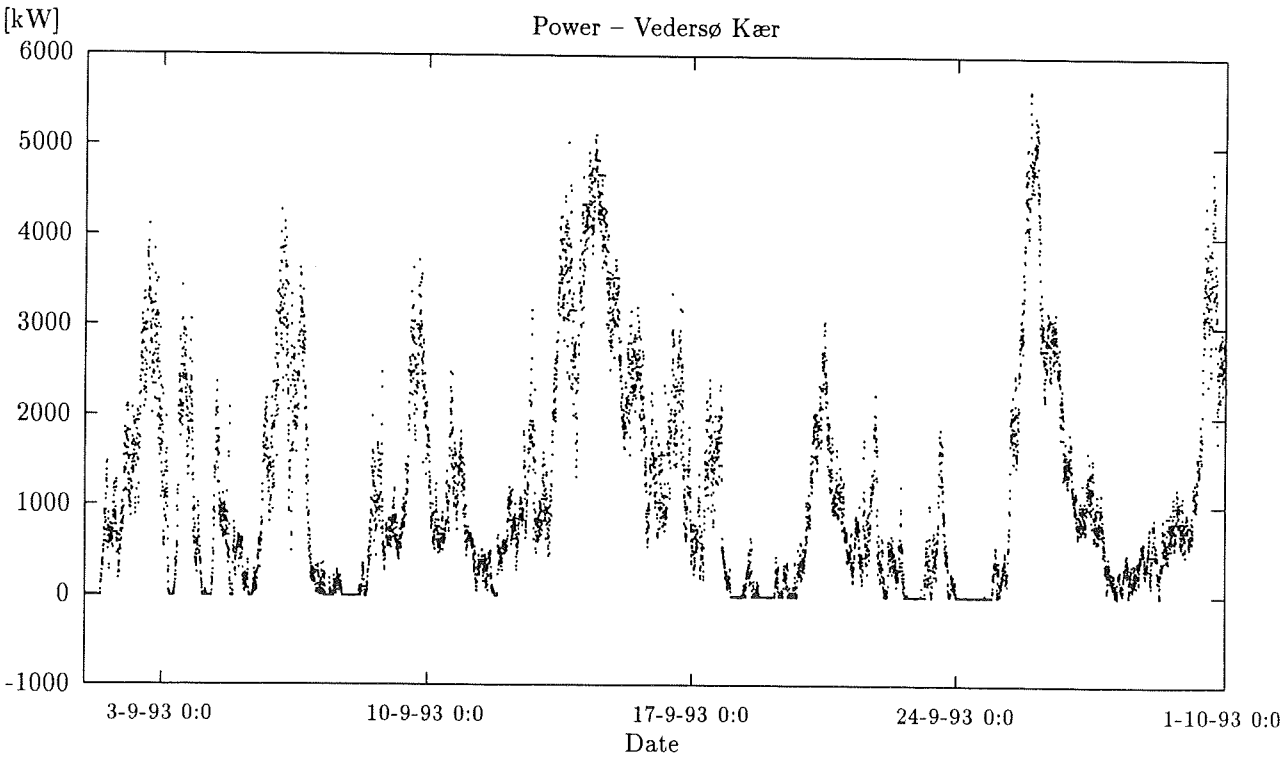


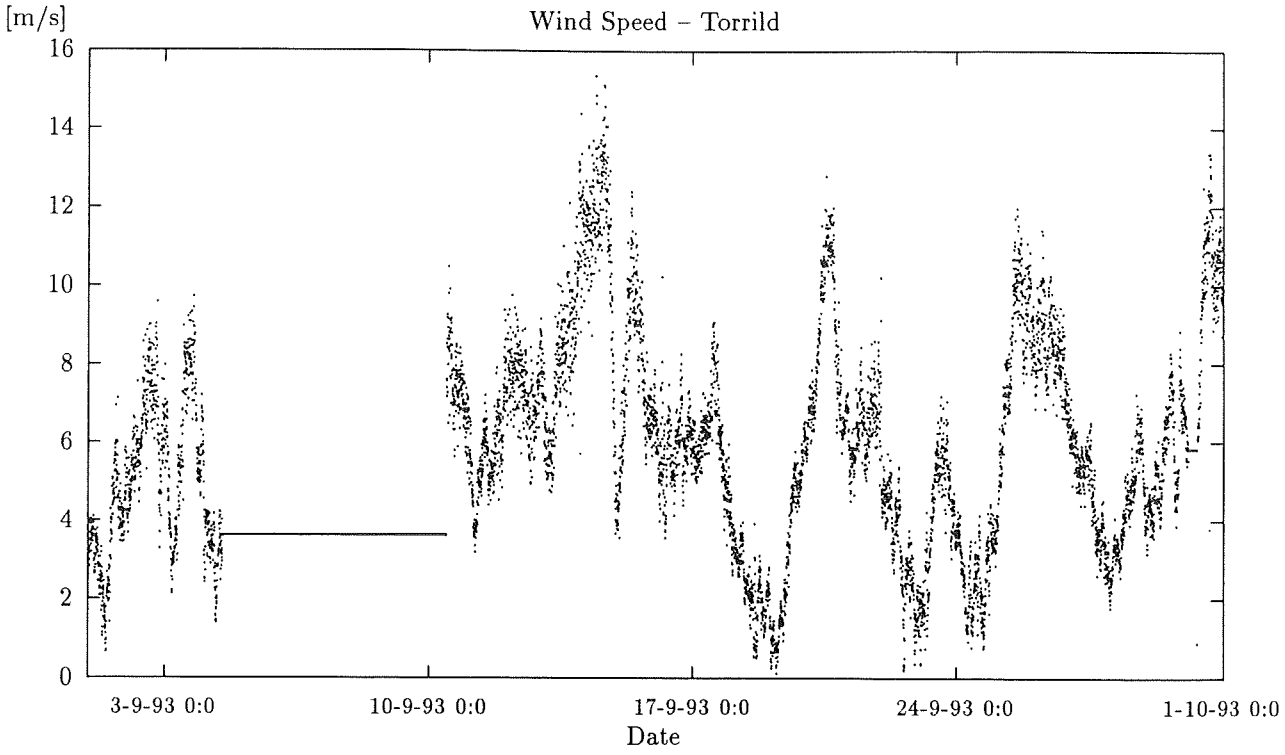
Figure 2.1: Key map of the ELSAM supply area with the location of the seven reference wind farms.

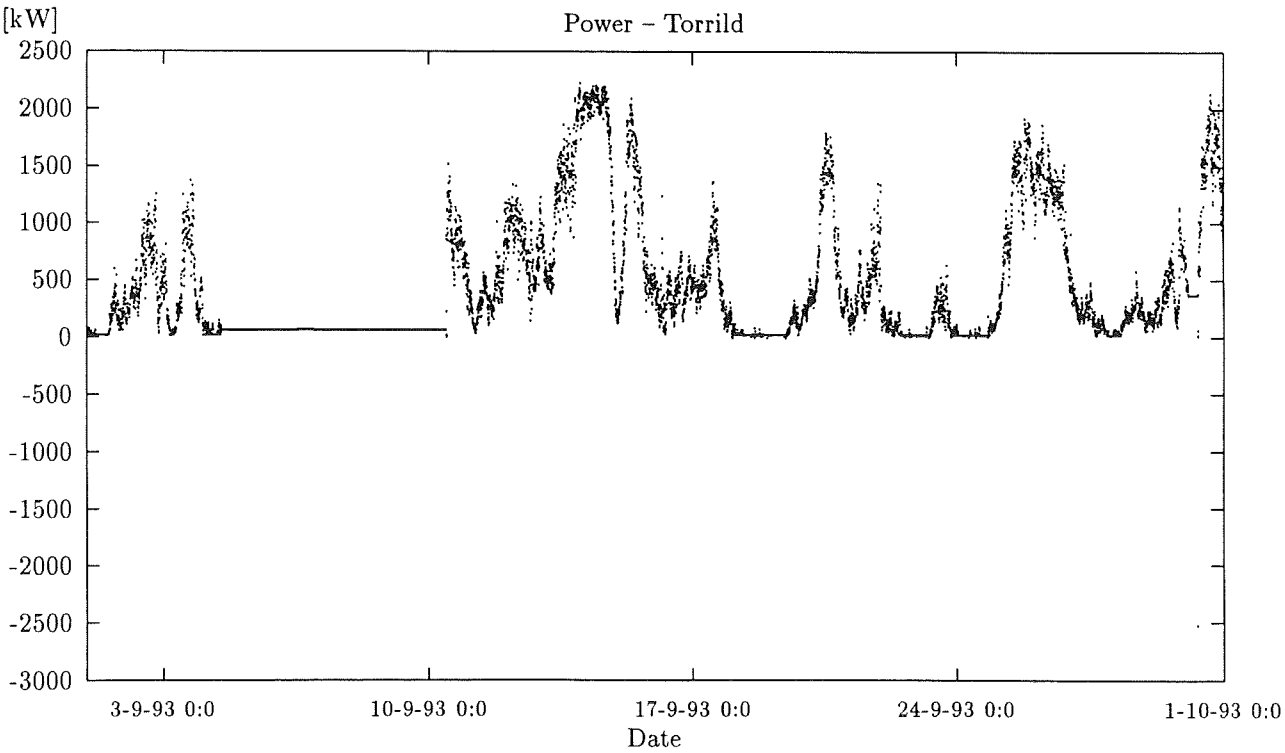

[m/s]
Wind Speed – Nørrekær Enge
18
16
14
12
10
8
6
4
2
0
3-9-93 0:0
10-9-93 0:0
17-9-93 0:0
24-9-93 0:0
1-10-93 0:0
Date


[kW]
Power – Nørrekær Enge
14000
12000
10000
8000
6000
4000
2000
0
-2000
-4000
-6000
3-9-93 0:0
10-9-93 0:0
17-9-93 0:0
24-9-93 0:0
1-10-93 0:0
Date


[m/s]
Wind Speed – Vedersø Kær
16
14
12
10
8
6
4
2
0
-2
3-9-93 0:0
10-9-93 0:0
17-9-93 0:0
24-9-93 0:0
1-10-93 0:0
Date


[kW]
Power – Vedersø Kær
6000
5000
4000
3000
2000
1000
0
-1000
3-9-93 0:0
10-9-93 0:0
17-9-93 0:0
24-9-93 0:0
1-10-93 0:0
Date


[m/s]
Wind Speed – Torrild
16
14
12
10
8
6
4
2
0
3-9-93 0:0
10-9-93 0:0
17-9-93 0:0
24-9-93 0:0
1-10-93 0:0
Date


[kW]
Power – Torrild
2500
2000
1500
1000
500
0
-500
-1000
-1500
-2000
-2500
-3000
3-9-93 0:0
10-9-93 0:0
17-9-93 0:0
24-9-93 0:0
1-10-93 0:0
Date

constant at 0 kW.

Periods of erroneous constant wind speed and/or power can easily be detected. However, it is difficult to reconstruct the lacking measurements. Reconstruction of the wind speed may be carried out by using both the time-wise correlation of the wind speed at one wind farm and the spatial correlation between the wind speed at different farms. From reliable wind speed measurements lacking power measurements may be reconstructed by use of a suitable power-to-wind speed relation. These reconstructed measurements may, however, deviate considerably from the true values.

- Single extreme measurements occur. However, these errors are easily detectable, and they can also be substituted by reasonable values. That is, single errors are not crucial for estimation of models and prediction as they can be handled by suitable techniques.

- It is only rare in the shown set of data that measurements are lacking. However, in the on-line situation lacking observations may be seen more frequently.

  Lacking observations should be treated in the same way as detected erroneous measurements. That is, when single measurements are absent they can easily be substituted, but when they are lacking for longer periods it is more complicated and not without introduction of errors to make a reconstruction. Furthermore, if measurements are absent from all wind farms the reconstruction corresponds to the prediction situation, and accordingly is followed by the same uncertainty.

- For some of the wind farms (Nørrekær Enge, Hollandsbjerg and Torrild) the level of the power measurements corresponding to zero production is below 0 kW and not constant. This kind of error is not crucial for the estimation and prediction performance. However, if it is due to a general bias of the power measurements it will lead to an undesirable displacement of the estimated power curve.

- The wind speed at Dræby Fedsodde does in the first part of the shown period follow a discretized pattern, and the power at this farm does for a certain period in August seem to have the fault that it is shifted to 3000 kW, when it should be 0 kW. These errors have in the later part been removed, and they are not likely to appear again.

- The wind speed measurements at Brøns seems to be erroneous in the way that it is attracted to certain levels around 4, 7 and 14 m/s. This error in the measurements is present in the whole period of data, and it is not possible to correct these errors. It is likely that the errors are due to defective measuring equipment.

It is important for estimation and prediction of power that periods of lacking or erroneous data are as short as possible. Robust techniques are able to handle erroneous data occurring alone or in shorter periods. However, when correct measurements fail to appear for longer periods the robust techniques cannot carry out a reliable reconstruction of erroneous data. Consequently, it is important that the number and the lengths of periods of erroneous data are kept at a minimum.

The quality of the measurements have improved considerably since 1993 for most of the wind farms. This improvement is mostly due to an established weekly contact with the people

who are responsible for the data logging equipment at the individual farms. This means that a much longer period of data of a higher quality is available today.

Chapter 3

Frequency Analysis of Wind Behaviour

In this chapter a frequency analysis of the wind behaviour is carried out. The purpose is to investigate and uncover the frequency contents of wind speed variations under such climatic conditions that are prevalent in Denmark and the Netherlands.

The particular interest is how much dynamic information is lost when wind speed is sampled at a given frequency. Furthermore, it is investigated whether it is imperative to use a faster sampling frequency followed by prefiltering and subsampling to avoid unwanted consequences of the aliasing effect, which otherwise will occur in the data.

Finally, some investigations are carried out to find a reasonable sampling procedure. Using high frequency data various sampling procedures are investigated.

3.1 Wind Speed Measurements

The investigation is based on five separate periods of wind speed measurements collected in Tjæreborg close to Esbjerg in Denmark. These measurements are sampled at $f_s = 25$ Hz meaning that all wind speed variations at frequencies below 12.5 Hz are preserved at the proper location in the frequency domain.

The length of each period is one hour leading to 90,000 samples for each period. The time and date of the first sample in each of the five periods are

I	16.44.04 at January 16, 1991.	**IV**	13.35.39 at February 23, 1992.
II	08.25.37 at February 14, 1991.	**V**	07.39.20 at August 5, 1992.
III	12.28.09 at February 28, 1991.		

The measurements were collected at a height of sixty meters with an anemometer placed on a meteorology mast in the vicinity of the 2 MW wind turbine at Tjæreborg, Denmark.

The measurements are accurate to the first decimal corresponding to the measuring errors having a zero mean uniform distribution over the interval [-0.05,0.05]. It is assumed that the dynamics of the anemometer itself is negligible at the sample frequency of 25 Hz.

Time series plots of the wind speed for period II and IV are shown in Figures 3.1-3.2. These two periods represent the extremes of the variation of all five periods. It is obvious that the measurements for each period are specific to the weather condition prevailing at the very hour when the measurements have been collected. The five periods differ with respect to the level of the wind speed as well as to the variation of the wind speed.

The smoothed periodograms, which constitute estimates of the spectral densities, of the five series in Figure 3.3 also show that there are differences among the five series. It is relevant for these periodograms to know that the periodogram of a sequence of independent uniform random variables on the interval [-0.05,0.05] is about -31.5 dB in the full frequency range. It can be seen that for all series the periodogram approaches a level close to -31.5 dB for increasing frequency, meaning that for such frequencies there is no frequency content in the series apart from the independent random variation due to the measuring/round-off errors.

The extreme series is the second one, the periodogram of which reaches the noise level (~ -31.5 dB) at a frequency near 2.5 Hz ($0.1 \times f_s$), and the fourth series, which seems to have frequency contents up to 10 Hz. The reason that the second series does not seem to have any substantial variation above 2.5 Hz might just be that the signal-to-noise ratio is so low that the wind speed variations are covered by noise. Actually it is difficult to tell whether the differences in the periodograms are caused by differences in the variances of the series alone or if there are additional differences in the characteristics of the series.

The investigation is carried out for all the series, but only the results for the fourth series are shown in the following.

3.2 Sampling Wind Speed for Prediction Purposes

Having the purpose of predicting wind speed and wind power production from 1 to 36 hours ahead in mind, it is not appropriate to base prediction models on 25 Hz measurements. Intuitively it is obvious that variations above a certain frequency will not be of influence to the production of the wind turbine. The reason is that such fast wind variations will not influence the energy transfer from wind to turbine, i.e., the turbine has a low-pass filtering characteristic.

In order to be able to match standard meteorological data, which comes as 10-minute-registrations, and production data (15-minute-registrations), it has been decided to prepare the wind measuring equipment for 5-minute-registrations. The question is how the 5-minute-registrations should be obtained to constitute the optimal basis for the prediction purpose.

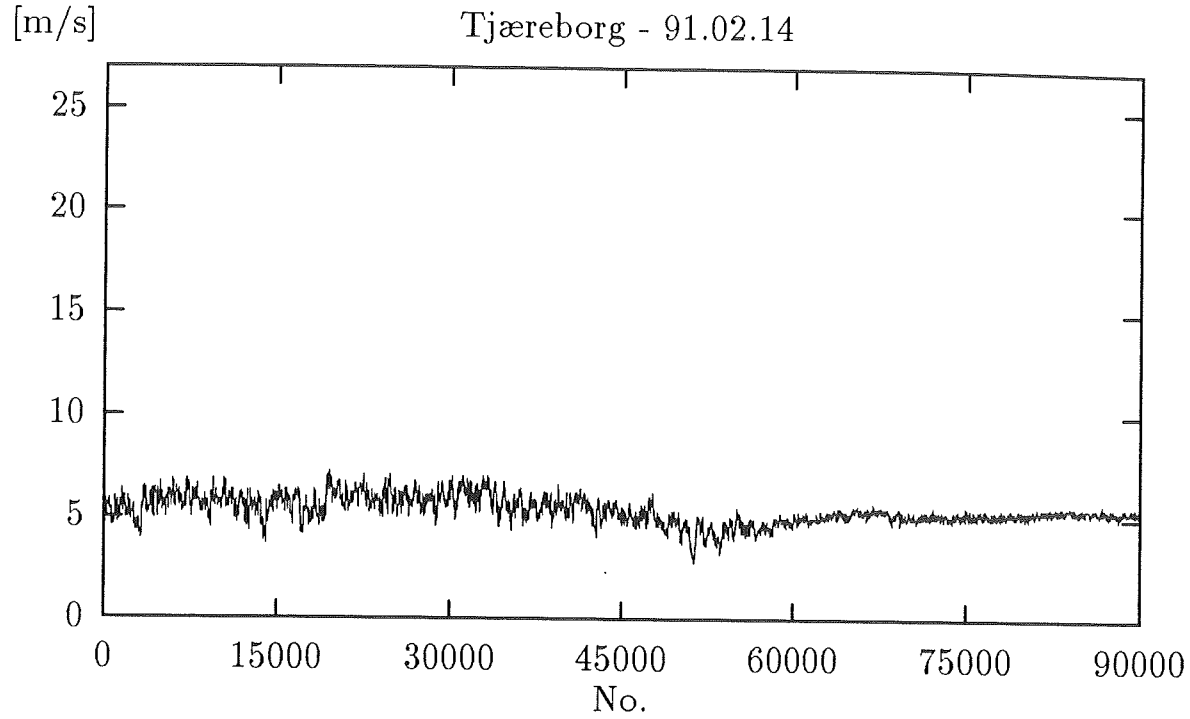


Figure 3.1: Data II.

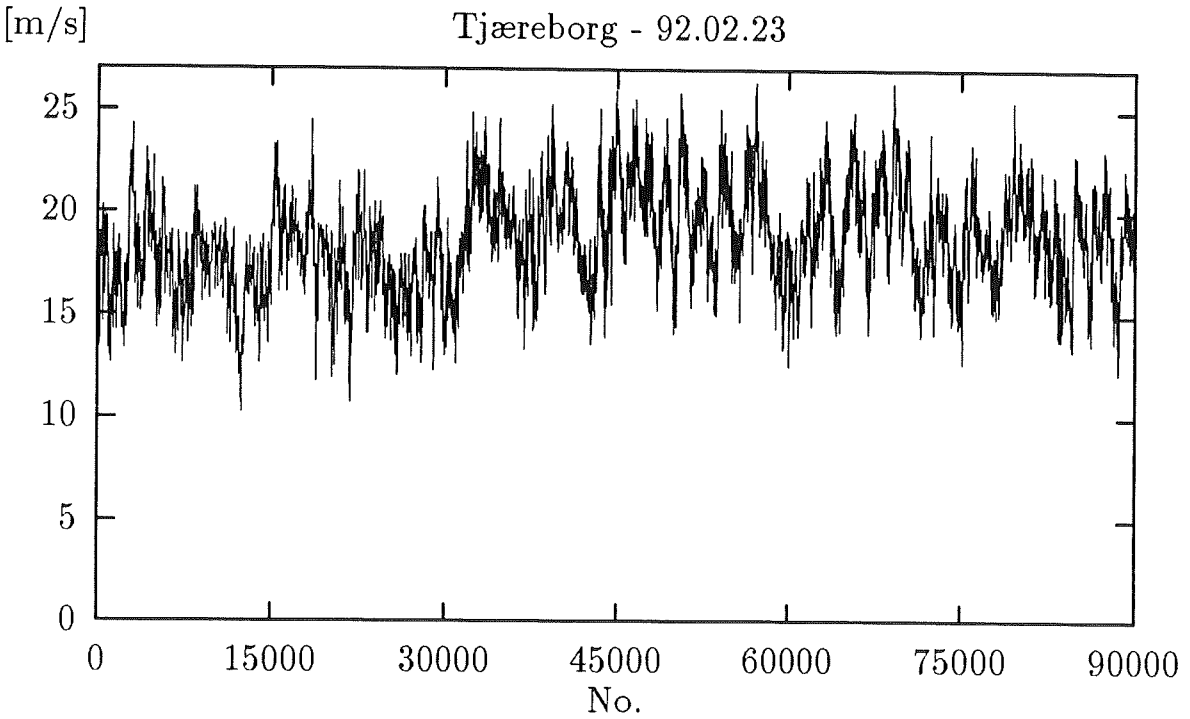


Figure 3.2: Data IV.

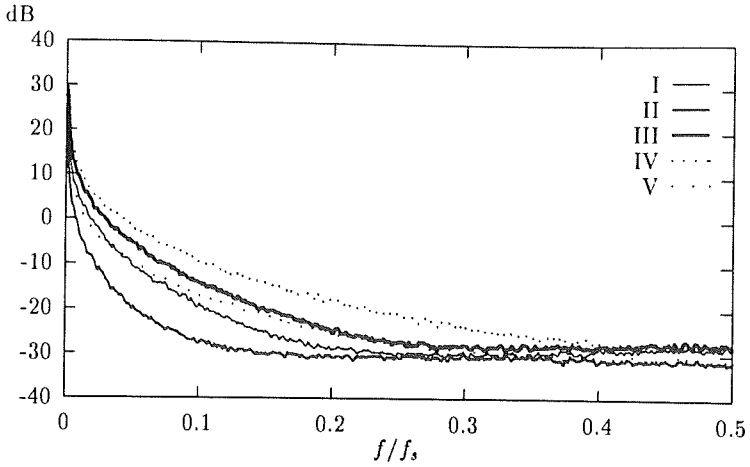


Figure 3.3: Smoothed periodograms for the five series.

3.2.1 Aliasing

As seen from the periodograms of the previous section, it is given that wind speed measurements may have frequency contents up to 10 Hz. When the wind speed is represented by 5-minute values, the high-frequency variations cannot be described. As a matter of fact for 5-minute values it is the case that only variations at frequencies below .00167 Hz ($\sim 1/600$ s^{-1}) can be observed. The upper limit on the frequency variations that can be observed is known as the *Nyquist frequency*, see e.g. Bloomfield (1976), and it is given by $1/2T$, where T is the length of the sampling interval. Below is illustrated why this is the case and what happens with frequency contents above the Nyquist frequency.

Suppose that the sampling interval is T, and that data consist of a pure cosine wave at angular frequency $\omega = 2\pi f$. Hence the observations at tT for integer values of t will be

$$x_t = \cos \omega t T. \tag{3.1}$$

When increasing ω from zero the wave oscillates more and more rapidly until at the Nyquist frequency $\omega = \pi/T$, where the observations are

$$x_t = \cos \pi t = (-1)^t. \tag{3.2}$$

This corresponds to the shown sampling of the lower wave in Figure 3.4. Increasing the wave frequency further to a value, satisfying $\pi/T < \omega < 2\pi/T$, gives

$$\begin{aligned} x_t &= \cos \omega t T \\ &= \cos\left(\frac{2\pi}{T} - \omega'\right) tT \quad , \quad \omega' = \frac{2\pi}{T} - \omega \\ &= \cos\left(2\pi t - \omega' t T\right) \\ &= \cos \omega' t T. \end{aligned} \tag{3.3}$$

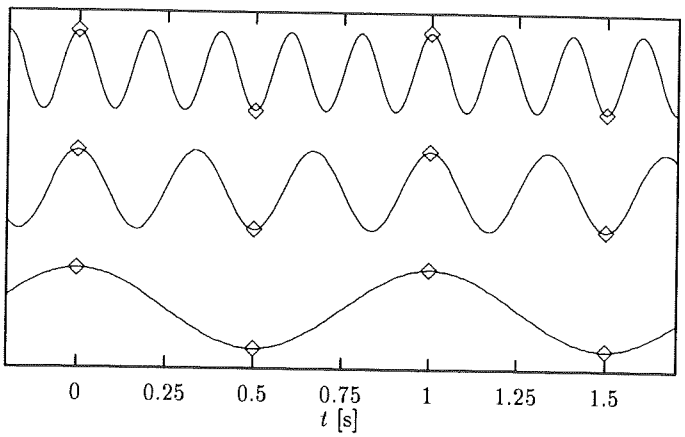


Figure 3.4: 2.0 Hz sampling of sinusoids with frequencies 1, 3 and 5 Hz.

