SHORT–TERM WIND POWER PREDICTION

Alfred Joensen

Department of Mathematical Modelling
Technical University of Denmark
Ph.D. Thesis No. 108
Kgs. Lyngby 2002

IMM

ISSN 0909-3192

This document was prepared with LaTeX and printed by IMM, DTU, Lyngby.

Preface

This thesis has been written during my work as a Ph.D. student at the Department of Wind Energy and Atmospheric Physics, Risø National Laboratory and the Department of Informatics and Mathematical Modelling, Technical University of Denmark in partial fulfillment of the requirements for acquiring the Ph.D. degree in Engineering.

The aim of the Ph.D. study has been to develop short-term wind power prediction models, and to implement these models in an on-line software application. In the model development the emphasis is on combining physical knowledge and statistical models and methods. As the models are to be implemented in an on-line application, the focus is on solutions which are reliable and practically feasible.

The thesis consists of a summary report and a collection of 10 research papers written during the period 1997–2002, and elsewhere published.

Lyngby, May 2002

Alfred Joensen

Acknowledgements

In carrying out the work described in this thesis I have received important assistance from many people. First of all I want to address my sincere gratitude to my two supervisors, supervisor Prof. Henrik Madsen from the Department of Mathematical Modelling, Technical University of Denmark and co-supervisor Ph.D. Lars Landberg from the Department of Wind Power Meteorology, Risø National Laboratory. Thanks for your invaluable help and guidance.

Thanks are also due to my colleagues and the administrative staff at the above mentioned departments for their invaluable cooperation, help, and discussions. Especially, I would like to thank the Ph.D. students Gregor Giebel, Peter Thyregod, Klaus Andersen, Harpa Jonsdottir and Henrik Oejlund, with whom I have shared office during longer and shorter periods during the course of my Ph.D.

I wish to thank Torben Skov Nielsen and Henrik Aalborg Nielsen, both working at the Department of Mathematical Modelling. Several of the papers in this thesis have been prepared in cooperation with them, and this has provided many interesting and invaluable discussions.

Also I wish to thank Prof. Edgar O'Hair, and the rest of the staff at the Department of Electrical Engineering, Texas Tech University, where I spent six months as visiting researcher. Thanks for making this period such a pleasent one.

Furthermore, I am truly indebted to the Danish Research Academy, for supporting my work financially.

Finally, a huge thanks goes to my wife, Oluva, our two daugthers, Beinta and Anna, and our son, Andreas, for putting up with a husband and father, who spent most of the last three years in a book or in front of a computer, and still provided an endless amount of love and patience.

Papers included in the thesis

[A] Alfred Karstin Joensen, Henrik Madsen, Henrik Aalborg Nielsen and Torben Skov Nielsen. Tracking time-varying parameters with local regression. *Automatica*, Vol **36**, pages 1199–1204. 2000.

[B] Henrik Aalborg Nielsen, Torben Skov Nielsen, Alfred Karstin Joensen, Henrik Madsen and Jan Holst. Tracking time-varying coefficient-functions. *Int. J. of Adaptive Control and Signal Processing*, 2000. Accepted.

[C] Torben Skov Nielsen, Alfred Karstin Joensen, Henrik Madsen, Lars Landberg and Gregor Giebel. A new reference for wind power forecasting. *Wind Energy*, Vol **1**, pages 29–34, 1999.

[D] Alfred Karstin Joensen, Torben Skov Nielsen and Henrik Madsen. Statistical methods for predicting wind power. In *Wind Energy for the Next Millenium*, European Wind Energy Conference, pages 784–788, Dublin, Ireland, October 1997.

[E] Alfred Karstin Joensen, Gregor Giebel, Lars Landberg, Henrik Madsen and Henrik Aalborg Nielsen. Model output statistics applied to wind power prediction. In *Wind Energy for the Next Millenium*, European Wind Energy Conference, pages 1157–1161, Nice, France, March 1999.

[F] Lars Landberg and Alfred Karstin Joensen. A model to predict the output from wind farms – an update. In *proceedings from BWEA 20*, British Wind Energy Conference, pages 127–132, Cardiff, UK, 1998.

[G] Lars Landberg, Alfred Karstin Joensen, Gregor Giebel, Henrik Madsen and Torben Skov Nielsen. Short-term Prediction towards the 21st Century. In *proceedings from BWEA 21*, British Wind Energy Conference, pages 127–136, UK, 2000.

[H] Lars Landberg, Alfred Joensen, Gregor Giebel, Simon Watson, Henrik Madsen, Torben Nielsen, Leif Laursen, J. U. Jørgensen, Dimitrios Lalas, Maria Trombou, S. Pesmajoglou, John Tøfting, Hans Ravn, Ed MacCarty, Earl Davis and Jamis Chapman. Implementation of Short-term Prediction. In *Wind Energy for the Next Millennium*, European Wind Energy Conference, pages 52–57, Nice, France, March 1999.

[I] Alfred Joensen. HIRLAM - Analysis of vertical model levels. *Submitted for publication.*

[J] Simon J. Watson, Gregor Giebel and Alfred Joensen. The Economic Value of Accurate Wind Power Forecasting to Utilities. In *Wind Energy for the Next Millennium*, European Wind Energy Conference, pages 1009–1012, Nice, France, March 1999.

Summary

The present thesis consists of 10 research papers published during the period 1997–2002 together with a summary report. The objective of the work described in the thesis is to develop models and methods for calculation of high accuracy predictions of wind power generated electricity, and to implement these models and methods in an on-line software application. The economical value of having predictions available is also briefly considered.

The summary report outlines the background and motivation for developing wind power prediction models. The meteorological theory which is relevant for the thesis is outlined and the background for the models and methods which are proposed in the various papers is described. The software system, Zephyr, which has been developed is also described in the summary report.

The main part of the papers have been written in conjunction with two research projects where the Department of Informatics and Mathematical Modelling and the Department of Wind Energy and Atmospheric Physics have been two major participants. The first project entitled 'Implementing Short-term Prediction at Utilities', founded by the European Commission under the JOULE programme. The second project is founded by the Danish Ministry of Energy under the Energy Research Programme, and is entitled (in Danish) 'Vindmølleparks Produktions Prediktor'. Both projects have now finished.

The papers A and B are related to general issues in modelling and estimation. Paper A considers on–line estimation of linear models, where the parameters to be estimated exhibit smooth time variations. An estimation method derived from local polynomial regression is suggested, using local polynomials in the direction of time to approximate the parameters locally. The results presented in the paper indicate that the method is superior to the classical Recursive Least Squares (RLS) method, if the parameter variations are smooth. In paper B a method for on-line and adaptive estimation of Conditionally Parametric Auto-Regressive eXtraneous (CPARX) models is derived, and some of the properties of the method are analyzed. This method can be interpreted as recursive local regression. Essentially it is a combination of the RLS method with exponential forgetting and local polynomial regression. Furthermore, the paper suggests a modification of the exponential forgetting scheme of the RLS method, to cope with the added complexity, which is introduced by allowing the parameters to be functions of other variables than just time.

The Papers C to I are all related to short-term prediction of wind power. In Paper C a new reference for short-term prediction models is proposed, and it is argued that the new reference model is more suitable than the often used persistence predictor, especially if the prediction horizon is above a few hours. The new reference model is almost as simple as the persistence predictor, basically, it is a prediction horizon dependent weighting between the persistence and the mean of the power, where the weighting is determined by the auto-correlation of the wind power time series. In Paper D conditional parametric models estimated using local regression are used to identify important explanatory variables in short-term prediction models. These models are estimated using off-line techniques. In paper E similar models are considered, but in this paper the new recursive estimation method described in Paper B is used to estimate the functional shapes of the coefficient functions in the conditionally parametric models. This paper also presents some models where physical relations are incorporated in the models, and the performance of these models is compared to the performance of models where no direct physical relations are used. The result from this comparison is that it is not advantageous to use the physical relations. Paper I compares wind speed predicted by the numerical weather prediction model HIRLAM, to measured wind speeds at different heights. The purpose of this comparison is to analyse the influence of the turbulence intensity and how the

turbulence is handled by HIRLAM.

Paper F gives an overview of the short-term prediction system, Prediktor, developed at Risø National Laboratory. Some simple statistical correction models are tested, which correct the predictions from the physical relations used by the Risø system. In this paper a change in the properties of the predictions from the numerical weather prediction model is observed. Usually prediction performance is measured by e.g. the Root Mean Square error (RMS). This paper gives a more direct picture of the performance, by using combined time series plots of measurements and corresponding predictions. From these plots it can be verified that the overall flow is predicted well by the numerical weather prediction model.

Paper G briefly outlines the reason for considering combined statistical and physical models. The architecture of the software application, Zephyr, is also briefly outlined. This paper also proposes models for the purpose of calculating short-term predictions covering a larger area. The background for the models which are proposed, is that it is expected that power measurements from a larger area covering several wind farms, will be smoother than the measurements from a single wind farm, due to spatial averaging effects. The models have not been tested, as no measurements of total wind power have been available for this work.

Paper H describes the results from the 'Implementing Short-term Prediction at Utilities' project. In this paper both the prediction system developed at The Department of Informatics and Mathematical Modelling and the system developed at Risø National Laboratory are outlined, and the performance of the systems is evaluated. Experience from the use of these two systems at utilities is also provided, and the utilities find that both systems are very useful.

The last paper addresses the economical value of short-term predictions. Predictions from several prediction models are used as input to a model of the England/Wales electrical grid, and it is found that for low penetration of wind energy, predictions have little value. As the penetration increases the predictions and their accuracy of the become more important, and it is also shown that confidence limits for the predictions can increase the economical value of the predictions.

Resumé

Nærværende afhandling består af ti forskningsartikler publiceret i perioden 1997-2002, samt af et sammendrag heraf.

Formålet med arbejdet, der er beskrevet i denne afhandling, har været at udvikle modeller og metoder til beregning af kortfristede prognoser af vindmølleproduceret elektricitet. Derudover har det været at implementere de udviklede modeller og metoder i en on-line softwareapplikation. Den økonomiske værdi af disse forudsigelser er også blevet undersøgt.

Afhandlingen indeholder en resumé rapport, der beskriver baggrunden og motivationen for at udvikle prognosemodeller for vindenergi. Den meteorologiske teori, der er relevant for afhandlingen, er beskrevet og baggrunden for de modeller og metoder der er foreslået i artiklerne, er gennemgået. Softwareapplikationen, Zephyr, der er udviklet som en del af projektforløbet, er også beskrevet.

Hovedparten af artiklerne er skrevet i forbindelse med to forskningsprojekter, hvor Institut for Informatik og Matematiske Modellering, samt Afdelingen for Vindenergi og Atmosfærefysik har været de største deltagere. Det første projekt, kaldet 'Implementing Short-term Predictions at Utilities', er finansieret af Europa Kommissionen, under JOULE programmet. Det andet projekt er finansieret af det danske energiministerium under energiforskningsprogrammet og kaldes 'Vindmølleparks

Produktions Prediktor'. Begge projekter er nu afsluttede.

Artiklerne A og B omhandler generelle forhold i forbindelse med modellering og estimation. Artiklen A omhandler on–line estimation af lineære modeller, hvor de estimerede parametre udviser langsomme tidsvariationer. En estimationsmetode, udledt fra lokal polynomieregression er foreslået, der anvender lokale polynomier i retning af tidsvariablen til lokal tilnærmelse af parametrene. Resultaterne præsenteret i artiklen indikerer, at denne metode giver bedre resultater end den klassiske adaptive rekursive mindste kvadraters metode (RLS), forudsat at parametervariationerne er glatte. I artiklen B undersøges en metode til on–line og adaptiv estimation af betingede parametriske autoregressive modeller, hvor der indgår eksterne forklarende variable. Derudover er nogle af metodens egenskaber undersøgt. Denne metode kan fortolkes som rekursiv lokal regression, specielt som en kombination af den klassiske mindste kvadraters metode med eksponentiel glemsel og lokal polynomie regression. Endvidere foreslås der i artiklen, at den eksponentielle vægtfunktion i den traditionelle adaptive mindste kvadraters metode modificeres. Dette for at håndtere den øgede kompleksitet, der introduceres ved at parametrene tillades at være funktioner af flere variable end tidsvariablen.

Artiklerne C til I omhandler alle kortfristede prognoser af vindenergi. I artikel C præsenteres en ny referencemodel for kortfristede prognosemodeller, og der argumenteres for, at den nye referencemodel er mere velegnet end den ofte anvendte persistent-prediktor. Dette specielt i situationer, hvor prediktionshorisonten er over et par timer. Den nye referencemodel er næsten lige så enkel som den hyppigt anvendte persistentprediktior. I princippet består den nye referencemodel af en vægtning mellem persistent-prediktor og gennemsnitsværdien af energien, hvor vægtningen for en given prediktionshorisont bestemmes af autokorrelationen i den givne vindenergitidsserie. I artiklen D er betingede parametriske modeller, estimeret ved lokal regression, anvendt til at identificere de vigtigste forklarende variable i kortfristede prognosemodeller. Disse modeller er estimeret med off-line metoder. I artiklen E undersøges lignende modeller, men her estimeres disse ved anvendelse af de rekursive metoder beskrevet i artikel B. Denne artikel præsenterer også modeller, der indeholder fysiske relationer, og resultaterne ved anvendelse af disse modeller sammenlignes med de statistiske modeller. Resultatet

af sammenligninger viser i dette tilfælde, at det ikke er en fordel at anvende fysiske relationer. Artikel I sammenligner prognoser af vindhastigheder fra en numerisk vejrprognosemodel, HIRLAM, med målte vindhastigheder fra forskellige højder. Formålet med denne sammenligning er at undersøge indflydelsen af turbulens intensiteten, og hvorledes denne håndteres i HIRLAM.

Artikel F gennemgår det kortfristede prognosesystem, Prediktor, er udviklet på Risø. Modeller, hvor der er foretaget enkle statistiske korrektioner af prognoserne fra de fysiske relationer, er afprøvede. I denne artikel er der vist, at prognoserne fra den numeriske vejrprognosemodel ændrer egenskaber. Typisk måles kvaliteten af prognoser ved statistiske mål, så som summen af afvigelseskvadraterne. I denne artikel er disse mål suppleret med en direkte sammenligning af målte og predikterede tidsserier. Af disse sammenligninger fremgår det klart, at den numeriske vejrprognosemodel i høj grad er i stand til at prediktere det generelle flow i atmosfæren.

Artikel G beskriver kort årsagen til at der udvikles modeller, der bygger på en kombination af statistik og fysik. Arkitekturen i opbygningen af softwareapplikationen, Zephyr, er også gennemgået. Denne artikel foreslår også modeller til beregning af prognoser af vindenergi produceret i større geografiske områder. Baggrunden for disse modeller er, at det forventes at variationen i målingerne fra et større område, vil være mindre end variationerne i målingerne fra en enkelt vindmøllepark. Dette som følge af spatielle udligningseffekter. Modellerne er ikke afprøvede, idet der ikke har været målinger af den samlede produktion fra et større område til rådighed for dette projekt.

Artikel H beskriver resultaterne fra projektet 'Implementing Short-term Predictions af Utilites'. I denne artikel beskrives prognosesystemet udviklet ved Institut for Informatik og Matematisk modellering, samt prognosesystemet udviklet på Risø. Desuden er kvaliteten af disse systemer evalueret. Erfaringer fra anvendelse af begge systemer hos el-selskaber viser, at begge systemer er anvendelige.

Den sidste artikel omhandler den økonomiske værdi af kortfristede prognosemodeller. Prognoser fra flere prognosemodeller bruges som inddata til en model for det elektriske netværk i England og Wales. Resultaterne

viser, at ved lave penetrationer af vind energi har prognoserne ingen eller lille værdi, men i takt med at penetrationen stiger, bliver prognoserne vigtigere. Desuden vises at, konfidensintervaller for prognoserne kan være med til at øge værdien.

Contents

Chapter 1

Introduction

This thesis deals with the issue of making short-term predictions of wind-power-generated electricity. The first section describes the background for the thesis and the motivation for developing short-term prediction models for wind power. Section 1.2 contains bibliographic notes to previous research within the field of short-term prediction of wind power. Section 1.3 describes the objectives of the study and in Section 1.4 a brief outline of this summary report is provided.

1.1 Background and motivation

Electrical utilities all over the world are beginning to realize the need for reliable wind power predictions, as the penetration of electricity generated by wind farms in the electrical grids is increasing. As the industry is approaching maturity, the market is shifting from heavily subsidised technology demonstration plant to capital-driven shareholder value. From the description in the following sections it will become apparent, that in order for the wind power industry to survive in the future, new methods that facilitate the reliability of wind power are necessary.

1.1.1 Increasing penetration

Figure 1.1 shows the development in the penetration of electricity produced by wind farms. The numbers shown are from Denmark. In 1999 the raw share of electricity from wind power is close to 9 %, for the normalized wind production values the share is more than 10 %. The values are from (Krohn 2000).

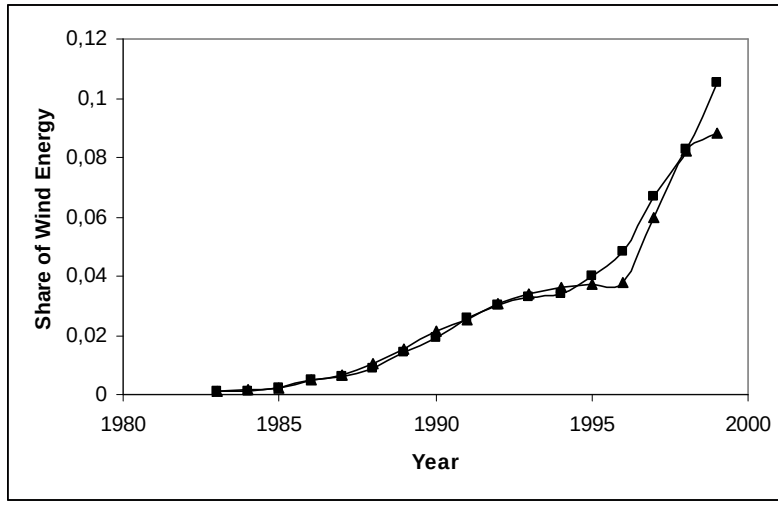


Figure 1.1: Share of electricity produced by wind energy in Denmark. Rectangles correspond to values where the wind power production has been normalized to an average wind year.

If the exponential growth in the share seen in Figure 1.1 continues, one could be lead to expect that all electricity will come from wind power after only a few years. This is not the case. It is important to note that the share value shown in the figure is not the same as savings in fossil fuels used by conventional power plants. This is illustrated in the following example.

An example – A storm coincides with low load

In Paper H the power production set-up in the Western part of Denmark is outlined. The power production set-up consists of 6 primary stations with a total capacity of 4.3 GW, a large number of local CHP (Combined

Heat and Power) units with a total capacity of 1.4 GW and, finally, wind turbines with a total rated capacity of approximately 1 GW. The production from the local CHP units and the wind turbines is treated as priority production, which means that this production has to be accepted in the electrical grid. The annual variation in the load in the Western part of Denmark is in the range 1.2–3.7 GW.

From these numbers it is seen that if a storm coincides with a low load situation, then the wind turbines alone will be able to cover the demand, resulting in overproduction of power if the primary power plants are not shut down. This could be the case if the storm peaks during the night-hours where the load is particulary low. On the other hand, if the storm is predicted, actions can be taken in due time to e.g. shut down primary power plants. As demonstrated by this example, wind power predictions are necessary for optimal dispatch and scheduling of the total power production, and the importance of the predictions increases as the penetration from wind power increases.

In the worst case, this means that if the wind power is completely unpredictable, it will not save any fossil fuels at all. In principle there are three ways to prevent this. Accumulation of wind energy, e.g. using batteries, wind power predictions and/or spatial distribution of wind turbines. The first method is evident, but has not yet proved practically feasible, the second method is the subject of this thesis. Predictions can be used for creating an optimal combination of wind power production with other power sources, like hydro power and/or fossile fuel power plants. The final method is a feature of the atmosphere. More specifically, from the assumption that the wind always blows somewhere, distributing the wind turbines over a larger area stabilises the wind power production.

Germany, Spain and Denmark are the leading countries in Europe with regard to installed wind power capacity. In the beginning of 2001 the installed wind power capacity in Europe was close to 13000 MW, in Germany 6100 MW, in Spain 2400 MW and in Denmark 2300 MW. This means that the three leading countries account for more than 83 % of the installed capacity in Europe. Only two other countries have wind power capacities that are in the same order of magnitude as the leading countries in Europe, these are USA with an installed capacity of 2500 MW and India with an installed capacity of 1200 MW.

The global installed capacity in the beginning of 2001 was 17700 MW, which means that the 5 countries mentioned above account for approximately 82 % of the world wide installed wind power capacity.

1.1.2 Liberalization

The electricity sector is currently subject to liberalization, and as a consequence of this, a new structure is emerging. The sector is being divided into three independent groups, which are production, transmission and distribution.

The production companies are both the owners and operators of the conventional power plants and some of the wind farms. The transmission companies are the owners and operators of the high voltage transmission network, and the distribution companies are responsible for the low voltage distribution network supplying the individual consumers.

In this set-up, when fully implemented, the dealings between the operators will be based on short-term contracts, typically day to day contracts. Any deviations from the reported demand or production will be subject to economic penalty. This means that operators with a considerable amount of wind power will be highly dependent on wind power predictions.

The same is the case for the players in emerging energy trade markets. Nord Pool, The Nordic Power Exchange, is an energy trade marked established in Norway in 1993. In 1995 the national authorities in Denmark, Finland and Sweden agreed to establish a common Nordic energy trade marked. As a result of this, Sweden joined Nord Pool in 1996, Finland in 1998, and, finally, Denmark joined the marked in 1999.

On this marked the value of wind power depends on the availability and accuracy of wind power predictions, as in this market the dealings of power is also based on short-term contracts. One of the key concepts of the short-term contracts, is that the dealings of power for the following day have to be settled at noon the day before. This means that the wind power prediction horizon has to be between 12–36 hours in order to be useful for trading on Nord Pool.

1.2 Previous research

Developing models for short-term prediction of wind power production is by no means a trivial task, as the underlying system covers everything from the large scale atmospheric flow, influence by local topography, vegetation and atmospheric conditions, the wind farm layout and the single turbine. This system, including each single component, is by nature non-linear and non-stationary.

Short-term prediction of wind farm power production has already been the subject of extensive research prior to this study. The approach used in this research can be distinguished by the type of input-data used in the prediction models. In principle there are three categories; models based on local measurements, models based on numerical weather predictions, and finally, models based on a combination of both local measurements and numerical weather predictions. The following sections provide an overview of these types of models and bibliographic notes.

As mentioned previously, the research presented in this thesis has been performed in a collaboration between the Department of Informatics and Mathematical Modelling at the Technical University of Denmark and the Research Programme Wind Power Meteorology at Risø National Laboratory. These departments have been working within the short-term wind power prediction field for a long time, Risø since 1989 and the Technical University of Denmark since 1992. Both departments have participated, and still do, in international research projects related to this field. This research has resulted in two on-line software prediction systems, implementing short-term prediction models which today are considered as state-of-the-art.