Similarly it can be shown that $\sin \omega t T = -\sin \omega' t T$. This means, given the sampling interval T, the two signals

$$A\cos(\omega t) + B\sin(\omega t) \quad \text{and} \quad A\cos(\omega' t) - B\sin(\omega' t),$$

apart from a phase shift, are indistinguishable, and they are said to be aliases of each other. The same is said about the angular frequencies ω and ω'. The same argumentation can be carried through for any positive frequency, and it can be concluded that every frequency not in the range $0 \leq \omega \leq \pi/T$ has an alias in that range. Figure 3.4 shows how the two upper waves are aliases of the lower wave, when sampled as marked on the figure.

In the frequency domain the variation of the stationary continuous signal $x(t)$ is described by the spectral density, given as the Fourier transform of the auto covariance function, i.e.,

$$f_x(\omega) = \int_{-\infty}^{\infty} \gamma_x(\tau) e^{-i\omega\tau} d\tau, \quad -\infty < \omega < \infty \tag{3.4}$$

where the auto covariance function is

$$\gamma_x(\tau) = \mathrm{Cov}\left[x(t), x(t+\tau)\right]. \tag{3.5}$$

The spectral density of the sampled signal can be shown to be (Bloomfield, 1976)

$$f_x^T(\omega) = \sum_{m=-\infty}^{\infty} f_x(\omega - m\frac{2\pi}{T}), \quad -\frac{\pi}{T} < \omega < \frac{\pi}{T}. \tag{3.6}$$

It is clear that the spectral density of the sampled signal has contributions from the spectral density of the continuous signal, which correspond to frequencies above π/T. Consequently, if the continuous signal has variations of frequencies above π/T, these will after sampling be interpreted as contributions at lower frequencies.

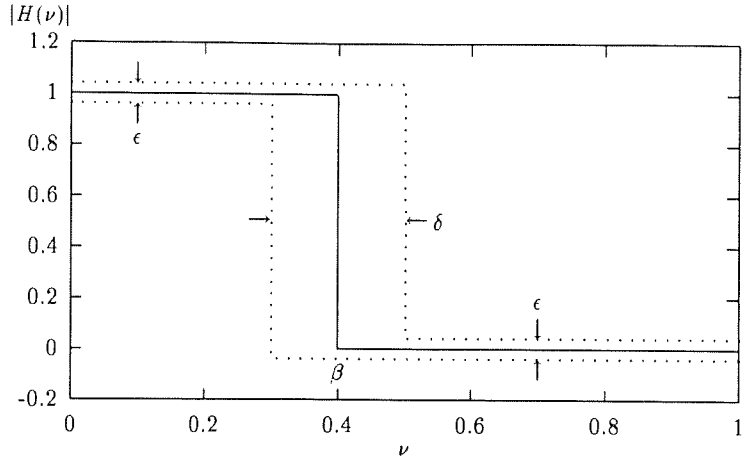


Figure 3.5: Definition of δ, ϵ and β used in design of Nearly Equal Ripple filters.

3.2.2 Prefiltering

One way of avoiding the undesired down-folding of high-frequency aliases into the Nyquist interval is to carry out the sampling at a sufficiently high frequency, considering the frequency contents of the signal, and subsequently apply a presampling filter followed by subsampling of the series. The purpose of the prefiltering is to remove the frequency contents that otherwise would be folded into the interval $-\pi/T \leq \omega \leq \pi/T$, T being the sampling interval after subsampling.

Kaiser & Reed (1977) describes algorithms for designing digital low-pass filters that comply with given tolerances on the deviation from ideal filters. The output according to a given set of filter weight coefficients b_k is given by

$$y_t = \sum_{k=-N_p}^{N_p} b_k x_{t-k}. \tag{3.7}$$

The frequency response function of the filter is

$$H(\omega) = \sum_{k=-N_p}^{N_p} b_k e^{-i\omega T k}, \tag{3.8}$$

and the spectral density of the filtered signal is

$$f_y^T(\omega) = |H(\omega)|^2 f_x^T(\omega), \quad -\frac{\pi}{T} < \omega < \frac{\pi}{T}. \tag{3.9}$$

For symmetric weights, $b_k = b_{-k}$, the phase shift of the filter is zero, and expressing the filter in terms of a normalized frequency $\nu = \omega/(\pi/T)$, for which the Nyquist interval $0 < \omega < \pi/T$

maps to the interval $0 < \nu < 1$, gives

$$H(\nu) = b_0 + \sum_{k=1}^{N_p} 2b_k \cos \pi k\nu. \tag{3.10}$$

An ideal low-pass filter removes all variation at frequencies above a band width β, i.e.

$$H_i(\nu) = \begin{cases} 1 & , \quad |\nu| \leq \beta \\ \\ 0 & , \quad \beta < |\nu| \leq 1 \end{cases} . \tag{3.11}$$

This ideal filter requires an infinite number of weight terms $\{b_k\}$. Kaiser & Reed (1977) remark that truncating the sequence of weights $\{b_k\}$ results in a bad approximation to the ideal filter with considerable deviation at the band width frequency β. Instead they propose a filter, denoted Nearly Equal Ripple (NER) approximation, that has the effect of spreading out the large approximation error, found at the band edge, over the whole of the pass and stop bands. See Kaiser & Reed (1977) for the algorithm for designing the filter as well as the Fortran code for its implementation. As input to the design algorithm must be given the tolerances δ and ϵ on the deviation from the ideal filter along with the location of the band width frequency β, see Figure 3.5.

3.2.3 Prefiltering the Wind Speed Measurements

In this section it is investigated whether the NER prefiltering has a valuable effect on the wind speed measurements compared to either simple subsampling or average calculation.

The fact that the relation between the sampling interval of 0.04 s of the original measurements and the desired 300 s of the final data is so considerable sets up heavy demands on the filter, if it is desired that all variation above the Nyquist frequency, for $T = 300$ s, ought to be removed in one single filtering step. This would require $\beta = 7500^{-1}$, and consequently if the stationary gain at $\omega = 0$ is desired to be within the ϵ-tolerance, the δ-tolerance must be below $2/7500$. To meet this demand, which indeed is a reasonable requirement, the filter length N_p should be above 6800.

A more reasonable procedure seems to be to divide up the prefiltering and subsampling in a number of stages, where each stage consists of prefiltering followed by subsampling. This could be implemented in three stages as follows

1. Apply the NER filter with $(\epsilon, \delta, \beta) = (0.02, 0.02, 0.04)$ on the 25 Hz measurements, and then subsample every twenty of the filtered measurements. Hereby the resulting 1.25 Hz measurements contain effectively no frequency contents of the original series from above 0.625 Hz. The reason that 'the frequency of the subsampled series is placed above 1 Hz, corresponding to the filter band width, is to avoid aliasing of frequency contents from just above the filter band width, having in mind that the filter is not ideal.

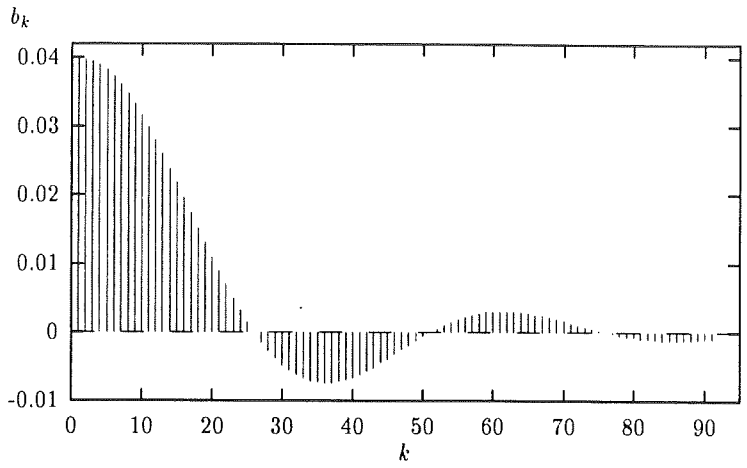


Figure 3.6: Time domain anti-aliasing filter $(\epsilon, \delta, \beta) = (0.02, 0.02, 0.04)$.

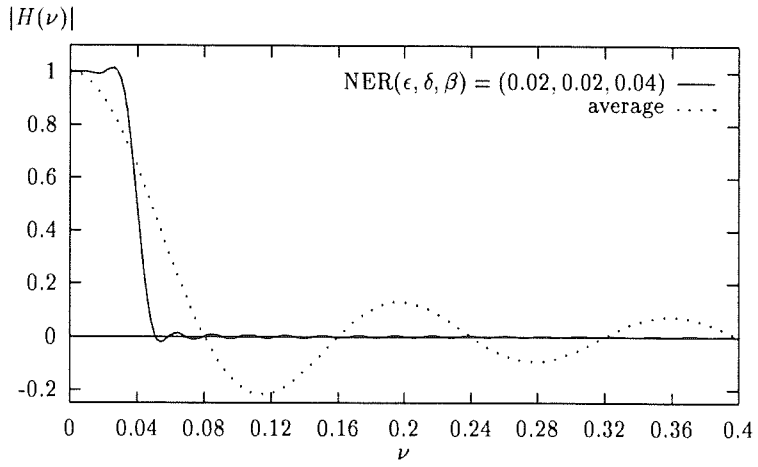


Figure 3.7: Frequency response characteristic of anti-aliasing filter $(\epsilon, \delta, \beta) = (0.02, 0.02, 0.04)$ and of a 25 observation equally weighted average.

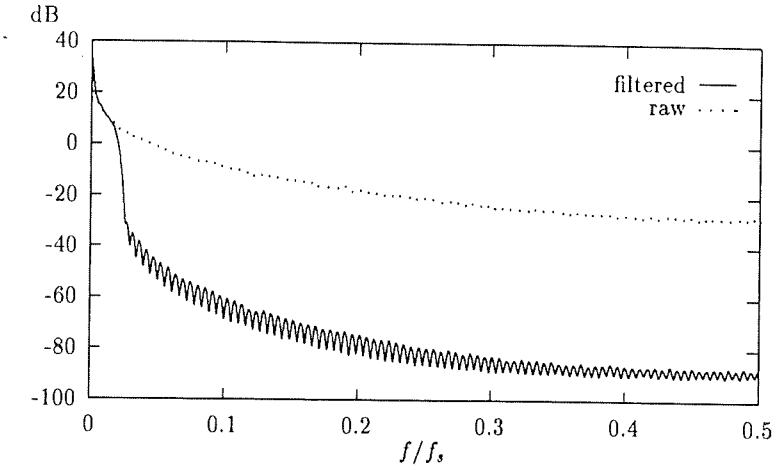


Figure 3.8: Smoothed periodogram of the fourth series, filtered and raw.

2. Apply the NER filter with $(\epsilon, \delta, \beta) = (0.02, 0.02, 0.05)$ on the 1.25 Hz measurements, and then subsample every fifteen of the filtered measurements to give measurements with sampling interval 12 s.

3. Finally apply the NER filter with $(\epsilon, \delta, \beta) = (0.02, 0.02, 0.04)$, and subsample every twenty five of the filtered measurements. In this last stage the frequency of the subsampled series matches the filter band width.

Figure 3.6 shows the weight sequence of the prefilter with $(\epsilon, \delta, \beta) = (0.02, 0.02, 0.04)$, and in Figure 3.7 is shown the amplitude of the frequency response of the filter. Similarly is shown the frequency response of an average calculation corresponding to an equally weighted smoothing of 25 neighbouring measurements. It is seen that the average indeed is a low-pass filter, which, however, shows heavy deviations from the ideal filter with $\beta = 0.04$.

Figure 3.8 shows the smoothed periodogram of the fourth series before and after the series has been run through the NER filter. A simular figure could have been shown for the second series. It is seen that the filter has the desired effect of down-scaling the spectral density with at least a factor 100 (~ -40 dB) for f/f_s above 0.025. This frequency corresponds to $\beta + \delta/2$ of the filter specification.

Figure 3.9 shows the smoothed periodogram of the filtered series after subsampling to 1 Hz measurements. In the same figure is also shown the smoothed periodogram of the raw series transferred to 1 Hz measurements using either pure subsampling or pure average calculation. It is assumed that the periodogram of the filtered and subsampled series is the one closest to the true spectral density of the signal in the Nyquist interval. The periodogram of the pure subsampled series show considerable deviations from those of the filtered series. The deviations are reasonably explained as down-folded frequency aliases of the raw series from above the Nyquist frequency. The deviations tend to increase with frequency corresponding to, as the scale is logarithmic, the aliased frequency contributions being additive effects.

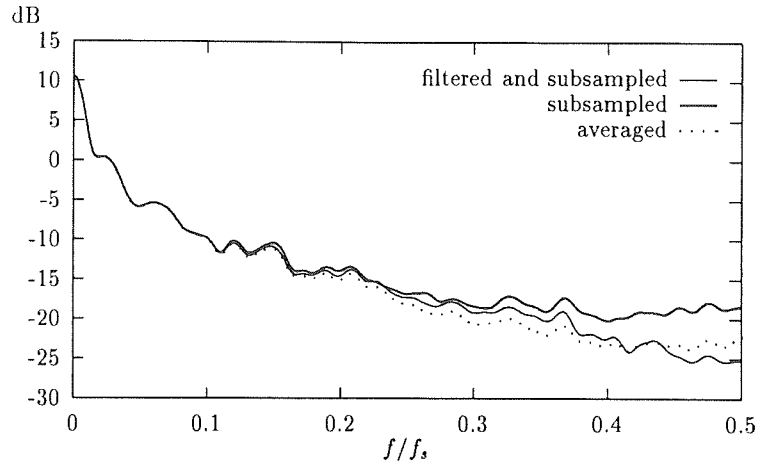


Figure 3.9: Smoothed periodogram of the fourth series (1 Hz).

For the average of the raw series it is seen that the periodogram is below the periodograms of the filtered series in the middle of the Nyquist interval. However, for frequencies approaching the Nyquist frequency the two periodograms intersect, and the averaged series show to have more spectral variation just below the Nyquist frequency. The explanation to this result is probably that the equally weighted average is a low-pass filter, which deviates from the ideal filter (3.11) in such a way that aliasing of variations above the Nyquist frequency is not totally avoided, and variations below the Nyquist frequency are down-scaled – see the frequency response in Figure 3.7. In Figure 3.10 is shown the periodogram of the fourth series after filtering with $\beta = 0.04$, subsampling to 1.25 Hz, filtering with $\beta = 0.05$ and finally subsampling to 1/16 Hz. In the filtering stages is used $(\epsilon, \delta) = (0.02, 0.02)$. The periodogram is compared to the periodogram for subsampling of raw data and equally weighted averages. The same comparison as for the 1 Hz measurements can be made. That is, the pure subsamples have aliased contributions in the full range of the Nyquist interval, and averaging has the effect of removing some spectral information for frequencies approaching the Nyquist frequency and folding variation from above the Nyquist frequency down below.

It would be relevant to pursue the same comparisons for sampling interval $T = 5$ m. However, for the one hour sequences there would only be twelve or less observations left, and this is considered to be insufficient for the spectral analysis.

3.3 Conclusions

The spectral analysis presented in this chapter of five sequences of wind speed measurements of one hour length sampled at 25 Hz has shown that the spectral density of 5-minute registrations of wind speed may have aliased contributions from frequencies up to 10 Hz if no appropriate prefiltering is used.

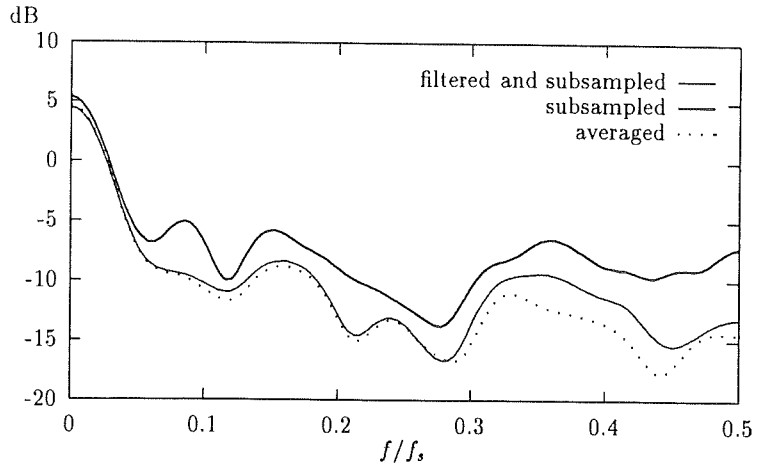


Figure 3.10: Smoothed periodogram of the fourth series (1/16 Hz).

The best reduction of aliased contributions is obtained by sampling fast (20-25 Hz) and applying a filtering-subsampling procedure that makes use of a Nearly Equal Ripple filter with a characteristic close to an ideal filter. Alternatively equally weighted averages could be used. However, the low-pass filtering of the average calculation has clear deviations from the ideal filter, having the unwanted consequence that relevant spectral information is down-scaled and higher frequency contents are folded down into the Nyquist interval.

Considering predictions it is clear that the presence of aliased spectral variations contributes to the unpredictable part of the variations. It therefore has a negative effect on the prediction quality leading to the conclusion that prefiltering is beneficial with respect to prediction of wind speed.

Chapter 4

Models for Predicting Wind Speed and Power

This chapter presents the main results concerning linear and non linear parametric modelling and prediction of the wind speed and the power at a wind farm. In Chapter 5 an alternative method, namely the neural network approach, is considered.

The modelling will focus on models that are to be used for predicting the power output in a wind farm on a relatively short horizon, say up to 36 hours ahead. In order to obtain a prediction of the total wind power production in some geographical region (e.g. the ELSAM supply area – see Figure 2.1), the predictions for the individual farms have to be extrapolated to a prediction for the whole region. This problem, which is called the upscaling problem, is discussed in Chapter 6.

Three conceptually different approaches for the *formulation of the model* that is to be used for obtaining the predictions are investigated in the present chapter. The three different approaches are

1. A model for the wind speed is used to calculate predictions of the wind speed. Another model, describing the relation between wind speed and power for the wind farm in consideration, is used to convert the wind speed predictions into power predictions. That is, the power predictions are obtained by using two models estimated separately.

2. One model holding a description of both the correlation structure of the wind speed and the relation between the wind speed and the power is used to obtain the power predictions. That is, the model is a composition of the two models referred to above with all the parameters being estimated simultaneously.

3. One model describing the correlation structure of the power and the cross correlation between power and wind speed are used for calculating the predictions. This leads to models which are linear in the parameters.

When it comes to *estimation of the parameters* of a model there are two possibilities. One possibility is to *estimate the model parameters off-line* on a limited set of data. By doing

this it is possible to estimate models being nonlinear in the parameters, and subsequently carry out an evaluation of the model estimate. The disadvantage of this way of estimating the model parameters is that when new data is collected and one wants to base the model estimate also on these data, the model has to be re-estimated on the full set of data. The off-line methods and models are described in Sections 4.1 to 4.4.

Alternatively, *the model parameters can be estimated by recursive methods*, e.g. the recursive least squares algorithm, see Ljung (1987) and Section 4.5.1. These methods carry out the estimation sequentially in such a way that the parameter estimate is updated stepwise as data is sampled. This implies considerable savings of computer time, compared to the off-line estimation techniques, especially when it is a demand that the influence of on-line data is to be reflected in the model estimates immediately as they have been sampled. The recursive estimation method is often called *on-line estimation method.*

Furthermore, by minor modifications of the recursive algorithms it is possible to obtain that the estimation becomes adaptive. Hereby, the model estimate will be able to follow slow changes in the system, for instance, changes in the number of wind mills being active in the wind farm or even changes in the characteristics of the wind speed variations as the seasons change.

The drawback of the recursive methods is, however, that it is problematic to estimate the parameters of nonlinear models recursively, meaning that it cannot be guaranteed that parameter estimates will converge to the desired values. Therefore, for practical purposes the possibility of using recursive estimation of models being nonlinear in the parameters is considered not to be preferable. The on-line – or recursive – models and methods are described in Section 4.5.1.

The main contributions to the concept of modelling wind speed and power production given in this chapter are

- *Models for the dependence of power on wind speed and direction* are formulated in Section 4.2 using data from wind farms at Mors and Ærø, respectively. For these farms simultaneous measurements of power, wind speed and direction are available, and therefore they can be used for carrying out an investigation of how the power curve depends on the wind direction.
- *Models for the relation between power and wind speed* without having measurements of the wind direction are found in Section 4.3 using data from the wind farm at Vedersø Kær. The lacking measurements of the wind direction are compensated for by modelling the correlation of the deviances of the observations from the power curve.
- Different linear as well as nonlinear *models for the variation of the wind speed* are estimated on the same set of data from Vedersø Kær as used for the power curve analysis. This is described in Section 4.4.
- *Recursive estimation of linear models for wind speed* is found in Section 4.5.2 and 4.5.3. These models are used for predicting the wind speed, which then is transformed into a prediction of the power.
- *Recursive estimation of models for power* which include a description of the dependency

on wind speed is considered in Section 4.5.1. These models are formulated as explicit predictors for the power and estimated adaptively and evaluated on the data from Vedersø Kær, Dræby Fedsodde and Brøns.

4.1 Least Squares Estimation

This section describes the least squares (LS) estimation method, which is used below for off-line (non-recursive) estimation of the parameters θ of linear as well as nonlinear models. The process to be described by the model is denoted $\{y(t)\}$, and the prediction conditional on given values of the parameters and on an appropriate set of observations (dependent on the model) is denoted $\hat{y}_\theta(t)$. Accordingly the estimation is carried out by a minimization of the criterion

$$V(\theta) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\left(y(t) - \hat{y}_\theta(t)\right)^2 = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\hat{e}^2(t), \tag{4.1}$$

where N is the number of observations used for the estimation.

In this investigation the estimation is carried out by numerical minimization of the criterion (4.1). The procedure for minimization is described in Melgaard & Madsen (1991). The minimization routine provides the numerical approximation to the Hessian, $\hat{H}(\hat{\boldsymbol{\theta}})$, obtained in the minimum of the criterion,

$$\hat{H}(\hat{\theta}) = \frac{\partial^2 V(\theta)}{\partial\theta^2}\Big|_{\theta=\hat{\theta}}. \tag{4.2}$$

Based on the assumption of normal distributed white noise the covariance matrix for the estimate of θ is estimated by

$$\widehat{\mathrm{Cov}}\left\{\hat{\theta}\right\} = \frac{\tilde{\sigma}_e^2}{N}\left[\hat{H}(\hat{\theta})\right]^{-1}, \tag{4.3}$$

where

$$\tilde{\sigma}_e^2 = \frac{1}{N}\sum_{i=1}^{N}\hat{e}^2(i). \tag{4.4}$$

The unbiased estimate of the standard deviation of the prediction errors is given by

$$\hat{\sigma}_e^2 = \frac{1}{N-p}\sum_{i=1}^{N}\hat{e}^2(i) \tag{4.5}$$

with p being the number of estimated parameters.

Bayes Information Criterion (BIC) can be used as an indication of the optimal model choice in the model giving the minimum value, see e.g. Harvey (1989). It is given by

$$BIC = N\log\tilde{\sigma}_e^2 + p\log N. \tag{4.6}$$

For more information on estimation of linear dynamic models see Abraham & Ledolter (1983), Box & Jenkins (1976), Ljung & Söderström (1983), Söderström & Stoica (1989) and Madsen (1989). In the following also some non-linear models are considered. Information on modelling non-linear systems can be found in Tong (1990), Madsen (1985), Priestley (1988), Robinson (1983) and Bard (1974).

4.2 Models for the Dependence of Power on Wind Speed and Direction

For a wind farm it is reasonable to expect that the power depends on both wind speed and direction. First of all, it is clear that the the wind speed at each mill is affected by the existence of other wind mills since the wind mills give shelter to each other. Furthermore, the wind speed also depends on the surrounding landscape, e.g. vegetation and distance to open see.

The purpose of the investigation in this chapter is to analyze the influence of the wind direction. This investigation is carried out by using data from two wind farms at Mors and Ærø, respectively. These data have been collected previous to this project and contain measurements of power along with wind speed and direction. *The sampling interval is one minute*, and the data are used without sub-sampling or averaging. The data from Mors are sampled between December 30, 1989 and February 25, 1990 and contain 81093 observations. Data from Ærø are sampled between January 15, 1990 and April 21, 1990 and contain 123350 observations.

The dependence of power $p(t)$ on wind speed $w(t)$ is in this section assumed to be suitably described by the basic model

$$p(t) = L \exp\left[-b \exp\left[-kw(t)\right]\right] + e(t). \tag{4.7}$$

with three parameters L, b and k. $e(t)$ is the discrepancy between $p(t)$ and the power value prescribed by the model, and it is assumed to be an independent and identically distributed random variable. The curve described by the first term on the right hand side of (4.7), i.e. $p^*(t) = L \exp\left[-b \exp\left[-kw(t)\right]\right]$, is called the *Gompertz curve.*

This model corresponds to a power-wind relation, where no wind direction dependence is present. A reasonable way of introducing a wind direction dependence is by assuming that each of the three parameters actually is a smooth function of the wind direction as indicated by the following model formulation

$$p(t) = L(d(t)) \exp\left[-b(d(t)) \exp\left[-k(d(t))w(t)\right]\right] + e(t), \qquad d(t) \in [0, 360], \tag{4.8}$$

where $d(t)$ is the direction of the wind. The direction dependence is modelled by trigonometric functions. Hereby it is insured that the parameter functions are continuous at all wind directions.

$$\begin{aligned} L(d) \;= L_0 \;& + L_1 \sin\left(\frac{2\pi d}{360}\right) + L_2 \cos\left(\frac{2\pi d}{360}\right) + L_3 \sin\left(\frac{4\pi d}{360}\right) + L_4 \cos\left(\frac{4\pi d}{360}\right) \\ & + L_5 \sin\left(\frac{6\pi d}{360}\right) + L_6 \cos\left(\frac{6\pi d}{360}\right) + L_7 \sin\left(\frac{8\pi d}{360}\right) + L_8 \cos\left(\frac{8\pi d}{360}\right) \\ & + L_9 \sin\left(\frac{10\pi d}{360}\right) + L_{10} \cos\left(\frac{10\pi d}{360}\right), \end{aligned} \tag{4.9}$$

$$b(d) \;= b_0 \;+ b_1 \sin\left(\frac{2\pi d}{360}\right) + b_2 \cos\left(\frac{2\pi d}{360}\right) + b_3 \sin\left(\frac{4\pi d}{360}\right) + b_4 \cos\left(\frac{4\pi d}{360}\right)$$

$$+b_5 \sin\left(\frac{6\pi d}{360}\right) + b_6 \cos\left(\frac{6\pi d}{360}\right) + b_7 \sin\left(\frac{8\pi d}{360}\right) + b_8 \cos\left(\frac{8\pi d}{360}\right)$$
$$+b_9 \sin\left(\frac{10\pi d}{360}\right) + b_{10} \cos\left(\frac{10\pi d}{360}\right) \tag{4.10}$$

$$\begin{aligned} k(d) = k_0 \; & +k_1 \sin\left(\frac{2\pi d}{360}\right) + k_2 \cos\left(\frac{2\pi d}{360}\right) + k_3 \sin\left(\frac{4\pi d}{360}\right) + k_4 \cos\left(\frac{4\pi d}{360}\right) \\ & +k_5 \sin\left(\frac{6\pi d}{360}\right) + k_6 \cos\left(\frac{6\pi d}{360}\right) + k_7 \sin\left(\frac{8\pi d}{360}\right) + k_8 \cos\left(\frac{8\pi d}{360}\right) \\ & +k_9 \sin\left(\frac{10\pi d}{360}\right) + k_{10} \cos\left(\frac{10\pi d}{360}\right) \end{aligned} \tag{4.11}$$

Table 4.1 shows the results for the model estimates with and without trigonometric wind direction dependence. Clear reductions in BIC are seen for both wind farms, when the direction dependence is modelled, and only few parameter estimates are not significant. Models with more terms have been tried out, but no serious improvements were found. Furthermore, the variance reductions are 24.5% for Mors and 42.6% for Ærø. The variations with wind direction of the power curve estimates are illustrated on Figure 4.1.

The estimation results in this section have demonstrated a clear dependence on the wind direction in the relation between wind speed and power at two wind farms. Moreover, it has also been shown that the degree of direction dependence is different for the two wind farms considered here. This is not unreasonable since the degree of direction dependence must depend on the number of wind mills, the configuration of the wind farm and landscape surrounding the wind farm. Accordingly, there may be considerable differences in the direction dependence for different wind farms. However, it is most likely that all wind farms have some direction dependence.

4.3 Models for the Relation between Power and Wind Speed without Direction Dependence

The models describing the relation between the power in a wind farm and the measured wind speed are in this section assumed to be of the following form

$$p(t) = g\left(w(t)\right) + v(t), \tag{4.12}$$

where the error $v(t)$ is modelled by

$$\begin{aligned} A(q^{-1})v(t) &= C(q^{-1})e(t) \Leftrightarrow \\ (1 - a_1 q^{-1} - \cdots a_{n_a} q^{-n_a})v(t) &= (1 + c_1 q^{-1} + \cdots c_{n_c} q^{-n_c})e(t) \Leftrightarrow \\ v(t) - a_1 v(t-1) - \cdots - a_{n_a} v(t-n_a) &= e(t) + c_1 e(t-1) + \cdots + c_{n_c} e(t-n_c) \end{aligned} \tag{4.13}$$

$g\left(w(t)\right)$ represents a steady state relation between the wind speed and the power for the wind farm. $v(t)$ represents the difference between the steady state relation and the observed power at time t (noise), and $C(q^{-1})/A(q^{-1})$ describes the correlation structure of the noise.

Statistic	Mors		Ærø	
	Constant	Trigonometric	Constant	Trigonometric
$\hat{L}_0$	814.0	893.6	448.2	350.7
$\hat{L}_1$		-240.2		-64.7
$\hat{L}_2$		287.3		-100.9
$\hat{L}_3$		-206.2		-2.0
$\hat{L}_4$		396.7		-49.8
$\hat{L}_5$		-155.7		7.5
$\hat{L}_6$		309.4		-62.6
$\hat{L}_7$		-90.6		9.8
$\hat{L}_8$		205.3		-36.2
$\hat{L}_9$		-4.7$\star$		12.4
$\hat{L}_{10}$		82.1		-44.9
$\hat{b}_0$	6.70	-0.2$\star$	15.8	47.8
$\hat{b}_1$		16.1		6.9
$\hat{b}_2$		-17.5		35.5
$\hat{b}_3$		22.7		-3.1
$\hat{b}_4$		-19.4		28.7
$\hat{b}_5$		22.9		2.0
$\hat{b}_6$		-16.0		22.7
$\hat{b}_7$		17.3		-3.3
$\hat{b}_8$		-12.0		23.3
$\hat{b}_9$		5.6		3.3
$\hat{b}_{10}$		-6.9		28.1
$\hat{k}_0$	0.277	0.01$\star$	0.352	0.518
$\hat{k}_1$		0.36		0.042
$\hat{k}_2$		-0.65		0.196
$\hat{k}_3$		0.39		-0.017
$\hat{k}_4$		-0.73		0.135
$\hat{k}_5$		0.28		-0.0003$\star$
$\hat{k}_6$		-0.58		0.122
$\hat{k}_7$		0.16		-0.015
$\hat{k}_8$		-0.34		0.123
$\hat{k}_9$		0.02		0.008
$\hat{k}_{10}$		-0.13		0.143
BIC	647760	625302	920334	852243
σ_e^2 [kW^2]	54.25^2	47.14^2	41.70^2	31.60^2

Table 4.1: Estimation results. Estimates not significant on 5% -level of t-test are indicated by a $\star$.

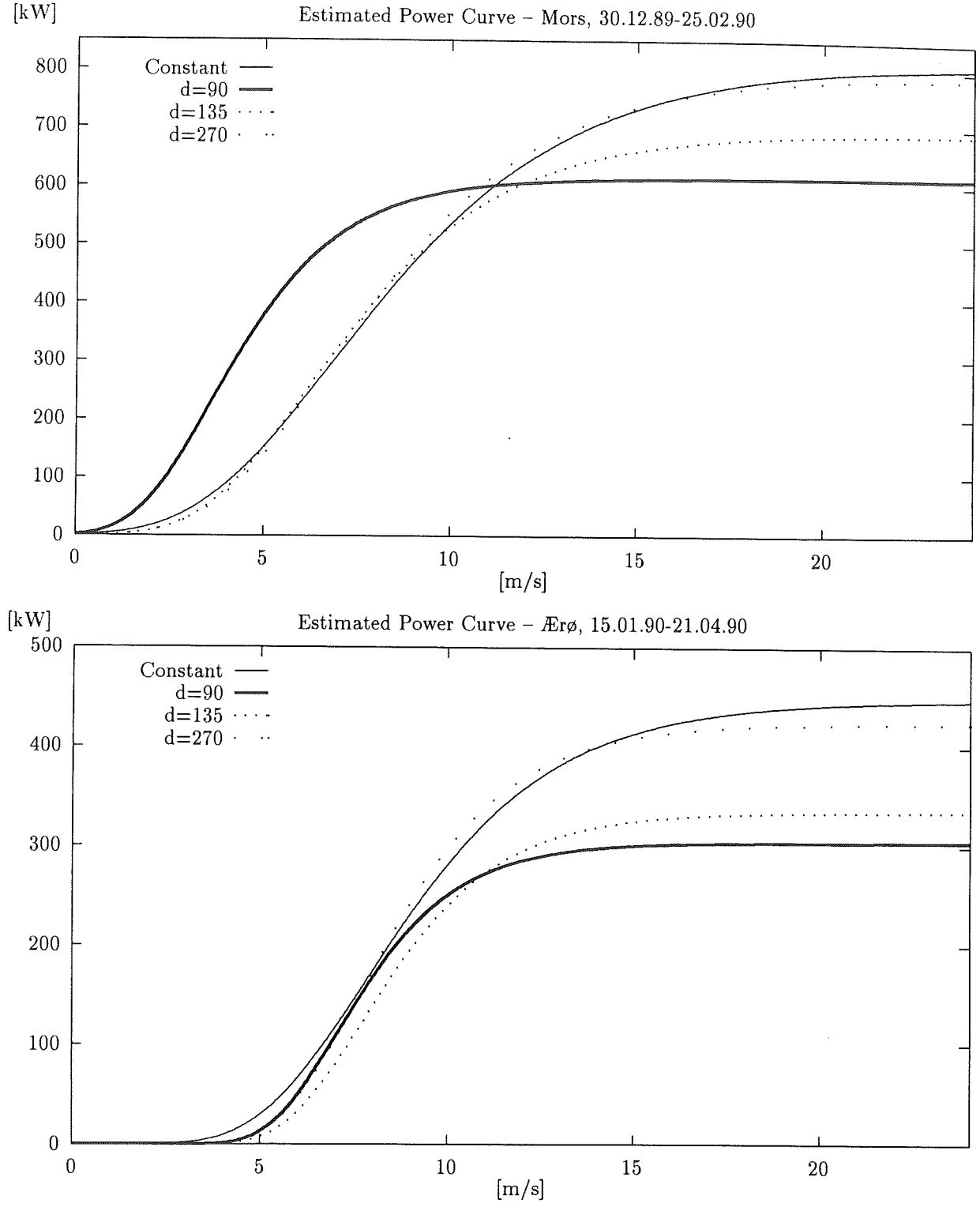


Figure 4.1: Power curve estimates for different values of the wind direction.

As a consequence of not having access to the wind direction the differences between the power curve and the actually observed power will turn out to be correlated to the same extent as the wind direction is correlated. Consequently the lacking observations of wind direction is partly compensated by modelling the correlation of the discrepancies between a power curve, applying for all wind directions, and the actual power values.

Figure 4.2 shows the data used in the present investigation. *The data consists of half-hour averages of both wind speed and power* from the wind farm at Vedersø Kær. The data until September 6 at 7.00 a.m. is used for estimation, and the rest of the data is used for validation, i.e., the estimated models are tested on this part of the data. The separation of the data into an estimation set and a validation set is shown by the dotted line.

The wind farm at Vedersø Kær has been chosen for this first part of the investigation as it seems to be the wind farm with least contamination of errors in data.

The wind farm at Vedersø Kær consists of 27 wind mills with a total power capacity of 6075 kW. All of the wind mills are of the same brand (Vestas V27-225 kW), and they are given to reach maximum power for wind speeds above 12 m/s.

Table 4.2 shows the results for the estimation of various choices of $g(w(t))$, $A(q^{-1})$ and $C(q^{-1})$. The least squares estimation method outlined in Section 4.1 is used. It is clear from this table that the modelling of the correlation structure of the noise certainly pays off. The only reasonable explanation to this result must be that there is a clear dependence on the wind direction at the wind farm. The further conclusion from this table is that when the noise correlation is modelled there is almost no difference between the results for the three different power curve models. The logistic power curve (at the bottom) is slightly superior to the other models.

4.3.1 Comparison of Models with and without Wind Direction Dependence

An imperative question is now whether modelling of the correlation of the power curve error is capable of replacing explicit modelling of the wind direction dependence. Therefore the set of data from Mors and Ærø, described in Section 4.2, is used for comparing modelling results both with and without explicit modelling of direction dependence and with and without modelling of error correlation. That is, in (4.12) $g(w(t))$ is either a simple Gompertz parameterization (4.7) or it is the trigonometric version given by (4.8)-(4.11), and $v(t)$ is either not modelled or it is modelled by the ARMA(1,1) model

$$v(t) - a_1 v(t-1) = e(t) + c_1 e(t-1). \tag{4.14}$$

This choice of model structure for $v(t)$ is based on the previous results shown in Table 4.2, which indicates that it is sufficient. The model estimation is carried out on half hour averages of the measurements from Mors and Ærø (contrary to the estimation in Section 4.2).

Table 4.3 compares the prediction error variances for the four outlined modelling possibilities and the two wind farms, respectively. From this table it is seen that both the explicit modelling of the direction dependence and the modelling of the error correlation significantly

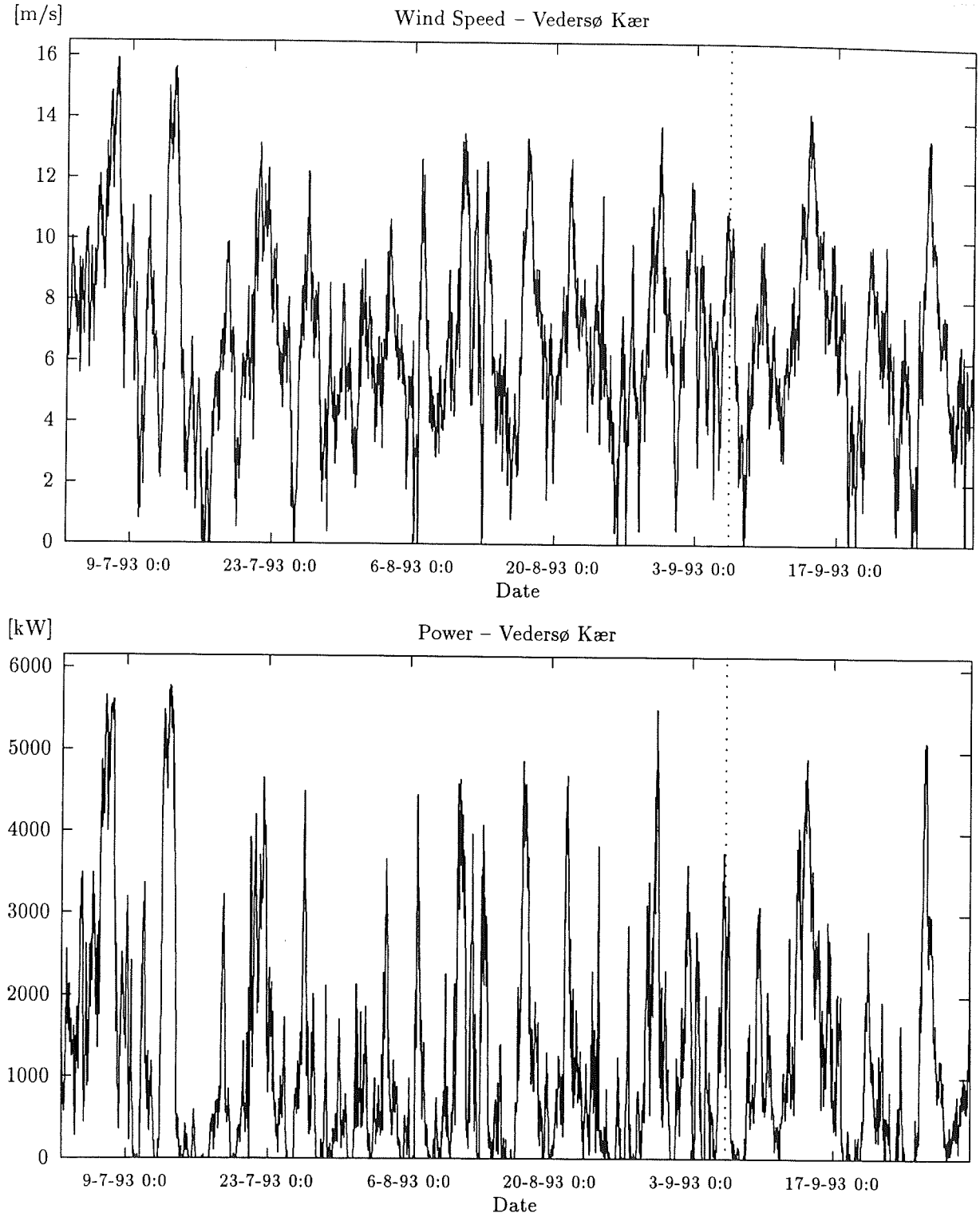


Figure 4.2: Data used for estimation and validation.

$g(w(t))$	$A(q^{-1})$	$C(q^{-1})$	$\hat{\sigma}_e$ [kW]	BIC	$\tilde{\sigma}_e$ [kW]
$\overset{(152)}{8061}\exp\left[-\overset{(.2)}{10.8}\exp\left[-\overset{(.004)}{0.227}\,w(t)\right]\right]$	-	-	299.7	35929	344.1
$\overset{(223)}{5723}\exp\left[-\overset{(.6)}{10.5}\exp\left[-\overset{(.009)}{0.247}\,w(t)\right]\right]$	$1-\overset{(.01)}{0.89}\,q^{-1}$	-	181.7	32787	158.0
$\overset{(206)}{5986}\exp\left[-\overset{(.6)}{11.0}\exp\left[-\overset{(.008)}{0.250}\,w(t)\right]\right]$	$1-\overset{(.008)}{0.93}\,q^{-1}$	$1-\overset{(.02)}{0.28}\,q^{-1}$	176.8	32622	156.6
$\overset{(206)}{5978}\exp\left[-\overset{(.6)}{11.0}\exp\left[-\overset{(.008)}{0.250}\,w(t)\right]\right]$	$1-\overset{(.06)}{0.97}\,q^{-1}+\overset{(.05)}{0.03}\,q^{-2}$	$1-\overset{(.06)}{0.31}\,q^{-1}$	176.8	32632	156.5
$\overset{(208)}{5982}\exp\left[-\overset{(.6)}{11.0}\exp\left[-\overset{(.008)}{0.249}\,w(t)\right]\right]$	$1+\overset{(.01)}{0.05}\,q^{-1}-\overset{(.01)}{0.92}\,q^{-2}$	$1+\overset{(.02)}{0.71}\,q^{-1}-\overset{(.02)}{0.28}\,q^{-2}$	176.7	32634	157.5
$\overset{(57)}{308}-\overset{(19)}{286}\,w(t)+\overset{(3)}{71}\,w(t)^2-\overset{(.1)}{2.1}\,w(t)^3$	$1-\overset{(.008)}{0.93}\,q^{-1}$	$1-\overset{(.02)}{0.25}\,q^{-1}$	182.2	32818	161.3
$\overset{(50)}{292}-\overset{(19)}{284}\,w(t)+\overset{(3)}{71}\,w(t)^2-\overset{(.1)}{2.1}\,w(t)^3$	$1-\overset{(.03)}{1.85}\,q^{-1}+\overset{(.02)}{0.86}\,q^{-2}$	$1-\overset{(.03)}{1.20}\,q^{-1}+\overset{(.02)}{0.29}\,q^{-2}$	181.7	32818	162.0
$-\overset{(.03)}{0.1}-\overset{(.5)}{3.8}\,w(t)-\overset{(.2)}{15.5}\,w(t)^2+\overset{(.3)}{7.2}\,w(t)^3-\overset{(.01)}{0.32}\,w(t)^4$	$1-\overset{(.008)}{0.93}\,q^{-1}$	$1-\overset{(.02)}{0.28}\,q^{-1}$	176.9	32641	156.7
$\overset{(50)}{61}+\overset{(29)}{39}\,w(t)-\overset{(8)}{30.3}\,w(t)^2+\overset{(.8)}{8.8}\,w(t)^3-\overset{(.03)}{0.37}\,w(t)^4$	$1-\overset{(.02)}{1.85}\,q^{-1}+\overset{(.02)}{0.87}\,q^{-2}$	$1-\overset{(.03)}{1.23}\,q^{-1}+\overset{(.03)}{0.32}\,q^{-2}$	176.0	32627	156.4
$\overset{(108)}{4835}\Big/\left(1+\overset{(11)}{121}\exp\left[-\overset{(.01)}{0.49}\,w(t)\right]\right)$	$1-\overset{(.008)}{0.93}\,q^{-1}$	$1-\overset{(.02)}{0.27}\,q^{-1}$	176.7	32618	155.2
$\overset{(108)}{4836}\Big/\left(1+\overset{(11)}{122}\exp\left[-\overset{(.01)}{0.49}\,w(t)\right]\right)$	$1-\overset{(.13)}{0.12}\,q^{-1}-\overset{(.13)}{0.76}\,q^{-2}$	$1+\overset{(.14)}{0.54}\,q^{-1}-\overset{(.04)}{0.21}\,q^{-2}$	176.7	32627	155.3

Table 4.2: Modelling results for the relation between power and wind speed.

Modelling of $v(t)$	Mors		Ærø	
	$g(w(t))$		$g(w(t))$	
	Constant	Trigonometric	Constant	Trigonometric
-	43.76^2	30.36^2	32.22^2	24.08^2
ARMA(1,1)	26.52^2	23.24^2	21.57^2	18.37^2

Table 4.3: Prediction error variance for model of the relation between power and wind speed.

improves the description of the relation between wind speed and power. Modelling of the error correlation is obviously the most beneficial to the model fit. The table also shows that even when the error correlation is modelled, there is still something to gain by modelling the wind dependence by the trigonometric parameter profiles. This means that even though modelling of error correlation is beneficial, and possibly compensates to some extent for lacking wind direction values, the wind direction would still be beneficial, and consequently it is a drawback for the modelling of the relation between wind and power, when wind direction values are not available.

4.4 Models for Wind Speed

This section describes the formulation and off-line estimation of both linear and non-linear models for wind speed. In the first place the simple linear models, namely the ARMA models, are considered. These models are of the structure

$$\begin{aligned} A(q^{-1})w(t) &= C(q^{-1})e(t) \Leftrightarrow \\ (1 - a_1 q^{-1} - \cdots a_{n_a} q^{-n_a})w(t) &= (1 + c_1 q^{-1} + \cdots c_{n_c} q^{-n_c})e(t) \Leftrightarrow \\ v(t) - a_1 w(t-1) - \cdots - a_{n_a} w(t-n_a) &= e(t) + c_1 e(t-1) + \cdots + c_{n_c} e(t-n_c), \end{aligned} \quad (4.15)$$

where $\{e(t)\}$ is assumed to white noise.

Furthermore, nonlinear extensions to the ARMA class of models are considered. An overview of various nonlinear model structures can, for instance, be found in Tong (1990), Priestley (1988) and in Madsen & Holst (1995). The nonlinear models considered in this investigation seem to be best choices for describing the variations of wind speed. The nonlinear models are the bilinear models of the following general structure

$$w(t) + \sum_{i=1}^{p} a_i w(t-i) = \alpha + \sum_{j=1}^{r} \sum_{k=1}^{s} b_{jk} w(t-j)e(t-k) + e(t), \quad (4.16)$$

and the smooth threshold autoregressive models

$$w(t) = a_0 + \sum_{i=1}^{p} \left(h_i\left(y(t)\right) w(t-i)\right) + e(t), \quad (4.17)$$

where

$$h_i(y) = a_i + \frac{b_i}{1 + \exp\left[c_i - d_i y\right]}. \tag{4.18}$$

$\{y(t)\}$ is a process determining the actual values of the autoregressive coefficients through the smooth threshold h_i. In this case this process is chosen as the first order filtered wind speed

$$y(t) = w_f(t) = f w_f(t-1) + (1-f)w(t), \tag{4.19}$$

of which the parameter f is estimated in the same step as the other parameters of the model.

Table 4.4 shows the results for a selection of the estimated models belonging to the classes mentioned above. The figures in parenthesis above the estimated parameters are the estimated standard deviations of the parameter estimates.

The immediate conclusion for the results in Table 4.4 is that only very little can be gained in performance when increasing the model complexity beyond the first order AR model (seen as the uppermost model in the list). The bilinear models as well as the smooth threshold models do give a slightly better fit in the estimation, but when validating the estimated models on a different set of data only the bilinear models are better than the AR(1) model.

Actually these results are rather disappointing as it seems to be obvious that the variations of the wind speed cannot be sufficiently well described by models within the ARMA class. However, it must be concluded that the nonlinear modelling approaches tried in this section only result in a very slight improvement for the considered data.

4.5 Power Prediction using Recursive Estimation

This section describes some models and methods for power prediction which are well suited for on-line purposes. The models are estimated using adaptive recursive estimation. The adaptivity gives a possibility to account for (slow) variations in time of the parameters. Hence, such methods can be used for taking into account the annual variation.

More specifically the following approaches are considered:

1. **An overall model for wind speed followed by a transformation to power**: The wind speed is described by a model belonging to the class of ARMA models. This model is estimated recursively using the one-step prediction errors and in an adaptive way. At each time step the model is applied iteratively so a sequence of wind speed predictions, reaching from one to 42 steps ahead, is obtained. The power predictions are obtained by transferring the sequence of wind speed predictions into a sequence of power predictions by using a fixed dynamic model for the relation between wind speed and power. This approach is described in Section 4.5.2.
2. **k-step prediction models for wind speed followed by a transformation to power**: Another approach, which is described in Section 4.5.3, is to estimate separate prediction models for the wind speed for each prediction horizon. The estimation is done adaptively and recursively. Again the power predictions are obtained using the fixed dynamic model for the relation between wind speed and power.