1.2.1 Local measurements

The methodologies that have been applied to local measurements are within the field of time series analysis, regression analysis and neural networks. One of the easiest prediction models is the persistence model. In this model, the prediction for all prediction horizons is set to the most recent measurement value. This means, by definition, that the error for

the now cast, i.e. zero prediction horizon, is zero. Furthermore, for short prediction horizons, i.e. on the order of minutes or a few hours, the error is relatively small compared to the errors for predictions from numerical weather prediction models or more sophisticated time series models. This is because the atmosphere is quasi-stationary, the time scales in the atmosphere are in the order of days (at least in Europe). It takes about one to three days for a low-pressure system to cross the continent, high-pressure systems can be more stationary. As the pressure systems are the driving force for the wind, the changes in the wind have time scales of the same order. Therefore, the persistence model has been used as a comparative model for other prediction models. For longer prediction horizons, i.e. more than a few hours, the persistence model is not adequate as a comparative model. In Paper C it is shown that a first order auto-regressive model is more adequate.

In the research described by (Bossanyi 1985) a Kalman Filter with the last 6 values (1 minute averaged data) as input is used to predict the next step. This gave 10 % improvement in the RMS error compared to the error of the persistent predictions for the next time step. This improvement decreased for longer averages, and disappeared completely for 1-hourly averages.

In (Dutton, Kariniotakis, Halliday & Nogaret 1999) an autoregressive model and an adaptive fuzzy logic based model for the cases of Crete and Shetland showed minor improvements over persistence for 2-hour horizons. For longer horizons significant improvements were found, i.e. for the 8 hour horizon an 20 % improvement in the RMS error was found. However, as described in Paper C, the persistence model is not adequate as reference for these horizons. Furthermore, the fact that most of the likely wind speeds were contained in the 95 % confidence band for the longer horizons, means that using the mean value for all times as the predictand would provide almost the same RMS error results compared to the models in this paper.

The early models (Madsen 1996) developed at the Department of Informatics and Mathematical Modelling, were based on local measurements only. Like the models described above, these models did not perform well for prediction horizons above 6–12 hours.

In (Beyer, Degner, Hausmann, Hoffmann & Ruján 1994) neural networks are used for next-step prediction of either 1-minute or 10-minute averaged data. In both cases they find 10 % improvement over the persistence model. This is achieved with a rather simple topology, while more complex neural network structures did not improve the results further. In (Tande & Landberg 1993) it is found that neural networks used to predict 10-second values using 1-second averages perform only marginally better than the persistence model. In (Bechrakis & Sparis 1998) neural networks are used to utilise wind direction information, but no performance measures over persistence are presented.

1.2.2 Measurements and Numerical Weather Predictions

Based on the methodology developed for the European Wind Atlas (Troen & Petersen 1989), Risø National Laboratory has developed short-term prediction models based on physical reasoning (Landberg 1999, Landberg & Watson 1994). These models are primarily based on numerical weather predictions as input, and are more thoroughly described in Chapter 2.

The University of Oldenburg (Beyer, Heinemann, Mellinghoff, Mönnich & Waldl 1999) has developed models similar to those developed at Risø. The main difference is that the models developed at Oldenburg use numerical weather predictions from the Deutschlandmodell of the German Weather Service DWD instead of HIRLAM.

Vitec AB from Sweden is working on a model based on meteorological forecasts from Swedish Meteorological and Hydrological Institute SMHI. So far, nothing is published (Giebel 2000).

In (Martin, Zubiaur, Moreno, Rodriguez, Cabre, Casanova, Hormigo & Alonso 1993) a prediction tool for the rather special case of Tarifa/Spain is described. Due to the special topological situation for the wind farms in Strait of Gibraltar, they could predict the power output from the pressure difference between measurements at Jerez and Malaga airports, with the additional use of Spanish HIRLAM. The founding for this project was stopped, and the project therefore ended half way through.

As mentioned previously, the early models developed at the Department

of Informatics and Mathematical Modelling, were only taking local measurements as input. Therefore, models which included meteorological forecast were developed (Nielsen & Madsen 1996), and it is a statistical model which takes both measurements and meteorological forecasts as input, which is implemented in the on-line prediction system, WPPT (Wind Power Prediction Tool), developed at this department. This system is briefly outlined in Paper H.

The models developed at Risø National Laboratory, also make use of local measurements. This is mainly for calibration purposes, also described as MOS (Model Output Statistics) (Landberg & Joensen 1998). Therefore it is important to note that on-line measurements are not used in this model to calculate the actual predictions.

EWind is an US-American model by TrueWind, Inc (Bailey, Brower & Zack 1999). Instead of using a once-and-forall parameterization for the local effects, like the Risø approach does with WAsP, they run the ForeWind numerical weather model as a meso-scale model using boundary conditions from a regional weather model. Due to the enhanced resolution in the meso-scale model more physical processes are captured, and the predictions can be better tailored to the local site. Nevertheless, they use adaptive statistics to remove the final systematic errors. No performance results are presented.

In (Shuhui, Wunsch, O'Hair & Giesselmann 2001) regression and neural network methodology is compared in the aim of modelling the wind turbine power curve. From the models tested it is concluded that the neural network approach is superior to regression. The power curve is estimated using local measurements of meteorological variables and power. This power curve is then supposed to be used for the transformation of numerical weather predictions to predictions of the power production. This approach is not sound as the properties of the numerical weather prediction are not necessarily the same as the properties of the measurements, i.e. properties like the statistical metrics mean and variance. This problem is further described in (Jonsson 1994), which argues that if errors are present in the regressors the use of the true system for prediction will not result in optimal predictions. No results from using numerical weather predictions are presented in this paper.

1.3 Objectives

The main objective of the research presented in this thesis has been to develop models and methods for short-term prediction of wind farm power production. As described in the previous section, short-term prediction has already been the subject of extensive research. The purpose of the research described in this thesis, is different in the way that the purpose has been to find out how the physical and statistical approaches taken previously, can be combined and further refined in order to improve the prediction quality.

Furthermore, the emphasis has been on the development of practically applicable models and methods, as the final objective has been to develop an on-line software application, which implements the developed models and methods.

All considered models are taking numerical weather predictions as input, i.e. the objective of the thesis has not been to develop models for the description of the large scale atmospheric flow. The maximum prediction horizon which has been considered is 36 hours. The horizon is limited by the prediction horizon of the weather forecasts from the numerical weather prediction model.

An issue which is closely related to predictions is the economical value of the predictions, and specifically how the economic value depends on the accuracy of the predictions. This issue is also briefly considered in this thesis.

1.4 Brief outline

The thesis consists of 10 research papers, which have been written during the Ph.D study, and a summary report.

The purpose of the summary report is to give an overview of the included papers, and to give a description of the theoretical background for the thesis. The summary report also includes description of work which has

not been published elsewhere.

Chapter 2 outlines some of the meteorological theory which is relevant for the thesis. In Section 2.2 the numerical weather prediction model, HIRLAM, which has provided input variables to all the considered prediction models, is briefly described. A somewhat more detailed description of the equations and physical parameterizations used in HIRLAM is outlined in Appendix A. In Section 2.3 the short-term prediction system developed by the Department of Wind Energy and Atmospheric Physics at Risø National Laboratory, and the physical models applied in this system are described. Chapter 3 briefly describes the statistical models and methods that have been considered in this research, and provides bibliographic notes.

Chapter 4 links the included papers together, and the goal of this chapter is to give a unified view of the obtained results and bring the papers into context. Some of the topics addressed in the various papers are described in more detail, and some general remarks on the statistical models and methods considered in the papers will be provided.

In Chapter 5 the developed on-line software application, called Zephyr, will be described.

Finally, in Chapter 6, the overall conclusions of the thesis are stated.

Chapter 2

Meteorology

This chapter gives a brief introduction to meteorology in general, and secondly it goes into some more detail about the meteorological theory which is relevant for the thesis.

2.1 Basic concepts

Air flow, or wind, can be divided into three broad categories: mean wind, turbulence and waves. Each can exist separately or super-imposed onto each other. Transport of quantities such as moisture, heat and momentum is dominated in the horizontal by the mean wind, and in the vertical by turbulence. A large number of phenomena can be observed in the atmosphere, which are driven by highly complex processes, and, consequently, the theory which exists to describe these phenomena is very comprehensive and complex. This chapter can therefore only give a brief introduction to meteorology, and the emphasis is on phenomena, which are relevant for the objectives of this thesis.

2.1.1 Basic equations

The behaviour of the atmosphere is well described by seven variables: pressure, temperature, density, moisture, two horizontal velocity components, and the vertical velocity; all functions of time and position. The behaviour of these seven variables is governed by seven equations: the equation of state, the first law of thermodynamics, three components of Newton's second law and the continuity equations for mass and water substance. Motions in the atmosphere are slow enough compared to the speed of light that the Galilean/Newtonian paradigm of classical physics applies. These equations, collectively known as the equations of motion, contain time and space derivatives that require initial and boundary conditions for their solution.

The complete set of equations is so complex that no analytical solution is known. In a particular meteorological field or application, like boundary-layer meteorology or in a numerical weather prediction model, these equations are simplified and parameterizations and approximations are utilized which are valid in the particular field.

2.1.2 Turbulence

Turbulence, the gustiness super-imposed on the mean wind can be visualized as consisting of irregular swirls of motion called eddies. Usually turbulence consists of many different sized eddies super-imposed on each other. Much of the turbulence is generated by forcings from the ground. For example, solar heating of the ground during sunny days causes thermals of warmer air to rise. These thermals are just large eddies. Frictional drag on the air flowing over the ground causes wind shears to develop which generates turbulence (*Kelvin-Helmholtz* waves). The largest size eddies can be 100-3000m in diameter, these are the most intense eddies because they are produced directly by the forcings described previously. Smaller size eddies are apparent in the swirls of leaves and in the wavy motions of the grass. These eddies feed on the larger ones. The small eddies, on the order of a few millimeters in size, are very weak because of the dissipating effects of molecular viscosity.

2.1.3 Turbulent flow

Although the equations mentioned in Section 2.1.1 could be applied directly to turbulent flow, this is not possible in practice. The scales of motion in the atmosphere cover the range from thousands of kilometers down to the scale of the smallest eddies described in the previous section, therefore, direct application of the equations would require observations with one millimeter spatial and a fraction of a second temporal resolution.

Instead, some cut-off scale is selected below which the influence of turbulence is only treated statistically. The selected cut-off scale depends on the current application, in a numerical weather prediction model the cut-off is on the order of 10 to 100km, while for some boundary-layer models known as large eddy simulation models the cut-off is on the order of 100m (Stull 1988).

Therefore the dependent variables in the basic equations are expanded into mean and turbulent (perturbation) parts, i.e. $U = \overline{U} + u'$, where the bar above the variable signifies that it is a mean value and the prime signifies that it is the departure from the mean. Reynolds averaging (Stull 1988) is then applied to get equations for the mean variables within a turbulent flow. After this procedure the equations contain variables of the form $\overline{u'v'}$, which represent the turbulent motions statistically, i.e. these variable can be interpreted as the covariance between the variables u' and v'.

One unfortunate feature of the set of equations which are derived by this procedure, is that it is not possible to derive as many equations as there are unknown variables (Stull 1988), i.e. the equation system cannot be closed. When equations for the $\overline{u'v'}$ covariance terms described above are derived, these equations contain new $\overline{u'v'w'}$ terms, and this pattern continues when equations for $\overline{u'v'w'}$ are derived. At some point, the process of deriving new equations must be stopped, and the unknown variables need to be parameterized in terms of other known variables. If the unknown covariance or higher order statistical moments are parameterized using spatial derivatives of other known variables, this is called nth-order local closure, where n is the order of the statistical moments which are retained in the equations. These and other closure techniques

are described in (Stull 1988). The unknown covariance of the higher order statistical moments can not be neglected as these terms correspond to energy.

2.1.4 The boundary layer

Figure 2.1 illustrates how the boundary-layer in a high pressure region over land evolves during the day. In this case the boundary layer has a well defined structure. The three major components of this structure are the mixed layer, the residual layer and the stable boundary layer.

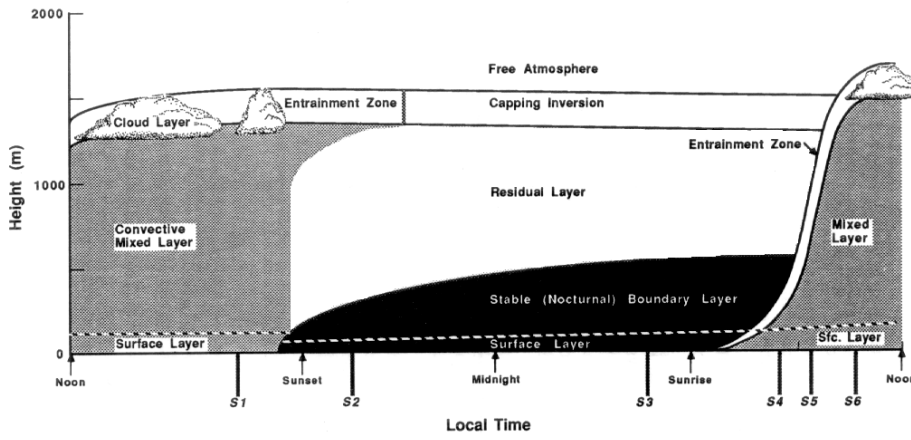


Figure 2.1: Illustration of how the boundary layer evolves with time and height. From (Stull 1988). For explanation see text.

The surface layer is defined as the region at the bottom of the boundary layer where turbulent fluxes and stress vary by less than 10% of their magnitude.

The turbulence in the mixed layer is usually convectively driven. The convective sources include heat transfer from a warm ground surface, and radiative cooling from the top of the cloud layer. Even when convection is the dominant mechanism, there is usually wind shear across the top of the mixed layer that contributes to the turbulence generation. The mixed layer grows in height by mixing down into it the less turbulent air

from above, and the maximum height is reached in the late afternoon. A stable layer at the top of the mixed layer acts as a lid to the rising thermals. It is called the entrainment zone because the entrainment into the mixed layer occurs here. At times this layer is strong enough to be classified as a temperature inversion, which means that the absolute temperature increases with height.

About an half hour before sunset the thermals cease to form (in the absence of cold air advection), allowing the turbulence intensity to decay in the formerly well mixed layer. This layer is usually called the residual layer because its initial mean state variables are the same as those of the recently decayed mixed layer.

As the night progresses, the bottom portion of the residual layer is transformed by its contact with the ground into a stable boundary layer. This layer is characterized by statically stable air with weaker, sporadic turbulence.

In low pressure regions the upward motions carry boundary-layer air away from the ground to large altitudes. In this case the boundary layer has a less well defined structure.

2.1.5 Vertical profiles

As mentioned in the introduction to this chapter the transport of atmospheric constituents in the vertical is mainly driven by turbulence. Therefore, the closure techniques applied to the governing equations depends on adequate parameterizations of how the vertical profiles for the atmospheric constituents depend on the turbulence intensity. A large number of such parameterizations have been proposed in the litterature (Stull 1988), and some examples are shown in Appendix A. The purpose of this section is to describe the qualitative behaviour of these parameterizations.

The atmospheric stability is usually classified in the range from stable, over neutral to unstable. In the stable case there is no or only little vertical mixing, and in this case the flow in different vertical layers is more or less decoupled. This means that there can be large differences

in the atmospheric state at different heights. This can lead to low wind speeds close to the ground and high wind speed just above the ground, i.e. low level jets. It should be noted though, that the shear between high and low wind speed layers leads to formation of *Kelvin-Helmholtz* waves, which consequently creates turbulence. Therefore, the atmosphere will only remain stable if the wind speed is low in all vertical layers.

As the turbulence intensity increases the vertical mixing increases correspondingly. This means that the difference in the atmospheric state becomes less dependent on the height, i.e. the wind speed and other variables are now more or less constant in the vertical.

2.2 The numerical weather prediction model

The numerical weather prediction model which has provided the weather forecast variables, is the HIgh Resolution Limited Area Model (HIRLAM), run by the Danish Meteorological Institute (DMI). The development of this model was started in 1985 as a joint project between the national meteorological institutes in Denmark, Finland, Norway, Sweden, Spain and The Netherlands. HIRLAM is subject to continues development and the latest news on the model can be found at `www.dmi.dk`

2.2.1 General features

Numerical weather prediction (NWP) can be described as the simulation of the processes in the atmosphere on a computer, with the purpose to predict the future state of the atmosphere based on the actual state. For a good overview of the historical development of numerical weather prediction models see (Kalnay, Lord & McPherson 1998). The assessment of the state of the atmosphere is called data assimilation and is of crucial importance for the accuracy of the predictions. The state is estimated from measurements from synoptic stations all over the world and satellites. The data assimilation procedure validates the measurements, discards erroneous observations and fills in the gaps between the stations (e.g. over the oceans). The last step is accomplished by multivariate

statistics in HIRLAM. A description of the 3-dimensional data assimilation system used in HIRLAM can be found in (Lorenc 1981, Lonnberg & Shaw 1987).

A somewhat detailed description of the equations used to simulate the atmosphere in HIRLAM is given in Appendix A, where the emphasis is on how the turbulence is taken into account. A thorough description of HIRLAM is given in (Sass, Nielsen, Jørgensen & Amstrup 1999), where also the numerical methods for the integration of the model and how the boundary conditions are applied is described.

Theoretical estimates limit the predictability of the weather by NWP to about 72 hours; this demands however, that the model domain is global (Haltiner & Williams 1970). Early chaos theory predicts a total divergence of weather patterns from virtually identical starting points after 14–20 days (Lorenz 1963). Using ensemble forecasts, this limit can be extended somewhat, since the ensemble members have some of the possible variation already built in (Kalnay et al. 1998).

2.2.2 Integration domains

HIRLAM is a limited area model (LAM), which means that lateral boundary conditions have to be specified. DMI is running four nested HIRLAM models with individual integration areas illustrated in Figure 2.2. The lateral boundary conditions for the model applied to the largest area, denoted by 'G' in Figure 2.2, are supplied by ECMWF (The European Center for Medium Range Weather Forecasting). The 'N' and 'E' models use lateral boundary conditions from the 'G' model, while the very high resolution model 'D' around Denmark, uses boundary conditions from the 'E' model.

The numerical weather predictions which have been used in this thesis are from the 'D' model. The horizontal resolution used in this model is 5.5 km. The time step used in the model integration is 36 s, and the influence of the physical parameterizations is applied every third time step.

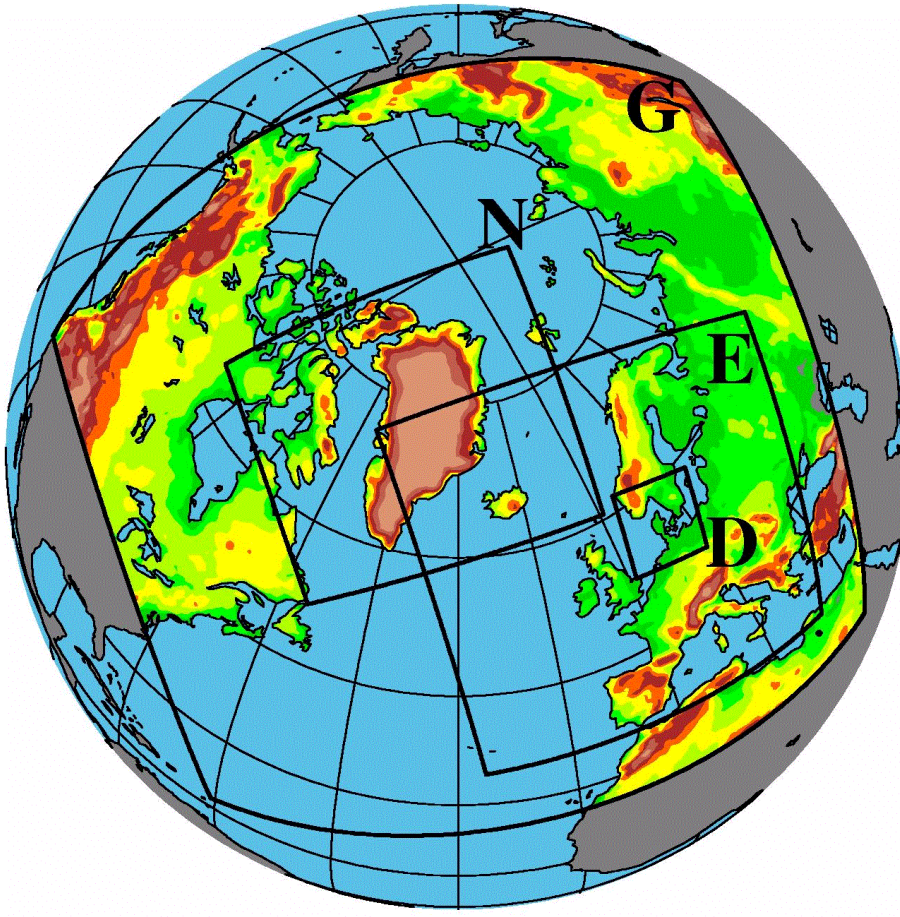


Figure 2.2: The operational setup of the HIRLAM model. From (Sass 1998).

2.2.3 The data

The variables which have been available from HIRLAM for use in this study are the following:

At the surface:

- The u and v components of the wind at 10 m agl.
- Surface friction velocity u_*.
- Sensible heat flux H_s.
- Latent heat flux H_L.
- Pressure p_s

At the vertical model levels $l = 31, \ldots, 25$ ($l = 31$ is the lowest):

- The u_l and v_l components of the wind.
- The U_l and V_l components of the geostrophic wind.
- Temperature T_l.
- Height above mean sea level h_l.

The model run frequency of HIRLAM has not been constant. In one period the predictions were received twice a day, corresponding to the initial times 00:00 and 12:00 (UTC). While in a second period the predictions were received four times a day, at 00:00, 06:00, 12:00 and 18:00 (UTC). The predicted variables are given in 3 hourly steps 36 hours ahead.

The HIRLAM levels l correspond to constant pressure surfaces, therefore the levels do not correspond to a fixed height above the surface. A more detailed description of the individual variables and how they are calculated can be found in Appendix A.

2.3 The Risø system

This chapter describes the physical models, which the short-term prediction model developed by the Department of Wind Energy and Atmospheric Physics at Risø National Laboratory, is constructed from.

In Paper F the procedure for the calculation of the power predictions is illustrated. The input wind is taken from a numerical weather prediction model. Due to the resolution of the numerical weather prediction model, the predictions from the model does not include local effects, i.e. on scales less than 5–10 km. Therefore, a model layer is selected at which the wind is assumed to be approximately geostrophic. The geostrophic drag law is then used to calculate the surface stress u_*, which subsequently is used as input to the logarithmic wind profile to calculate the wind at the hub height of the turbines. To take the effects of the local topography into account the results from a WAsP (Mortensen, Landberg, Troen & Petersen 1993) analysis of the particular site is used to correct the wind calculated from the logarithmic wind profile. Finally, the WAsP corrected wind is folded through the wind farm power curve determined by the PARK (Sanderhoff 1993) application, using empirical power curves supplied by the manufacturer.

In the following sections the steps and sub-models used in the above procedure will be described. In the last section the sub-models will be analyzed in some detail. The purpose of the analysis is to find out how the physical models can be used in combined statistical and physical models, and if it at all is possible to develop such combined models.