Model	$\hat{\sigma}_e$ [kW]	BIC	$\tilde{\sigma}_e$ [kW]
$w(t) = \overset{(.004)}{0.97}\, w(t-1) + e(t)$	0.7080	-2166	0.6654
$w(t) = \overset{(.02)}{0.98}\, w(t-1) - \overset{(.02)}{0.003}\, w(t-2) + e(t)$	0.7081	-2157	0.6654
$w(t) = \overset{(.004)}{0.97}\, w(t-1) + e(t) + \overset{(.02)}{0.004}\, e(t-1)$	0.7081	-2157	0.6654
$w(t) = \overset{(.004)}{0.97}\, w(t-1) + e(t) + \overset{(.02)}{0.009}\, e(t-1) - \overset{(.006)}{0.02}\, w(t-1)e(t-1)$	0.7068	-2161	0.6638
$w(t) = \overset{(.004)}{0.97}\, w(t-1) + e(t) + \overset{(.01)}{0.003}\, w(t-1)e(t-1) - \overset{(.01)}{0.03}\, w(t-2)e(t-1) + \overset{(.005)}{0.02}\, w(t-2)e(t-2)$	0.7050	-2169	0.6628
$w(t) = \overset{(.004)}{0.97}\, w(t-1) + e(t) - \overset{(.006)}{0.03}\, w(t-2)e(t-1) + \overset{(.005)}{0.02}\, w(t-2)e(t-2)$	0.7049	-2177	0.6629
$w(t) = \overset{(.004)}{0.97}\, w(t-1) + e(t) - \overset{(.005)}{0.02}\, w(t-2)e(t-1) + \overset{(.005)}{0.02}\, w(t-2)e(t-2) - \overset{(.02)}{0.03}\, e(t-2)$	0.7048	-2171	0.6625
$w(t) = \overset{(.004)}{0.97}\, w(t-1) + e(t) - \overset{(.006)}{0.02}\, w(t-2)e(t-1) + \overset{(.005)}{0.02}\, w(t-2)e(t-2) - \overset{(.02)}{0.01}\, e(t-1) - \overset{(.02)}{0.03}\, e(t-2)$	0.7049	-2162	0.6624
$w(t) = \overset{(.09)}{1.53}\, w(t-1) - \overset{(.08)}{0.54}\, w(t-2) + e(t) - \overset{(.006)}{0.02}\, w(t-2)e(t-1) + \overset{(.005)}{0.03}\, w(t-2)e(t-2) - \overset{(.08)}{0.58}\, e(t-1)$	0.7035	-2174	0.6628
$w(t) = \left(\overset{(.004)}{0.98} - \frac{\overset{(.01)}{0.05}}{1+\exp\left[\overset{(20000)}{12593} - \overset{(5000)}{2771}\, w_f(t-1)\right]} \right) w(t-1) + e(t)$ $w_f(t) = b w_f(t-1) + (1-b) w(t), \qquad \hat{b} = \overset{(.005)}{0.44}$	0.7065	-2147	0.6658
$w(t) = \left(\overset{(.004)}{0.97} + \overset{(.001)}{0.002}\, w_f(t-1) \right) w(t-1) + e(t)$ $w_f(t) = b w_f(t-1) + (1-b) w(t), \qquad \hat{b} = \overset{(.4)}{0.5}$	0.7076	-2154	0.6651
$w(t) = a(t) w(t-1) - \overset{(.006)}{0.02}\, w(t-2) + e(t) + \overset{(.005)}{0.03}\, w(t-2)e(t-1) - \overset{(.08)}{0.53}\, w(t-2)e(t-2) - \overset{(.07)}{0.50}\, e(t-1)$ $a(t) = \left(\overset{(.2)}{0.93} - \frac{\overset{(.2)}{0.6}}{1+\exp\left[\overset{(2)}{2.6} - \overset{(18)}{23}\, w_f(t-1)\right]} \right)$ $w_f(t) = b w_f(t-1) + (1-b) w(t), \qquad \hat{b} = -\overset{(.02)}{0.84}$	0.6984	-2188	0.7070

Table 4.4: Modelling results for the wind speed.

3. **k-step prediction models for power**: A direct approach, where the power is described by a linear model, which may or may not include a description of the relation to the wind speed, is described in Section 4.5.4. The model is specific for the prediction horizon, meaning that a number of models to be used for different prediction horizons are estimated and evaluated. The models are estimated recursively and adaptively.

The adaptive recursive estimation technique is described in Section 4.5.1.

4.5.1 Adaptive Recursive Least Squares Estimation

In those cases, where the estimation is carried out recursively and adaptively, the recursive least squares estimation method with exponential forgetting (Ljung, 1987) is used. This method requires that the model can be formulated linearly in the parameters

$$y(t) = \varphi^{\mathrm{T}}(t)\theta + e(t), \tag{4.20}$$

where $\varphi(t)$ is a vector of regressors, and θ is the parameters of the model.

A Model for Predicting the Wind Power

Let us illustrate how a model, which in a subsequent section is suggested as a very reasonable overall model for predicting the wind power (see (4.38), is written in the linear form given in (4.20).

The model is an autoregressive model of the wind power production ($p(t)$), with wind speed ($w(t)$) as an input variable. Furthermore, the model contains a deterministic harmonic with period 24 h. The model is:

$$\begin{aligned} \sqrt{p(t+k)} &= \theta_0 + \theta_1\sqrt{p(t)} + \theta_2\sqrt{w(t)} + \theta_3 w(t) \\ &+ \theta_4 \sin(2\pi h(t+k)/24) + \theta_5 \cos(2\pi h(t+k)/24) + e(t+k), \end{aligned} \tag{4.21}$$

where $\theta_0, \ldots, \theta_5$ are the parameters for the horizon k. The running time index is t, k is the prediction horizon, $h(t)$ is the time of the 24 hour clock cycle at time t, and $e(t)$ is white noise with variance σ_e^2. Note, that although (4.21) is referreed to as the model, the equation actually corresponds to one model for each prediction horizon.

The k-step prediction model (4.21) can be written in the linear form

$$y(t) = \varphi_k^T(t)\theta + e(t), \tag{4.22}$$

where $y(t) = \sqrt{p(t)}$, and

$$\varphi_k(t) = \begin{bmatrix} 1 \\ \sqrt{p(t-k)} \\ \sqrt{w(t-k)} \\ w(t-k) \\ \sin(2\pi h(t)/24) \\ \cos(2\pi h(t)/24) \end{bmatrix}, \tag{4.23}$$

Using this formulation it is clear that the model is linear in the parameters, which are given by

$$\theta = [\theta_0, \ldots, \theta_5]^T \tag{4.24}$$

The Recursive Least Squares Method

Conventionally the 1-step prediction errors are used for estimation using the recursive least squares method, see e.g. (Ljung, 1987). Later on in this section we discuss how the method can be extended to k-step prediction methods.

Using the method of adaptive recursive least squares with constant forgetting factor a weighted sum of the squared prediction errors up to the present time is minimized. The adaptivity is obtained by assigning the largest weight to the most resent observations. In mathematical terms the estimates at time t ($\hat{\theta}(t)$) are chosen to minimize

$$V_t(\hat{\theta}(t)) = \sum_{s=1}^{t} \lambda^{t-s}(y(s) - \varphi_1^T(s)\hat{\theta}(t))^2, \tag{4.25}$$

where λ is the forgetting factor and $\varphi_1^T(s)\hat{\theta}(t)$ is the prediction using the parameter estimates at time t. A comparison with the least squares estimation method considered previously in Eq. 4.1 shows that, apart from some constants without any significance, the difference is the exponential weighting of older observations by the term λ^{t-s}.

The choice of the forgetting factor λ is determined by a trade-off between the needed ability to track time-varying parameters and the noise sensitivity of the estimate. A low value of λ results in a system with a good ability to track time-varying parameters but a higher sensitivity against noise or errors in the data. A typical choice of λ is in the range $0.95 \leq \lambda \leq 0.999$. The number of effective observations (or the number of observations kept in "memory") is given as

$$N_{eff} = \frac{1}{1-\lambda} \tag{4.26}$$

For $\lambda = 1$ all previous observations contributes to each set of estimates with equal weight, i.e. the final estimates correspond to the estimates that would have been obtained using a non-recursive least squares estimation procedure.

The adaptive recursive least squares algorithm is given by the following steps at time t

1. Calculation of prediction error using old estimate at time $t-1$:

$$\tilde{y}(t|t-1) = y(t) - \varphi_1^T(t)\hat{\theta}(t-1) \tag{4.27}$$

2. An update of the covariance matrix of the parameter estimates is obtained using:

$$P(t) = \frac{1}{\lambda}\left(P(t-1) - \frac{P(t-1)\varphi_1(t)\varphi_1^T(t)P(t-1)}{\lambda + \varphi_1^T(t)P(t-1)\varphi_1(t)}\right), \tag{4.28}$$

The matrix $P(t)$ constitutes, except from the factor σ_e^2, an estimate of the covariance matrix of the parameter estimates at time t.

3. Update of the parameter estimates:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\varphi_1(t)\tilde{y}(t|t-1) \tag{4.29}$$

The algorithm has to be initialized. The initial estimates may be chosen quite arbitrarily – often zero is used. The initial covariance matrix has to be chosen such that the variance of the initial estimates is large – often selected as a diagonal matrix with all elements on the diagonal set to 100 or 1000.

The 1-step prediction of $y(t+1)$ at time t is calculated as

$$\hat{y}(t+1|t) = \varphi_1^T(t+1)\hat{\theta}(t). \tag{4.30}$$

In the application of the recursive least squares method for estimating the parameters of the model (4.21) the predictions may be negative although they are predictions of the square-root of the wind power production. In this case the predictions are set to zero. Note however that the predictions used in the recursive estimation should be allowed to be negative. Otherwise the estimation procedure may get unstable.

The Recursive Least Squares Method and k-step Predictions

If a prediction horizon larger than one is used for estimation purposes a choice between two alternative ways of updating the estimates must be made.

- The estimates $\hat{\theta}(t-k)$ and the regressors $\varphi_k(t)$ are used instead of $\hat{\theta}(t-1)$ and $\varphi_1(t)$ in the algorithm described above, or
- pseudo prediction errors are used in the update of estimates. The pseudo prediction error at time t is calculated as

  $$\tilde{y}_{pseudo}(t|t-k) = y(t) - \varphi_k^T(t)\hat{\theta}(t-1), \tag{4.31}$$

  from this equation it is seen that the pseudo prediction error corresponds to variables known at time $t-k$ (i.e. $\varphi_k^T(t)$) and the most recent estimates (i.e. $\hat{\theta}(t-1)$).

In both cases the true k-step prediction is calculated as

$$\hat{y}(t+k|t) = \varphi_k^T(t+k)\hat{\theta}(t). \tag{4.32}$$

Using the true k-step prediction error in the update of the most recent estimates will result in highly inappropriate estimates. This is due to the fact that the prediction error will give a feed-back not corresponding to the estimates that are to be updated.

In the software, called the Wind Power Prediction Tool, which is described in Chapter 7, the algorithm using pseudo prediction errors is implemented.

4.5.2 Wind Speed Prediction Followed by Transformation to Power

In this section the model for wind speed is estimated by the adaptive recursive least squares method and by considering a minization of the one-step predictions. Using the model, predictions of wind speed reaching from one to e.g. 42 steps (21 hours) ahead can be obtained.

With the purpose of finding the optimal model for predicting the wind speed a large number of models have been investigated. The models tried belong to the following model class

$$\begin{aligned} w(t) \;=\; & a_1 w(t-1) + \cdots + a_{n_w} w(t-n_w) + \\ & b_1 w(t-48) + \cdots + b_{n_{w_{48}}} p(t-n_{w_{48}}-47) + \\ & c_1 e(t-1) + \cdots + c_{n_e} e(t-n_e) + \\ & f_1 e(t-48) + \cdots + f_{n_{e_{48}}} e(t-n_{e_{48}}-47) + \\ & \sum_{i=1}^{n_F} (h_{2i-1} \sin(2\pi t i/48) + h_{2i} \cos(2\pi t i/48)) + \\ & l + e(t) \end{aligned} \tag{4.33}$$

These models contain the possibility of describing an autoregressive variation, a random diurnal variation, a moving average variation and a deterministic diurnal variation. The level-parameter l is always contained in the model meaning that if all the other model orders are zero the model consists of just a level ($w(t) = l + e(t)$). Each polynomial is only contained in the model if its corresponding order is greater than zero.

The sequence of predictions $\hat{w}(t+k|t), k \in [1;42]$ obtained at each time step t by iterative application of the model (4.33) is transferred into power prediction by using the model

$$p(t) = 5986 \exp\left[-11.0 \exp\left[-0.250 w(t)\right]\right] + \frac{1 - 0.28q^{-1}}{1 - 0.93q^{-1}} e(t). \tag{4.34}$$

This model has previously been estimated (and found suitable) on a set of data from Vedersø Kær. The estimation and evaluation is in this section carried out on a set of data, which includes the set used previously, but is prolonged until October 11 at 7.00 a.m. Table 4.5 shows the obtained prediction performance for the model having the optimal performance at each prediction horizon among a large number of model configurations. The table indicates the values of the model orders which have been tried for each of the five model orders. All combinations have been tried, meaning that the performance has been evaluated for 96 different wind speed models.

Also given in the table is the evaluated standard deviation of the prediction error (S.E.), and the standard error (in percentage) scaled by the power capacity of the wind farm ((S.E.$_{\%}$). The table also contains the standard deviation of the prediction errors obtained by the *naive predictor*

$$\hat{p}(t+k|t) = p(t). \tag{4.35}$$

The forgetting factor, which turned out to give the lowest average squared prediction error, is also shown in the table. A number of different forgetting factor were tried, and it turned

k	λ	n_w (0-3)	$n_{w_{48}}$ (0-2)	n_e (0-1)	$n_{e_{48}}$ (0-1)	n_F (0-1)	S.E. [kW]	S.E.%	Naive S.E. [kW]
1	0.9998	3	2	1	1	1	285.9	4.71	285.5
2	0.9998	2	2	1	0	1	406.6	6.69	412.1
4	0.9998	2	2	1	0	1	562.9	9.27	580.3
6	0.9998	3	1	1	1	1	680.6	11.20	714.0
12	0.9998	3	1	1	1	1	918.3	15.12	1009.7
24	0.9998	2	1	1	0	1	1168.6	19.24	1364.8
42	0.9998	0	0	1	1	1	1303.6	21.46	1580.4

Table 4.5: Prediction performance for Vedersø Kær (early period) - Prediction of wind speed using a one-step prediction model followed by transformation to power. The standard deviation of the power is 1335.7 kW.

out that the optimal results in all cases were obtained for the largest of the tried forgetting factors $\lambda = 0.9998$. This means that it is not optimal, for this data set, to have any adaptivity of importance in the estimation. One likely explanation is that the data set covers only a minor part of the annual cycle. If, as expected, the dynamic of the wind variations were different over the seasons it might be beneficial to have the adaptivity in the estimation. That is, a longer period of data would be necessary to confirm this hypothesis.

4.5.3 k-step Prediction Models for Wind Speed Followed by Transformation to Power

Instead of estimating just one model for the wind speed, which then is used for prediction of the complete sequence of wind speed variation 42 steps ahead, the wind speed at a specific prediction horizon k can be predicted by a k-step predictor by constructing a dedicated k-step prediction model. This means that a number of models are to be estimated and used for prediction in parallel to obtain a complete sequence of wind speed predictions.

The advantage could be that the parameters and even the model orders in (4.33), which are optimal for prediction a specific number of steps k ahead, are not the same for different values of k. If this is the case the best prediction result would be obtained if the model orders and parameters are determined and estimated in models which are used only for specific prediction horizons.

Although it is possible to write down explicitly how the k-step predictor is determined from the model (4.33), and hereby determine the parameters in (4.33) having the optimal performance for k-step prediction, it is not preferable to do this. The reason is that the

k-step predictor is not linear in the parameters of the original model, and due to this the use of recursive estimation methods would involve the introduction of approximations. The consequence is that it can not be assured that the recursive estimation will work well.

Instead the k-step predictor can be reformulated in such a way that it is linear in the parameters as given below

$$\begin{aligned} w(t+k) &= a_1 w(t) + \cdots + a_{n_w} w(t+1-n_w) + \\ & b_1 w(t+k-48) + \cdots + b_{n_{w_{48}}} p(t+k-n_{w_{48}}-47) + \\ & c_1 e(t) + \cdots + c_{n_e} e(t+1-n_e) + \\ & f_1 e(t-48) + \cdots + f_{n_{e_{48}}} e(t-n_{e_{48}}-47) + \\ & \sum_{i=1}^{n_F} (h_{2i-1} \sin(2\pi t i/48) + h_{2i} \cos(2\pi t i/48)) + \\ & l + e(t). \end{aligned} \tag{4.36}$$

This model is directly applicable for the calculation of predictions k step ahead, and it can be estimated recursively and adaptively by use of the recursive least squares algorithm with exponential forgetting.

k	λ	n_w (0-3)	$n_{w_{48}}$ (0-2)	n_e (0-1)	$n_{e_{48}}$ (0-1)	n_F (0-1)	S.E. [kW]	S.E.%	Naive S.E. [kW]
1	0.9996	3	2	1	1	1	286.1	4.71	285.5
2	0.9996	1	1	0	1	1	407.4	6.71	412.1
4	0.9996	1	2	1	1	1	564.2	9.29	580.3
6	0.9996	3	1	1	1	1	683.1	11.24	714.0
12	0.9996	1	2	1	1	1	922.9	15.19	1009.7
24	0.9996	0	0	1	1	1	1177.7	19.39	1364.8
42	0.9996	0	0	1	0	1	1310.6	21.57	1580.4

Table 4.6: Prediction performance for Vedersø Kær (early period) - Prediction of wind speed using separate k-step predictors followed by transformation to power. The standard deviation of the power is 1335.7 kW.

Table 4.6 shows the prediction performance results when the k-step predictors are used and transferred into power predictions via model (4.34). As previously, each polynomial is only contained in the model if its corresponding order is greater than zero. In this case the largest forgetting factor which was tried was $\lambda = 0.9996$, showing that the degree of adaptivity is not of so much importance.

Comparing the two tables 4.5 and 4.6 shows that the prediction performance is very much the same. Actually the first approach seems to give a little better prediction result, but this

may be due to the difference in the largest (and optimal) forgetting factor which was tried. It is also interesting to see that the standard deviation of the prediction errors obtained by the naive predictor is not so much worse than for the optimal model. However, it is also seen that the standard deviation of the prediction error for $k = 42$ is close to the standard deviation of power variation for the same period. Consequently, using the prediction error standard deviation as a measure of prediction performance, one might as well predict the power by the average of the power.

4.5.4 Prediction by Using a Linear Model for Power

In this section the approach for obtaining predictions of the power is to use k-step prediction models, which have the option of using the relation to wind speed or leave it out. All of the models tried out belong to the following class of models

$$\begin{aligned} p(t+k) \;=\; & a_1 p(t) + \cdots + a_{n_p} p(t+1-n_p) + \\ & b_1 p(t+k-48) + \cdots + b_{n_{p_{48}}} p(t+k-n_{p_{48}}-47) + \\ & c_1 w(t) + \cdots + c_{n_w} w(t+1-n_w) + \\ & d_1 w^2(t) + \cdots + d_{n_{w^2}} w^2(t+1-n_{w^2}) + \\ & g_1 w(t+k-48) + \cdots + d_{n_{w_{48}}} w(t+k-47-n_{w_{48}}) + \\ & h_1 e(t) + \cdots + h_{n_e} e(t+1-n_e) + \\ & \sum_{i=1}^{n_F} (h_{2i-1} \sin(2\pi t i/48) + h_{2i} \cos(2\pi t i/48)) + \\ & l + e(t+k) \end{aligned} \tag{4.37}$$

It is seen that the correlation between wind and power is described by the terms involving n_w, n_{w^2} and $n_{w_{48}}$. If all these quantities are equal to zero the wind speed does not enter the model. The squared wind speed is used as a possible explanatory variable to allow for a modelling of a nonlinear relation between wind speed and power.

Table 4.7 shows the results when this model for different model orders and different k is applied on the same set of data from Vedersø Kær used previously. The first row of the table shows, in parenthesis, which model orders are tried. Furthermore, the table contains the standard deviation of the prediction error for the optimal model orders and optimal forgetting factor. Note that in this case the optimal forgetting factor has been found by use of a numerical minimization for all the models tried out.

Comparing with the previous two tables, it is clear that the prediction performance is improved by this approach. The increase in S.E. from the adaptive to the naive predictor is between 2% and 23%.

It is interesting to note that for $k \leq 12$ it is optimal to use observations of wind speed (also squared) when predicting the power, whereas for $k \geq 16$ the wind speed observations should not be used. For all k the optimal forgetting factor is equal to or close to 1 meaning that

k	λ	n_p (0-2)	$n_{p_{48}}$ (0-1)	n_w (0-2)	n_{w^2} (0-1)	$n_{w_{48}}$ (0-2)	n_e (0-1)	n_F (0-1)	S.E. [kW]	S.E.%	Naive S.E. [kW]
1	0.9988	1	0	2	1	0	0	1	279.8	4.61	285.5
2	0.9988	1	0	1	1	0	0	1	399.7	6.58	412.1
4	0.9987	1	0	1	1	0	0	1	550.8	9.07	580.3
6	0.9989	1	0	1	1	0	0	1	666.0	10.96	714.0
12	0.9990	1	0	1	1	0	0	1	897.8	14.78	1009.7
24	1.0000	1	0	0	0	0	0	1	1141.6	18.79	1364.8
42	0.9996	0	0	0	0	0	1	1	1284.1	21.14	1580.4

Table 4.7: Prediction performance for Vedersø Kær (early period). The standard deviation of the power is 1335.7 kW.

the adaptivity is almost not beneficial to the prediction performance. The conclusion about the forgetting factor will probably not hold for a longer set of data.

The same investigation is carried out on three newer set of data from Vedersø Kær, Dræby Fedsodde and Brøns. These data were sampled in the period from December 15, 1993 to January 24, 1994.

k	λ	n_p (0-2)	$n_{p_{48}}$ (0-1)	n_w (0-2)	n_{w^2} (0-1)	$n_{w_{48}}$ (0-2)	n_e (0-1)	n_F (0-1)	S.E. [kW]	S.E.%	Naive S.E. [kW]
2	0.9996	1	0	1	0	0	0	0	544.6	8.96	550.1
6	0.9990	1	0	1	0	0	0	0	919.8	15.14	944.2
24	0.9990	0	0	1	0	0	1	0	1653.5	27.22	1839.0
42	1.0000	0	0	0	0	0	0	0	1774.1	29.20	2185.4

Table 4.8: Prediction performance for Vedersø Kær. The standard deviation of the power is 1769.9 kW.

The results are shown in Tables 4.8-4.10 (and only for $k = 2, 6, 24, 42$). It is clear that the standard deviations for Vedersø Kær have increased considerably, which, however, must be traced back to the higher degree of variation in wind speed and power for the later data period. Also the optimal model orders are different for the two set of data, and furthermore, there are differences in model orders for the three wind farms.

It is interesting to note that, since $n_F = 0$, the deterministic diurnal component does not

k	λ	n_p (0-2)	$n_{p_{48}}$ (0-1)	n_w (0-2)	n_{w^2} (0-1)	$n_{w_{48}}$ (0-2)	n_e (0-1)	n_F (0-1)	S.E. [kW]	S.E.%	Naive S.E. [kW]
2	0.9999	1	0	0	0	1	0	0	227.8	8.63	229.3
6	0.9980	2	0	0	0	0	1	0	367.1	13.91	371.7
24	1.0000	0	0	0	0	0	1	0	703.2	26.64	756.3
42	1.0000	0	0	0	0	0	1	0	766.5	29.03	916.8

Table 4.9: Prediction performance for Dræby Fedsodde. The standard deviation of the power is 774.1 kW.

k	λ	n_p (0-2)	$n_{p_{48}}$ (0-1)	n_w (0-2)	n_{w^2} (0-1)	$n_{w_{48}}$ (0-2)	n_e (0-1)	n_F (0-1)	S.E. [kW]	S.E.%	Naive S.E. [kW]
2	0.9982	1	0	0	0	1	0	0	307.7	9.62	311.8
6	0.9984	1	0	1	0	1	0	0	479.2	14.98	496.6
24	0.9990	1	0	0	0	0	0	0	817.8	25.56	881.9
42	1.0000	0	0	0	0	0	1	0	918.8	28.71	1071.4

Table 4.10: Prediction performance for Brøns. The standard deviation of the power is 929.5 kW.

enter the optimal models for these newer data. The reason is most likely that the systematic diurnal variation of the wind speed is missing (or week) during winter time.

4.5.5 A Candidate for the Best Overall Model

Based on the investigations above it is concluded that the following model leads to the best overall performance

$$\sqrt{p(t+k)} = a_1\sqrt{p(t)} + c_1\sqrt{w(t)} + d_1 w(t) + h_1 \sin\frac{2\pi t}{48} + h_2 \cos\frac{2\pi t}{48} + l + e(t+k), \quad (4.38)$$

where $e(t)$ is assumed to be Gaussian independent and identically distributed random variables. The square root transformation of power and wind speed is motivated by the skew density of power and wind speed. Investigations have shown that the square root transformation leads to a distribution of the prediction errors, which reasonably well can be approximated by the Gaussian distribution. This implies that confidence intervals for the predictions can be based on fractiles in the Gaussian distribution.

The model (4.38) is implemented in the Wind Power Prediction Tool (WPPT) described in

Chapter 7. Although it has been found that the optimal forgetting factor is close to 1, it has been decided to implement the forgetting $\lambda = 0.999$. First of all, some adaptivity in the estimation is desirable, and furthermore, the case that the optimal λ is so close to 1 may be due to not having a longer period of data.

	S.E. [kW]			
k	Vedersø Kær 1	Vedersø Kær 2	Brøns	Dræby Fedsodde
1	281.8	404.6	224.4	162.2
2	402.4	549.1	314.9	237.7
4	555.1	755.3	420.6	326.2
6	671.3	945.6	499.7	406.2
12	906.1	1384.4	656.3	640.3
24	1167.2	1752.1	952.8	852.5
42	1326.0	2011.8	978.0	824.5

Table 4.11: Prediction performance for the best overall model.

Table 4.11 shows the standard deviation of the prediction error (calculated after having transformed back to power values). This shows that for the chosen model and value of λ only a slightly worse standard deviation of the prediction error is obtained.

4.6 A Model using Meteorological Forecasts

Most of the models for wind power prediction are prepared for using meteorological forecasts whenever they are available. This section describes or illustrates how meteorological forecasts can be used as input to the model (4.38). More results on that subject are found in (Nielsen & Madsen, 1995f).

The data used is a full one year data set from Vedersø Kær starting May 1st, 1994. The meteorological forecasts are provided by the HIgh Resolution Limited Area Models-system (HIRLAM) (see (Landberg, Watson, Halliday, Jørgensen, & Hilden, 1994)). New HIRLAM forecasts are available every 12 hour for the next 36 hours and the resolution is 3 hours. The forecasts used in the study described in the following are delivered by the Danish Meteorological Institute (DMI). HIRLAM provides forecasts in the free atmosphere (approx. 1 km above the surface). This forecast is used for a transformation to the wind 10 m above the ground using the geostrophic drag law combined with the logarithmic wind profile ((Landberg et al., 1994)). However, this transformation does not take into account local conditions, as trees, buildings, etc. Therefore a simple (and preliminary) statistical model is used to take into account the local conditions. The model found for calculating the prediction of the

	Naive Pred.	Stat. Model	Improvement
k	S.E. [kW]	S.E. [kW]	%
6	889	771	13.3
12	1266	964	23.9
18	1526	1051	31.1
24	1703	1106	35.1
30	1819	1153	36.6
36	1889	1199	36.5
42	1928	1238	35.8
48	1954	1275	34.7

Table 4.12: Prediction performance for the statistical model using meteorological forecasts as input

wind speed $\hat{w}(t+k|t)$ at the farm is

$$\begin{aligned} w(t+k) &= a_0 w(t) \\ &\quad +(b_0 + b_1 \sin(\hat{\phi}_{HIR}) + b_2 \cos(\hat{\phi}_{HIR}))\hat{w}(t+k|t)_{HIR} \\ &\quad +m + e_2(t+k), \end{aligned} \tag{4.39}$$

where w is the wind speed in the farm, and $\hat{\phi}_{HIR}$ and $\hat{w}(t+k|t)_{HIR}$ are the transformed HIRLAM forecasts for the direction and the wind speed, respectively. The term m is accounting for the level of the wind, and e_2 is the prediction error. It is obvious that the model can easily be improved by e.g. introducing a better angular dependency.

The predicted wind speed at the farm $\hat{w}(t+k|t)$ is used as input in the model for finally predicting the wind power. As a simple extension of (4.38) the following models are found to give reasonable predictions

$$\begin{aligned} \sqrt{p(t+k)} &= a_1\sqrt{p(t)} + c_1\sqrt{w(t)} + d_1 w(t) + h_1 \sin\frac{2\pi t}{48} + h_2 \cos\frac{2\pi t}{48} \\ &\quad +v_1\sqrt{\hat{w}(t+k|t)} + q_1\hat{w}(t+k|t) + l + e_1(t+k), \end{aligned} \tag{4.40}$$

The performance of the model (4.40) is shown in Table 4.12. In (Nielsen & Madsen, 1995f) several other models are studied. For the same period, and hence the same data, the improvement of the model without meteorological forecasts (i.e. model (4.38)) compared to the naive predictor is between 6% for $k = 6$ and 20% for $k = 48$.

4.7 Potential Improvements of the Models

The results shown in the previous section give a lower value of the performance of simple statistical models. The models found are the one which were possible to formulate using the rather limited amount of data available. However, it is clear that the models can be

improved. This section is devoted to a description of the main areas, where improvements are judged to have the largest impact on the quality of the predictions.

Compared to the implemented tool at ELSAM one key area is the use of meteorological forecasts. The implemented system is capable of giving reasonable predictions 6 to 12 hours ahead without using meteorological forecasts, but it is clear, as illustrated above, that on-line forecasts from a meteorological service station can be used to improve the predictions for the longer horizons. Forecasts up to 36 hours ahead would be useful in the production planing.

Another key activity is an optimization of the prediction models based on the data collected during the trial period of the wind power prediction tool (in the following denoted "prediction tool"). Most of the models presented in this book were found using only about 3 months of data and furthermore the actual 3 months were a late summer period with a nearly constant weather situation. With more data it would be of interest to find a model which is globally optimal, i.e. a model which demonstrates optimal prediction performance all over the year.

The following subjects are of special interest in further model investigations:

- *Optimal prediction models without using wind speed measurements.* It has been experienced that the quality of the wind power production measurements is good, whereas the quality of the wind speed measurements very often is less good. Since the wind speed is included in the prediction models used today, a drop out of the wind speed measurements means that the total prediction for that farm also drops out. For production planning it is crucial that reliable predictions are available all the time and it would therefore be beneficial to investigate the possibility of using models without requirements for wind speed measurements in the prediction tool. Another incentive for investigating such models is that the predominant part of the costs by establishing on-line measurements for a wind farm are related to the measurement of wind speed.

- *Identification of an optimal self-learning rate.* Due to the short period of data available, it has not been possible to find an optimal self-learning or adaption rate. The need for the adaptation of the models is mainly due to the annual variation in the dynamical behavior of the wind speed. Since more than one year of data now is available, it will be possible to identify an optimal or reasonable self-learning rate (also called forgetting factor).

- *Multivariate models.* As mentioned above, the model used in the prediction tool today is based on data from a 3 months period with stable weather conditions and furthermore only data from a few farms were available during the model investigation. Hence, it was not possible to construct a multivariate model, i.e. a model which takes into account the cross correlation between wind mill parks. However, it is clear that for instance a change in the weather situation would parse gradually over the whole ELSAM supply area. This means that changes in one region are likely to occur - with some time delay - in the neighboring regions. Therefore it is obvious that a multivariate model will improve the predictions.

- *Input from meteorological forecasts.* In this book some preliminary results about using meteorological forecasts as input to the models are shown. The HIRLAM forecasts

used as input were available only every 12 hour, and in 3 hour steps. More frequent forecasts and a better resolution would most likely improve the prediction performance. Furthermore, the models for transferring the output from HIRLAM to the wind farms can be improved.

Another issue that must be considered in building wind power prediction tools is the number of available measurement points. Today the prediction tool at ELSAM uses measurements from 7 wind farms in the ELSAM supply area. It has been clear that more measurement points, especially in the western part of Jutland, could give better forecasts.

4.8 Conclusion

In this chapter linear as well as a number of linear and nonlinear dynamic models have been applied primarily on wind speed data from Vedersø Kær. The results showed that none of the models tested gave any significant performance improvement compared to the models belonging to the linear AR class.

Also different models for the relation between wind speed and power have been estimated on data from Vedersø Kær. It turned out to be very important for the fit of the model that the differences between the estimated power curve and the actual power observations were modelled by an ARMA model. This is most likely due to these differences being highly correlated as a consequence of a substantial wind direction dependence in the wind farm. The different types of models for the power curve did, however, not show any clear differences in results.

A class of linear models for the wind speed were estimated recursively and adaptively, and subsequently the wind speed predictions were transferred to power predictions via an off-line estimated power model. Similarly linear k-step predictors were estimated in the same way, and the predictions were transferred into power predictions. The two approaches turned out to have almost the same prediction performance.

A class of linear k-step predictors for the power were estimated recursively and adaptively. Both models with and without the wind speed as explanatory variable were estimated. For the shortest prediction horizons it seemed to be important to have the wind speed in the vector of explanatory variables, whereas for prediction horizons exceeding something around 8-12 hours the wind speed observations should not be used in the power prediction.

The performance of a proposed reasonable overall model, which is selected for implementation in the wind power prediction tool (see Chapter 7), is evaluated. This model is based on square root transformations of power and wind speed. The evaluation is done using the standard deviation of the prediction error criterion, and on this criterion the overall model turns out to have a slightly worse performance compared to non-transformed observations. The implemented model is prepared for taking into account meteorological forecasts of the wind speed.

Chapter 5

Neural Network for Wind Power Prediction

In this chapter it is investigated whether simple neural networks can be used for k-step predictions of half-hour averages of wind power production and how they perform compared to the adaptive k-step predictors based on models of the form (4.38). Furthermore comparisons with the naive (or persistent) predictor are made. Comparisons are carried out for prediction horizons $k = 1, \ldots, 6$, corresponding to 1/2 to 3 hour.

5.1 Neural Networks

Type of Neural Network

A feed-forward neural network with one hidden layer and without connections directly from input to output is used, see e.g. (Ripley, 1994).

Suppose that observations (indexed by i) of the independent variables (indexed by j) x_{ij} and the dependent variable y_i are present. The dependence of y on x can then be modelled by a neural network of the above type as follows:

$$y_i = \phi_o \left(\alpha_o + \sum_{h=1}^{n_h} w_{ho} \, \phi_h \left(\alpha_h + \sum_{j=1}^{n_j} w_{jh} x_{ij} \right) \right) + e_i \tag{5.1}$$

where $w_{\cdot o}$ are the weights on the connections from the hidden layer to the output layer, $w_{\cdot h}$ are the weights on the connections from the input layer to unit h in the hidden layer, α_o is the bias on the output unit, and α_h are the biases on the hidden units. n_h and n_j are the No. of hidden units and inputs, respectively. It is seen that the weights and the biases are just parameters of the model. Considering (5.1) as a statistical model one would assume that the e_i's are independently identically distributed.

The functions $\phi_h(\cdot)$ and $\phi_o(\cdot)$ are predefined functions associated with the units in the hidden

and output layer, respectively. Most frequently these functions are sigmoid[1], i.e. the output of the network is restricted to the interval $]0,1[$. This is not desirable in this application since the future output of the network is then somehow restricted to the range of observations in the data set used for estimating the parameters. For this reason the output unit is chosen to be linear ($\phi_o(z) = z$).

Estimation of Parameters

Given the observations $(y_i, x_{i1}, x_{i2}, \ldots, x_{in_j})$, $i = 1, 2, \ldots, N$ and based on the formulation (5.1) an obvious way to estimate the parameters is to choose

$$\theta = [\alpha_0, \ldots, \alpha_{n_h}, w_{1o}, \ldots, w_{n_h o}, w_{11}, \ldots, w_{n_j,n_h}]^T \tag{5.2}$$

so that

$$V(\theta) = \sum_{i=1}^{N} \hat{e}_i^2(\theta), \tag{5.3}$$

where

$$\hat{e}_i(\theta) = y_i - \phi_o\left(\alpha_o + \sum_{h=1}^{n_h} w_{ho}\ \phi_h\left(\alpha_h + \sum_{j=1}^{n_j} w_{jh} x_{ij}\right)\right), \tag{5.4}$$

is minimized. Since the number of parameters is equal to $n_h(n_j + 2) + 1$ it is seen that when the number of inputs and the number of hidden units are large the model will contain a large No. of parameters. In this case the minimization of (5.3) may lead to a model which fit the data used for estimation too well, i.e. the model adapts to the random part of the observations.

The initial values of the estimates are rather important since the optimization problem may contain local minima due to the fact that the model is non-linear in the parameters. For this reason each model should be estimated several times using different initial parameter estimates. Since it is rather difficult to suggest appropriate values it seems reasonable to select these values at random. In this case the data is scaled to the interval $[0,1]$ (see below) and according to the documentation on the software used (see Section 5.1, and (Ripley, 1994) or (Nielsen & Madsen, 1995c)) it should be sufficient to sample from the $U(-1,1)$ distribution. In spite of this the $U(-5,5)$ distribution is used in order to cover a wider interval of initial parameter estimates.

Selection of Network Size

As mentioned earlier in this section the class of networks used is restricted to networks with one hidden layer and the function corresponding to the output unit is chosen to be linear.

To use a neural network it remains to decide upon the independent variables to include in the model and on the No. of hidden units. This may be done by using some kind of information

[1] Also called logistic functions.

criteria. In this case the Bayes Information Criterion (BIC) is used, see (Schwarz, 1978). With L^*, n_p, and N being the value of the likelihood function (see e.g. Rao (1965)) in the optimum, the No. of parameters, and the No. of observations used in the estimation, respectively, the criteria corresponds to chose the model so that

$$\log L^* - \frac{n_p}{2} \log N, \tag{5.5}$$

is maximized. For a large class of linear time series models and other linear models with the residuals being normally distributed and with constant variance the criteria is equivalent to minimizing

$$BIC = N \log \tilde{\sigma}_e^2 + n_p \log N, \tag{5.6}$$

where

$$\tilde{\sigma}_e^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{e}_i^2(\hat{\theta}), \tag{5.7}$$

is the maximum likelihood (ML) estimate of the variance of e_i. Note that $\hat{e}_i$ must be based on ML estimates also. The prediction errors ($\hat{e}_i(\hat{\theta})$) based on data scaled to the interval $[0, 1]$ are used, see below.

In this case the procedure used for estimation of the parameters (see above) is not an ML procedure. Hence the above procedure must be regarded as an approximation. Furthermore, in (Schwarz, 1978) the derivation of (5.5) is based on the assumption that the observations come from a Koopman-Darmois family.

In order not to investigate an excessive No. of models the criterion (5.6) is used to select an appropriate No. of hidden units only. The independent variables are considered fixed, see below.

Software

The software used is written by Professor of Applied Statistics, B.D. Ripley, University of Oxford. The software can be obtained from StatLib by anonymous ftp from `lib.stat.cmu.edu`[2]. The software is written for S-Plus and is briefly described in (Ripley, 1994).

The adaptive predictions are calculated by use of the software Off-line Wind Power Prediction Tool, version 1.0, see (Nielsen & Madsen, 1995b) and (Nielsen & Madsen, 1995a).

5.2 Variables in the Models

As mentioned above the aim is to obtain k-step predictors using a neural network. The following independent variables are used: The present wind power production (p_t) scaled to approximately $[0, 1]$, the present wind speed (w_t) scaled to approximately $[0, 1]$, $\frac{1}{2}sin(2\pi h_{t+k}/24)+\frac{1}{2}$, and $\frac{1}{2}cos(2\pi h_{t+k}/24)+\frac{1}{2}$. The wind power production k step ahead (p_{t+k}) scaled to approximately $[0, 1]$ is used as the dependent variable. Compare with (4.38).

[2]The software can also be obtained from `markov.stats.ox.ac.uk`

5.3 Validation

The models are validated using a different data set than the one on which the selection of the No. of hidden units and the estimation of parameters is based. This data set is called the validation set.

The neural network model selected for each prediction horizon k is compared with the naive k-step predictor (4.35). Furthermore the neural network predictors are compared with a predictor based on the model (4.38) estimated adaptively with a forgetting factor of 0.999. This corresponds to the method recommended in Chapter 4. The estimation and the validation set are just two parts of one time series. Therefore it is possible to allow the adaptive predictions to settle before the validation is initiated. This method is chosen since this corresponds to the real application.

Based on the validation set the k-step residuals (or prediction errors) are calculated on the original scale and based on these the Root Mean Square (RMS) is calculated. For the residuals $(r_1, r_2, \ldots, r_N)$ the RMS of the residuals is defined as $\sqrt{\frac{1}{N}\sum_i r_i^2}$.

5.4 Data

The data used in this investigation has been collected in the Vedersø Kær wind farm in the ELSAM area during the period 2 July, 1993, 5.30 p.m. until 11 October, 1993, 7 a.m.. The data until 6 September at 7 a.m. is used for estimation whereas the remaining data is used for validation. The original sampling time is 5 minutes. Based on these values half-hourly averages are calculated.

Some negative values of the half-hourly averages of wind speed and wind power production occur. The minimum wind speed observed is -0.01674 m/s and the minimum wind power production observed is -28.86 kW. All negative values are set to zero.

5.5 Results

Estimation

For prediction horizons (k) 1 to 6 (1/2 to 3 hours), neural networks with 1 to 5 hidden units are considered. Each estimation is performed 20 times with initial values of the parameters sampled from the uniform distribution covering the interval $[-5, 5]$.

In Figure 5.1 the resulting values of BIC are shown. Extremely high values clearly corresponding to local minima are excluded from the plots. It is seen that for $k = 1, 2, 3$ the lowest value of BIC is observed for a network with three hidden units. For $k = 4, 6$ a network with four hidden units results in the lowest observed BIC, and for $k = 5$ a network with five

hidden units results in the lowest BIC[3]. The network size for which the minimum values of BIC is observed and the actual minimum values are shown in Table 5.1.

k	1	2	3	4	5	6
n_{min}	3	3	3	4	5	4
BIC_{min}	-18807	-16680	-15571	-14730	-14104	-13638

Table 5.1: Minimum value of BIC observed for each horizon (k) and the corresponding network size n_{min}.

In conclusion, a neural network with three hidden units is expected to be close to optimal among the networks investigated for all horizons ranging from 1 to 3 steps. For prediction horizons 4 and 6 a neural network with four hidden units is expected to be nearly optimal, whereas for the 5-step predictions a network with five hidden units gives a marginal lower minimum value of BIC than the network with four hidden units.

It is seen from Figure 5.1 that for $k = 3, \ldots, 6$ it is not evident, which size of the neural network is truly optimal, i.e. a new investigation might change the size of the network for which the minimum BIC is observed. This could be investigated with statistical tests but since the interest is on the minimum rather than the mean or median this would require a very large No. of estimations with random initial weights, making the investigation almost infeasible (the user time on an HP9000/735 for investigating one horizon is up to two hours).

Validation

For all prediction horizons the neural network with the lowest BIC (see Table 5.1) is validated as described in Section 5.3. The program used and the output of the program can be found in (Nielsen & Madsen, 1995c). The results of the validation are shown in Table 5.2.

It is seen that the neural network models investigated are all inferior to both the naive and the adaptive predictors. It is noted that the naive predictor is slightly better than the adaptive for $k = 1, 2, 3$. For $k = 5, 4, 6$ the adaptive predictor is better than the naive. However, the ratio between RMS_{adap} and RMS_{naive}, as shown in Table 5.3, reveals that the difference is minor.

Networks with One Hidden Unit

From the validation of the network of optimal size it is seen that the naive predictor performs well compared to the other methods investigated. It is therefore peculiar that the selection procedure does not lead to a selection of the most simple network; a network with one hidden unit only. Therefore this kind of network is compared with the network selected according to BIC. Results are indexed by *opt* and 1 for the optimal and the simple network, respectively.

[3] Confirmed by an inspection of the actual values.

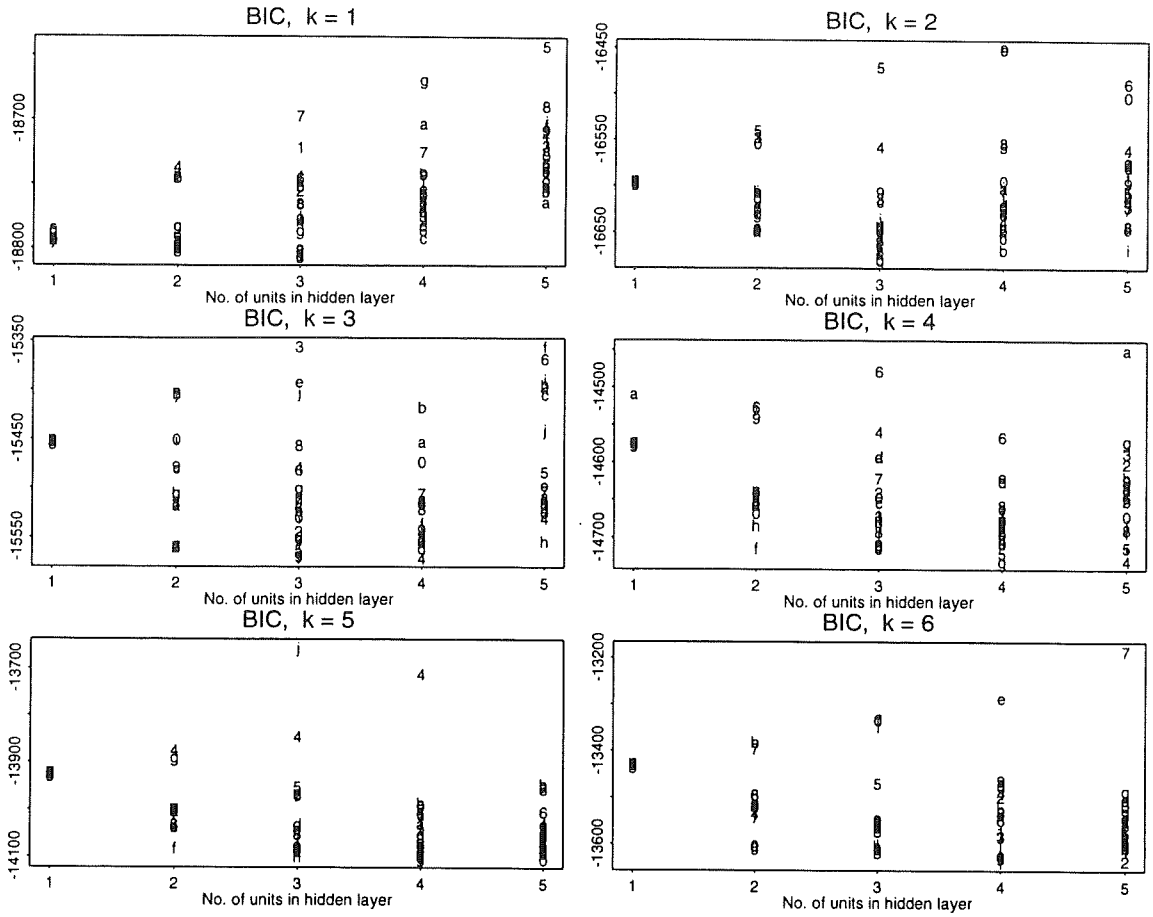


Figure 5.1: Values of BIC, extreme (high) values excluded.

k	RMS_{nn} (kW)	RMS_{naive} (kW)	RMS_{adap} (kW)
1	297.6	261.7	262.8
2	448.6	375.7	377.0
3	534.9	440.9	442.0
4	621.2	507.6	507.5
5	678.4	569.2	565.1
6	705.8	623.1	614.1

Table 5.2: Validation set; RMS of prediction errors of the best neural network (see Table 5.1), naive, and adaptive predictor.

k	1	2	3	4	5	6
$\frac{RMS_{nn}}{RMS_{naive}}$	1.14	1.19	1.21	1.22	1.19	1.13
$\frac{RMS_{adap}}{RMS_{naive}}$	1.00	1.00	1.00	1.00	0.99	0.99

Table 5.3: RMS-ratio when comparing with the naive predictor.

The parameters of each model are estimated 20 times with random initial parameter estimates and the estimates corresponding to the lowest value of BIC are selected, i.e. the same procedure as for the optimal network is used.

The results based on the estimation and on the validation set are shown in Table 5.4, together with the results corresponding to the network selected according to BIC.

	Estimation set				Validation set		
k	BIC_{opt}	BIC_1	RMS_{opt} (kW)	RMS_1 (kW)	k	RMS_{opt} (kW)	RMS_1 (kW)
1	-18807	-18793	284.7	289.7	1	297.6	268.8
2	-16680	-16597	398.8	410.4	2	448.6	411.6
3	-15571	-15452	475.2	491.8	3	534.9	523.3
4	-14730	-14575	538.6	565.0	4	621.2	622.8
5	-14104	-13923	590.0	626.2	5	678.4	699.9
6	-13638	-13430	639.9	676.9	6	705.8	758.3

Table 5.4: Estimation and validation set; comparison of optimal network with a network with one hidden unit (BIC of models and RMS of prediction errors).

From the table it is seen that for $k = 1, 2, 3$ the neural network with one hidden unit actually performs better on the validation set than the network selected according to BIC. However, comparing Table 5.4 with Table 5.2 it is seen that the neural network with one hidden unit is inferior to the naive and the adaptive predictor.

5.6 Conclusion

The type of neural networks investigated is inferior in prediction performance to both the adaptive predictor and the simple naive predictor for the prediction horizons investigated (1/2 to 3 hours).

For the prediction horizons investigated the naive predictor performs better than the adaptive predictor for the short prediction horizons (up to $1\frac{1}{2}$ hour). However, the difference between

the two predictors is minor. For horizons larger than 2 hours the adaptive predictor is better than the naive.

It is seen that the estimation of parameters in neural networks is complicated, since the estimates are dependent on the initial estimates. For this reason it is recommended to perform the estimation several times with initial parameter estimates chosen at random.

5.7 Discussion

For one particular set of parameters the autoregressive model corresponds to the naive predictor. It is therefore peculiar that the adaptive predictor does not perform strictly better than the naive predictor.

One obvious reason why this possibility exists is that when using the adaptive predictor six parameters have to be estimated. Another possible reason is that the correlation structure in the validation data set differs from the correlation structure of the estimation data set. This hypothesis is consistent with the results in Chapter 4, in which the comparison of the adaptive and naive predictor is performed on the union of the estimation and validation data set.

Apart from the non-linear response of the hidden units, a neural network predictor also includes the naive predictor. The reason why the neural network predictor performs considerably worse than the naive and the adaptive predictors is probably that: (i) The estimation of the parameters in the neural network is not adaptive, (ii) the No. of parameters in the neural network is large (seven or larger), and/or (iii) the non-linear response of the hidden units is inappropriate for wind power predictions.

For the low horizons investigated the naive predictor performs slightly better than the adaptive predictor based on the autoregressive model. For the larger horizons the adaptive predictor is slightly superior. For very large horizons a simple profile[4] will probably be the best predictor. The adaptive predictor processes the characteristic of being able to interpolate between the extremes. For this reason the adaptive predictor is attractive.