2.3.1 The geostrophic drag law

The geostrophic drag law is a result of merging the wind in two layers in the atmosphere. In the derivation the atmosphere is divided into three vertical layers: the free atmosphere, the mixed layer and the surface layer (see Figure 2.1).

A derivation of the geostrophic law is provided in (Landberg 1994), and for neutral atmospheric conditions the law is given by the following set

of equations

$$\frac{G}{u_*} = \frac{1}{\kappa}\sqrt{\left[\ln\left(\frac{u_*}{f z_0}\right) - A\right]^2 + B^2}, \tag{2.1}$$

and

$$\tan\alpha = \frac{-B}{\ln\left(\frac{u_*}{f z_0}\right) - A}, \tag{2.2}$$

where $f = 2\Omega\sin\phi$ is the Coriolis parameter, Ω the angular velocity of the Earth, ϕ is the latitude, u_* is the surface friction velocity, $\kappa = 0.4 \pm 0.01$ is the von Karman constant, G is the magnitude of the geostrophic wind, z_0 is the surface roughness length, and α is the angle between the geostrophic and surface wind direction. A and B are empirical constants. A great deal of experiments have been carried out to determine the values of these constants, and there is quite some scatter in the values which have been proposed. In the Risø system (Landberg 1994) $A = 1.8$ and $B = 4.5$ is used.

For non-neutral conditions a large variety of geostrophic draw laws have been proposed and derived (Landberg 1994). Common for these derivations, is that the dependency is modeled by letting the parameters A and B be functions of a stability parameter μ. In (Landberg 1994) μ is defined as $\mu = \kappa u_*/fL$, where $L = u_*^3/\kappa|B_s|$ is the Monin-Obukhov length, and B_s is the near-surface value of the vertical buoyancy flux, defined as

$$B_s = \beta\frac{H_s}{c_p\rho} + 0.608\frac{gH_L}{L_c\rho}, \tag{2.3}$$

where $\beta = g/\theta$ is the buoyancy parameter, H_s is the sensible heat flux, c_p the heat capacity of air at constant pressure, ρ the density of the air, g the gravitational acceleration, H_L the latent heat flux, L_c the latent heat of vaporization, and θ is the potential temperature.

As the Risø system only uses the neutral geostrophic drag law, the functional shapes of A and B are not provided here.

2.3.2 The logarithmic wind profile

The stability dependent logarithmic wind profile is shown in Appendix A.2. In the Risø system it is the logarithmic profile for neutral conditions

which is used, and in this case the expression simplifies to

$$u(z) = \frac{u_*}{\kappa} \ln\left(\frac{z}{z_0}\right), \tag{2.4}$$

where z is the height above the surface. It is seen that the surface friction velocity u_* can be used in this expression to calculate the wind at height z.

2.3.3 WAsP

The Wind Atlas Application and Analysis Program WAsP (Mortensen et al. 1993) is a program to make wind atlases. A wind atlas is a generalized wind climate for an area. The idea behind the program is to take measurements from a specific site (e.g. a meteorological mast at an airport) and calculate the sector-wise distribution (Weibull) of the wind. This distribution is then 'cleaned' for local effects in the following order:

- Shelter from obstacles in the vicinity of the site.
- Changes in the roughness of the surface.
- Orography.

This procedure correspond to the upwards arrow in Figure 2.3. The corrected distribution is called a wind atlas, and corresponds to the sector-wise wind speed distribution of an area which is completely flat, where the surface is smooth and there are no obstacles. This atlas can now be extrapolated horizontally to other locations within the area, and used to calculate the expected wind climate at another location. This is done by applying the procedure described above in the reverse order, corresponding to the downwards arrow in Figure 2.3, using data describing the obstacles, roughness and orography at the new site.

To accomplice these tasks WAsP uses three sub-models, and it is assumed that the effect of these models can be applied independently of each other (Mortensen et al. 1993).

To take the effect of the orography into account, WAsP uses a simplified

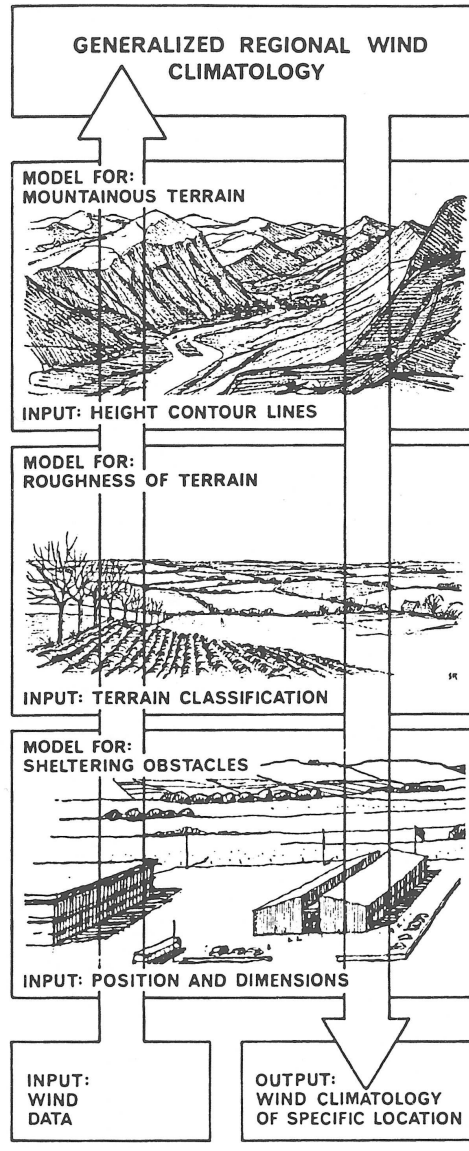


Figure 2.3: Illustration of the steps used in WAsP, for explanation see text.

analytical solution to the same governing equations which the numerical weather prediction model is based on. When the solution for these equations is derived, the atmosphere is assumed neutral and it is assumed that flow separation does not occur. The last assumption implies that the model is only valid for gentle to medium complex terrain, i.e. corresponding to a maximum terrain slope of 0.3. For the derivation see (Landberg 1994) Chapter 5. The result of applying this model to a specific site, which is then extrapolated to a site nearby, is that the wind at the new site can be written as (Mortensen et al. 1993)

$$\omega_s = a_{c,\theta_r}\omega_r, \quad \theta_s = \theta_r + b_{c,\theta_r}, \tag{2.5}$$

where subscript r refers to the reference site, i.e. the site which the wind atlas is generated from, and s refers to the site which the atlas is extrapolated to, ω and θ is wind speed and direction, respectively, and a_{c,θ_r}, b_{c,θ_r} are wind direction dependent constants.

Similarly, the effect of the roughness and obstacles leads to the following corrections (Mortensen et al. 1993)

$$\omega_s = a_{z,\theta_r}\omega_r, \quad \theta_s = \theta_r + b_{z,\theta_r} \tag{2.6}$$

for the roughness, and

$$\omega_s = a_{o,\theta_r}\omega_r, \quad \theta_s = \theta_r + b_{o,\theta_r} \tag{2.7}$$

for the effect of obstacles. Thus the total correction from WAsP can be written as

$$\omega_s = a_{t,\theta_r}\omega_r, \quad \theta_s = \theta_r + b_{t,\theta_r}, \tag{2.8}$$

where

$$a_{t,\theta_r} = a_{c,\theta_r}a_{z,\theta_r}a_{o,\theta_r} \tag{2.9}$$

and

$$b_{t,\theta_r} = b_{c,\theta_r} + b_{z,\theta_r} + b_{o,\theta_r}. \tag{2.10}$$

Subscript t refers to the total correction, c to the correction due to orography, z to the roughness correction and o to the correction due to obstacles. For a derivation of the relations used to take the effect of roughness and obstacle into account see (Landberg 1994).

Note that the constants described above are also functions of the height above the surface. This has not explicitly been pointed out here, because

when these constants are used in the Risø system, the height is fixed at the wind turbine hub height. Furthermore, WAsP uses a special vertical wind profile to transfer the wind between different heights. This feature is not used in the Risø prediction system, which uses its own vertical profile. The profile which is used in WAsP takes the effect of the stability into account in a mean sense, i.e. the effect of the stability on the annual mean of the wind speed is modeled by this profile (Troen & Petersen 1989*b*). This profile can not be used in the Risø prediction system, as this system predicts instantaneous values.

2.3.4 PARK

The PARK program (Sanderhoff 1993) is calculating a mean efficiency for the wind turbines in a wind farm. It models the the reduction of the wind speed behind the turbines due to wake effects. The wake propagation is illustrated in Figure 2.4.

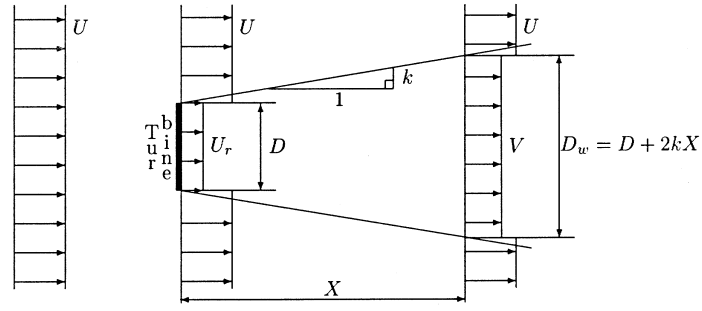


Figure 2.4: Illustration of the wind turbine wake calculation use in PARK.

The model is based on the assumption that the wake expands linearly behind the turbine. The only parameters in the models are the initial velocity deficit at the start, and the wake decay constant describing the expansion of the wake.

The necessary input to the program is therefore the coordinates of the turbines, the power and thrust curves, the hub height and the rotor diameter and meteorological data for the site. The program is limited

to wind farms consisting of identical turbines and identical wind turbine hub heights.

2.3.5 From input to output

Now we are ready to take a closer look at what is happening in the relation which are used in the Risø system, especially what the shape of the resulting relation looks like, when all the relations described in the previous section have been applied in the order described in Paper F. Note that the analysis is this section is closely related to the analysis described in (Landberg 1999*a*).

First let us take a look at the geostrophic drag law. The relations shown in Section 2.3.1 are not in a closed form. This means that the surface friction velocity u_* and the difference between the geostrophic wind direction and the wind direction in the surface layer have to be found by numerical methods. Furthermore, note that the roughness length $z_0 = z_0(\theta)$ is a function of the wind direction, θ, in the surface layer. This leaves three coupled equations which have to be solved by a combined iterative and root solving method.

In Figure 2.5 the value of the surface wind speed ω_s as a function of the geostrophic wind speed G, and various values of the roughness length is shown.

The figure has been constructed by solving the above mentioned equations for u_*, where for simplicity the roughness has been assumed uniform for all wind directions. Subsequently, the wind speed has been extrapolated to 30 m above ground level using the neutral logarithmic wind profile.

From this figure it is seen, that to a very close approximation, the wind at a given height in the surface layer is linearly related to the geostrophic wind speed, for a given roughness length. If this linearity is utilized and the effect of the corrections from the WAsP analysis are added, it is seen that the wind speed at the wind turbine hub-height can be written as

$$\omega_h = a(\theta_s)G \tag{2.11}$$

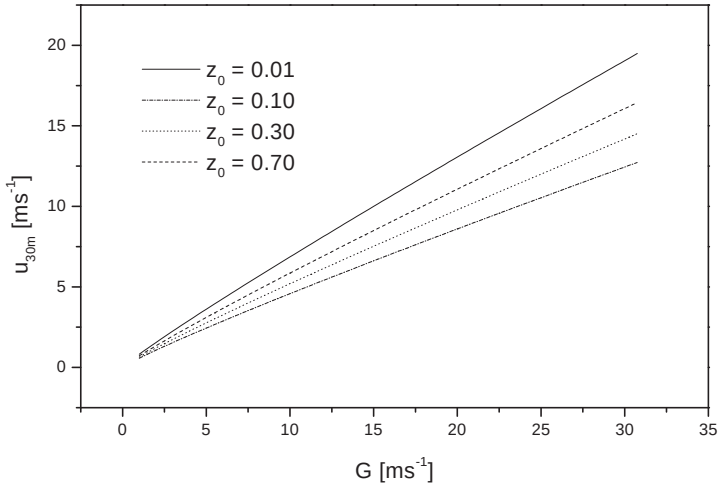


Figure 2.5: Surface wind speed (30 m agl) as a function of geostrophic wind speed and various roughness lengths.

where θ_s is the surface wind direction calculated from (2.2). The wind direction dependency used in the Risø system is given by $a(\theta_s) = a_{\theta_s}$, i.e. a wind direction dependent constant. This is accomplished by dividing the circular direction into a number of sectors of equal size. The standard size used in WAsP is 30 Deg (Mortensen et al. 1993).

The relation for calculating the wind direction in the surface layer from the geostrophic drag law, is slightly more complicated. This relation is shown in Figure 2.6 and it is seen that the surface wind direction depends non-linearly on the roughness and the magnitude of the geostrophic wind.

It is seen that the curvature is most substantial for low geostrophic wind speeds, $G < 5$ ms^{-1}, and especially where low speeds coincide with rough surfaces $z_0 > 0.8$. These value are the ones which are most unlikely to be representative for the place where a wind farm is located. A roughness value $z_0 > 0.8$ corresponds to a very rough surface like a larger city. Wind speeds below 5 ms^{-1} are more or less irrelevant, as this in most cases will result in surface wind speeds which are below the cut-in wind speed (i.e. the speed for which the turbine starts producing power) of

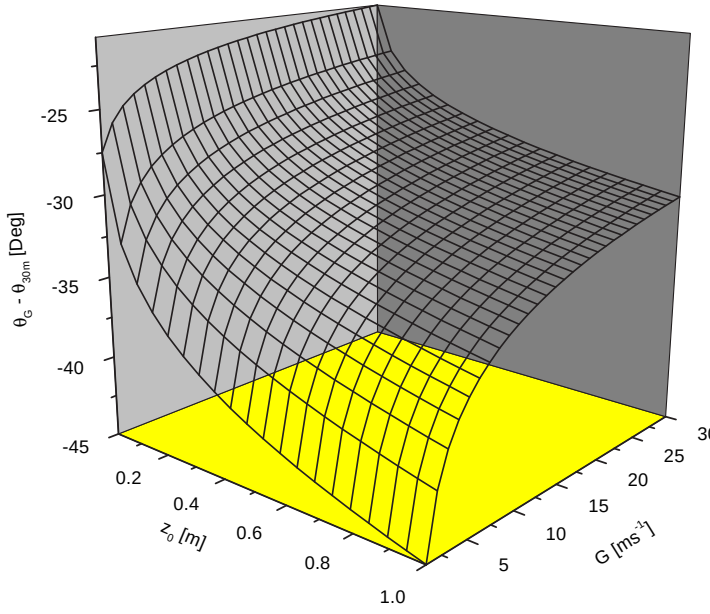


Figure 2.6: Geostrophic and surface layer (30m agl) wind direction difference as a function of roughness length and geostrophic wind speed.

the most common wind turbines.

If these values are cut out of the relation shown in Figure 2.6, it is seen that the remaining relation is relatively constant, i.e. -30 ± 10 Deg. This variations should be compared to the standard resolution used in the Risø system and the WAsP application, which as mentioned above is 30 Deg. It can therefore be argued that the wind direction, independent of the geostrophic wind speed, in most cases will map to the same sector. These findings will be further commented on later when the statistical models used will be described.

Now, the final step is to transform the wind speed to power. The wind farm power curve which is used in the Risø system is derived by the PARK application described in Section 2.3.4, using wind turbine power curves supplied by the manufacturer. A typical power curve is shown in Figure 2.7, which is seen to be highly non-linear.

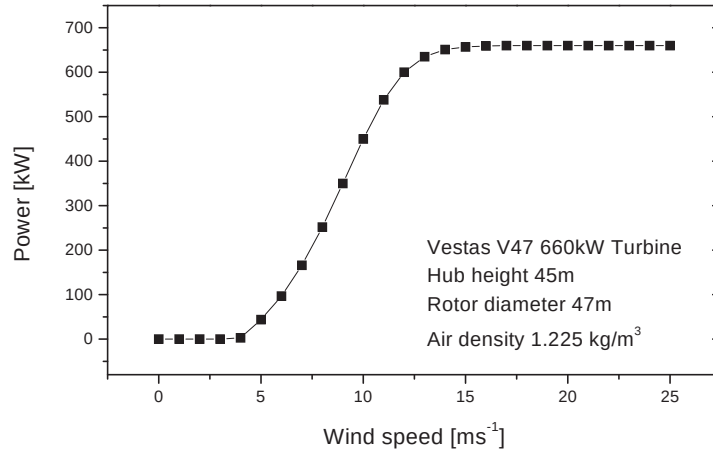


Figure 2.7: The Vestas V47 660kW wind turbine power curve. Cut-off at 25 ms^{-1} not shown.

The point to be noted in the analysis in this section, is that with regard to the wind speed, there is no non-linearity in the relations used by the Risø system, before at the final stage where wind speed is transformed into power.

Chapter 3

Statistics

The purpose of this chapter is to define the statistical models and the basic properties of the models that have been considered in this thesis, and to provide bibliographic notes. The considered models are also described in the more theoretical papers in this thesis, i.e. - the Papers A and B. These papers also describe the new recursive estimation methods that have been developed. The emphasis in this chapter is therefore not on describing the estimation methods that have been applied, instead the reader is referred to the Papers A and B and the bibliographic notes. Furthermore, Section 3.3 outlines the considerations for the choice of estimation methods that have been developed. As the objectives of this project has been to develop models and methods for on-line prediction, it is important that the estimation methods are suited for on-line estimation.

3.1 Statistical models

Three types of models have been considered in this thesis. Linear, nonlinear and conditionally parametric models. The last model type is ac-

tually a non-linear model where a special model structure has been imposed. This section briefly outlines these models.

The basic model which relates a response or dependent variable y_i to a vector $\boldsymbol{x}_i$ of explanatory or independent variables is written

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1 \dots N, \tag{3.1}$$

where N is the number of observations and $f(\cdot, \cdot)$ is some general function of the explanatory variables and the parameters in $\boldsymbol{\theta}$. The residual ε_i is assumed to be an sequence of identically distributed variables, which are independent of $\boldsymbol{x}_i$.

In a linear model a special structure has been imposed on $f(\cdot, \cdot)$ in (3.1)

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\theta} + \varepsilon_i, \quad i = 1 \dots N, \tag{3.2}$$

The definition of a non-linear model can thus be considered as model, which can not be written in the same form as (3.2).

It should be emphasized that the term linear relates to the parameters, i.e. – the linear model is linear in the parameters. Non-linear relations between the explanatory variables can thus be included in the linear model as extra variables in $\boldsymbol{x}_i$ and transformations of $\boldsymbol{x}_i$.

The conditionally parametric model, also know as a varying-coefficient model, is written

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\theta}(\boldsymbol{u}_i) + \varepsilon_i, \quad i = 1 \dots N, \tag{3.3}$$

where $\boldsymbol{u}_i$ is a vector of explanatory variables. It is seen that this model can either be interpreted as a linear model where the parameters have been replaced by unspecified functions of some explanatory variables, or a more general model where general functionals and linear effects have been separated. Furthermore, notice that $\boldsymbol{\theta}(\boldsymbol{u}_i)$ may be linear, i.e. written in the same form as (3.2). This is illustrated in the Papers A and B. In this case (3.3) reduces to a linear model of the same form as (3.2).

In time series analysis the index i refers to a time index and the vector of explanatory variables might contain previous values of the dependent variable. In this case the error sequence ε_i is required to be independent.

3.2 Multi-step prediction models

Generally, a multi-step prediction model can be considered as a model where the prediction horizon $k = 1, \ldots, k_{max}$ is added to the vector $\boldsymbol{x}_i$ of explanatory variables in the models described in Section 3.1, and where k_{max} is the maximum prediction horizon considered. Therefore, the multi-step prediction models are contained in the model definitions given in Section 3.1.

In time series analysis, where the index i is a time index, there are k_{max} observations at each discrete time point i of the dependent and independent variables. In this case it is convenient to write the model as

$$y_{i,k} = f(\boldsymbol{x}_{i,k}, \boldsymbol{\theta}, k) + \varepsilon_{i,k}, \quad i = 1, \ldots, N, \quad k = 1, \ldots, k_{max}, \tag{3.4}$$

where the number of observations now is Nk_{max}. In this definition $y_{i,1} = y_{i-k,k}$, $i > k$, i.e. the same observation of the dependent variable is used k_{max} times. This is not necessarily the case for $\boldsymbol{x}_{i,k}$, as the explanatory variables themselves might be predictions from some other multi-step prediction model.

The definition (3.4) of a multi-step prediction model allows for a wide variety of parameterizations for the prediction horizon dependency. Some examples are outlined below.

Consider the first order auto-regressive AR(1) model, using the same notation as in (3.4)

$$y_{i,1} = a y_{i-1,1} + \varepsilon_{i,1}. \tag{3.5}$$

The k-step prediction, $\hat{y}_{i,k}$, from this model is calculated recursively, yielding

$$\hat{y}_{i,k} = a^k y_{i-1,1}, \tag{3.6}$$

and it is seen that the AR(1) model could be formulated explicitly as a k-step model, i.e.

$$y_{i,k} = a^k y_{i-1,1} + \varepsilon_{i,k}, \tag{3.7}$$

which is a non-linear model. Note also that this model is of the same form as (3.3), i.e. it can be interpreted as a conditionally parametric or varying-coefficient model. It is easily verified, although this will not be done here, that if more general time series models, i.e. the ARMAX

model class, were considered, then the k-step prediction model formulation derived from this class would result in non-linear models, which in general not will be of the same form as (3.3).

If a system is only approximately described by a AR(1) model, it can no longer be assumed that the estimate of the parameter from the one-step prediction model will be optimal for all prediction horizons. In a case like this, it would make more sense to use the non-linear k-step model instead, as it is more likely that the parameter function a^k would yield better predictions averaged over all prediction horizons.

As mentioned previously, it is not feasible to use non-linear models in on-line applications. Another alternative is to consider conditionally parametric models of the form

$$y_{i,k} = \boldsymbol{x}_{i,k}^T \boldsymbol{\theta}(k) + \varepsilon_{i,k}, \tag{3.8}$$

and use the simple parameterization $\boldsymbol{\theta}(k) = \boldsymbol{\theta}_k$, i.e. assume on set of parameters for each prediction horizon k. This approach has been adopted in this thesis, although, in Paper E local regression is used to estimate the prediction horizon dependency of the parameter functions.

3.3 On-line and off-line estimation

Two techniques exist for estimation of the parameters in the models outlined in the previous section, on-line and off-line. In the off-line method the parameters are estimated using the full set of data, which means that when new data becomes available, the parameters have to be re-estimation on the full set of data, which now also includes the new data.

In the on-line method, usually referred to as recursive estimation, information about the parameter estimates from a previous estimation step is utilized when the parameters are to be re-estimated. Compared to off-line techniques, this leads to considerable savings in computational effort. Furthermore, the on-line techniques are relatively easy made adaptive, which enables the model to track slow changes in the underlying system.

The drawback of the recursive methods is, however, that they are not

suited for estimation in non-linear models. The reason for this is twofold, first of all it can not be guaranteed that the parameter estimates will converge to the true values for non-linear models. Furthermore, most often the non-linear estimation equations have no closed form solutions, therefore the estimates have to be found by numerical optimization.

Non-linear models have therefore only been considered briefly, an example is demonstrated in Paper E. As the conditionally parametric models outperformed the non-linear model, this approach was not pursued further. Instead of using non-linear models conditionally parametric models estimated via local regression have been considered. These models and a recursive estimation technique is described in the Papers A and B. In this model class the non-linearity is not explicitly parameterized, the shape of a specific relation is determined via the estimation method.

3.4 Bibliographics notes

A general introduction to time series analysis and estimation of time series models can be found in (Box & Jenkins 1976). An introduction to non-linear time series models, and a description of the Kalman filter approach applied for the non-linear model in Paper E, can be found in (Madsen & Holst 1999).

The conditional parametric model defined in Section 3.1 is similar to the varying-coefficient model defined in (Hastie & Tibshirani 1993). The varying-coefficient model defined in (Hastie & Tibshirani 1993) considers coefficients that are functions of time only, while in (3.3) the coefficients may be functions of several variables. In (Anderson, Fang & Olkin 1994) (3.3) is denoted a conditional parametric model, because when $\boldsymbol{u}_i$ is constant the model reduces to an ordinary linear model as in (3.2).

The models and estimation methods described in the Papers A and B are based on a combination of the recursive methods described in (Ljung & Söderström 1983) and local regression. Early work on local regression includes (Stone 1977, Cleveland 1979, Cleveland 1981), although, as described by (Cleveland & Loader 1996), it dates back to the 19'th century. A comprehensive overview of local regression can be found in (Cleveland

& Devlin 1988, Cleveland, Devlin & Grosse 1988).

In the estimations methods used in this thesis, selection of smoothing parameters has not been a real issue. The recursive nature of the methods implies that the data used to calculate the prediction errors is never used in the estimation. In traditional local regression cross-validation and related methods have traditionally been used for selection of smoothing parameters (Hastie & Tibshirani 1990). Also leave-one-out cross-validation is used, although (Breiman & Spector 1992, Shao 1993) argue that this is not optimal. Specifically for time series (Hart 1996) considers the subject of selection of smoothing parameters.

Chapter 4

Models and methods

The objective of this thesis has been to develop models for short-term prediction of wind power, with special emphasis on the integration of statistical and physical models and methods. Before considering such combined models, it should first be considered what statistics and physics is all about.

4.1 Initial considerations

Basically, a physical model and a statistical model are similar, in the way that both models describe a relation between some input and some output to and from a system.

The physical approach is to consider the nature of the system, and based on which laws can be determined to govern the system, a relation is derived. If the derivation was sound, the relation can be verified on measurements of the input, the system states and the output from the system.

The statistical approach works the other way around. It starts with the measurements, and based on what relations can be seen in the measurements, a model is derived.

In reality, the distinction is not as separate as stated above. Both physicists and statisticians would argue, that their approaches consists of an iterative use of both the physical and statistical approach as stated above. If the statistician has knowledge about which laws govern the system, then he knows what to look for in the data. Furthermore, in practice there is rarely an sufficient amount of data available as to describe the system response to all combinations of input data.

In the physical approach there has to be observations at some point, otherwise the physical model would just be guesswork. Also, the verification process of a physical model typically leads to identification of week points in the model derivation, and as a consequence of this, a modified model is derived.

Another important issue with regard to the physical, or deductive, approach is pointed out by (Hasselmann 1981). As described in Chapter 2, the meteorological system represents a very complex structure of coupled subprocesses of which each subprocess represents a detailed discipline in its own right. These subprocesses interact across a wide spectra of space and time scales in a complicated manner, which ultimately determines the dynamics of the complete system. If detailed models of the dynamics of each subprocess existed, and the necessary computing power was available, one could try to obtain a total model by coupling the individual subprocesses together in a very comprehensive numerical model. It is, however, questionable whether such a deductive approach would be successful (Hasselmann 1981)

It is argued (Hasselmann 1981) that a larger probability for finding a useful model is obtained by the opposite approach, i.e. the inductive approach. This approach contains an attempt to identify the governing interactions in the system by statistical methods. Once such a model is obtained for the most important variations, the model can be iteratively improved by a combination of more detailed comparison with data and use of well-established physical facts.

These considerations are also in accordance with the findings in (Jonsson 1994), which argues that if errors are present in the regressors the use of the true system for prediction will not result in optimal predictions.

4.2 What are the options?

The purpose of this section is to identify the most important variables in developing prediction models for wind power, and outline ideas for how statistics and physics can be combined. Based on the description of the meteorological theory in the previous chapter, the variables which can be assumed to be most important in the description of the wind speed and consequently the power production at a given location are: the wind speed and direction from some model level from the numerical weather prediction model and the turbulence intensity. The energy content in the air flow is related to the density of the air, this variable is therefore also of importance when wind power is considered. As no measure of this variable has been available, this dependency has not been considered.

4.2.1 Direct use of statistics in HIRLAM

The numerical weather prediction model, HIRLAM, has not been the subject of development in this study. All developed models and methods start from the variables which have been delivered by HIRLAM.

This does not mean, though, that statistical models could not be used in the numerical weather prediction model. One way to include a statistical modelling approach would be to use stochastic differential equations directly in HIRLAM. Such an approach could for instance be applied to the HIRLAM soil model described in Section A.4. Stochastic versions of ordinary differential equations are easily derived, see e.g. (Melgaard 1994), while it is more difficult to work with stochastic versions of partial differential equations. Partial differential equations are usually transformed into ordinary differential equations using the same approach as in HIRLAM for the soil temperature prognostic equations, where the soil is divided into three layers. These equations could therefore easily be replaced by stochastic versions, and statistical estimation

methods could be used to estimate the parameters in these equations.

The potential advantage of integrating this approach in the numerical weather prediction model is that the drag imposed on the flow in the atmosphere is directly related to the turbulence intensity as described in Appendix A, and the surface temperature and wetness are necessary in the calculation of the turbulence intensity.

Stochastic versions of the soil model equations could also be used outside the numerical weather prediction model. This approach could be used to predict e.g. the surface temperature locally at the wind farms. The temperature could then be used in calculation of the surfaces fluxes of the sensible and latent heat fluxes, which in turn could be used in the stability dependent wind speed profile. This approach has not been tested in this thesis, the main reason is that some of the necessary variables were not available, and that the three hour time resolution which the available data is given in, would not be sufficient for this approach.

Another area for potential improvement of the numerical weather predictions, is to consider the data assimilation procedure. The 3-dimensional data assimilation procedure is briefly described in (Sass et al. 1999) and in more detail in (Lorenc 1981, Lonnberg & Shaw 1987). The current assimilation procedure used in HIRLAM consists of several steps. Bilinear interpolation is used in the vertical, tension spline interpolation is used in the vertical, subsequent averaging is applied followed by a covariance dependent weighting between predicted values and corresponding observations. Nevertheless, in the description of the assimilation procedure, there does not seem to be any feedback from the prediction errors to the assimilation procedure. The introduction of the feedback concept in the assimilation procedure is beyond the scope of this thesis, but is definitely worth pointing out as an area of potential improvement.

4.2.2 Intermediate models

Potentially statistical models possess the ability to outperform any physical model in performance, but it should be noted that the development of statistical models is based on finite data samples, and it is possible that some relations exist which are not visible in such a sample. Generally,

as the number of explanatory variables increases, the noise in the data might severely contaminate the relations which are found by statistical methods. In local regression this problem has be dubbed the "curse of dimensionality", and in this context the problem is particularly difficult to cope with, because very weak assumptions are imposed on the relation between the explanatory variables and the response. In practice this means that one needs to select the explanatory variables which provide the highest degree of explanation.

Therefore, it might be advantageous to incorporate known statistical relations in statistical models, such relations can be viewed as intermediate models, linking one or more explanatory variables to one single variable which can be used as input to a subsequent relation. This approach can therefore effectively reduce the dimension of the problem. The parameters of a known relation might be estimated using statistical estimation methods, or only a subset of the parameters could be estimated. Known relations could also be used directly to model some parts of the relation between the explanatory variables and the response. Physical systems are often described by differential equations, and, from the description in the previous section it is clear that such relations can also be combined with statistics.

4.2.3 The Risø model

From the considerations in the previous section it is clear, that one way to combine physics and statistics in short-term prediction models, is e.g. to use the relations in the Risø system to calculate the wind speed at the wind turbine hub-height.

In Section 2.3.5 it was shown that the wind speed ω_h at the wind turbine hub height calculated by the Risø system could be written as

$$\omega_h = a(\theta_s)G \tag{4.1}$$

where G is the geostrophic wind speed and θ_s is the surface wind direction. It is seen that this corresponds to a conditionally parametric model if an error term is added. In Paper E models of this type have been examined. In this paper it is not pointed out, that in a statistical model we need to make a choice with regard to the wind direction.

In Section 2.3.5 it was shown that the surface wind direction depends non-linearly on the geostrophic wind speed and direction, and therefore the surface wind direction should be calculated by the geostrophic drag law. It turned out that it made no difference in the performance results whether the geostrophic wind direction was used directly instead of the calculated surface wind direction. Furthermore, the results in Paper E indicate that the physical wind direction dependency used in the Risø system is not optimal.

When a conditionally parametric model where the parameters are considered to be functions of the wind direction is used, it should be noted that the wind direction dependency does not necessarily correspond to the dependency found by the WAsP application. In (Troen & Petersen 1989*b*) wind roses for a large number of locations in Denmark are shown, and from these wind roses it can be verified that the average wind speed depends on the wind direction, e.g. the wind speed is usually higher when coming from South-East compared to North or West. Therefore, this must be a feature of the overall atmospheric flow not local conditions. As long as the predictions from the numerical weather prediction model are not perfect, this dependency is automatically included in the conditionally parametric model. This means also, that if a WAsP analysis finds that local effects prescribe a correction of the wind which is not in accordance with average wind direction dependency, then the WAsP corrected wind might lead to worse performance than the uncorrected wind.

The reason for considering the non-linear model in Paper E is also based on the considerations in the previous sections. The non-linearity is introduced at the final stage in the Risø system, where the wind speed in transformed into power. The results from this model were also rather disappointing, as a simple linear model, where the power curve was approximated by a polynomial in the wind speed, performed just as well. The final conclusion of these findings is therefore that the physical relations used in the Risø system are of little use when there is data available, and statistical models can be used. It should be noted though, that if no data is available, then the physical relations can be used, also, the statistical models need one to three months of data before the parameter estimates are fully reliable, therefore the physical relations can be used until the statistical models are fully reliable.

4.2.4 Turbulence intensity

From the description in the previous chapter it is clear that the relation between the wind speed at two different heights above the ground depends on the turbulence intensity. If the numerical weather prediction model handles this dependency properly, then the model level which is closest to the wind turbine hub-height can be selected. The approach taken here has been, first of all to find which model level gives the best performance, the results from this examination are described in Paper I. Secondly a closer look is taken at the winds at different model levels in HIRLAM, and how these winds depend on the Buoyancy flux, which is a measure of turbulence intensity. The results from this examination are described in Paper I. The purpose of the analysis is to find out if it is necessary to include the turbulence intensity dependency in the statistical models.

Chapter 5

The implementation – Zephyr

The aim of this chapter is to describe the software application Zephyr. Zephyr is implemented in the Java^{TM} programming language, and it is assumed that the reader has some basic knowledge of object oriented programming and terms used in object oriented design. A comprehensive description of the Java programming language and the Java Application Programming Interface (API) is available at `www.java.sun.com`

The first version of Zephyr is planned to go into operation in May 2000. This beta version will be evaluated by the two Danish utilities Elkraft and Elsam.

The reason for choosing the Java programming language, is that programs written in this language are not limited to run on one computer platform. Java programs can run on all platforms, which have a Java Virtual Machine (JVM) available. The list of platforms for which the JVM is available is long, and includes Windows95/98/NT, Linux, OS/2 and Sun.

The level of detail in this description of the Zephyr system is rather low, as a complete description of all objects and object methods would be

far too comprehensive. The objective of the description is to provide an overview of the system and outline the system architecture. A description of some of the object patterns used in the implementation of Zephyr can be found in (Reese 1997) and a more genereal description of design patterns can be found in (Gamma, Helm, Johnson & Vlissides 1994).

The description of the server part of the Zephyr system focuses on the software architecture and some level of detail in the implementation. The aim of the description of the client is to illustrate the graphical user interface; no implementation details are provided on this part of the system.

5.1 Requirements to the application

The requirements to the Zephyr application are based on experience gained by the Danish utilities Elkraft, Eltra and Elsam, by the use of the Risø and the WPPT on-line systems. These two systems are described briefly in Paper H.

There has been no formal requirement specification to the system, although, as a minimum, all the tools which are available in the applications mentioned above, need to be available in the new system. Zephyr has been developed in a process, where project meetings with participants from the utilities has been the forum where the requirements to the system have been formulated. The following description summarizes the requirements obtained from this process:

- Any prediction application, which is to be a useful tool in the daily dispatch and control in the utility control room, requires that the application is highly automated and easily customizable.
- In general there are three sources of data to the system, numerical weather predictions from a national weather service, measurements of meteorological variables and power production from the wind farm sites. The system needs to maintain a database of this data, validate the data and update the prediction models and calculate predictions as soon as new data becomes available.

- One or more users will be connected to the system simultaneously, observing the predictions. Also, the system needs to show information related to the validity of the predictions and how the farms connected to the system are currently operating.
- Wind farms may need to be activated/deactivated, and as new wind farms are constructed, these farms will need to be added to the system.

5.2 System architecture outline

Client/server is an application architecture that divides processing among two or more processes, often on two or more machines. The idea behind a client/server application architecture in a database application is to provide multiple users with access to the same data, and this is exactly what is required by the short-term prediction system. The data, which consists of measurements from the wind farms, numerical weather predictions from the weather service, the short-term power predictions, and so on, has to be accessible from several clients.

In a two-tier system, which is the simplest one, the processing is divided into a data storage layer and a presentation layer. The short-term prediction system is not suiteable for implementation as a two-tier system. A middle tier which handles the business logic rules is necessary. The system is divided into three main layers as in a three-tier system; the data storage, the object server and clients. The object server is the middle-tier, and this is where the business objects reside. To allow for extra flexibility and scaling several object services can exist in the system. This requires a naming service where the object services are registered, which can be used by clients and/or other services to look up the object service for a particular type of business objects. A business object in the Zephyr system is e.g. an object which represent a wind farm or a table which contains wind power predictions. Some of the Zephyr business objects will be described in Section 5.4.

Some operations involve several business objects, e.g. when the predictions are calculated input data is taken from several data tables, and the results are put into some other tables. Therefore it is necessary to

have a mechanism that groups several operations into one atomic operation, which either succeeds or fails. This is handled by introducing transactions and object locking. The locking mechanism is required in a multi-user environment in order to avoid inconsistent object states. This can happen if multiple users are modifying the same objects simultaneously.

The general framework is outlined in Figure 5.1. Due to the complexity in the information flow between the services, this flow has not been illustrated in detail in the figure.

The services shown in Figure 5.1 will be describe in more detail in the following sections. As illustrated in the figure the Remote Method Invocation (RMI) API (`www.java.sun.com`) is used for the communication between objects residing in different services. The services themselves might reside on different JVM and on different computers.

All communication from the clients to the business objects residing in the object services goes through a special caching mechanism. Objects residing in the clients never communicated directly with the business objects, the communication goes through the cache/proxy. Some operations involving manipulation of the business object state might be executed inside the proxy, not inside the business object. The details of this concept will be described in Section 5.4. The key point of this concept, is that a JVM running the object service should not be burdened with to many operations. The reason is that the business objects potentially are being accessed by several clients and services simultaneously. Therefore this concept increases the responsiveness of the object services.

5.3 Services

The most important services in Figure 5.1 are the object services, the other services can be thought of as utility services, which are necessary for performing operations on the business objects which are residing in the object services.

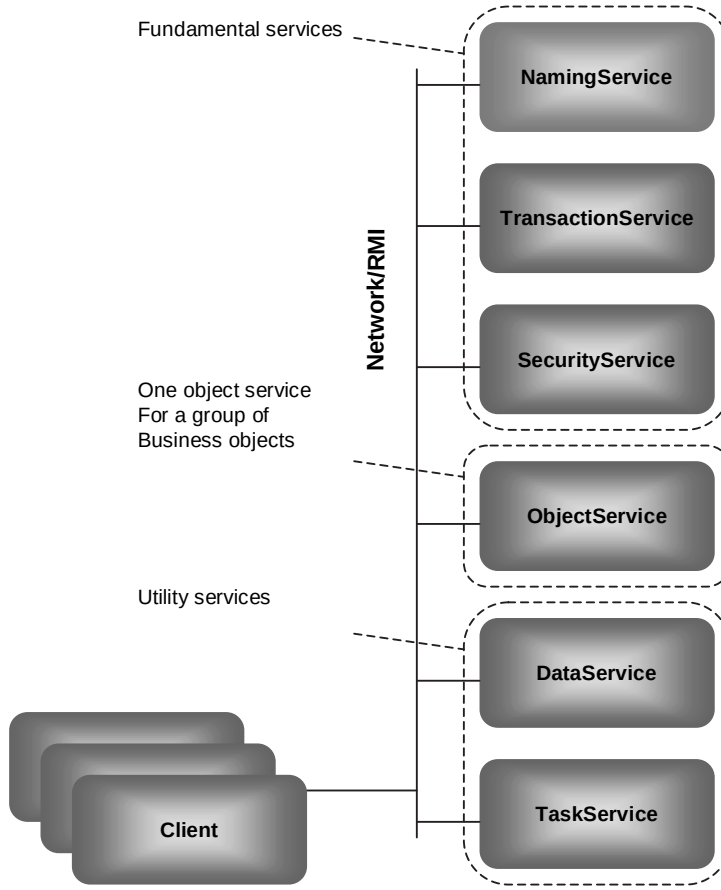


Figure 5.1: Overview of the basic components in the distributed system

Two special services are also shown in the figure, which are the data and the task service. These services are more or less Zephyr specific, and the distinction between a service or a client is not so clear for these two services. In one sense the data and the task services can be considered as clients, as they are being served by some business objects.

5.3.1 Naming service

The naming service is simple. Each object service registers itself with the naming service, and when this happens, the object service informs the naming service about which business objects the object service is serving. Clients or other services can then use the naming service to look up the object service for a particular group of business objects.

In any well designed object oriented system an object instance should represent a unique identity. There should never exist two or more instances of an object which represent the same identity. This might seem obvious and easy to implement, but this is not as simple as it seems. For instance, consider an object which represents a specific wind farm. If two instance were created of this object, and two users were making changes to the state of these two instances, then we would end up with two different states representing the same wind farm. Therefore, each business object needs to be identified with a unique id. As the naming service is the registry of all object services, the naming service has the potential to generate unique id's. The id is represented by a long value, which means that there are roughly $2 \cdot 10^{19}$ unique id's available.

The algorithm which is used to generate id's is simple. The naming service does not generate the id itself, it generates seeds which are used by the object services to generated id's for business objects that are created in a particular object service.

5.3.2 Object service

An object service is a business object manager. The problem with object uniqueness was mentioned in the previous section. It is the object service

responsibility to keep track of the business objects. If a request is made for a business object with a particular id, this id is compared to a master list of business object instances. If the instance is in the master list, a reference to the business object is returned. If the id is not in the list, the object service creates a new instance of the business object. The state of the object is restored from the object store, the id of this object is added to the master list and a reference to the object is returned. The object service also manages the creation of new business objects and generation of id's for new business objects.

The master list is implemented as a hashtable with a special kind of weak references to the business objects. This has to do with the Java implementation of garbage collection. In the distributed garbage collection, the garbage collector thread has to poll the remote references which exist to the objects in the JVM. The master list request information from the garbage collector about references to business objects. When only one references exist to a business object, then this must be the reference in the master list. When this happens the object reference is removed from the master list, and the object is now free to be garbage collected.

The reason for using this implementation has to do with performance. There is no need to fill up memory with business objects that are not being used, therefore, business objects which are not referenced should be garbage collected.

The object service also has some utility methods for browsing the business objects in the service, like information on how many objects of a give type are in the object store and methods for retrieving references to all objects of a given type. Furthermore, lookup queries can be executed on the descriptor property of business objects. The matching semantics and the descriptor property will be described in Section 5.4.

5.3.3 Security service

The role of the security service is to authenticate users. Based on a user name and a password the security service generates a unique digital signature. The signature is created using the SecureRandom class in the Java API, and consist of 32 bytes.

Users are authenticated from a user database, where the user names and passwords are stored. A signature is only created if the user name and password match. All access operations on business objects are validated with the security service. As it is the security service which creates the signatures, it is the only instance which knows the particular combination of user name, password and signature. This concept therefore provides a high level of security.

5.3.4 Transactions and Transaction service

As mentioned previously, there is a need to group a multitude of operations into a single atomic operation in a distributed system.

Three things happen when a user makes the first write request on a business object. The first step is to create a lock on the object. The lock is identified by the users signature, if the object is currently locked, and the signature holding the lock is not the same as the one requesting the lock, an exception is thrown. The next step has to do with the transaction. The transaction contains a factory method, when a write request is made on a business object, this transaction factory method is called by the business object, passing on the signature of the user who made the write request. If a transaction for the signature does not exist a new transaction is created and this transaction asks the transaction service to register this transaction. The transaction is responsible for making a copy of the business object state on the storage media, before the first write actually occurs. If the two first steps succeed, a snapshot is taken of the business objects state.

The transaction keeps track of all modified business objects which are being modified by a user with a particular signature. Each time a write request is made on a business object, the business object uses the transaction factory method to get the transaction for this signature. The business object then informs the transaction that the object has been modified.

One single user might have several transactions, this is the case if the user modifies business objects in several object services. As shown in Figure 5.1 the transaction instances reside in the object services, and the

transaction service holds a registry of signatures to transactions. Once the user has finished modifying business objects, he only needs to make a single call to the transactions service with his signature, requesting that all his transactions should be saved.

The transactions service now calls the save methods in the transactions, which is turn call the save methods in all the business objects which have been modified as a part of the transactions. If the save operations succeed, the transactions service call commit on all the transactions, which is turn call commit on all the just save business objects.

Several things can happen during a transaction. A network error might occur, and if the user is only half way through his operation, the safest option is to restore the business objects back to the original state before the operation started. The user can then call the transaction service, requesting a roll back of his transactions. This results in a roll back or restore of all objects which have been modified so far as a part of the transaction.

If an error occurs between the save and commit phase of the transaction, the transaction is responsible for restoring the copy of the business object state on the storage media. The business object itself restore the state from the snapshot which was taken of the object before the first write occurred on the object.

If the user disappears, i.e. in the case of a network error, then the lock will not be renewed and the lock takes care of asking the business objects to restore themselves.

5.3.5 Data service

The data service is responsible for feeding data from the external sources into business objects residing in object services. These business objects, which in general are tables and columns, will be described later.

How the external data is provided to the data service depends on the particular utility. The data can for instance be fetched from a ftp server. This is how the data from the Danish Meteorological Institute is provided

to the system, WPPT, running at the Danish utilities Elsam, Elkraft and SEAS.