The largest No. of hidden units investigated is five. In most cases the optimal network size is found to be less than five. Since the sum of the squared prediction errors for the estimation data set is a non-increasing function of the No. of hidden units BIC will have one minimum only. Therefore the maximum size of the networks investigated is sufficient.

For further discussion see (IMM et al., 1995).

[4] The profile will probably contain harmonics corresponding to daily and yearly periods.

Chapter 6

Predicting the Wind Power in a Region – The Upscaling Problem

The problem of estimating the total power in the ELSAM supply area when knowing the power in seven wind farms is referred to as the upscaling problem. In this chapter it is investigated how to carry out the transformation from values of wind and power at the seven wind farms to an estimate of the total wind power contribution.

Different aspects are to be taken into account, or at least consideration, in the investigation and decision on the approach for carrying out the upscaling. First of all, it is evident that the wind speed locally is influenced by the surrounding landscape, and also by the location in relation to the dominating wind direction and the North Sea. This means that the wind speed variations at one of the wind farms cannot be used as a representation of the wind speed at a specific wind farm/mill located away from the seven wind farms without introducing a source of error.

Furthermore, the relation between wind speed and power (the power curve) is specific for a given brand and type of mill. This implies that the ratio of the total wind power for the ELSAM system to the wind power at the seven wind farms depends on the wind speed (for an assumption of an overall valid wind speed). The consequence of this fact is that the upscaling must somehow depend on the wind speed.

Finally, the configuration of the mills in wind farms implies that the wind speed at each of the mills is influenced by the other mills for which reason the shape of the power curve for a wind farm, consisting of a number of identical mills, is different from the shape of the power curve for one of such mills.

In this chapter the relationship between power and wind speed is investigated for different wind farms. This makes it possible to evaluate whether the scaled power production from one wind farm can be used as an estimator for the power production for nearby sited wind farms. Later an alternative upscaling method using a central wind speed measurement and low frequency energy readings from the windmills sited in an area is suggested but not investigated.

6.1 Relations Between Nearby Wind Farms

Previous analyses have exposed that there are considerable differences in the utilization time for different wind farms. The utilization time is given as the amount of energy for a given period divided by the power capacity. There may be different explanations to this fact, some of which are mentioned above.

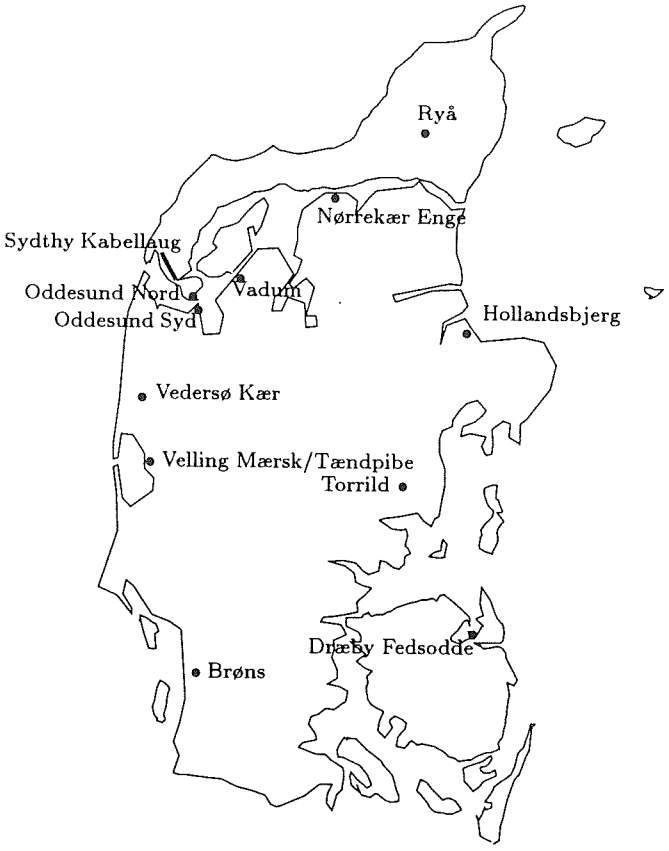


Figure 6.1: A key map of the ELSAM supply area with marking of the seven reference wind farms as well as the wind farms considered in this chapter.

The purpose of the present investigation is to determine the relation between the power utilization and the wind speed (the power utilization curve) for a number of different wind farms. The power utilization is defined as the actual power outlet divided by the power capacity of the wind farm, and is introduced to make the results for the different wind farms comparable. Having estimated the power utilization curves it may be concluded whether the different utilization times can be explained by differences in wind energy conditions or by differences in the efficiency of the wind farms.

The present analysis is based on power measurements from a number of wind farms all located in the north-western part of Jutland, including one of the seven wind farms, along with wind speed measurements at two of the reference wind farms located in this area.

The wind farms in question are

Wind farm	Number of Mills	Power Capacity [kW]
Vedersø Kær	27	6075
Velling Mærsk	65	9985
Vadum	4	1650
Sydthy Kabellaug	25	5625
Oddesund Nord	20	1100
Tændpibe	30	2250

Vedersø Kær is one of the seven reference farms from which the upscaling is to be made. Furthermore, Velling Mærsk and Vadum are owned by power companies owned by ELSAM whereas the three other farms are privately owned. The special thing about Sydthy Kabellaug is that the windmills are placed on a straight line, parallel and close to the bank of Limfjorden, with such intermediate distances that the wind speed at one mill is not influenced by the other mills.

Figure 6.1 shows the location of the wind farms.

6.1.1 Data

The data used in the present investigation consists of 15 minute average values of the wind speed at Vedersø Kær and Nørrekær Enge as well as the power utilization at Vedersø Kær, Sydthy Kabellaug, Velling Mærsk, Tændpibe, Vadum and Oddesund Nord. Data is sampled in the period from 6 August at 10.15 a.m. to 8 September at 1.15 a.m. 1993.

6.1.2 Model

The model used to describe the relation between wind and power utilization is chosen as

$$P_r(t) = M \exp\left(-b \exp\left(-k \cdot w(t)\right)\right) + e(t) \tag{6.1}$$

where $P_r(t)$ is the power utilization, and $w(t)$ is the wind speed. It is a problem that it is only at Vedersø Kær that locally collected wind speed observations are available. However, the wind speed observations at Vedersø Kær and Nørrekær Enge are used to represent the wind speed according to the following weighting scheme

$$\begin{aligned} w(t) \;=\; & \nu_1 w_{\mathrm{VK}}(t-1) + \nu_2 w_{\mathrm{VK}}(t) + \nu_3 w_{\mathrm{VK}}(t+1) + \\ & \nu_4 w_{\mathrm{NE}}(t-1) + \nu_5 w_{\mathrm{NE}}(t) + \nu_6 w_{\mathrm{NE}}(t+1) \end{aligned} \tag{6.2}$$

where $w_{VK}(t)$ and $w_{NE}(t)$ are the wind speeds at Vedersø Kær and Nørrekær Enge, respectively, and where the following restriction is used

$$\nu_1 + \nu_2 + \nu_3 + \nu_4 + \nu_5 + \nu_6 = 1. \tag{6.3}$$

By estimating this expression for the wind speed a weighting in time and a weighting between the two wind speed observations are introduced. Hereby an approximation to the delays in the distribution of the wind speed variations is possible. Furthermore, representing the wind speed by a weighted average may imply a smoothing and down-weighting of noise in the wind speed variation, which is advantageous for the relation to power utilization.

6.1.3 Estimation

The parameters of the model

$$\theta = \begin{bmatrix} M & b & k & \nu_1 & \nu_2 & \nu_3 & \nu_4 & \nu_5 & \nu_6 \end{bmatrix} \tag{6.4}$$

are estimated by the least squares method, i.e. the parameter estimates $\hat{\theta}$ are found as the parameters minimizing the criterion

$$V(\theta) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\left(P_r(t) - \hat{P}_r(t)\right)^2 = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\hat{e}^2(t), \tag{6.5}$$

where N is the number of observations used in the estimation.

An approximation to the variance of the estimated power utilization curve based on a linearization of (6.1) around the parameter estimates is

$$\widehat{Var}[P_r(w)] \simeq \left(\frac{\partial P_r(w)}{\partial \theta}\right)^T \cdot \widehat{Cov}\left\{\hat{\theta}\right\} \cdot \left(\frac{\partial P_r(w)}{\partial \theta}\right) \tag{6.6}$$

where $\widehat{Cov}\left\{\hat{\theta}\right\}$ is estimated as described in Section 4.1. Under the assumption of uncorrelated estimates (6.6) is reduced to

$$\widehat{Var}[P_r(w)] \simeq \left(\frac{\partial P_r(w)}{\partial M}\right)^2 Var[\hat{M}] + \left(\frac{\partial P_r(w)}{\partial b}\right)^2 Var[\hat{b}] + \left(\frac{\partial P_r(w)}{\partial k}\right)^2 Var[\hat{k}] \tag{6.7}$$

6.1.4 Results

Table 6.1 lists the estimation results for the six different wind farms. Clearly the best fit is obtained for Vedersø Kær, and most of the weight on wind speed measurements is on the simultaneous wind speed measured at Vedersø Kær. For the rest of the wind farms the fit of the model is not as good as for Vedersø Kær. This is reasonable as no local wind speed measurements have been available at these farms. It also seems to be the case that the standard deviation of the prediction error increases with the distance from the wind farm to the nearest wind speed measurement.

					Weighting of Wind Speed					
					Vedersø Kær			Nørrekær Enge		
Wind Farm	$\hat{\sigma}_e$	$\hat{M}$	$\hat{b}$	$\hat{k}$	$t-1$	t	$t+1$	$t-1$	t	$t+1$
Vedersø Kær	0.048	1.31	10.5	0.227	0.14	0.64	0.13	0.04	0.03	0.02
Sydthy Kabellaug	0.119	0.98	11.9	0.358	0.26	0.13	0.37	0.09	0.02	0.13
Velling Mærsk	0.083	1.32	10.8	0.248	0.51	0.19	0.29	0.03	0.00	-0.02
Tændpibe	0.090	1.13	11.2	0.273	0.53	0.17	0.31	0.02	-0.02	-0.01
Vadum	0.117	0.89	10.9	0.279	0.15	-0.06	-0.09	0.56	0.24	0.20
Oddesund Nord	0.122	0.93	17.9	0.421	0.29	0.11	0.33	0.15	0.05	0.07

Table 6.1: Modeling results.

Apart from Vadum it is the case that most weight is given to the wind speed measurement at Vedersø Kær. This may be explained by the assumption that the dominating wind speed comes from the western direction for all of the wind farms except for Vadum, where the wind speed may be dominated from the northern direction and consequently is better represented by the wind speed measurement at Nørrekær Enge.

The upper part of Figure 6.2 shows the estimated power utilization curves for the six wind farms, and for comparison is shown the specification for the windmill, which is the only type of mill used at Vedersø Kær and Sydthy Kabellaug. Figure 6.3 and the lower part of Figure 6.2 show the curve estimates for Vedersø Kær, Sydthy Kabellaug and Velling Mærsk along with curves corresponding to plus/minus one standard deviation. The curves are shown for the interval where wind speed values actually have been observed.

First of all, considerable differences in the curve estimates are seen. The power utilization curve for Vedersø Kær is clearly below the curve corresponding to the specification of the type of mill which is the only type used in this wind farm. The most reasonable explanation to this result is that the windmills give shelter to each other, meaning that the wind speed at the mills which are sheltered by other mills is lower than the unaffected wind speed.

For Sydthy Kabellaug it is seen that the curve estimate is clearly above the specification when the wind speed is in the lower range. This could have the explanation that the actual wind speed at the wind farm is badly represented by the measured wind speed at Vedersø Kær and Nørrekær Enge. Looking at the location of Sydthy Kabellaug it is likely that the wind speed in general is higher, implying that the wind speed used in the model estimation has been too low compared to the real wind speed. If this is the case the estimated curve should be pulled to the right in the interval between zero and maximum utilization, which would imply the curve estimate to have a better agreement with the specification.

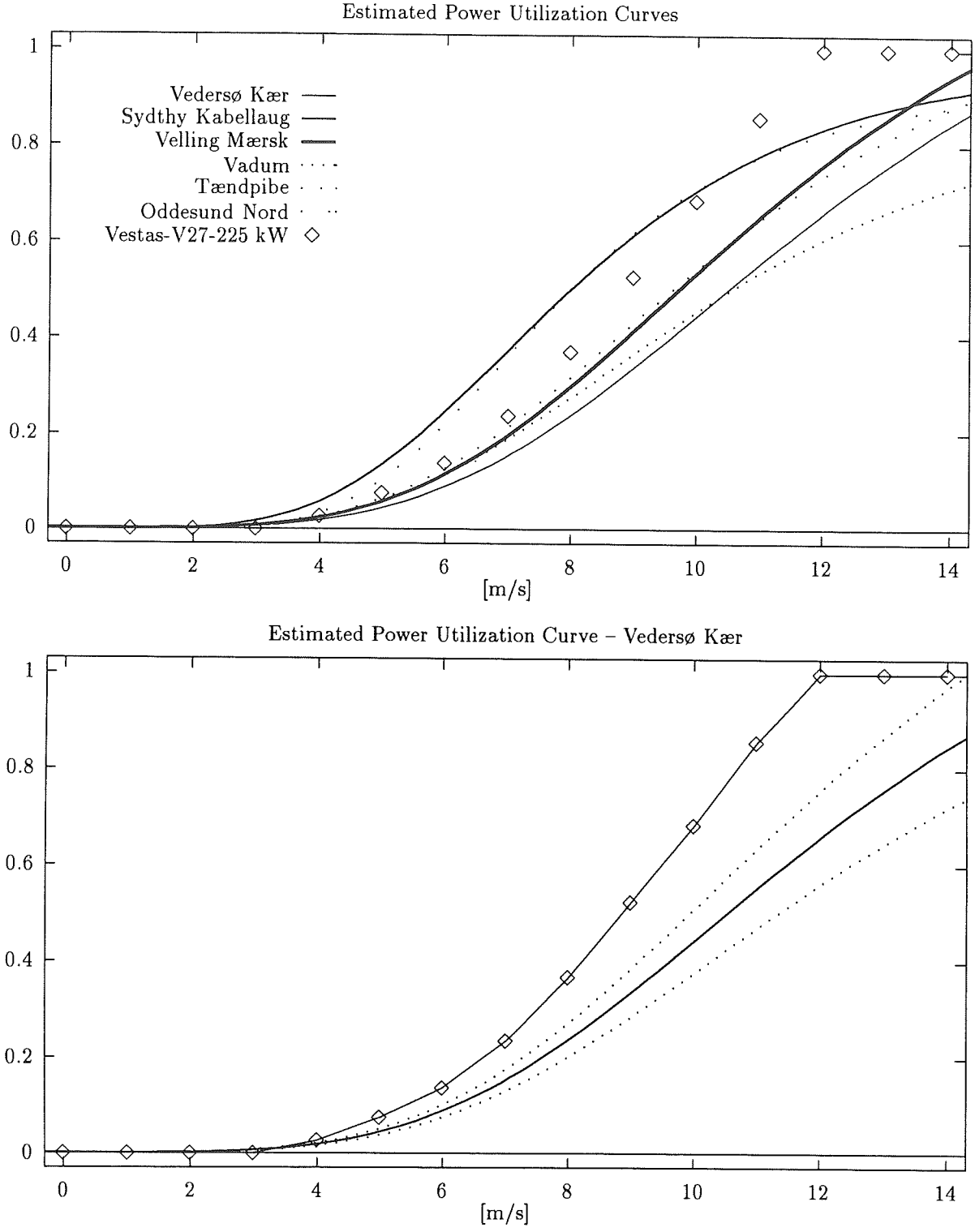


Figure 6.2: Estimate power curves for all wind farms (above) and for Vedersø Kær (below) with confidence bands (± one standard deviation).

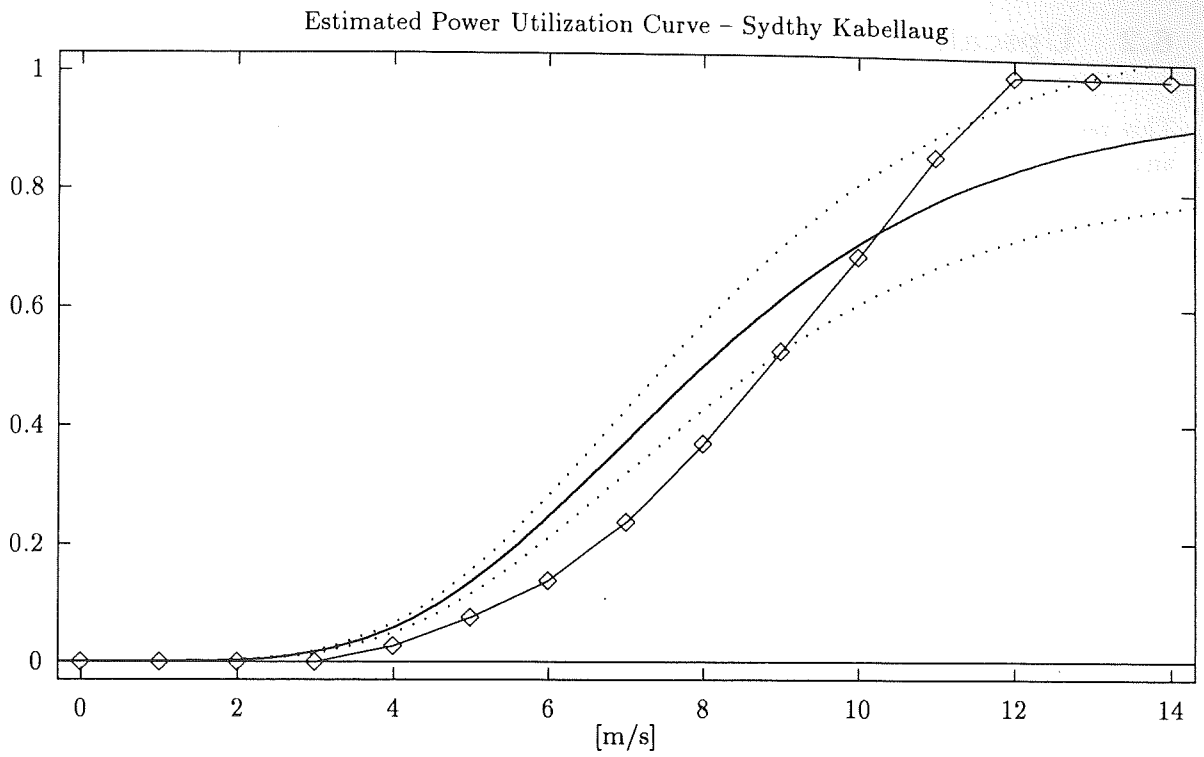


Figure 6.3: Estimated power curves for Sydthy Kabellaug with confidence bands (± one standard deviation).

6.1.5 Conclusion

The main conclusion to the results of the present analysis is that the power utilization at one wind farm cannot be used as a representation of the power utilization at other wind farms without appropriate rescaling in both the wind speed direction and the power direction. This can be necessary even if the wind farms are within close distance.

The estimates of the power utilization curve for different wind farms have shown to be significantly different. Possible explanations to these differences are

- The energy in the wind at the farm is influenced by the surrounding landscape and the distance to open sea.

- The number of windmills in a farm and the configuration of the farm are characteristics, which are determining for the power curve.

- The power curve for a single mill depends on brand and type.

6.2 Upscaling Using Area Energy Readings

This section draws up a simplified model for the relation between a wind speed measurement and the corresponding power production for windmills placed in the same area and suggests a solution to the power upscaling problem based on this simplified model.

A simplified model between a remote wind speed measurement and the power production from an individual windmill is derived. Subsequently this model is extended to describe the relationship between one wind speed measurement and the wind power production in an entire area.

6.2.1 Modeling a Single Windmill

The power curve for a specific windmill is often parameterized using the following functional relationship between power output and wind speed

$$P(t) = M \cdot \exp(-b \cdot \exp(-k \cdot w(t))) \tag{6.8}$$

where $P(t)$ is the power output at time t, $w(t)$ is the wind speed at time t, M is the power capacity for the considered windmill and b, k are form parameters. M, b and k are either supplied by the manufacturer of the windmill or easily found using the specified power curve for the windmill.

The wind speed experienced by the individual mills in a group at a given moment is different due to disturbances from the surroundings and the physical extent of the group. In a certain height (H_{ref}) over the ground the influence of the disturbances is negligible, and the low pass filtered wind speed at that height can in a (small) local area be regarded as being uniformly distributed. In the following such an area will be referred to as a sub area.

The wind in a local area is considered to be a turbulent flow over a rough surface. For a given wind direction the local wind speed as a function of height can be written as

$$w(h,t) = K_{loc}(h) \cdot w_{ref}(t) \qquad 0.0 \leq K_{loc}(h) \leq 1.0 \tag{6.9}$$

$$\simeq \begin{cases} \frac{h}{H_{ref}} \cdot w_{ref}(t) & \text{for } 0.0 \leq h < H_{ref} \\ w_{ref}(t) & \text{for } H_{ref} \leq h \end{cases} \tag{6.10}$$

where $w(h,t)$ is the local wind speed in height h, $K_{loc}(h)$ is the wind speed scaling factor, w_{ref} is the local wind speed in height H_{ref} and H_{ref} is the height, in which the influence from local disturbances on the wind speed can be regarded as negligible (see Figure 6.4). An important assumption here is that H_{ref} is constant and thereby independent of w_{ref}.

It should be noticed that this model does not cover the effect of changing wind directions. For a single windmill this is a serious shortcoming, whereas when considering all mills in an area, the effect of a change in wind direction to some extent will be averaged out[1] due to the different local conditions for each mill.

[1] As the windmills presumably are sited with consideration for the predominant wind direction, the averaging will only partly remove the effect of a change in wind direction.

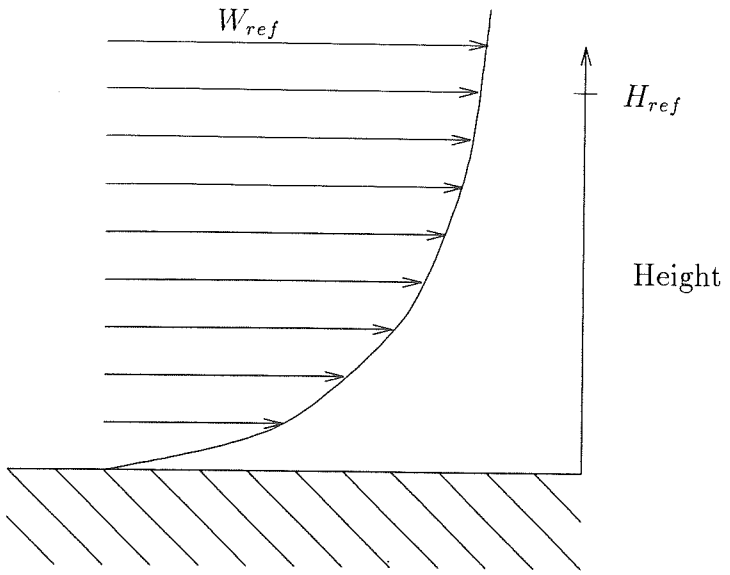


Figure 6.4: Wind Speed vs. Height

The power capacity for a windmill can with reason be assumed independent of the siting of the mill. Using (6.8) and (6.10) the power output from a single mill as a function of w_{ref} is then written as

$$P(t) = M \cdot \exp(-b \cdot \exp(-k \cdot \frac{h_{hub}}{H_{ref}} \cdot w_{ref}(t))) + e(t) \tag{6.11}$$

where h_{hub} is the height from ground to hub center (assumed less than H_{ref}) and $e(t)$ is a noise term. The relation between a measured wind speed in the sub area and the power output from a mill placed on a different location in the same sub area can now be expressed as

$$\begin{aligned} P(t) &= M \cdot \exp(-b \cdot \exp(-k \cdot \frac{h_{hub}}{H_{ref1}} \cdot w_{ref}(t))) + e(t) \\ w_{mea}(t) &= \frac{h_{mea}}{H_{ref2}} \cdot w_{ref}(t) + e(t) \\ \Rightarrow P(t) &= M \cdot \exp(-b \cdot \exp(-\tilde{k} \cdot w_{mea}(t))) + e(t)\ ; \quad \tilde{k} = k \cdot \frac{h_{hub} \cdot H_{ref2}}{h_{mea} \cdot H_{ref1}} \end{aligned} \tag{6.12}$$

where H_{ref1} and H_{ref2} are the local reference heights at the windmill, respectively, the wind speed measurement, and h_{mea} is the height from ground to the wind speed instrument. It is noticed that separating mill and wind speed measurement only affects the k parameter from (6.8), whereas the M and b parameters supplied by the mill manufacture still are valid.

If related measurements of wind speed and mill power are available, it is a trivial task to estimate $\tilde{k}$ (see Section 6.1), but for most of the windmills in a sub area only readings of the total energy supplied to the grid over a specific period (eg. 1 month) will be available.

The relation between a (monthly) energy reading and a time series of wind speed measure-

ments can be expressed as

$$\begin{aligned} \tilde{E} &= \sum_{i=1}^{N_t} P(i) \cdot T_s \\ &= T_s \cdot M \cdot \sum_{i=1}^{N_t} (\exp(-b \cdot \exp(-\tilde{k} \cdot w_{mea}(i))) + e(t)) \\ \Leftrightarrow \frac{\tilde{E}}{T_s \cdot M} &= \sum_{i=1}^{N_t} \exp(-b \cdot \exp(-\tilde{k} \cdot w_{mea}(i))) + e^{'}(t) \end{aligned} \tag{6.13}$$

where $\tilde{E}$ is the monthly energy reading, N_t is the number of samples in the current period and T_s is the sampling period. An estimate for $\tilde{k}$ can be found using an estimation technique as described previously.

6.2.2 Modeling an Entire Area

In this section it is assumed that it is possible to divide ELSAM's supply area into 7 sub areas each including one of the 7 wind farms in such a way that the wind speed in H_{ref} over the local wind farm can be regarded as being representative for the wind speed in the sub area.

This assumption will only be valid during stable or slowly changing weather conditions due to the large area covered by the wind measurement from each wind farm.

Furthermore, it is assumed that the sum of power curves from several windmills of different types and brands can be sufficiently represented by the power curve model used for a single windmill (See Equation (6.8)). The validity of this assumption depends on the windmill population, eg. if the population mainly consists of two mill types with very different power curves, the sum of the power curves cannot be approximated by (6.8), whereas the assumption will be much more reasonable in a population with more evenly spread power curve characteristics.

When the scope is changed from modelling the dependence between a sub area wind speed measurement and power output from individual windmills in this area to modelling the dependence between a sub area wind speed and the total power output from all mills in the area, a few modifications have to be applied to the scheme suggested in Section 6.2.1, as none of the M, b and k parameters used when calculating the power curve for the area are known exactly anymore.

On-line estimation of all the parameters in the non-linear model (6.8) might prove to be cumbersome and if any of the estimated parameters within reason can be assumed known the convergence properties for the parameter estimation in the summed power curves are likely to benefit from a reduction in the number of estimated parameters. The power capacity from a windmill can, as previously stated, with reason be assumed to be independent of the siting of the mill. This means that the power capacity from a group of windmills can be

written as

$$M_s = K_a \cdot \sum_{i=1}^{n} M_i \qquad 0.0 \leq K_a \leq 1.0 \tag{6.14}$$

where M_s is the total power capacity for the group, M_i is the power capacity for farm i and K_a is a factor taking into account that some of the mills in the group will be stopped for maintenance, etc. It is here suggested that the value of K_a is based on ELSAM's experience from their own wind farms.

Initially the two form parameters b and k have to be found off-line using least square estimation with a criterion as described in Section 6.1. By fixing b to the value estimated off-line the on-line estimation of the power curve will remain as in Section 6.2.1.

The advantage of this estimation scheme compared to on-line estimation of all three power curve parameters is that the shape of the power curve will remain well defined during on-line operation of the WPPT program.

6.2.3 Conclusion

In this section it is suggested that the estimation and prediction of the total power in the ELSAM supply area, as a first approach, are based on the wind speed measurements from the 7 wind farms.

An upscaling based on a scheme as outlined below has been described

- The ELSAM supply area has to be divided into 7 sub areas each including a wind farm in such a way that the wind speed measurement from the individual wind farms is representative for the sub areas.
- The monthly energy data from all installed mills in the ELSAM supply area has to be summarized for each of the 7 sub areas.
- An off-line estimation of the power curve parameters for each sub area has to be carried out before the installation of the WPPT program on site. The data used for the estimation is all periods where both wind speed and total power production measurements are available.
- The on-line estimation of the power curve parameters for each sub area will be based on data feed into the WPPT program once a month. Only a subset of the power curve parameters will be estimated on-line.

Chapter 7

Tools for Wind Power Prediction

This chapter describes both an on-line and an off-line system for predictions of the power production. The main part of the chapter is, however, devoted to a description of the main principles used in the on-line program system *Wind Power Prediction Tool* (WPPT). This system is implemented in the ELSAM supply area, and the hardware used from the measuring points to the dispatch center, is also described. In Section 7.4 an off-line version of the program is briefly described.

The purpose of WPPT is to provide on-line calculations and presentations of predictions of the power production in a region (like ELSAM) with a horizon ranging from 0.5 up to 36 hours. The predictions should be prepared for use in the production planning at a central dispatch unit. All the examples given in this chapter concerns the ELSAM case, but the system is built such that it rather easily can be changed for operating in another system.

The presentation part of WPPT (in the following denoted WPPT-P) is completely separated from the calculation/numerical part of WPPT (in the following denoted WPPT-N). The *presentation part* of WPPT is implemented as a menu driven graphical interface. The interface enables the operators to get on-line presentations of predictions as well as the observed (measured) values from the reference wind farms used in the calculation of the predictions.

The *numerical part* of WPPT is based on a modular system, where the individual modules can be combined in different ways to suit the needs of the system in question. Most of this chapter is devoted to a description of the principles employed in WPPT-P (Sec. 7.1) and WPPT-N (Sec. 7.2). A more thorough description of this software system is given in a set of manuals for WPPT (See (Nielsen & Madsen, 1995e), and (Nielsen & Madsen, 1995d)).

7.1 The Presentation Module (WPPT-P)

This section gives a short description of WPPT-P. WPPT-P is developed in the programming language C under UNIX (HP-UX). The graphical user interface is based on *Motif 1.1* and *"Version 11.4 of the X Window System"* from MIT.

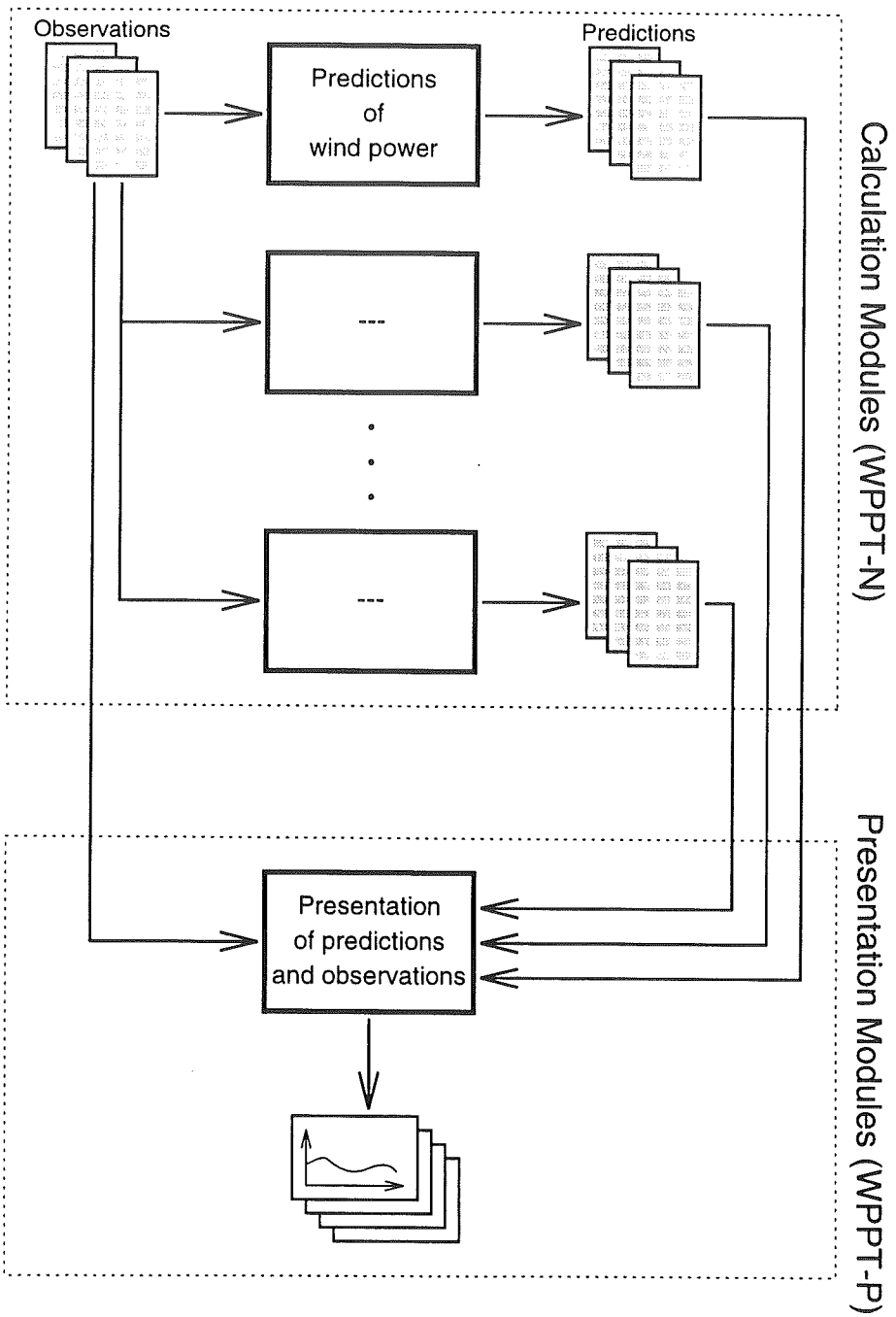


Figure 7.1: An overview of modules and the flow of data in WPPT.

Wherever it has been possible, the presentation module has been built using standard Motif elements, and it is assumed that the user of the program is familiar with the workings of a Motif user interface in general (menus, bottoms, labels, input-windows, etc.).

The observations of the wind speed and the power production at the seven wind farms are available in ASCII files as half hour averages. In WPPT-P it is possible simply to have a graphical on-line presentation of these observations as well as an on-line presentation of various prediction related plots. Fig. 7.2 shows the main window of the program with a short description of its different elements.

7.1.1 The Main Window

The important parts of the main window are the menu-bar and the value fields with observations and predictions. All other parts are permanent labels and descriptions. As it is seen from the figure, the following wind farms are considered: Brøns, Dræby, Vedersø Kær, Nørrekær Enge, Hollandsbjerg, Torrild og Ryå .

The Menu Structure

The menu bar provides the possibility of 3 alternative choices of menus: "Observations", "Predictions", and "Default-Display". Under these menus further sub-menus are found, as illustrated in Fig. 7.3. Sec. 7.1.2 shows window examples from all the sub-menus found in Fig. 7.3. The examples cover plot windows (output) as well as dialogue windows (input/output).

The Current Value Fields

For the purpose of a possible supervision, the observations are shown in value fields in the main menu. If a value is found to be invalid [1] the background of that value field changes from green to red. For all the wind farms both the wind speed and the power production are shown as indicated in Fig. 7.2. Furthermore, the predictions for selected horizons between 0.5 and 36 hours are shown. All the displayed values are updated every half hour.

7.1.2 Plots Provided by WPPT-P

The figures in this section contain examples of plot windows and dialogue windows in WPPT-P. The menu choice number in the figure captions corresponds to the numbered choice of menus in Fig. 7.3.

The depicted windows can be separated into two groups; the plot and dialogue windows associated with the observations from the reference wind farms (Fig. 7.4 to Fig. 7.6), and

[1] A value is declared invalid, if it violates some predefined limits, eg. a high limit

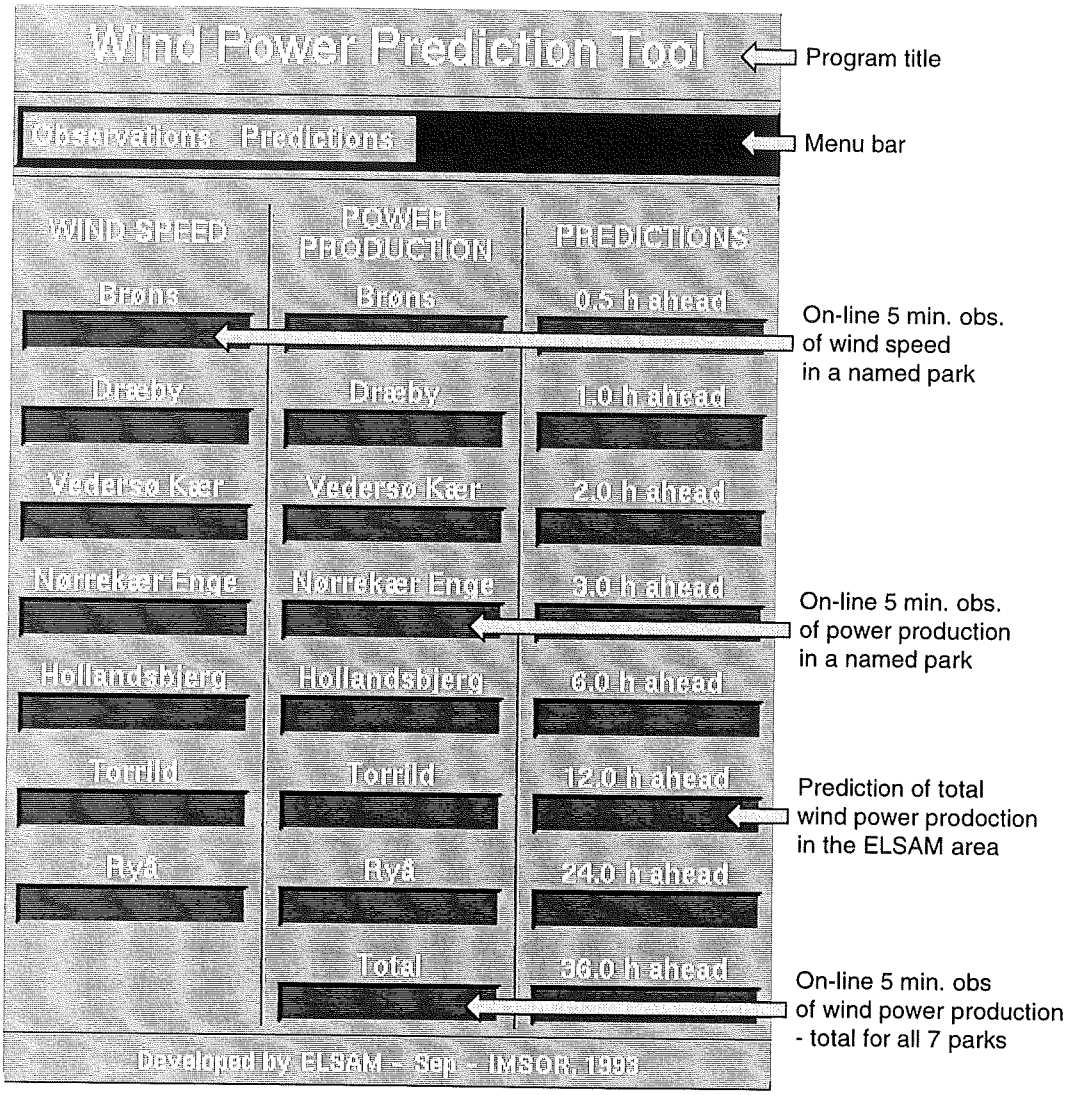


Figure 7.2: Main window in Wind Power Prediction Tool. (The arrows and the text on the right is not a part of the window – but serves as explanation.)

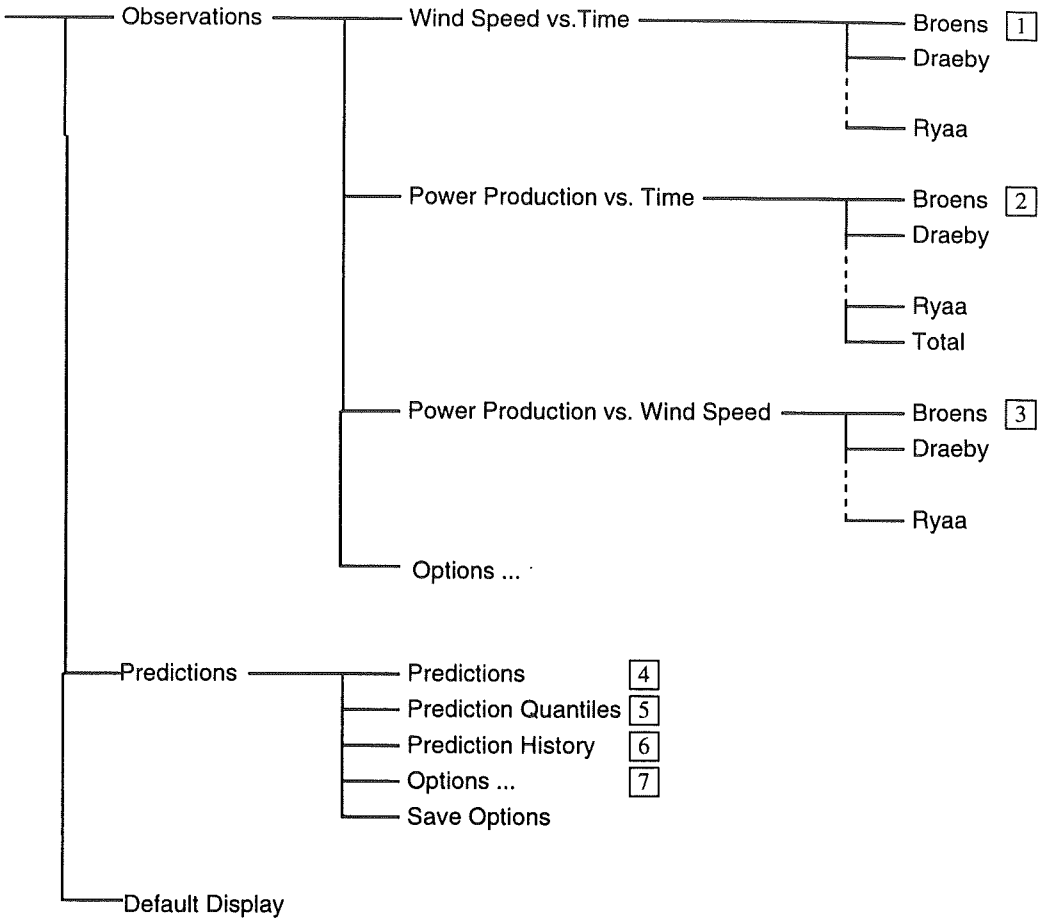


Figure 7.3: *Menutree in Wind Power Prediction Tool. The numbers given in* ☐ *are used for references.*

the plot and dialogue windows associated with the predictions (Fig. 7.7 to Fig. 7.10). The rest of this section contains a short description of the windows in each group.

Observation Windows

The wind observation windows (7 in total) plot wind speed versus time for each reference wind farm. The wind speed values displayed are 30 minutes average values and the plots cover the last 4 days of observations (192 observations). In the power observation windows (7 in total) the observed power and estimated expected power[2] are depicted versus time for each reference wind farm. The curves for observed and expected power are based on 30 minutes average values and the plots cover the last 4 days (192 observations).

Additionally the power observation plot windows contain 2 text fields below the actual plots

[2]The expected power is calculated using a gompertz representation of the representative power curve for each wind farm.

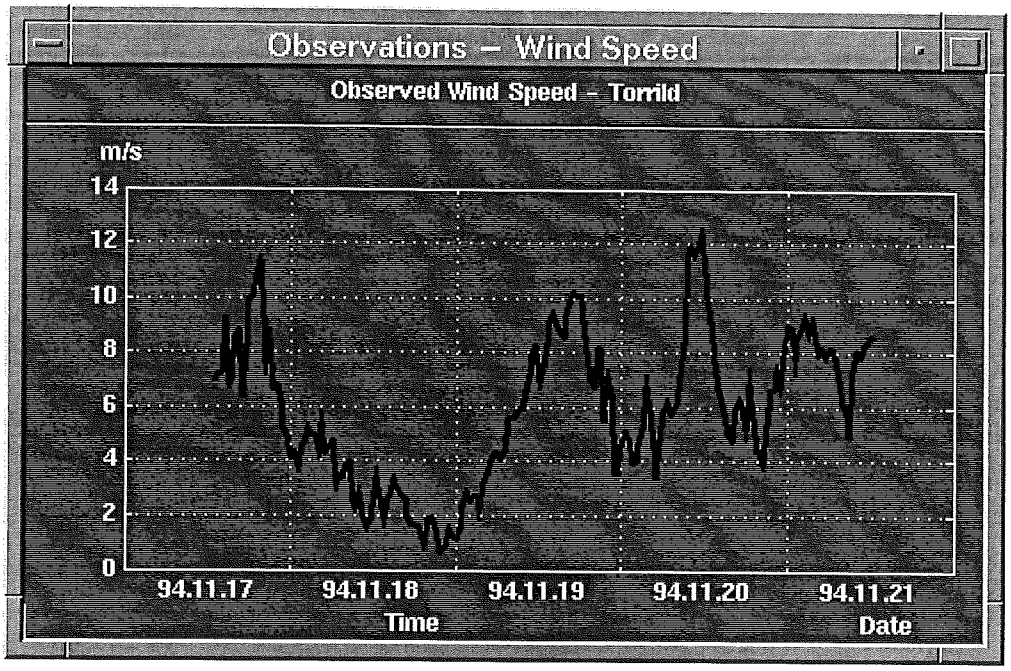


Figure 7.4: *Menu choice* [1] *in Figure 7.3.*

displaying the estimated currently running power (i.e. the number of running wind turbines as a percentage) in the wind farm as well as the number of MW represented by the wind farm in the up-scaling.

The total power window (not shown) plots the "observed" power for the ELSAM supply area versus time. The curve is based on an up-scaling of the 30 min. average values for the reference wind farms.

The power versus wind windows (7 in total) plot observed power and expected power versus observed wind speed for each reference wind farm. The plots are based on 30 minutes average values for power and wind displayed as dots and the plots cover the last 4 days of observation. The expected power is drawn as an estimated power curve using a gompertz representation of the power curve.

Some observations dialogue boxes (one for each farm) are introduced for changing the installed power for each of the reference wind farms. If the installed power is set to 0.0 kW the wind farm is disregarded in the up-scaling as described in (Nielsen & Madsen, 1995e).

Prediction Windows

The prediction window plots the power predictions for the ELSAM supply area versus time. The plot is updated every 30 minutes. The plot contains 4(3) curves:

- The "observed" power in the ELSAM supply area 3 hours back in time.

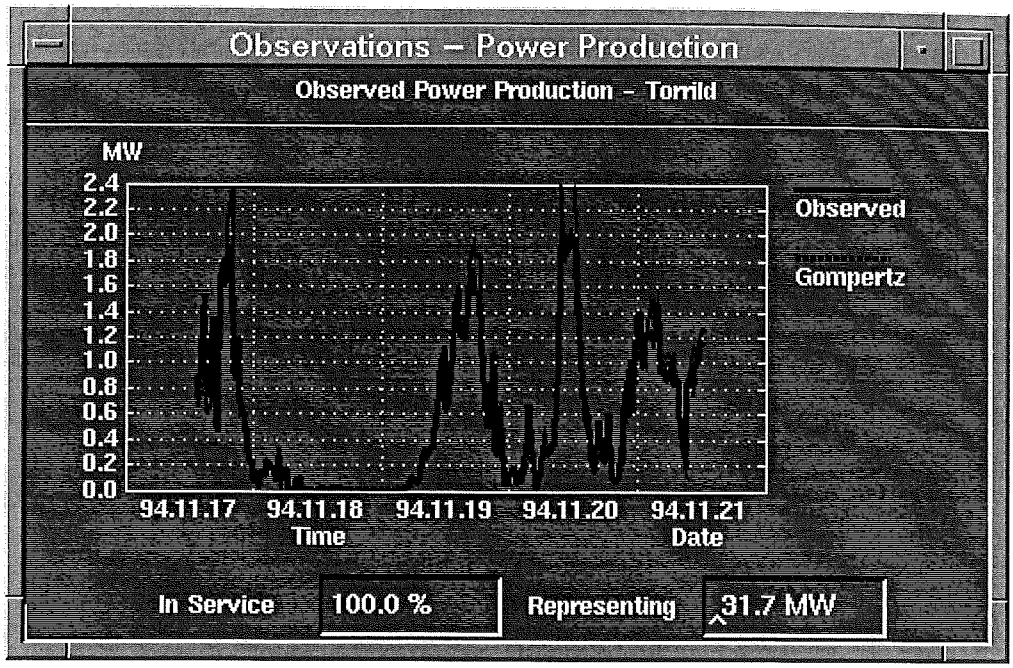


Figure 7.5: *Menu choice* [2] *in Figure 7.3.*

- The power prediction for the ELSAM supply area with a configurable prediction horizon (configurable between 30 minutes and 36 hours) given as the 50% quantile in the distribution for the predictions.
- The minimum expected power in the ELSAM supply area as a configurable quantile in the distribution of the predictions (configurable between 1% and 20%).
- The maximum expected power in the ELSAM supply area as a configurable quantile in the distribution of the predictions (configurable between 99% and 80%).

The historical prediction window shows the historical power predictions for the ELSAM supply area for a selected prediction horizon (configurable between 30 minutes and 36 hours) and the observed power in the ELSAM supply area versus time thereby enabling a visual evaluation of the quality of the predictions.