Without going into details, the data service cycle is the following: The data service is running at specific time intervals. At each run the service checks for new data. When new data is available, the first step is to validate the data. Zephyr uses the validation model from the WPPT system described in Paper G. Once the data has been validated, the data is put into the target business objects, this happens via the business object proxies. A task is then created which contains the necessary code for calling update and predict methods in the affected business object proxies. The task is then sent of to the task service which executes the task, and the data service is ready to receive new data.

The separation of data retrieval and model calculation is important. The measurements are receive at a rather high frequency, namely in the order of minutes, and it can not be guaranteed that the model updates will be finished within this time frame. Therefore, the data service cannot wait, it has to be ready to receive new data before the model calculations are finished. The data which is received at high frequency does not need to trigger model updates, it is only when new weather predictions are received that the models need to be updated, in the current setup this is each 6 hour. This way the models can be quite complex, as they have a considerable time to finish the calculations.

5.3.6 Task service

The task service is quite simple. It executes tasks which are handed to it from e.g. other services or clients. A task is an object which implements a specific interface. The interface only requires that the task object implements a run method. The task service starts the task by calling this run method.

When a task is handed to the task service from some source service or client, the task service requires that the source specifies a reference to a remote object which implements a specific interface. This interface requires that the remote object implements a finish method. Once the task run method has finished running in the task service, the task service

calls the finish method in the remote object. This way the source of the task is notified when the task has finished.

Several task services can be running on different computers, and e.g. the data service has the option to spawn of tasks on several different computers. This can be important if Zephyr is used by a large utility owing hundreds of wind farms, as this might lead to calculation times which cannot be performed within a reasonable time on a single computer. This way wind farm models which do not depend on each other, can perform their calculations simultaneously on different computers.

5.4 The business object super-classes

5.4.1 Object relations

Figure 5.2 shows a simplified overview of the relations between the various super-classes for the business objects and the client to server mapping between the business objects and the business object proxy. A business object is an object which inherits from the Persistent object, the issues related to persisting the business object and the fact that the object can be shared by several other objects, is surfaced in the SharedServerObject and the Persistent classes.

Each business object has two objects and one interface related to it. The business object proxy or the ClientCache has already been mentioned in Section 5.2. Each specific business object which inherits from Persistent has a corresponding proxy which inherits from ClientCache. The second object is the business object persistence handler or peer. This object surfaces storage details for different storage media. The interface is a remote interface, which specifies which methods in the business object can be invoked from objects in other JVM.

The business object itself resides in the object service, but other objects residing in e.g. - a client or in another service should never communicate directly with the business object. This communication should be surfaces through the proxy.

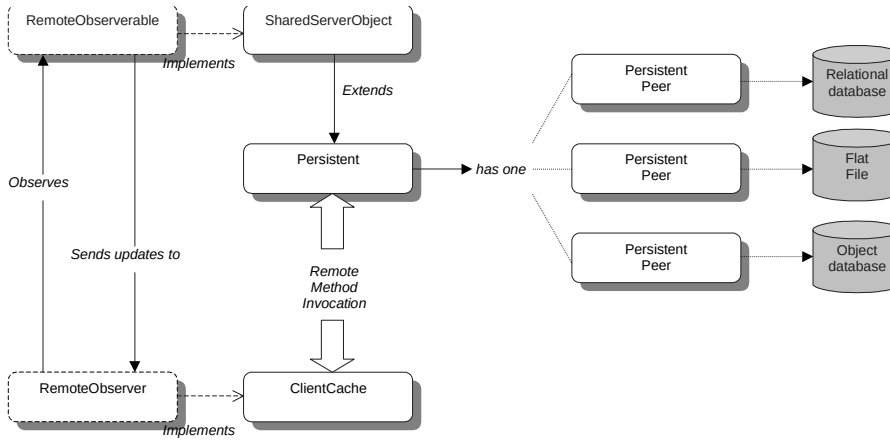


Figure 5.2: Simplified object model for the business object super-classes and related objects.

The state of the business object is persistent, i.e. - if the machine where an object service is running on goes down, the business objects in that service can all be restored. Furthermore, if a business object is not being used by any client or other service, this object is removed from the object service memory, and only restored when requested.

5.4.2 Distributed event model notification

A synchronization mechanism exist between the SharedServerObject and the ClientCache. The purpose of this mechanism is to provide a convenient way for a business object and the business object proxy to synchronize their states. The mechanism is similar to the event model pattern used in user interface components, but special considerations needs to be taken into account when the event model is transferred to a fragile network connection.

The synchronization mechanism is simple, and the reason for this has to do with the fact that it is not feasible to use complex event semantics over a network. The methods which are involved in the synchronization are defined by the two interfaces RemoteObserverable and RemoteObserver shown in Figure 5.2.

The RemoteObserverable takes two methods addObserver and deleteObserver. The RemoteObserver only requires an update method to be implemented. The add/delete methods manage a registry of ClientCaches which are interested in knowing about business object state changes. When a state change occurs in a business object, the business object notifies all registered ClientCaches by calling the update method. A special object called RemoteEvent is passed to the ClientCache via the update method. This object describes the type of change which occurred in the business object.

Some key points in this process are of importance. It can not be guaranteed that the calls to the update methods are performed chronologically, therefore one of the property in the RemoteEvent is a sequence number. This is an ever increasing number, and it is guaranteed that the sequence number increases between two subsequent state changes in a business object, the sequence number can therefore be used to order the state changes chronologically. Performance is also important, the RMI implementation in Java implies that a remote call does not finish before the remote object method finishes. Therefore, the notification from a business object to different proxies in different JVM is performed in threads. Similarly, the consumption of the event in the proxy is handled in a thread, which means the call to the proxy finishes as fast as possible. The thread which is created in the proxy, is then pushed into a thread queue, and the treads are executed one at a time as it can not be guaranteed that the proxy environment is thread safe. For instance the graphical user interface API, Swing, which is a part of the Java Foundations Classes (JFC) API is not thread safe.

5.4.3 Object locking

As shown in Figure 5.2 the persistent object inherits from the SharedServerObject. This object handles the logic involved with the fact that the object potentially is accessed by several other objects simultaneously. Write or update access to the business object results in a lock on the object. Before the first write occurs on a business object, a snapshot is taken of the object. The snapshot is not discarded before the write accesses are explicitly committed.

The lock is associated with a digital signature created by the security service, which uniquely identifies the user holding the lock. If the lock is not renewed within a fixed amount of time, the lock is released and the object is restored back to its original state. The state is restored using the snapshot which was taken of the object before the first write was performed on the object.

5.4.4 Persistence

As shown in Figure 5.2 the persistent object is associated with a persistence handler/peer. This handler is assigned to the business object when a new instance is created, and a particular type of business objects share the same handler. The type of handler which is assigned depends on a configuration parameter in the system, and different handlers can therefore be selected based on how the business object is actually made persistent. In the current system flat file persistence handlers are used, the handlers are capable of reading the state of a business object and assigning a state to a business object; the state itself is saved/loaded in/from a binary file. This architecture implies that the storage details are transparent to the business object, changing the storage media means that only the handler needs to be replaced, the business objects stay unchanged.

5.4.5 The descriptor

All business objects are associated with a descriptor property. This property is a simple object, which only contains fields of primitive types, i.e. String, Integer, Long, Float, Double. The idea of this property is to facilitate fast and easy business object searching based on intuitive criteria. A descriptor object can e.g. contain fields like manufacturer, make, model, name, etc.

A query can be executed via methods in the object service managing the business objects of interest. The query itself is represented by a query object, which is similar to the descriptor object, i.e. - the query object only contains fields of primitive types. The matching semantics is the

following:

- If a query object field is null this field is interpreted as a wildcard and matches all fields with the same name in the descriptor object.
- If the query holds a named field not defined in the descriptor or vice versa, these fields are ignored, i.e. do not influence the match.
- All non-null identically named fields and identical field values are considered as a match.

This architecture implies that instances of the business objects do not need to be created as a part of a query, only the descriptor object is needed. Potentially this means faster query responses, and the descriptor/query object information pattern can be defined in a uniform way; several different business objects can use the same implementation of a general descriptor object.

5.4.6 Security

All access to a business object is validated in the security service. Access to a business object is categorized as either create, read, update, delete or observe. A user has the privilege to perform a set of these access types on the business objects. The security service holds a user database, where the information on the user privileges is stored. In the case of illegal access the business object throws an exception.

All business object implementations need to follow this convention in the business object accessory methods. The SharedServerObject has methods that business object implementations can use to facilitate this convention.

5.5 Zephyr business objects

The Zephyr system is based on a hierarchical structure of a utility owning several wind farms, where the wind farms are placed in a larger

geographical area.

The objects which are outlined in the following sections are simply the basic building block, which are used to create a software model of the wind farm setup of a particular utility. Some objects are omitted in this description, as they are not important for understanding the basic structure. The overall software system is very flexible, there is no limit on adding new types of business objects. Such objects could e.g. be representations of primary power plants, district heating systems, models for automated scheduling of conventional power plants and so on.

5.5.1 Region

The region corresponds to a geographical area. A region can have an arbitrary size, and regions can be nested, i.e. - a region can contain several other regions, which are inside the boundary of the geographical area represented by the region. It is the utility who defines the region, and the region is not static, it can vary in size during time.

The reason for using regions is to have a way to group wind farms geographically. The electrical utilities are interested in predictions on several geographical levels, i.e. asking questions like what is the expected production in the North-east of Jutland?

The way the region is implemented is quite general. Actually no geographical area checking is done by the region object, it is the responsibility of user to define the regions in an consistent way.

The region object is a list of arbitrary business object references. The implementation has methods for adding, removing and retrieving business objects. This object can thus be used to define regions consisting of e.g. - wind farms, prediction models calculating prediction for the region, tables of data holding the predictions and sub-regions.

5.5.2 Farm

The wind farm object is modeled after a general wind farm. The wind farm is mainly a collection of references to other business objects. These objects are tables containing the measured data at the wind farm, tables with numerical weather predictions of variables like wind speed, direction and temperature for the location of the wind farm, and tables of the power predictions and related uncertainty fractiles for the predictions.

The wind farm also contains a model, the model is not a business object, it is a property of the wind farm. The model which is assigned to the wind farm depends on the data which is available for the farm. A user has the option to select between several models for a wind farm, if for instance no on-line measurements are available for the wind farm, the user can select a model which is identical to the model used in the Risø system. The generic model is an implementation of the model which was found as the best one in Paper E, and this model requires on-line measurements of the power production at the wind farm.

As mentioned previously, the update of the model parameters and the calculation of the predictions is triggered by the data service. When new data arrives at the data service, these data are put into tables of each wind farm. The data service the sends a task to the task service, and when the task is executed in the task service, it calls the relevant methods in the farm object to inform the farm that new data is available.

A user can change data in the tables related to a wind farm. This can happen if the automatic data validation system has not captured some erroneous data. Wind farm models which are based on recursive statistical methods need to be re-calibrated as a consequence of such a change. This does not happen automatically, the user needs to trigger an event which asks the wind farm model to re-calibrate. Such events can be generated via the user interface. It could be argued that the model should be consistent with the data at all times, therefore, model updates should be triggered automatically every time data is changed. This is not a feasible solution in the Zephyr system, as a model update may take a considerable amount of time, performance issues require that this logic is handled outside the Farm object.

5.5.3 RegionModel

The structure of the region model is closely related to the wind farm object. This object contains tables of data and a model for calculation of predictions for a particular group of wind turbines in a particular region. The model update and prediction is triggered in the same way as for the wind farm object.

The main difference lays in the model and which data is stored in the data tables. The model which is used is very simple, it is based on upscaling constants as in the WPPT system described in Paper G. No meteorological data exist to represent a region model, as such data corresponds to singular points, therefore the region model object only contains predictions of power and the related upscaled measurements of total power produced by a group of wind turbines in the region.

5.5.4 Columns

The column business object represents a collection of simple objects, where each object is referenced by an index. The are several types of columns, depending on the indexing mechanism which is used. No distinction is made between columns containing different types of objects, instead the column has a parser and a formatter property, which can be used to transform between textual and internal representation of the objects contained in the column. The generic column is the time column. In this column the indexing is based on time, and without going into details, this approach allows for easy access to instantaneous time indexed data.

5.5.5 Tables

All columns are contained in tables, and a table is simply a collection of columns. The table contains methods for accessing the columns. All columns have a name property which can be used to retrieve columns from a table, columns can also be retrieved based on simple integer indexing, columns can be added and removed from a table and so on.

A special type of table which is used in Zephyr, is the prediction table. This table has a special indexing which is related to the time step between subsequent predictions and the time step between each retrieval of new predictions. If the prediction horizon is 36 hours and the time step between each prediction is 3 hours, then this table will contain 13 columns. This table also has another interpretation, which is based on only two columns, a timestamp column and a data column, where the timestamp column contains the valid time for the predictions contained in the data column. This representation is obtained by using the prediction table as input to a filter table, which transforms the input table into a table only containing two columns, a timestamp column and a data column. This filter table is e.g. used to construct prediction plots.

5.6 Clients and user interface

The system architecture as presented in the previous sections, is ideally suited for a multitude of small clients, each dedicated to a particular task. E.g. depending on which information is wanted by a particular user, or a group of users, dedicated clients can be developed. Some users might only be interested in knowing the predictions for one larger area, while other users might be interested in knowing the current running conditions at a particular farm, e.g. the maintenance crew. When a new wind farm is constructed, this farm has to be added to the system and the system has to be re-configured, this task can also be handle by a specific client.

The idea behind the client and the client user interface which is presented here, is to provide a way for the user to access and perform all the above mentioned tasks. The user can configure the client to his or hers particular needs.

5.6.1 Main user interface

Figure 5.3 shows a screen shot of the graphical user interface (GUI) implemented in the Zephyr system. The GUI layout is project oriented, where a project is illustrated by a project tree. The system operates

with two project definitions, and these two projects are illustrated by graphical tree components on the left hand side of the GUI. The lower left tree illustrates the project defining the system setup of a particular utility, and the tree in the upper left corner illustrates the project of a particular user.

The system setup of a particular utility is organized hierarchically. The basic objects in the system setup tree are regions, wind farms and tables of data, where data related to the regions and the wind farms are stored. These objects or the tree nodes are directly related to the business objects described previously. The region object defines the properties of a larger region, containing one or more wind farms. The properties of a region are e.g. - the wind farms in the region, the model which is used to calculate total power predictions for the region and tables of measured and predicted total power production for the region. Similarly, the wind farm object defines the properties of a wind farm, like the model for calculating the wind farm power predictions, and tables of data containing measured variables at the wind farm. Furthermore, the wind farm object contains the state of the wind farm, i.e. - if the farm is currently running or not and the current wind farm power production. The system setup is customizable via the system setup tree and node specific popup menus, allowing super-users to customize the system setup, i.e. wind farms can be added or removed, regions can be redefined, the models connected to the wind farms and the regions can be changed and so on. These changes are applied at runtime, and the system does not need to be restarted. Such a change will in general correspond to adding, removing or changing a business object in an object store. Therefore, these changes will be visible to all clients, and, as this is the system setup or utility wind farm setup model, only super-users are allowed to perform such changes.

The user project tree holds the structure and definition of the windows which are shown in the right hand side of the GUI, each window is illustrated by a node in this tree. The user has the option to customize how the windows are organized to fit his or her's particular needs, by inserting, moving and deleting the position of the window nodes in the user project tree. Apart from the default windows defined by the system, like the wind farm window, windows showing plots of predictions and measurements, the user can create customized windows either using scripts or via dialog windows. When a new window is created, the user

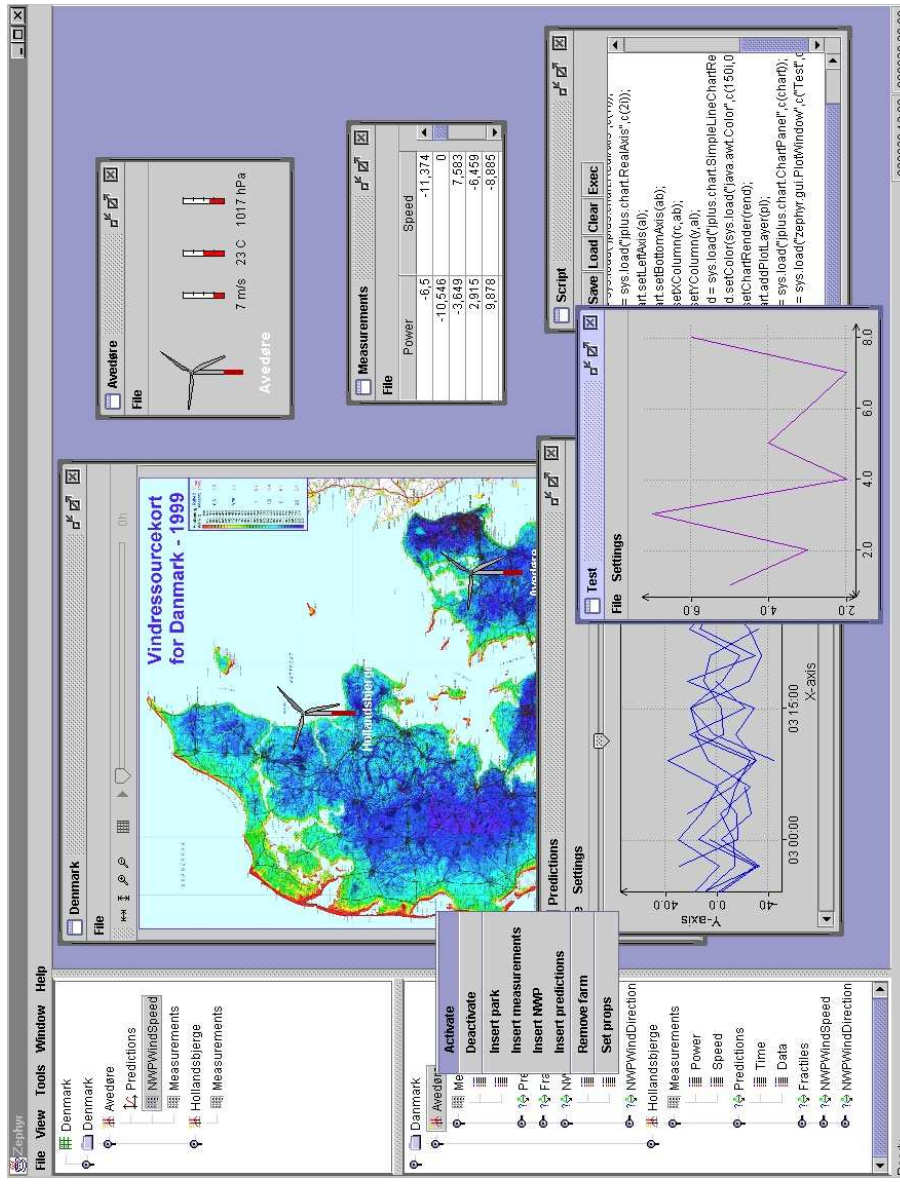


Figure 5.3: Screen-shot of the Zephyr graphical user interface

has to select the node in the tree where the new window will be placed as a sub-node. Each user is not limited to having only one project tree, the user project tree can be saved and reloaded, allowing the users to operate with several trees, i.e. the user can define several user trees, where each tree is fitted to one particular operational situation. Each window node in this tree links to one or more business objects, illustrating the information contained in the business objects.

5.6.2 Internal windows

The right hand side of the GUI screen shot in Figure 5.3 shows some of the windows in the system, such windows are referred to as internal windows. The windows which are shown in Figure 5.3 are: the map window, a table window, a prediction plot, the wind farm status window and the script window. The map window shows the area which the system is operation on. This map contains wind farm symbols, which illustrate the current running conditions at the wind farm, i.e. the wind farm status. Furthermore, the map window servers as background for weather animations created from the numerical weather predictions. The current implementation is capable of showing the predicted meteorological variables as contour animations.

Some of the internal windows in the Zephyr client are shown below. Figure 5.4 shows a time series plot window, and Figure 5.5 shows a combined scatter and line plot of a power curve.

The slider on the top and the scroll bar on the bottom of the time series plot can be used to specify the visible time period of the time series. The slider specifies the time window width, and the scroll bar specifies the start and end times.

Figure 5.6 shows the dialog window for adding a layer to a plot window. A layer defines the x and y data, the axis settings and the plot type, i.e. scatter or line. The data is selected by dragging a node, which represents a column, from the tree to the fields in this window.

Data can be edited via the table window, an example of a table window is shown in Figure 5.7. Editing a number in a table from a client auto-

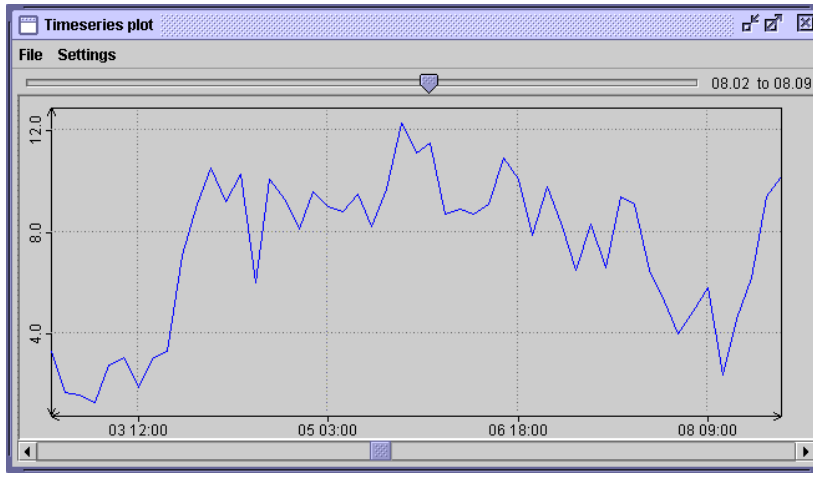


Figure 5.4: Time series plot window.

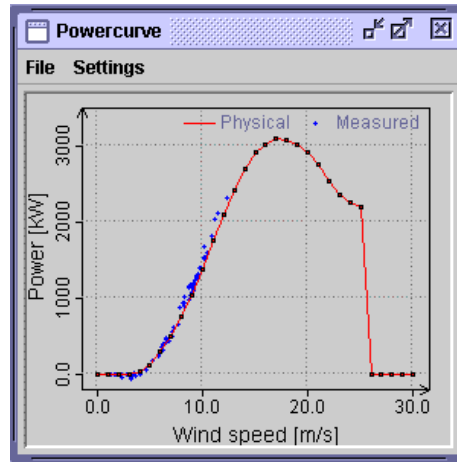


Figure 5.5: Power curve plot window.

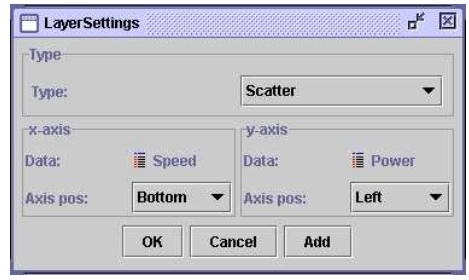


Figure 5.6: Plot layer window.

matically updates all relevant plots, tables, etc. in all connected clients. Such an actions does not automatically force a model update. If e.g. data related to a wind farm is modified then a wind farm model update has to be requested via the wind farm node popup menu.

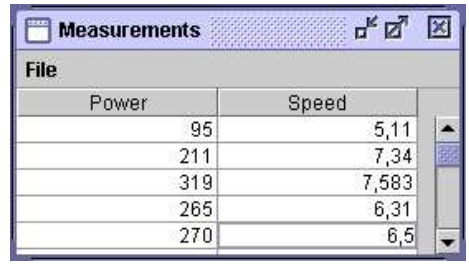


Figure 5.7: Table window.

Figure 5.8 shows the wind farm status window. This window illustrates graphically the current state of a wind farm. The state is defined as the current production, shown as the dark fraction of the wind turbine tower, the direction, shown by the dark gray turbine blade and other variables shown by the meters in the right hand side of the window.

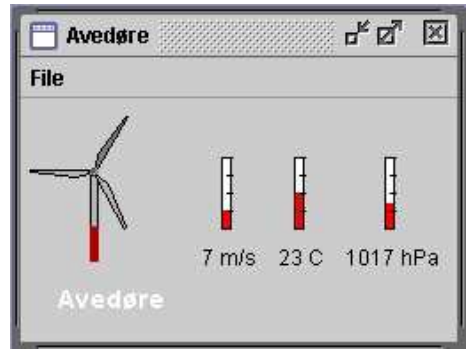


Figure 5.8: Farm window.

Chapter 6

Conclusions

This thesis deals with a number of aspects related to short-term prediction of wind power. The thesis is structured as a summary report and a collection of ten research papers. In the summary report the background and the motivation for the Ph.D. study are outlined. Bibliographic notes to previous research within the field of short-term wind power prediction are provided. The meteorological theory, which is particularly relevant for wind power prediction is described and the physical approach used in the system developed by Risø National Laboratory is analysed and described. A short introduction to the statistical models that have been considered is provided together with bibliographic notes to the literature for a more detailed description of the models and in particular the estimation methods for the models. Some general consideration with regard to statistical versus physical modelling is outlined, and possible areas of combined statistical and physical models for short-term wind power prediction are described. Finally, the summary report describes the client/server software application that has been developed. The included papers can be categorized into two groups, development of general statistical models and methods, and development of dedicated short-term wind power prediction models.

The conclusions will be structured according to the main areas considered in this thesis and in the Ph.D. study. In short, these areas are the development of general statistical models and methods for on-line estimation and prediction, development of models and methods for short-term prediction of wind power and implementation of the prediction models in a client/server software application.

6.1 Statistical models and methods

The Papers A and B describe new estimation methods for statistical models. Paper A considers on-line estimation of linear models, where the parameters to be estimated exhibit smooth variations in time. An estimation method for derived from local polynomial regression is suggested, using local polynomials in the direction of time to approximate the parameters locally. The results presented in the paper indicate that the method is superior to the classical Recursive Least Squares (RLS) method, if the parameter variations are smooth.

In paper B a method for on-line and adaptive estimation of Conditionally Parametric Auto-Regressive eXtraneous (CPARX) models is derived, and some of the properties of the method are analyzed. This method can be interpreted as recursive local regression. Essentially it is a combination of the RLS method with exponential forgetting and local polynomial regression. Furthermore, the paper suggests a modification of the exponential forgetting scheme of the RLS method, to cope with the added complexity, which is introduced by allowing the parameters to be functions of other variables than just time.

6.2 Short-term wind power prediction models

In Paper C a new reference for short-term prediction models is proposed, and it is argued that the new reference model is more suitable than the often used persistence predictor, especially if the prediction horizon is above a few hours. The new reference is almost as simple as the persistence predictor, basically it is a prediction horizon dependent weighting

between the persistence and the mean of the power, where the weighting is determined by the auto-correlation of the wind power time series.

The physical relations used in the Risø short-term prediction system have been analyzed and validated. This analysis is provided in the summary report. The purpose of the analysis has been to find out how these relations can be incorporated in statistical models. It turned out that the behaviour of the physical relations was very simple as far as to calculate the wind at the wind turbine hub height. This relation was found to be very well approximated by sector-wise linear functions. Based on these findings some models have been proposed and tested in Paper E, some including physical relations and some pure statistical models, and it turned out that it is not advantageous to use any of the considered physical relations. The reason for why the physical relations did not prove to be useful has to do with the fact that the relations are derived from idealized assumptions and these assumptions are simply not valid when the input to the relations is variables from a numerical weather prediction model.

It should be noted, though, that if no measurements of the power production is available, then the physical relations can be used. Also, the statistical models which have been proposed need one to three months of data before these models are fully calibrated, therefore, the physical models can be used until the statistical models are calibrated.

An examination of how the dependency on the turbulence intensity is handled by the numerical weather prediction model, HIRLAM, is provided in Paper I. In this paper it is showed that the turbulence intensity dependency is handled in a slightly unrealistic manner. Nevertheless it is found that by selecting an appropriate model level, then the dependency is properly taken into account.

Conditionally parametric models for short-term wind power prediction have been considered. In Paper D weighted least squares is used for estimation of the model parameters, while in Paper E the new estimation method described in Paper B is used. When the special recursive estimation method described in Paper B is used, then this model class is an obvious choice for on-line applications. The explanatory variables which were found to give the best results in the conditionally parametric

models were: the wind speed and direction from the numerical weather prediction model, lagged values of the power production, the prediction horizon, the time of day and the time of year. The advantage of using conditionally parametric models is that variables can be included for which no specific functional relations are known. This has been used for the wind direction dependency in the models, as it was found that the physically determined dependency could not be used. The conditionally parametric model, which was found to be most adequate, outperformed both simple linear models and non-linear models.

Paper J addresses the economical value of short-term predictions. Predictions from several prediction models are used as input to a model of the England/Wales electrical grid, and it is found that for low penetration of wind energy, predictions have little value. As the penetration increases the predictions and their accuracy of the become more important, and it is also shown that confidence limits for the predictions can increase the economical value of the predictions.

6.3 Client/server software application

A client/server software application has been developed, which is capable of handling the data flow and model calculations related to a utility with several wind farms. The software application is described in the summary report. The application is based on a flexible architecture, in principle there is no limit on the prediction model modules, which can be added to the application. The application is based on a distributed programming model, i.e. distributed services, where data handling and model calculations can be performed on different computers. The first version of the system is planned to go into operation in May 2000, where it will be evaluated by the two Danish utilities Elkraft and Elsam.

References

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

Bailey, B., Brower, M. C. & Zack, J. (1999), 'Short-term wind forecasting', Proceedings of the European Wind Energy Conference. pp 1062–1065, Nice, France.

Bechrakis, D. A. & Sparis, P. D. (1998), 'Wind speed prediction using artificial neural networks', *Wind Engineering* **22**, 287–295.

Beyer, H. G., Degner, T., Hausmann, J., Hoffmann, M. & Ruján, P. (1994), 'Short term prediction of wind speed and power output of a wind turbine with neural networks', Proceedings of the EWEC '94 in Thessaloniki. pp 349–352, Thessaloniki, Greece.

Beyer, H. G., Heinemann, D., Mellinghoff, H., Mönnich, K. & Waldl, H. P. (1999), 'Forecast of regional power output of wind turbines', Proceedings of the European Wind Energy Conference. Nice, France.

Bossanyi, E. A. (1985), 'Short-term wind prediction using kalman filters', *Wind Eng.* **9**, 1–8.

Box, G. E. P. & Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, revised edn, Holden-Day, San Francisco.

Breiman, L. & Spector, P. (1992), 'Submodel selection and evaluation in regression. The X-random case', *International Statistical Review* **60**, 291–319.

Businger, J. A., Wyngaard, J. C., U, I. & Bradley, E. F. (1971), 'Flux-profile relationships in the atmospheric boundary layer', *Journal of Atmospheric Science* **28**, 181–189.

Cleveland, W. S. (1979), 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association* **74**, 829–836.

Cleveland, W. S. (1981), 'LOWESS: A program for smoothing scatterplots by robust locally weighted regression', *The American Statistician* **35**, 54.

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

Cleveland, W. S., Devlin, S. J. & Grosse, E. (1988), 'Regression by local fitting: Methods, properties, and computational algorithms', *Journal of Econometrics* **37**, 87–114.

Cleveland, W. S. & Loader, C. (1996), Smoothing by local regression: Principles and methods, *in* W. Härdle & M. G. Schimek, eds, 'Statistical Theory and Computational Aspects of Smoothing. Proceedings of the COMPSTAT '94 Satellite Meeting', Physica-Verlag, Heidelberg, pp. 10–49. (Discussion: pp. 80-127).

Dutton, A. G., Kariniotakis, G., Halliday, J. A. & Nogaret, E. (1999), 'Local and wind power forecasting methods for the optimal management of isolated power systems with high wind penetration', *Wind Eng.* **23**, 69–87.

Gamma, E., Helm, R., Johnson, R. & Vlissides, J. (1994), *Design Patterns*, Addison-Wesley, Reading, Massachusetts.

Giebel, G. (2000), On the Benefits of Distributed Generation of Wind Energy in Europe, PhD thesis, Carl von Ossietzky Universität, Oldenburg, Germany.

Haltiner & Williams (1970), *Numerical Prediction and Dynamic Meteorology.*

Hart, J. D. (1996), 'Some automated methods of smoothing time-dependent data', *Journal of Nonparametric Statistics* **6**, 115–142.

Hasselmann, N. N. (1981), 'N n', *N N* **79**, 2253–2769.

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

Hastie, T. & Tibshirani, R. (1993), 'Varying-coefficient models', *Journal of the Royal Statistical Society, Series B, Methodological* **55**, 757–796.

Jonsson, B. (1994), 'Prediction with a linear regression model and errors in a regressor', *International Journal of Forecasting* **10**, 549–555.

Kalnay, E., Lord, S. J. & McPherson, R. D. (1998), 'Maturity of operational numerical weather prediction: Medium range', *Bullentin of the American Meteorological Society* **79**, 2253–2769.

Krohn, S. (2000), 'Danish wind energy 1999', *Wind Power Note* **24**, 1–3.

Landberg, L. (1994), Short-term prediction of local wind conditions, Technical Report Risø - R - 702 (EN), Department of Wind Power Meteorology, Risø National Laboratory, Roskilde, Denmark.

Landberg, L. (1999*a*), 'A mathematical look at physical power prediction model', *Wind Energy* **1**, 23–28.

Landberg, L. (1999*b*), 'Short-term prediction of power production from wind farms', *Jour. of Wind Engineering and Industrial Aerodynamics* **80**, 207–220.

Landberg, L. & Joensen, A. (1998), 'A model to predict the output from wind farms – an update', proceedings from BWEA 20, British Wind Energy Conference. pp 127–132, Cardif, UK.

Landberg, L. & Watson, S. (1994), 'Short-term prediction of local wind conditions', *Boundary-Layer Meteorology* **70**, 171–195.

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

Lonnberg, P. & Shaw, D. (1987), 'ECMWF data assimilation: Scientific documentation', *ECMWF Research Manual, second revised edition* **1**.

Lorenc, A. C. (1981), 'A global three-dimensional multivariate statistical interpolation scheme', *Mon. Weather Rev.* **109**, 701–721.

Lorenz, E. N. (1963), 'Deterministic nonperiodic flow', *Journal of Atmospheric Sciences* **20**, 130–141.

Machenhauer, B., ed. (1988), *HIRLAM Final Report*, Danish Meteorological Institute, Copenhagen, Denmark.

Madsen, H. (1996), Models and methods for predicting wind power, Technical report, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Madsen, H. & Holst, J. (1999), *Modelling non-linear and non-stationary time series*, IMM, DTU, Denmark.

Martin, F., Zubiaur, R., Moreno, P., Rodriguez, S., Cabre, M., Casanova, M., Hormigo, A. & Alonso, M. (1993), Operational tool for short term prediction model of energy production in wind power plants at tarifa (spain), *in* 'Proceedings of the 1993 ECWEC in Travemünde', Travemünde, Germany, pp. 802–803.

Melgaard, H. (1994), Identification of Physical Models, PhD thesis, The Technical University of Denmark, Department of Mathematical Modelling, Lyngby, Denmark.

Mortensen, N. G., Landberg, L., Troen, I. & Petersen, E. L. (1993), 'Wind atlas analysis and application program (WAsP)', RisøNational Laboratory user guide: Risø-I-666(EN)(v.1) , Roskilde, Denmark.

Nielsen, T. S. & Madsen, H. (1996), Using meteorological forecasts in on-line predictions of wind power, Technical report, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Reese, G. (1997), *Database Programming with JDBC and Java*, O'Reilly, Cambridge, UK.

Sanderhoff, P. (1993), 'PARK - User's Guide. A PC-program for calculation of wind turbine park performance', RisøNational Laboratory, Roskilde, Denmark. Risø-I-668(EN).

Sass, B., Nielsen, N., Jørgensen, J. & Amstrup, B. (1999), The operational HIRLAM system at DMI, Technical report, Danish Meteorological Institute, Copenhagen, Denmark.

Shao, J. (1993), 'Linear model selection by cross-validation', *Journal of the American Statistical Association* **88**, 486–494.

Shuhui, L., Wunsch, D. C., O'Hair, E. & Giesselmann, M. (2001), 'Comparative analysis of regression and artificial neural network models for wind turbine power curve estimation', *Journal of Solar Energy Engineering* **123**, 327–332.

Stone, C. J. (1977), 'Consistent nonparametric regression', *The Annals of Statistics* **5**, 595–620. (C/R: pp. 620-645).

Stull, R. B. (1988), *An Introduction to boundary layer meteorology*, Kluwer Academic Publishers, Dordrecht, Boston, London.

Tande, J. O. & Landberg, L. (1993), 'A 10 sec. forecast of wind turbine output with neural networks', Proceedings of the 1993 ECWEC in Travemünde. pp 774–777, Germany.

Troen, I. & Petersen, E. L. (1989*a*), 'The european wind atlas', Published for the CEC by Risø National Laboratory, Roskilde,Denmark. 656 pp.

Troen, I. & Petersen, E. L. (1989*b*), 'The European Wind Atlas', Published for the CEC by Risø National Laboratory, Roskilde,Denmark. 656 pp.

Appendix A

HIRLAM equations

A.1 Model dynamics

As the model is hydrostatic, the hydrostatic relation between increments of geopotential and pressure is utilized. The model is a limited area model (LAM), which implies that not only the upper and lower boundary conditions need to be specified, but also the lateral boundary conditions have to be specified. The atmospheric forecast variables defined in tree dimensions are the horizontal wind components u and v, surface pressure p_s, temperature T, specific humidity, specific cloud condensation q_c and turbulent kinetic energy E.

The vertical coordinate used in the HIRLAM equations is $\eta(p, p_s)$, where p is pressure and p_s is the surface pressure. The vertical coordinate follows the terrain, and the boundary conditions for this variable are, at the surface $\eta(p_s, p_s) = 1$ and at the top of the atmosphere $\eta(0, p_s) = 0$.

HIRLAM is derived from a spherical coordinate system, but in the formulation two metric coefficients, h_x and h_y, have been introduced. For a short distance $\delta X, \delta Y$ on the earth with radius a, this yields $\delta X = a h_x \delta x$

and $\delta Y = ah_y \delta y$.

The terms F_γ in the equations below represent forces for the variable γ, which are due to other processes than dynamics.

The equations of motion (momentum conservation) utilized in HIRLAM now read

$$\frac{\partial u}{\partial t} = (f+\xi)v - \dot{\eta}\frac{\partial u}{\partial \eta} - \frac{R_d T_v}{ah_x}\frac{\partial \ln(p)}{\partial z} - \frac{1}{ah_x}\frac{\partial(\Phi + E_m)}{\partial x} + F_u, \quad \text{(A.1)}$$

$$\frac{\partial v}{\partial t} = -(f+\xi)u - \dot{\eta}\frac{\partial v}{\partial \eta} - \frac{R_d T_v}{ah_y}\frac{\partial \ln(p)}{\partial z} - \frac{1}{ah_y}\frac{\partial(\Phi + E_m)}{\partial y} + F_v. \quad \text{(A.2)}$$

Where

$$\xi = \frac{1}{ah_x h_y}\left(\frac{\partial(h_y v)}{\partial x} - \frac{\partial(h_x u)}{\partial y}\right) \quad \text{(A.3)}$$

is the vorticity,

$$E_m = \frac{1}{2}(u^2 + v^2) \quad \text{(A.4)}$$

is the is the kinetic energy of the mean horizontal motion, $\dot{\eta}$ is the vertical velocity in the η-coordinate system and Φ is the geopotential, $f = 2\Omega \sin\phi$ is the Coriolis parameter, Ω the angular velocity of the Earth, u_* is the surface friction velocity, $\kappa = 0.4 \pm 0.01$ is the von Karman constant

For temperature the equation is

$$\frac{\partial T}{\partial t} = -\frac{u}{ah_x}\frac{\partial T}{\partial x} - \frac{v}{ah_y}\frac{\partial T}{\partial y} - \dot{\eta}\frac{\partial T}{\partial \eta} + \frac{\delta_s T_v \omega}{(1-(\delta_c - 1)q)p} + F_T, \quad \text{(A.5)}$$

where δ_s is the ratio between the specific gas constant and the specific heat capacity, δ_c is the ratio between the specific heat capacity of water vapour and the corresponding value for dry air (at constant pressure), and ω is the pressure vertical velocity.

For the remaining variables ($\gamma = q, \gamma = q_c, \gamma = E$) the following equations apply

$$\frac{\partial \gamma}{\partial t} = -\frac{u}{ah_x}\frac{\partial \gamma}{\partial x} - \frac{v}{ah_x}\frac{\partial \gamma}{\partial y} - \dot{\eta}\frac{\partial \gamma}{\partial \eta} + F_\gamma \quad \text{(A.6)}$$

The hydrostatic equation takes the form

$$\frac{\partial \Phi}{\partial \eta} = -\frac{R_d T_v}{p}\frac{\partial p}{\partial \eta}, \tag{A.7}$$

and the continuity equation

$$\frac{\partial^2 p}{\partial \eta \partial t} + \nabla\left(\boldsymbol{V}\frac{\partial p}{\partial \eta}\right) + \frac{\partial}{\partial \eta}\left(\dot{\eta}\frac{\partial p}{\partial \eta}\right) = 0. \tag{A.8}$$

The definition of the divergence operator is

$$\nabla \boldsymbol{V} = \frac{1}{a h_x h_y}\left(\frac{\partial}{\partial x}(h_y u) + \frac{\partial}{\partial y}(h_x v)\right) \tag{A.9}$$

By integrating the continuity equation using the boundary conditions $\dot{\eta} = 0$ for $\eta = 0$ and $\eta = 1$ the equation for the surface pressure tendency is obtained

$$\frac{\partial p_s}{\partial t} = -\int_0^1 \nabla\left(\boldsymbol{V}\frac{\partial p}{\partial \eta}\right) d\eta \tag{A.10}$$

The equation for pressure vertical velocity is

$$\omega = -\int_\eta^1 \nabla\left(\boldsymbol{V}\frac{\partial p}{\partial \eta}\right) d\eta + \boldsymbol{V}\nabla p \tag{A.11}$$

and the equation for $\dot{\eta}$

$$\dot{\eta}\frac{\partial p}{\partial \eta} = \left(1 - \frac{\partial p}{\partial p_s}\right) + \int_\eta^1 \nabla\left(\boldsymbol{V}\frac{\partial p}{\partial \eta}\right) d\eta \tag{A.12}$$

A.2 Physical parameterizations

The physics compromises the process of radiation and subgrid scale transport of momentum, temperature and moisture variables down to the small scales associated with turbulence. In addition the thermodynamics associated with latent heat release (e.g. condensation, evaporation, sublimation and precipitation) must also be described. The boundary conditions at the ground need also to be taken into account.

In this section only the turbulence parameterizations for the transport of momentum, sensible heat and moisture used in HIRLAM will be considered. The treatment of the surface is described in the next section.

Originally first order local closure was used in HIRLAM, but by introducing the prognostic equation for the turbulence kinetic energy, HIRLAM has taken the first step towards second order closure. In (Sass et al. 1999) it is reported that various parameterizations schemes have been tested and used in HIRLAM, and the one described here is different from the one originally described in (Machenhauer 1988). One reason for including the prognostic equation for the turbulent kinetic energy, can be explained by the description in Section 2.1.4 of how the boundary layer evolves with time. The time derivative of the turbulent kinetic energy represent momentum or memory, and, therefore, as the mixed layer is transformed into the residual layer, the state of the formerly mixed layer is brought into the residual layer by this equation.

The equations for the mean variables used in HIRLAM are very similar to the general equations outlined in Section A.1. The effect on the mean variables caused by turbulence is obtained by replacing the variables in the equation in Section A.1 with mean variables $\overline{\gamma}$ and adding the term

$$\frac{\partial \overline{w'\gamma'}}{\partial z} \tag{A.13}$$

to each equation, where $\overline{\gamma' w'}$ is the vertical kinematic flux of γ. One assumption applied here is that the horizontal derivatives of the covariance is much smaller than the vertical derivatives. To the same level of approximation the prognostic equation for the turbulent kinetic energy is written

$$\begin{aligned} \frac{\partial \overline{E}}{\partial t} &= -\left[\overline{u'w'}\frac{\partial \overline{u}}{\partial z} + \overline{v'w'}\frac{\partial v}{\partial z} + \overline{w'w'}\frac{\partial \overline{w}}{\partial z}\right] \\ &\quad + \left[\frac{g}{\overline{\theta}_v}\overline{w'\theta'_v}\right] - \left[\frac{q}{\rho}\frac{\partial \overline{p'w'}}{\partial z}\right] - \left[\frac{\partial \overline{E'w'}}{\partial z}\right] - \epsilon, \end{aligned} \tag{A.14}$$

where $\overline{E} = \frac{1}{2}\left(\overline{u'^2 + v'^2 + w'^2}\right)$ is the turbulent kinetic energy and ϵ is the dissipation of $\overline{E}$. The first two terms on the right hand side of the equations are the horizontal sheer production of turbulent kinetic energy. The third term involving vertical velocity variance and the fifth term involving pressure correlations are neglected. The fourth term involves buoyancy generated turbulence and the sixth term describes the vertical convergence of subgrid scale vertical transport of $\overline{E}$.

As mentioned previously in Section 2.1.3, the equation system describing

the mean variables in a turbulent flow has to be closed by parameterizations. HIRLAM uses first order local closure, i.e. the covariances or the vertical kinematic fluxes are parameterized by assuming relations of the form

$$\overline{\gamma' w'} = K_\gamma \left(\frac{\partial \overline{\gamma}}{\partial z} \right); \quad \gamma = u, v, \theta, q, q_c, E. \tag{A.15}$$

In (A.15) K_γ is an eddy exchange coefficient analogous to the molecular viscosity and the diffusivity coefficients. The eddy exchange coefficients depend on $\overline{E}$ via the following relation

$$K_\gamma = c_\gamma K_u \phi(R_s), \tag{A.16}$$

where c_γ is a non-dimensional constant and $K_u = l\sqrt{\overline{E}}$ is the eddy exchange coefficient for momentum. $\phi(\cdot)$ is a function of the dry Redelsperger number R_s given by $\phi(R_s) = (1 + 0.139R_s)^{-1}$, where

$$R_s = \frac{g}{\overline{\theta}_v} \frac{l^2}{\overline{E}} \left(1 + 0.61\overline{q} \frac{\partial \overline{\theta}}{\partial z} + 0.61\overline{\theta} \frac{\partial \overline{q}}{\partial z} \right). \tag{A.17}$$

The dissipation term is handled by the expression

$$\epsilon = c_\epsilon \frac{\overline{E}^{3/2}}{l}. \tag{A.18}$$

In the equations above l is a diagnostic mixing length. It is computed from $l = \sqrt{l_u l_d}$, where l_u and l_d are the distances an air parcel must be displaced upward or downward, respectively, before its turbulent kinetic energy has been consumed by buoyancy.