The quantile window plots quantiles in the distribution for the current power predictions in the ELSAM supply area versus time where the curve for the quantile selected as the minimum expected power in the prediction plot is drawn in a different colour. The plot covers the same prediction horizon as selected for the prediction plot and is updated every 30 minutes. The prediction dialogue box is used to change the configurable settings in the prediction displays. The dialogue box manages the different display settings through the following controls:

- A push button activated menu labelled "Main Prediction Quantile" used for selecting the quantile in the distribution of the power predictions used as minimum expected

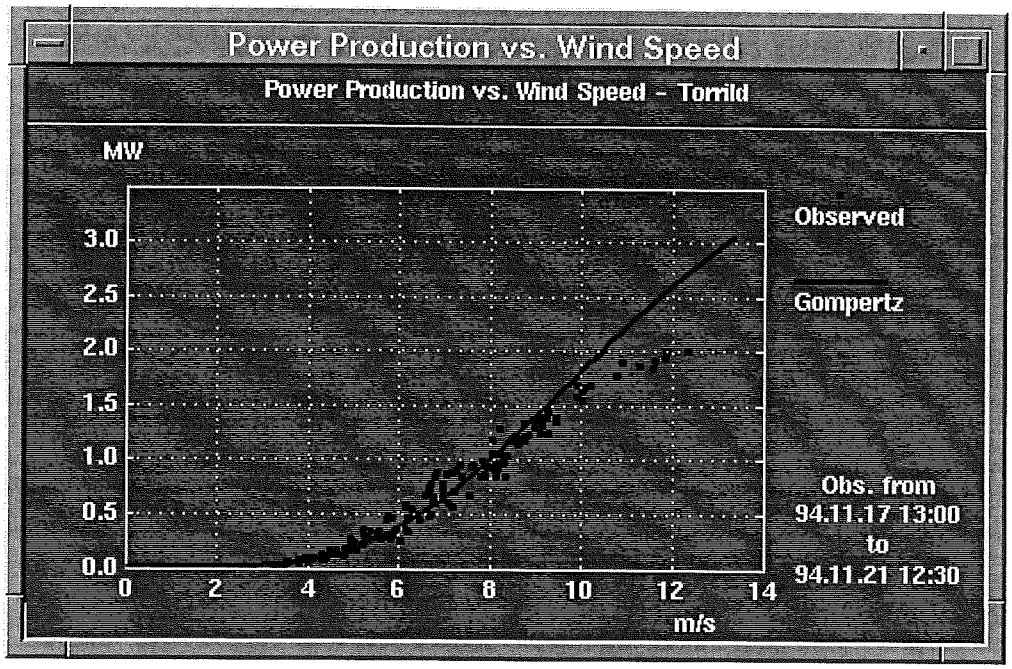


Figure 7.6: *Menu choice* [3] *in Figure 7.3.*

power in the prediction plot. The following percentages are selectable: 1.0%, 5.0%, 10.0%, 20.0%, and 50.0%.

- A slider labelled "Maximum Prediction Horizon" used to select the prediction horizon in the prediction plot. The selectable prediction horizon range between 1 step ($\approx$ 30 minutes) and 72 steps ($\approx$ 36 hours).
- A slider labelled "Horizon in Prediction History" used to select the prediction horizon displayed in the historical prediction plot. The select-able prediction horizon range between 1 step ($\approx$ 30 minutes) and 72 steps ($\approx$ 36 hours).

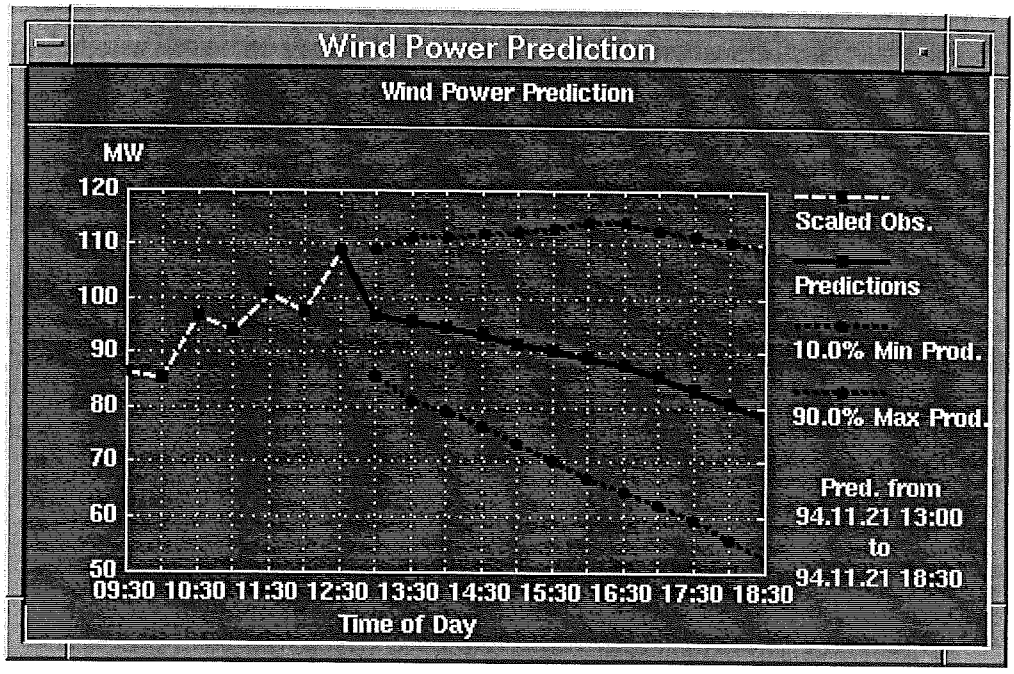


Figure 7.7: *Menu choice* [4] *in Figure 7.3.*

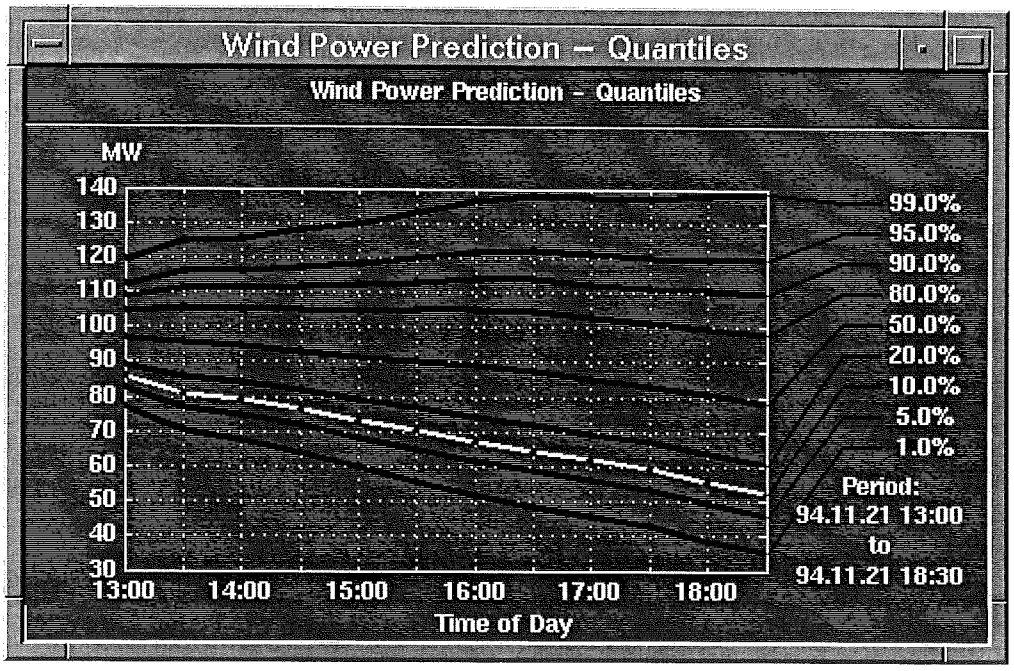


Figure 7.8: *Menu choice* [5] *in Figure 7.3.*

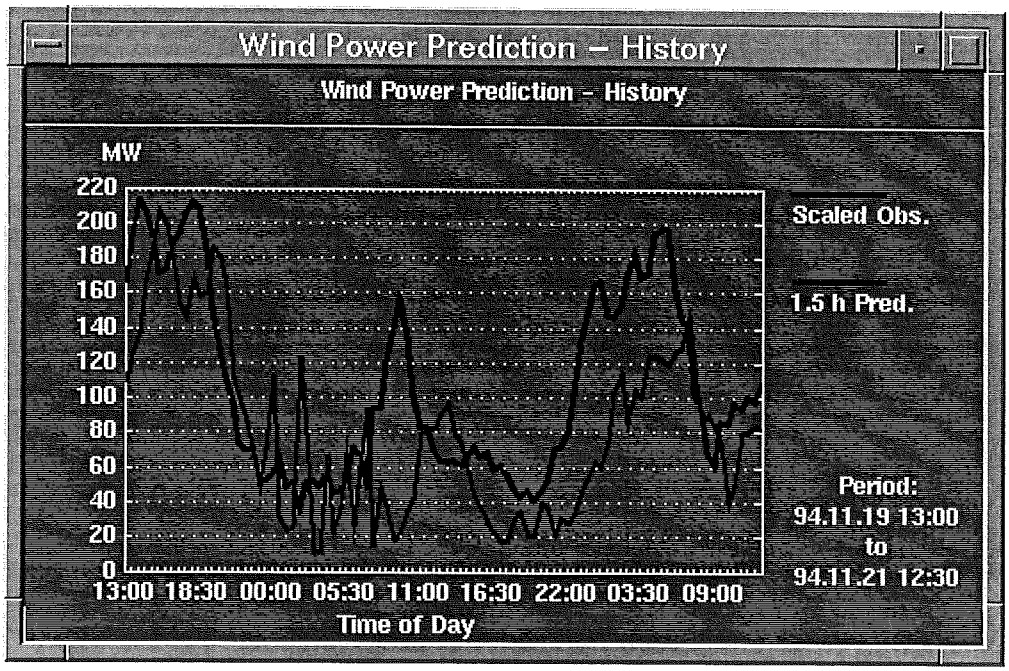


Figure 7.9: *Menu choice* [6] *in Figure 7.3.*

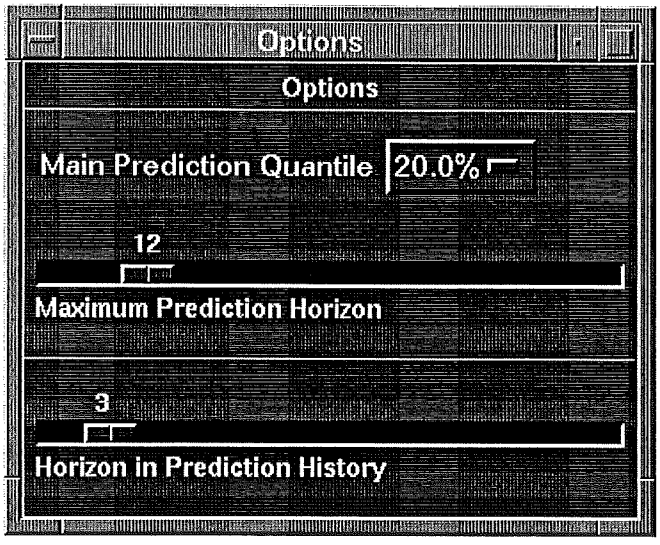


Figure 7.10: *Menu choice* [7] *in Figure 7.3.*

7.2 The Calculation Module (WPPT-N)

7.2.1 The Basic Program Layout

The calculation module (WPPT-N) has been implemented using a modular programming technique. Each functionality has been placed in its own module and all data exchange between the modules is done through a database. The program structure employs two

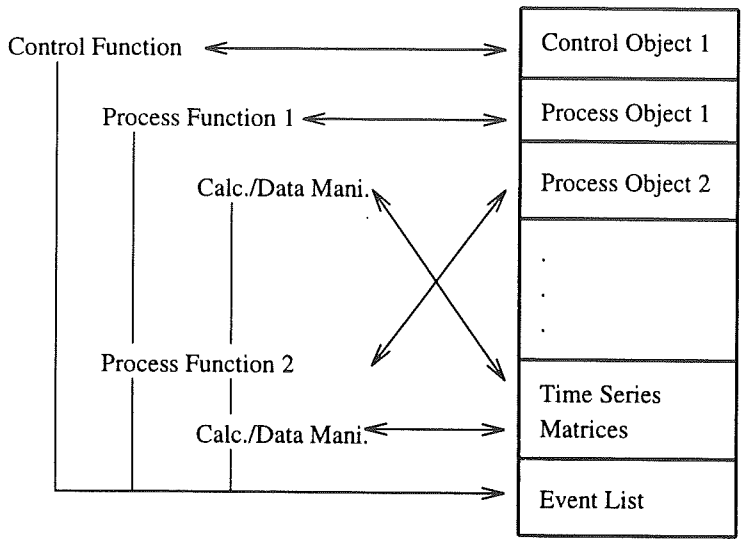


Figure 7.11: Main program structure

function templates, a control function template and a process function template, both with associated data structures (see Fig. 7.11).

The control function takes care of the time handling necessary in a real time system with a fixed execution frequency. Furthermore, the execution order for the process functions is determined by the control function. The associated data structure is called a control object.

All calculations and data manipulations are carried out by the process functions. Before any data manipulations take place a local copy of the necessary data objects [3] is made from the database. All data manipulations are then carried out on the local copies and when finished, the changed data objects are transferred back into the database. The data structure associated with the process function is called a process object.

Both control functions and process functions have access to a common event message list where all incidents of interest to the operators are logged thereby enabling the operators to trace the operation of the system.

[3] Data objects are of type time series or matrix

7.3 Hardware Used from the Measuring Points to the Load Dispatch Centre

This section describes briefly the hardware used in the ELSAM implementation of the wind power prediction tool. It describes the hardware used from the measuring points at the wind farms to the hardware selected at the load dispatch centre.

At the seven reference wind farms in the ELSAM supply area (see Figure 2.1) the wind speed and the power production are measured on-line.

7.3.1 Measurement and Data Transmission Equipment

A data collection device has been installed locally in each wind farm, and the device has been connected to the public alarm network. All data are transmitted via this network to the ELSAM load dispatch center.

Wind speed and wind power production are measured continuously in the wind farms. Both measurements are given as analogous signals (0 - 20 mA). The signals are fed into the local data collection device. The device prepares the data as five-minute integrated values which are transmitted to ELSAM. The integrated values are saved locally for the next 30 hours. It means that data can be retransmitted, if necessary.

7.3.2 Hardware in the Load Dispatch Centre

The transmitted data are collected at the ELSAM load dispatch centre. The data collection system is a Pentium PC running OS/2. The communication between the PC and the public alarm network is established as a 1200 Bd modem connection. The integrated five-minute values are packed and saved at the ELSAM data network.

The wind power prediction tool itself is running on a VAX computer using the data saved on the ELSAM network for the final prediction calculations.

7.4 Off-line Wind Power Prediction Software.

An off-line version of the software, which only includes the estimation and prediction part of the on-line version, has been developed. Fore more detailed information about the program the reader is referred to (Nielsen & Madsen, 1995b) and (Nielsen & Madsen, 1995a).

The software is written in ANSI C and has been tested under the UNIX systems: HP/HP-UX, DEC/OSF1, IMB/AIX, SGI/IRIX, and on a 486 PC running Linux. Furthermore, the program has been tested under DOS. Due to the limitations of this system the software cannot handle very large amounts of data when running under DOS. The program is running from the command line of a standard ASCII-oriented terminal or window.

Chapter 8

Practical Experience

In this chapter some practical experience of the implementation of the on-line version of the Wind Power Prediction Tool (WPPT) at ELSAM is stated. A separate section describes the practical value of the tool for load dispatching. Finally, a section describing some practical experience of the use of the off-line version of WPPT at Sep is included. The figures of this section are illustrative for the general behavior of the predictions.

8.1 On-line Wind Power Prediction

The separation of the software in a calculation module and a presentation (graphics) module has been very useful. WPPT was originally developed on a UNIX computer; but at ELSAM it was decided to implement WPPT on a VAX computer. The implementation of the calculation module on the VAX computer went well while only parts of the presentation module are implemented.

It has been quite difficult to obtain an acceptable quality of the on-line measurements of power production and wind speed in the seven reference wind farms. This is primarily due to the quality of the measuring equipment. However, also organisational matters play a role, since it has been experienced that maintenance and repair of the wind farms done by the wind farm operators are of vital importance to a prediction tool.

The problems with the on-line measurements were partly foreseen. Therefore methods for taking into account missing data and errors in the data are developed during the project. Software modules are implemented in WPPT so that errors are automatically detected and corrected. This is further described in the manual (Nielsen & Madsen, 1995d).

It is concluded that new equipment must be installed in the future and that the measurement points should be moved to the accounting points in the electrical network. The power production can easily be measured here, but more reliable wind speed equipment must be installed close to the accounting points.

8.1.1 Value for Load Dispatching

Due to the relatively high amount of energy produced by wind mills in the ELSAM supply area a tool is needed for short term predictions of the wind power production at the load dispatch center. Furthermore, the need will be even more prominent in the future due to the decision that the amount of wind energy should be more than doubled during the next five years (see Table 1.1).

Untill now it has been difficult to judge the economic value due to the previously mentioned equipment problems. It is clear that high quality equipment must be installed once the wind power prediction tool is to go into normal commercial operation.

The WPPT tool has improved the knowledge of the operators concerning the behaviour of wind power due to the systematic collection and evaluation of wind speed data and power production data. This information is continuously available to the operators.

For the moment, when on-line meteorogical forecasts are not available, it seems as if the 36 hour prediction horizon is too ambitious, because the information value of the prediction 36 hours ahead is very small. So it has been decided to concentrate on prediction with a 12 hour horizon. The evaluation from the operations department points to the fact that the weekly load dispatching will not change due to the use of an even very good prediction model.

The economic benefit will come from the situation with normal operation conditions where a good wind power prediction with a e.g. 12 hours horizon will enable the operators to take into account the wind production on beforehand instead of regulating the running units as a consequence of an experienced wind production. Hence the economic value of a prediction tool is found on the short term horizon.

Economic benefits are also coming from failure situations where a power plant unit is lost, e.g. Friday afternoon. If the prediction of wind power is sufficient and stable for the next 6 or 8 hours the operators might decide not to start up a reserve unit.

The operators conclude that the development of the WPPT system must be continued and that meteorological forecasts should be included into the WPPT model.

8.2 Off-line Software Used on Dutch Data

The off-line version of the wind power prediction tool was applied to data collected at Sep's experimental wind farm at Sexbierum in the Northern part of the Netherlands. The total installed capacity equals 5400 kW. The data were collected from March 2nd until June 1st 1992. The sampling interval is 10 minutes. Three values were considered corrupt, these are replaced by values found by interpolation. The range of the wind speed measurements is 0.1 – 25.5 m/s, with 50% of the values ranging from 5.3 to 10.1 m/s.

The off-line wind power prediction tool is used to calculate one and six hour predictions based on hourly averages. A forgetting factor of 0.999 is used. The prediction results are shown on Figure 8.1.

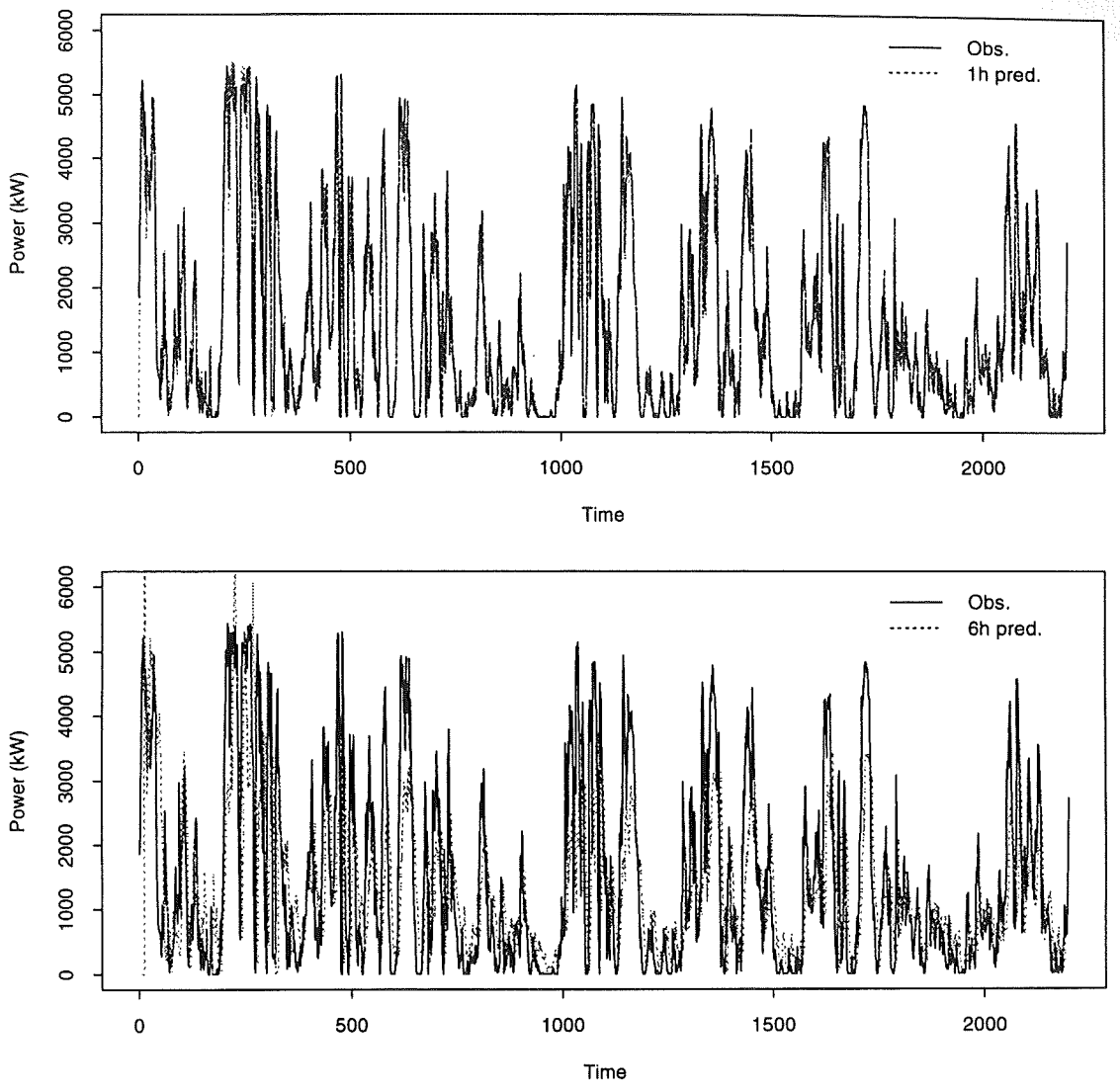


Figure 8.1: One and six hour prediction of wind power production for the period March 2nd until June 1st 1992, versus No. of hours from the start of March 2nd 1992.

It is noted that the six hour prediction shows large fluctuations in the beginning of the period - the maximum value predicted is 11114 kW. This phenomenon is expected since the parameter estimates have to reach reasonable values before the predictions become appropriate. On the figures it is clearly seen that the prediction procedure gives most weight to the present observation when predicting the future value of the wind power production.

The Root Mean Square (RMS) of the prediction error when skipping the first 500 values is calculated. For the one hour predictions an RMS of 412 kW is obtained compared with 417 kW for the persistent predictor. The values for the six hour predictions are 1042 kW and 1133 kW, respectively. For one hour predictions the persistent predictor performs about 1% worse than the implemented predictor, for six hour predictions the persistent predictor performs 9% worse than the implemented predictor. The corresponding values calculated from table 4.7 in Chapter 4 are 2% and 7%, respectively.

References

Abraham, B. & Ledolter, J. (1983). *Statistical Methods for Forecasting.* Applied Probability and Statistics. Wiley & Sons, New York.

Bard, Y. (1974). *Nonlinear Parameter Estimation.* Academic Press, London.

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction.* Applied Probability and Statistics. Wiley & Sons, New York.

Box, G. & Jenkins, G. (1976). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

ELSAM (1989). *Integration of Wind Power in the Danish Generation System.* EN3W-0057-DK. ELSAM, Denmark.

ELSAM (1992). *European Wind Power Integration Study.* JUR-0041-DK. ELSAM, Denmark.

Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman filter.* Cambridge University Press, Cambridge.

IMM, ELSAM, & SEP (1995). *Wind Power Prediction Tool in Control Dispacth Centres.* ELSAM, Project Joule II (JOU2-CT92-00XX) ISBN: 87-87090-42-2, Skærbæk, Denmark.

Kaiser, J. & Reed, W. (1977). Data Smoothing using Low-pass Digital Filters. *Rev. Sci. Instrum.*, **48**(11), 1447–1457.

Landberg, L., Watson, S., Halliday, J., Jørgensen, J., & Hilden, A. (1994). *Short-term Prediction of Local Wind Conditions.* JOUR-0091-C(MB). Risø, Roskilde, Denmark.

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

Ljung, L. (1987). *System Identification, Theory for the User.* Prentice-Hall, Englewood Cliffs, New Jersey.

Madsen, H. & Holst, J. (1995). *Modelling Non-Linear and Non-Stationary Time Series.* Lecture Notes. IMM, DTU, Lyngby, Denmark.

Madsen, H. (1985). *Statistically Determined Dynamical Models for Climate Processes.* Ph.D. thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Madsen, H. (1989). *Time Series Analysis* (Preliminary edition). Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Melgaard, H. & Madsen, H. (1991). The Mathematical and Numerical Methods used in CTLSM. Tech. rep. 7, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Nielsen, H. A. & Madsen, H. (1995a). *Adaptive Recursive Least Squares Algorithm Used in Wind Power Prediction Tool.* Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Nielsen, H. A. & Madsen, H. (1995b). *Off-line Wind Power Prediction Tool - Users Manual.* Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Nielsen, H. A. & Madsen, H. (1995c). *Wind Power Prediction Using Neural Networks.* Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Nielsen, T. S. & Madsen, H. (1995d). *Technical Manual for the program "Wind Power Prediction Tool".* Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Nielsen, T. S. & Madsen, H. (1995e). *User Manual for the program "Wind Power Prediction Tool".* Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Nielsen, T. & Madsen, H. (1995f). Using Meteorological Forecasts in On-line Predictions of Wind Power. Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Priestley, M. (1988). *Non-linear and Non-stationary Time Series Analysis.* Academic Press, London.

Rao, C. (1965). *Linear statistical inference and its applications.* John Wiley, New York.

Ripley, B. D. (1994). Software for Neural Networks.. Statlib Index (`lib.stat.cmu.edu`), S library, `nnet` package. Updated 24 January 1993, 19 September 1993, 13 February 1994.

Robinson, P. (1983). Nonparametric Estimators for Time Series. *Journal of Time Series Analysis*, **4**(3), 185–207.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6**(2), 461–464.

Söderström, T. & Stoica, P. (1989). *System Identification.* Prentice Hall, New York.

Tong, H. (1990). *Non-linear Time Series, A Dynamical System Approach.* Clarendon Press, Oxford.