A.3 Surface layer

In the surface layer, i.e. the layer between the HIRLAM lowest model layer and the surface, the fluxes are calculated using drag formulae relating the surface fluxes to the mean states of the surface and of the atmosphere at the observation height (in this case the lowest model level in HIRLAM). The drag formulae approximates the vertical flux $\overline{w'\gamma'}$ of γ by

$$\overline{w'\gamma'} = C_\gamma \Delta \overline{\gamma} |\overline{\boldsymbol{V}}_N|, \tag{A.19}$$

where C_γ is the drag coefficient, $\Delta\overline{\gamma} = \overline{\gamma}_s - \overline{\gamma}_N$, subscript s refers to the surface and N to the lowest model level values of $\overline{\gamma}$, and $|\overline{\boldsymbol{V}}_N|$ is the magnitude of the horizontal wind vector.

The drag coefficient is given by

$$C_\gamma = C_{MN}\left(1 + \ln\frac{z_{0M}}{z_{0H}}/\ln\frac{z}{z_{0M}}\right)^{-1}\Psi_\gamma\left(R_i, \frac{z}{z_{0H}}, \frac{z}{z_{0M}}\right) \qquad \text{(A.20)}$$

This parameterization is valid for the variables $\gamma = M, H_s, H_l$, where M corresponds to the momentum flux, H_s the sensible heat flux and H_l the latent heat flux. R_i is the surface bulk Richardson number (Stull 1988), which is a measure of turbulence intensity.

The relation for Ψ_γ is different for stable and unstable conditions. For unstable condition the relation is

$$\Psi_\gamma = 1 + \frac{a_\gamma R_i}{1 + b_\gamma C_{\gamma N}\left(R_i\frac{z}{z_{0M}}\right)^{0.5}}, \qquad \text{(A.21)}$$

where $a_M = 10, b_M = 75, a_{H_s} = a_{H_l} = 15$ and $b_{H_s} = b_{H_l} = 75$. The relation for stable conditions, and special modifications to handle the calculations over the sea are omitted here, see (Sass et al. 1999) for more details.

It should be noted that the surface roughness lengths for momentum z_{0M}, sensible heat flux z_{0H_s} and latent heat flux z_{0H_l} might bee different, but in HIRLAM these are assumed equal. Furthermore, it is reported in (Sass et al. 1999) that the values for the roughness lengths are slightly unrealistic. The reason for this is due the spatial resolution used, which means that effects that on a higher model 0resolution would be regarded as surface curvature, is considered as roughness in HIRLAM.

A.4 Surface energy budget

The energy and moisture budget of a land surface needs to be treated in a prognostic sense since the forecasting of a diurnal variation of meteorological variables close to the ground is vital. Only the equations for

the temperature will be given here, the equations for the moisture can be found in (Sass et al. 1999).

The equations for the temperature are based on a three layer soil model. The equation for the temperature in the surface layer T_s is

$$\frac{\partial T_s}{\partial t} = \frac{1}{\rho_s c_s D_1} \sum_i H_i + \frac{\kappa_0 (1 - k_{sn} F_{sn})(T_d - T_s)}{0.5 D_1 (D_1 + D_2)} \tag{A.22}$$

In this equation $\sum_i H_i = H_R + H_S + H_L$ are the net fluxes due to radiation, sensible and latent heat, respectively. $F_{sn} = \min(\frac{S}{S_t}, 1)$ is a snow fraction, with S being snow depth and $S_t = 0.015$ m a threshold snow depth in an equivalent height of water. T_d is the soil temperature in the intermediate layer, ρ_s is the soil density, c_s is the specific heat capacity of the soil, κ_0 is the heat diffusivity of soil without snow cover and k_{sn} is a constant used to reduce heat diffusivity if snow cover is positive.

The equation for the temperature in the intermediate layer T_d is

$$\frac{\partial T_d}{\partial t} = -\frac{\kappa_0 (T_d - T_s)}{0.5 D_2 (D_1 + D_2)} + \frac{\kappa_0 (T_{cli} - T_d)}{D_2 D_3} \tag{A.23}$$

where T_{cli} is the climatic deep soil temperature updated every month. $D_1 = D_2 = D_3/6 = 0.07$ m is the depth of the surface, intermediate and the deep soil layer, respectively.

A.5 Diagnostic output

HIRLAM calculates some special diagnostic output variables, i.e. variables which do not give any feedback to the integration of the model itself. For the list of variables see (Sass et al. 1999). Of special interest in this thesis is the wind corresponding to 10m above ground level. The calculation is performed for the u and v components of the wind separately. For the unstable boundary layer, the u component is calculated

by

$$u(z)_{z=10m} = u_N - \frac{u_*}{\kappa}\left(-\ln\left(\frac{z\left(1+X_N^2\right)}{z_N\left(1+X^2\right)}\right) - 2\ln\left(\frac{1+X_N}{1+X}\right) + 2\left(\tan^{-1}(X_N) - \tan^{-1}(X)\right)\right) \tag{A.24}$$

$$X = \left(1 - 15\frac{z}{L}\right)^{\frac{1}{4}}, \quad x_N = \left(1 - 15\frac{z_N}{L}\right)^{\frac{1}{4}}$$

where κ is the von Karman constant, L is the Monin-Obukov length scale and u_* is the surface friction velocity, see e.g. (Stull 1988) for definition. The relation for the stable boundary layer is a modified version of the profile suggested in (Businger, Wyngaard, U & Bradley 1971), which guarantees that the calculated wind speed is no larger than provided by the lowest model level. This relation is

$$u(z) = \frac{u_*}{\kappa}\ln\left(\frac{z}{z_0}\right) + u_N\left[1 - \exp\left(-\frac{4u_* z}{\kappa u_N L}\right)\right] \tag{A.25}$$

The relations for the v component, correspond to the above relations when u is replaced by v.

Papers

Paper A

Tracking time-varying parameters with local regression

Originally published in *Automatica*, Vol **36**, pages 1199–1204. 2000.

Tracking time-varying parameters with local regression

Alfred Joensen[1,2], Henrik Madsen[1],
Henrik Aa. Nielsen[1] and Torben S. Nielsen[1]

Abstract

This paper shows that the recursive least squares (RLS) algorithm with forgetting factor is a special case of a varying-coefficient model, and a model which can easily be estimated via simple local regression. This observation allows us to formulate a new method which retains the RLS algorithm, but extends the algorithm by including polynomial approximations. Simulation results are provided, which indicates that this new method is superior to the classical RLS method, if the parameter variations are smooth.

Keywords: Recursive estimation; varying-coefficient; conditional parametric; polynomial approximation; weighting functions.

1 Introduction

The *RLS* algorithm with forgetting factor (Ljung & Söderström 1983) is often applied in on-line situations, where time variations are not modeled adequately by a linear model. By sliding a time-window of a specific width over the observations where only the newest observations are seen, the model is able to adapt to slow variations in the dynamics. The width, or the bandwidth $\hbar$, of the time-window determines how fast the model adapts to the variations, and the most adequate value of $\hbar$ depends on how fast the parameters actually vary in time. If the time variations are fast, $\hbar$ should be small, otherwise the estimates will be seriously biased. However, fast adaption means that only few observations are used for the estimation, which results in a noisy estimate. Therefore the choice of $\hbar$ can be seen as a bias/variance trade off.

[1]Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

[2]Department of Wind Energy and Atmospheric Physics, Risø National Laboratory, DK-4000 Roskilde, Denmark

In the context of local regression (Cleveland & Devlin 1988) the parameters of a linear model estimated by the *RLS* algorithm can be interpreted as zero order local time polynomials, or in other words local constants. However, it is well known that polynomials of higher order in many cases provide better approximations than local constants. The objective of this paper is thus to illustrate the similarity between the *RLS* algorithm and local regression, which leads to a natural extension of the *RLS* algorithm, where the parameters are approximated by higher order local time polynomials. This approach does, to some degree, represent a solution to the bias/variance trade off. Furthermore, viewing the *RLS* algorithm as local regression, could potentially lead to development of new and refined *RLS* algorithms, as local regression is an area of current and extensive research. A generalisation of models with varying parameters is presented in (Hastie & Tibshirani 1993), and, as will be shown in this paper, the *RLS* algorithm is an estimation method for one of these models.

Several extensions of the RLS algorithm have been proposed in the literature, especially to handle situations where the parameter variations are not the same for all the parameters. Such situations can be handled by assigning individual bandwidths to each parameter, e.g. *vector forgetting*, or by using the *Kalman Filter* (Parkum, Poulsen & Holst 1992). These approaches all have drawbacks, such as assumptions that the parameters are uncorrelated and/or are described by a random walk. Polynomial approximations and local regression can to some degree take care of these situations, by approximating the parameters with polynomials of different degrees. Furthermore, it is obvious that the parameters can be functions of other variables than time. In (Nielsen, Nielsen, Madsen & Joensen 1999) a recursive algorithm is proposed, which can be used when the parameters are functions of time and some other explanatory variables.

Local regression is adequate when the parameters are functions of the same explanatory variables. If the parameters depend on individual explanatory variables, estimation methods for additive models should be used (Fan, Hardle & Mammen 1998, Hastie & Tibshirani 1990). Unfortunately it is not obvious how to formulate recursive versions of these estimation methods, and to the authors best knowledge no such recursive methods exists. Early work on additive models and recursive regression dates back to (Holt 1957) and (Winters 1960), which developed recursive estimation methods for models related to the additive models, where in-

dividual forgetting factors are assigned to each additive component, and the trend is approximated by a polynomial in time.

2 The varying-coefficient approach

Varying-coefficient models are considered in (Hastie & Tibshirani 1993). These models can be considered as linear regression models in which the parameters are replaced by smooth functions of some explanatory variables. This section gives a short introduction to the varying-coefficient approach and a method of estimation, local regression, which becomes the background for the proposed extension of the *RLS* algorithm.

2.1 The model

We define the varying-coefficient model

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

where y_i is a response, $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are explanatory variables, $\boldsymbol{\theta}(\cdot)$ is a vector of unknown but smooth functions with values in $\mathbf{R}$, and N is the number of observations. If ordinary regression is considered e_i should be identically distributed (i.d.), but if i denotes at time index and $\boldsymbol{z}_i^T$ contains lagged values of the response variable, e_i should be independent and identically distributed (i.i.d).

The definition of a varying-coefficient model in (Hastie & Tibshirani 1993) is somewhat different than the one given by Eq. 1, in the way that the individual parameters in $\boldsymbol{\theta}(\cdot)$ depend on individual explanatory variables. In (Anderson, Fang & Olkin 1994), the model given by Eq. 1 is denoted a conditional parametric model, because when $\boldsymbol{x}_i$ is constant the model reduces to an ordinary linear model

2.2 Local constant estimates

As only models where the parameters are functions of time are considered, only $\boldsymbol{x}_i = i$ is considered in the following. Estimation in Eq. 1 aims at estimating the functions $\boldsymbol{\theta}(\cdot)$, which in this case are the one-dimensional functions $\boldsymbol{\theta}(i)$. The functions are estimated only for distinct values of the argument t. Let t denote such a point and $\hat{\boldsymbol{\theta}}(t)$ the estimated coefficient functions, when the coefficients are evaluated at t.

One solution to the estimation problem is to replace $\boldsymbol{\theta}(i)$ in Eq. 1 with a constant vector $\boldsymbol{\theta}(i) = \boldsymbol{\theta}$ and fit the resulting model locally to t, using weighted least squares, i.e.

$$\hat{\boldsymbol{\theta}}(t) = \mathbf{arg}\min_{\theta} \sum_{i=1}^{t} w_i(t)(y_i - \boldsymbol{z}_i^T\boldsymbol{\theta})^2. \tag{2}$$

Generally, using a nowhere increasing weight function $W : \mathbf{R}_0 \to \mathbf{R}_0$ and a spherical kernel the actual weight $w_i(t)$ allocated to the ith observation is determined by the Euclidean distance, in this case $|i - t|$, as

$$w_i(t) = W\left(\frac{|i-t|}{\hbar(t)}\right). \tag{3}$$

The scalar $\hbar(t)$ is called the bandwidth, and determines the size of the neighbourhood that is spanned by the weight function. If e.g. $\hbar(t)$ is constant for all values of t it is denoted a fixed bandwidth. In practice, however, also the nearest neighbour bandwidth, which depends on the distribution of the explanatory variable, is used (Cleveland & Devlin 1988). Although, in this case where $\boldsymbol{x}_i = i$, i.e. the distribution of the explanatory variable is rectangular, a fixed bandwidth and a nearest neighbour bandwidth are equivalent.

2.3 Local polynomial estimation

If the bandwidth $\hbar(t)$ is sufficiently small the approximation of $\boldsymbol{\theta}(t)$ as a constant vector near t is good. This implies, however, that a relatively low number of observations is used to estimate $\boldsymbol{\theta}(t)$, resulting in a noisy estimate. On the contrary a large bias may appear if the bandwidth is large.

It is, however, obvious that locally to t the elements of $\boldsymbol{\theta}(t)$ may be better approximated by polynomials, and in many cases polynomials will provide good approximations for larger bandwidths than local constants. Local polynomial approximations are easily included in the method described. Let $\theta_j(t)$ be the jth element of $\boldsymbol{\theta}(t)$ and let $\boldsymbol{p}_d(t)$ be a column vector of terms in a d-order polynomial evaluated at t, i.e. $\boldsymbol{p}_d(t) = [t^d \; t^{d-1} \; \cdots \; 1]$. Furthermore, introduce $\boldsymbol{z}_i = [z_{1i} \; \cdots \; z_{pi}]^T$,

$$\boldsymbol{u}_{i,t}^T = \left[z_{1i}\boldsymbol{p}_{d_1}^T(t-i) \; \cdots \; z_{ji}\boldsymbol{p}_{d_j}^T(t-i) \; \cdots \; z_{pi}\boldsymbol{p}_{d_p}^T(t-i) \right], \tag{4}$$

$$\hat{\boldsymbol{\phi}}^T(t) = [\hat{\boldsymbol{\phi}}_1^T(t) \; \cdots \; \hat{\boldsymbol{\phi}}_j^T(t) \; \cdots \; \hat{\boldsymbol{\phi}}_p^T(t)], \tag{5}$$

where $\hat{\boldsymbol{\phi}}_j(t)$ is a column vector of local constant estimates at t, i.e.

$$\hat{\boldsymbol{\phi}}_j^T(t) = [\hat{\phi}_{jd_j+1}(t) \; \cdots \; \hat{\phi}_{j1}(t)] \tag{6}$$

corresponding to $z_{ji}\boldsymbol{p}_{d_j}^T(t-i)$. Now weighted least squares estimation is applied as described in Section 2.2, but fitting the linear model

$$y_i = \boldsymbol{u}_{i,t}^T\boldsymbol{\phi} + e_i; \quad i = 1, \ldots, t, \tag{7}$$

locally to t, i.e. the estimate $\hat{\boldsymbol{\phi}}(t)$ of the parameters $\boldsymbol{\phi}$ in Eq. 7 becomes a function of t as a consequence of the weighting. Estimates of the elements of $\boldsymbol{\theta}(t)$ can now be obtained as

$$\hat{\theta}_j(t) = \boldsymbol{p}_{d_j}^T(0)\hat{\boldsymbol{\phi}}_j(t) = [\underbrace{0 \; \cdots \; 0 \; 1}_{d_j+1}]\hat{\boldsymbol{\phi}}_j(t) = \hat{\phi}_{j1}(t); \quad j = 1, \ldots, p. \tag{8}$$

3 Recursive least squares with forgetting factor

In this section the well known *RLS* algorithm with forgetting factor is compared to the proposed method of estimation for the varying-coefficient approach. Furthermore, it is shown how to include local polynomial approximations in the *RLS* algorithm.

3.1 The weight function

The *RLS* algorithm with forgetting factor aims at estimating the parameters in the linear model

$$y_i = \boldsymbol{z}_i^T \boldsymbol{\theta} + e_i \tag{9}$$

which corresponds to Eq. 1 when $\boldsymbol{\theta}(\boldsymbol{x}_i)$ is replaced by a constant vector $\boldsymbol{\theta}$. The parameter estimate $\hat{\boldsymbol{\theta}}(t)$, using the *RLS* algorithm with constant forgetting factor λ, is given by

$$\hat{\boldsymbol{\theta}}(t) = \mathbf{arg}\min_{\theta} \sum_{i=1}^{t} \lambda^{t-i}(y_i - \boldsymbol{z}_i^T \boldsymbol{\theta})^2. \tag{10}$$

In this case the weight which is assigned to the ith observation in Eq. 10 can be written as

$$w_i(t) = \lambda^{t-i} = \left[\exp\left(\frac{i-t}{(\ln\lambda)^{-1}}\right)\right]^{-1} = \left[\exp\left(\frac{|i-t|}{-(\ln\lambda)^{-1}}\right)\right]^{-1} \tag{11}$$

where the fact that $i \leq t$ in Eq. 10 is used. Now it is easily seen that Eq. 11 corresponds to Eq. 3 with a fixed bandwidth $\hbar(t) = \hbar = -(\ln\lambda)^{-1}$, which furthermore shows how the bandwidth and the forgetting factor are related. By also comparing Eq. 9 and Eq. 1 it is thus verified that the *RLS* algorithm with forgetting factor corresponds to local constant estimates in the varying-coefficient approach, with the specific choice Eq. 11 of the weight function.

3.2 Recursive local polynomial approximation

The *RLS* algorithm is given by (Ljung & Söderström 1983)

$$\mathbf{R}(t) = \sum_{i=1}^{t} \lambda^{t-i} \boldsymbol{z}_t \boldsymbol{z}_t^T = \lambda \mathbf{R}(t-1) + \boldsymbol{z}_t \boldsymbol{z}_t^T, \tag{12}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{R}^{-1}(t)\boldsymbol{z}_t \left[y_t - \boldsymbol{z}_t^T \hat{\boldsymbol{\theta}}(t-1)\right], \tag{13}$$

with initial values

$$\mathbf{R}^{-1}(0) = \alpha \mathbf{I}, \quad \boldsymbol{\theta}(0) = \mathbf{0},$$

where α is large (Ljung & Söderström 1983). Hence, the recursive algorithm is only asymptotically equivalent to solving the least squares criteria Eq. 10, which on the other hand does not give a unique solution for small values of t.

In Section 2.3 it was shown how to include local polynomial approximation of the parameters in the varying-coefficient approach, and that this could be done by fitting the linear model Eq. 7 and calculating the parameters from Eq. 8. It is thus obvious to use the same approach in an extension of the *RLS* algorithm, replacing $\boldsymbol{z}_t$ by $\boldsymbol{u}_{i,t}$. However, the explanatory variable $\boldsymbol{u}_{i,t}$ is a function of t, which means that as we step forward in time,

$$\mathbf{R}(t-1) = \sum_{i=1}^{t-1} \lambda^{t-1-i} \boldsymbol{u}_{i,t-1} \boldsymbol{u}_{i,t-1}^T$$

can not be used in the updating formula for $\mathbf{R}(t)$, as $\mathbf{R}(t)$ depends on $\boldsymbol{u}_{i,t}$. To solve this problem a linear operator which is independent of t, and maps $\boldsymbol{p}_{d_j}(s)$ to $\boldsymbol{p}_{d_j}(s+1)$ has to be constructed. Using the coefficients of the relation

$$(s+1)^d = s^d + ds^{d-1} + \frac{d(d-1)}{2!} s^{d-2} + \cdots + 1. \tag{14}$$

it follows that

$$\begin{aligned} \boldsymbol{p}_{d_j}(s+1) &= \begin{bmatrix} 1 & d_j & \frac{d_j(d_j-1)}{2!} & \frac{(d_j-1)(d_j-2)}{3!} & \cdots & 1 \\ 0 & 1 & d_j - 1 & \frac{(d_j-1)(d_j-2)}{2!} & \cdots & 1 \\ & & 1 & d_j - 2 & \cdots & 1 \\ & & & 1 & & \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & & & \cdots & 1 \end{bmatrix} \begin{bmatrix} s^{d_j} \\ s^{d_j-1} \\ \vdots \\ 1 \end{bmatrix} \\ &= \mathbf{L}_j \boldsymbol{p}_{d_j}(s) \end{aligned} \tag{15}$$

Since $\mathbf{L}_j$ is a linear operator it can be applied directly to $\boldsymbol{u}_{i,t} = \mathbf{L}\boldsymbol{u}_{i,t-1}$,

where

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{L}_2 & 0 & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & \dots & & \dots & \mathbf{L}_p \end{bmatrix} \tag{16}$$

Which, when applied to the recursive calculation Eq. 12 of $\mathbf{R}(t)$, yields

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

and the updating formula for the parameters Eq. 13 is left unchanged. The proposed algorithm will be denoted *POLRLS* (Polynomial *RLS*) in the following.

Note that if the polynomials in Eq. 4 were calculated for the argument i instead of $t - i$, then $\boldsymbol{u}_{i,t} = \boldsymbol{u}_{i,t-1}$, and it is seen that the recursive calculation in Eq. 12 could be used without modification, but now there would be a numerical problem for $t \to \infty$.

4 Simulation study

Simulation is used to compare the *RLS* and *POLRLS* algorithms. For this purpose we have generated $N = 11$ samples of $n = 1000$ observations from the time-varying ARX-model

$$y_i = a y_{i-1} + b(i) z_i + e_i, \quad e_i \in N(0,1),$$

where

$$a = 0.7, \quad b(i) = 5 + 4\sin\left(\frac{2\pi}{1000} i\right), \quad z_i \in N(0,1).$$

The estimation results are compared using the sample mean of the mean square error (*MSE*) of the deviation between the true and the estimated parameters

$$MSE_a = \frac{1}{N-1} \sum_{j=2}^{N} \left\{ \frac{1}{n-s+1} \sum_{i=s}^{n} (a - \hat{a}(i))^2 \right\}$$
$$MSE_b = \frac{1}{N-1} \sum_{j=2}^{N} \left\{ \frac{1}{n-s+1} \sum_{i=s}^{n} (b(i) - \hat{b}(i))^2 \right\}$$

and the sample mean of the *MSE* of the predictions

$$MSE_p = \frac{1}{N-1}\sum_{j=2}^{N}\left\{\frac{1}{n-s+1}\sum_{i=s}^{n}(y_i - \hat{a}(i-1)y_{i-1} - \hat{b}(i-1)z_i)^2\right\}. \tag{18}$$

Only observations for which $i \geq s = 350 > max(\hbar_{opt})$, where $\hbar_{opt}$ is the optimal bandwidth, are used in the calculation of the *MSE*, to make sure that the effect of the initialisation has almost vanished. The observations used for the prediction in Eq. 18, has not been used for the estimation of the parameters, therefore the optimal bandwidth, $\hbar_{opt}$, can be found by minimizing Eq. 18 with respect to the bandwidth $\hbar$, i.e. forward validation. The optimal bandwidth is found using the first sample, $j = 1$, the 10 following are used for the calculation of the sample means.

The *POLRLS* method was applied with two different sets of polynomial orders. The results are shown in Figure 1 and Table 1. Obviously, knowing the true model, a zero order polynomial approximation of a and a second order polynomial approximation of b, should be the most adequate choice. In a true application such knowledge might not be available, i.e. if no preliminary analysis of data is performed. Therefore, a second order polynomial approximation is used for both parameters, as this could be the default or standard choice. In both cases the *POLRLS* algorithm performs significantly better than the *RLS* algorithm, and, as expected, using a second order approximation of a increases the *MSE* because in this case the estimation is disturbed by non-significant explanatory variables. In the figure it is seen, that it is especially when

Method	Pol. order	$\hbar_{opt}$	MSE_p	MSE_a	MSE_b
POLRLS	$d_1 = 2, d_2 = 2$	62	1.0847	0.0024	0.0605
POLRLS	$d_1 = 0, d_2 = 2$	57	1.0600	0.0005	0.0580
RLS	$d_1 = 0, d_2 = 0$	11	1.1548	0.0044	0.0871

Table 1: *MSE* results using the *RLS* and *POLRLS* algorithms.

the value of $b(i)$ is small, that the variance of $\hat{a}$ is large. In this case the signal to noise ratio is low, and the fact that a larger bandwidth can be used in the new algorithm, means that the variance can be significantly

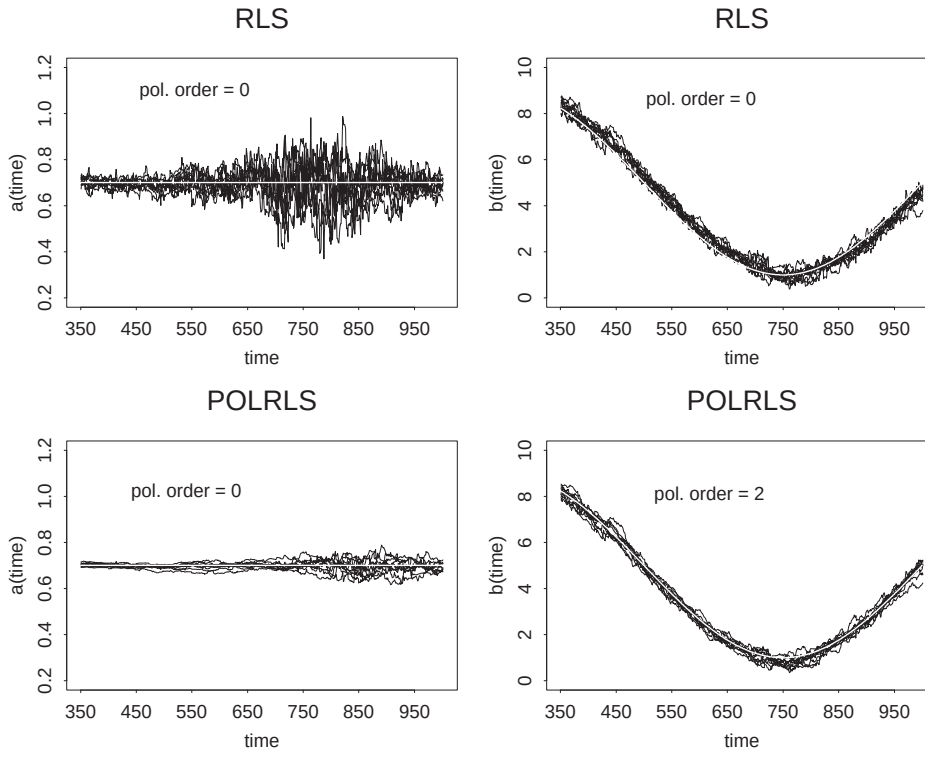


Figure 1: Estimated parameter trajectories. The first row shows the trajectories from the *RLS* algorithm, the second row shows the result from the *POLRLS* algorithm where a has been approximated by a zero order polynomial, and b by a second order polynomial.

reduced. Furthermore, it is seen that the reduction of the parameter estimation variance is greater for the fixed parameter than the time varying parameter. The reason for this is that the optimal bandwidth is found by minimising the *MSE* of the predictions, and bias in the estimate of b contributes relatively more to the *MSE* than variance in the estimate of a, i.e. the optimal value of $\hbar$ balances bias in the estimate of b and variance in the estimate of a. When a second order polynomial is used instead of a zero order polynomial, for the estimation of b, it is possible to avoid bias even when a significantly larger bandwidth is used.

5 Summary

In this paper the similarity between the varying-coefficient approach and the *RLS* algorithm with forgetting factor has been demonstrated. Furthermore an extension of the *RLS* algorithm, along the lines of the varying-coefficient approach is suggested. Using an example it is shown that the new algorithm leads to an significantly improvement of the estimation performance, if the variation of the true parameters is smooth.

References

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

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

Fan, J., Hardle, W. & Mammen, E. (1998), 'Direct estimation of low dimensional components in additive models', *The Annals of Statistics* **26**, 943–971.

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

Hastie, T. & Tibshirani, R. (1993), 'Varying-coefficient models', *Journal of the Royal Statistical Society, Series B, Methodological* **55**, 757–796.

Holt, C. (1957), 'Forecasting trends and sesonals by exponentially weighted moving averages', *O.N.R. Memorandum 52* . Carnegie Institute of Technology.

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

Nielsen, H. A., Nielsen, T. S., Madsen, H. & Joensen, A. (1999), 'Tracking time-varying coefficient-functions'. To be published.

Parkum, J. E., Poulsen, N. K. & Holst, J. (1992), 'Recursive forgetting algorithms', *Int. J. Control* **55**, 109–128.

Winters, P. (1960), 'Forecasting sales by exponentially weighted movings averages', *Man. Sci.* **6**, 324–342.

Paper B

Tracking time-varying coefficient-functions

Accepted for publication in *Int. J. of Adaptive Control and Signal Processing*. A version with more details is available as IMM technical report number 1999-9.

Tracking time-varying coefficient-functions

Henrik Aa. Nielsen[1], Torben S. Nielsen[1], Alfred K. Joensen[1], Henrik Madsen[1] and Jan Holst[2]

Abstract

A method for adaptive and recursive estimation in a class of non-linear autoregressive models with external input is proposed. The model class considered is conditionally parametric ARX-models (CPARX-models), which is conventional ARX-models in which the parameters are replaced by smooth, but otherwise unknown, functions of a low-dimensional input process. These coefficient-functions are estimated adaptively and recursively without specifying a global parametric form, i.e. the method allows for on-line tracking of the coefficient-functions. Essentially, in its most simple form, the method is a combination of recursive least squares with exponential forgetting and local polynomial regression. It is argued, that it is appropriate to let the forgetting factor vary with the value of the external signal which is the argument of the coefficient-functions. Some of the key properties of the modified method are studied by simulation.

Keywords: Adaptive and recursive estimation; Non-linear models; Time-varying functions; Conditional parametric models; Non-parametric method.

1 Introduction

The conditional parametric ARX-model (CPARX-model) is a non-linear model formulated as a linear ARX-model in which the parameters are replaced by smooth, but otherwise unknown, functions of one or more explanatory variables. These functions are called coefficient-functions. In

[1]Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

[2]Department of Mathematical Statistics, Lund University, Lund Institute of Technology, S-211 00 Lund, Sweden

(Nielsen, Nielsen & Madsen 1997) this class of models is used in relation to district heating systems to model the non-linear dynamic response of network temperature on supply temperature and flow at the plant. A particular feature of district heating systems is, that the response on supply temperature depends on the flow. This is modelled by describing the relation between temperatures by an ARX-model in which the coefficients depend on the flow.

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

Cleveland & Devlin (1988) gives an excellent account for non-adaptive estimation of a regression function by use of local polynomial approximations. Non-adaptive recursive estimation of a regression function is a related problem, which has been studied recently by Thuvesholmen (1997) using kernel methods and by Vilar-Fernández & Vilar-Fernández (1998) using local polynomial regression. Since these methods are non-adaptive one of the aspects considered in these papers is how to decrease the bandwidth as new observations become available. This problem do not arise for adaptive estimation since old observations are down-weighted and eventually disregarded as part of the algorithm. Hastie & Tibshirani (1993) considered varying-coefficient models which are similar in structure to conditional parametric models and have close resemblance to additive models (Hastie & Tibshirani 1990) with respect to estimation. However, varying-coefficient models include additional assumptions on the structure. Some specific time-series counterparts of these models are the functional-coefficient autoregressive models (Chen & Tsay 1993*a*)

and the non-linear additive ARX-models (Chen & Tsay 1993*b*).

The paper is organized as follows. In Section 2 the conditional parametric model is introduced and a procedure for estimation is described. Adaptive and recursive estimation in the model are described in Section 3, which also contains a summary of the method. To illustrate the method some simulated examples are included in Section 4. Further topics, such as optimal bandwidths and optimal forgetting factors are considered in Section 5. Finally, we conclude on the paper in Section 6.

2 Conditional parametric models and local polynomial estimates

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

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

where $\boldsymbol{\theta}(\cdot)$ is a vector of coefficient-functions to be estimated and e_s is the noise term. Note that $\mathbf{x}_s$ may contain lagged values of the response. The dimension of $\mathbf{x}_s$ can be quite large, but the dimension of $\mathbf{u}_s$ must be low (1 or 2) for practical purposes (Hastie & Tibshirani 1990, pp. 83-84). In (Nielsen et al. 1997) the dimensions 30 and 1 is used. Estimation in (1), using methods similar to the methods by Cleveland & Devlin (1988), is described for some special cases in (Anderson et al. 1994) and (Hastie & Tibshirani 1993). A more general description can be found in (Nielsen et al. 1997). To make the paper self-contained the method is outlined below.

The functions $\boldsymbol{\theta}(\cdot)$ in (1) are estimated at a number of distinct points by approximating the functions using polynomials and fitting the resulting linear model locally to each of these *fitting points*. To be more specific let $\mathbf{u}$ denote a particular fitting point. Let $\theta_j(\cdot)$ be the j'th element of $\boldsymbol{\theta}(\cdot)$ and let $\mathbf{p}_{d(j)}(\mathbf{u})$ be a column vector of terms in the corresponding d-order polynomial evaluated at $\mathbf{u}$, if for instance $\mathbf{u} = [u_1 \ u_2]^T$ then $\mathbf{p}_2(\mathbf{u}) = [1 \ u_1 \ u_2 \ u_1^2 \ u_1 u_2 \ u_2^2]^T$. Furthermore, let $\mathbf{x}_s = [x_{1,s} \dots x_{p,s}]^T$.

With

$$\mathbf{z}_s^T = \left[x_{1,s}\mathbf{p}_{d(1)}^T(\mathbf{u}_s) \dots x_{j,s}\mathbf{p}_{d(j)}^T(\mathbf{u}_s) \dots x_{p,s}\mathbf{p}_{d(p)}^T(\mathbf{u}_s) \right] \tag{2}$$

and

$$\boldsymbol{\phi}_u^T = [\boldsymbol{\phi}_{u,1}^T \dots \boldsymbol{\phi}_{u,j}^T \dots \boldsymbol{\phi}_{u,p}^T], \tag{3}$$

where $\boldsymbol{\phi}_{u,j}$ is a column vector of local coefficients at $\mathbf{u}$ corresponding to $x_{j,s}\mathbf{p}_{d(j)}(\mathbf{u}_s)$. The linear model

$$y_s = \mathbf{z}_s^T \boldsymbol{\phi}_u + e_s; \quad i = 1, \dots, N, \tag{4}$$

is then fitted locally to $\mathbf{u}$ using weighted least squares (WLS), i.e.

$$\hat{\boldsymbol{\phi}}(\mathbf{u}) = \underset{\boldsymbol{\phi}_u}{\operatorname{argmin}} \sum_{s=1}^{N} w_u(\mathbf{u}_s)(y_s - \mathbf{z}_s^T \boldsymbol{\phi}_u)^2, \tag{5}$$

for which a unique closed-form solution exists provided the matrix with rows $\mathbf{z}_s^T$ corresponding to non-zero weights has full rank. The weights are assigned as

$$w_u(\mathbf{u}_s) = W\left(\frac{||\mathbf{u}_s - \mathbf{u}||}{\hbar(\mathbf{u})} \right), \tag{6}$$

where $||\cdot||$ denotes the Euclidean norm, $\hbar(\mathbf{u})$ is the bandwidth used for the particular fitting point, and $W(\cdot)$ is a weight function taking non-negative arguments. Here we follow Cleveland & Devlin (1988) and use

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

i.e. the weights are between 0 and 1. The elements of $\boldsymbol{\theta}(\mathbf{u})$ are estimated by

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

where $\hat{\boldsymbol{\phi}}_j(\mathbf{u})$ is the WLS estimate of $\boldsymbol{\phi}_{u,j}$. The estimates of the coefficient-functions obtained as outlined above are called *local polynomial estimates*. For the special case where all coefficient-functions are approximated by constants we use the term local constant estimates.

If $\hbar(\mathbf{u})$ is constant for all values of $\mathbf{u}$ it is denoted a fixed bandwidth. If $\hbar(\mathbf{u})$ is chosen so that a certain fraction α of the observations fulfill $||\mathbf{u}_s - \mathbf{u}|| \leq \hbar(\mathbf{u})$ then α is denoted a nearest neighbour bandwidth. A

bandwidth specified according to the nearest neighbour principle is often used as a tool to vary the actual bandwidth with the local density of the data.

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

3 Adaptive estimation

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

3.1 The proposed method

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

$$\hat{\boldsymbol{\phi}}_t = \underset{\boldsymbol{\phi}}{\operatorname{argmin}} \sum_{s=1}^{t} \lambda^{t-s} (y_s - \mathbf{z}_s^T \boldsymbol{\phi})^2, \tag{9}$$

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

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

where $w_u(\mathbf{u}_s)$ is a weight on observation s depending on the fitting point $\mathbf{u}$ and $\mathbf{u}_s$, see Section 2.

In Section 3.2 it will be shown how the estimator (10) can be formulated recursively, but here we will briefly comment on the estimator and its relations to non-parametric regression. A special case is obtained if $\mathbf{z}_s = 1$ for all s, then simple calculations show that

$$\hat{\phi}_t(\mathbf{u}) = \frac{\sum_{s=1}^{t} \lambda^{t-s} w_u(\mathbf{u}_s) y_s}{\sum_{s=1}^{t} \lambda^{t-s} w_u(\mathbf{u}_s)}, \tag{11}$$

and for $\lambda = 1$ this is a kernel estimator of $\phi(\cdot)$ in $y_s = \phi(\mathbf{u}_s) + e_s$, cf. (Härdle 1990, p. 30). For this reason (11) is called an adaptive kernel estimator of $\phi(\cdot)$ and the estimator (10) may be called an adaptive local constant estimator of the coefficient-functions $\boldsymbol{\phi}(\cdot)$ in the conditional parametric model $y_s = \mathbf{z}_s^T \boldsymbol{\phi}(\mathbf{u}_s) + e_s$. Using the same techniques as in Section 2 this can be used to implement adaptive local polynomial estimation in models like (1).

3.2 Recursive formulation

Following the same arguments as in Ljung & Söderström (1983) it is readily shown that the adaptive estimates (10) can be found recursively as

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

and

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

It is seen that existing numerical procedures implementing adaptive recursive least squares for linear models can be applied, by replacing $\mathbf{z}_t$ and y_t in the existing procedures with $\mathbf{z}_t\sqrt{w_u(\mathbf{u}_t)}$ and $y_t\sqrt{w_u(\mathbf{u}_t)}$, respectively. Note that $\mathbf{z}_t^T\hat{\boldsymbol{\phi}}_{t-1}(\mathbf{u})$ is a predictor of y_t locally with respect to $\mathbf{u}$ and for this reason it is used in (12). To predict y_t a predictor like $\mathbf{z}_t^T\hat{\boldsymbol{\phi}}_{t-1}(\mathbf{u}_t)$ is appropriate.

3.3 Modified updating formula

When $\mathbf{u}_t$ is far from the particular fitting point $\mathbf{u}$ it is clear from (12) and (13) that $\hat{\boldsymbol{\phi}}_t(\mathbf{u}) \approx \hat{\boldsymbol{\phi}}_{t-1}(\mathbf{u})$ and $\mathbf{R}_{u,t} \approx \lambda\mathbf{R}_{u,t-1}$, i.e. old observations are down-weighted without new information becoming available. This may result in abruptly changing estimates if $\mathbf{u}$ is not visited regularly, since the matrix $\mathbf{R}$ is decreasing exponentially in this case. Hence it is proposed to modify (13) to ensure that the past is weighted down only when new information becomes available, i.e.

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

where $v(\cdot\,;\lambda)$ is a nowhere increasing function on $[0;1]$ fulfilling $v(0;\lambda) = 1/\lambda$ and $v(1;\lambda) = 1$. Note that this requires that the weights span the interval ranging from zero to one. This is fulfilled for weights generated as described in Section 2. In this paper we consider only the linear function $v(w;\lambda) = 1/\lambda - (1/\lambda - 1)w$, for which (14) becomes

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

It is reasonable to denote

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

the *effective forgetting factor* for point $\mathbf{u}$ at time t.

When using (14) or (15) it is ensured that $\mathbf{R}_{u,t}$ can not become singular because the process $\{\mathbf{u}_t\}$ moves away from the fitting point for a longer period. However, the process $\{\mathbf{z}_t\}$ should be persistently excited as for linear ARX-models. In this case, given the weights, the estimates define a global minimum corresponding to (10).

3.4 Nearest neighbour bandwidth

Assume that $\mathbf{u}_t$ is a stochastic variable and that the pdf $f(\cdot)$ of $\mathbf{u}_t$ is known and constant over t. Based on a nearest neighbour bandwidth the actual bandwidth can then be calculated for a number of fitting points $\mathbf{u}$ placed within the domain of $f(\cdot)$ and used to generate the weights $w_u(\mathbf{u}_t)$. The actual bandwidth $\hbar(\mathbf{u})$ corresponding to the point $\mathbf{u}$ will be related to the nearest neighbour bandwidth α by

$$\alpha = \int_{\mathbb{D}_u} f(\boldsymbol{\nu}) d\boldsymbol{\nu}, \tag{17}$$

where $\mathbb{D}_u = \{\boldsymbol{\nu} \in \mathbb{R}^d \mid ||\boldsymbol{\nu} - \mathbf{u}|| \leq \hbar(\mathbf{u})\}$ is the neighbour-hood, d is the dimension of $\mathbf{u}$, and $||\cdot||$ is the Euclidean norm. In applications the density $f(\cdot)$ is often unknown. However, $f(\cdot)$ can be estimated from data, e.g. by the empirical pdf.

3.5 Effective number of observations

In order to select an appropriate value for α the effective number of observations used for estimation must be considered. In Appendix A it is shown that under certain conditions, when the modified updating (15) is used,

$$\tilde{\eta}_u = \frac{1}{1 - E[\lambda^u_{eff}(t)]} = \frac{1}{(1-\lambda)E[w_u(\mathbf{u}_t)]} \tag{18}$$

is a lower bound on the effective number of observations (in the direction of time) corresponding to a fitting point $\mathbf{u}$. Generally (18) can be considered an approximation. When selecting α and λ it is then natural to require that the number of observations within the bandwidth, i.e. $\alpha\tilde{\eta}_u$, is sufficiently large to justify the complexity of the model and the order of the local polynomial approximations.

As an example consider $u_t \sim N(0,1)$ and $\lambda = 0.99$ where the effective number of observations within the bandwidth, $\alpha\tilde{\eta}_u$, is displayed in Figure 1. It is seen that $\alpha\tilde{\eta}_u$ depends strongly on the fitting point u but only moderately on α. When investigating the dependence of $\alpha\tilde{\eta}_u$ on λ and α it turns out that $\alpha\tilde{\eta}_u$ is almost solely determined by λ. In conclusion, for

the example considered, the effective forgetting factor $\lambda^u_{eff}(t)$ will be affected by the nearest neighbour bandwidth, so that the effective number of observations within the bandwidth will be strongly dependent on λ, but only weakly dependent on the bandwidth (α). The ratio between the rate at which the weights on observations goes to zero in the direction of time and the corresponding rate in the direction of u_t will be determined by α.

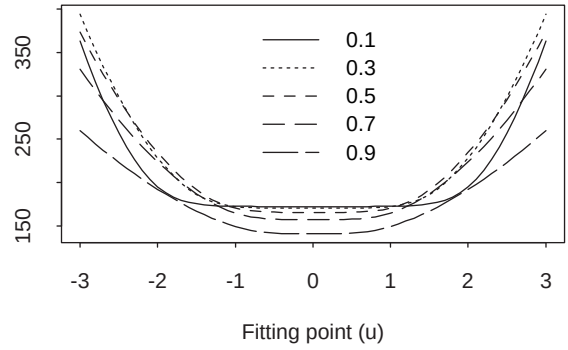


Figure 1: Effective number of observations within the bandwidth ($\alpha\tilde{\eta}_u(u)$) for $\alpha = 0.1, \ldots, 0.9$ and $\lambda = 0.99$.

As it is illustrated by Figure 1 the effective number of observations behind each of the local approximations depends on the fitting point. This is contrary to the non-adaptive nearest neighbour method, cf. Section 2, and may result in a somewhat unexpected behaviour of the estimates. If the system follows a linear ARX-model and if the coefficients of the system are estimated as coefficient-functions then both adaptive and non-adaptive nearest neighbour approaches will be unbiased. However, for this example the variance of local constant estimates will decrease for increasing values of $|u|$. This is verified by simulations, which also show that local linear and quadratic approximations results in increased variance for large $|u|$. Note that, when the true function is not a constant, the local constant approximation may result in excess bias, see e.g. (Nielsen et al. 1997).

If λ is varied with the fitting point as $\lambda(\mathbf{u}) = 1 - 1/(T_0 E[w_u(\mathbf{u}_t)])$ then $\tilde{\eta}_u = T_0$. Thus, the effective number of observations within the bandwidth is constant across fitting points. Furthermore, T_0 can be interpreted as the memory time constant. To avoid highly variable estimates

of $E[w_u(\mathbf{u}_t)]$ in the tails of the distribution of $\mathbf{u}_t$ the estimates should be based on a parametric family of distributions. However, in the remaining part of this paper λ is not varied across fitting points.

3.6 Summary of the method

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

Besides λ in (15), prior to the application of the algorithm a number of fitting points $\mathbf{u}^{(i)};\ i = 1, \ldots, n_{fp}$ in which the coefficient-functions are to be estimated has to be selected. Furthermore the bandwidth associated with each of the fitting points $\hbar^{(i)};\ i = 1, \ldots, n_{fp}$ and the degrees of the approximating polynomials $d(j);\ j = 1, \ldots, p$ have to be selected for each of the p coefficient-functions. For simplicity the degree of the approximating polynomial for a particular coefficient-function will be fixed across fitting points. Finally, initial estimates of the coefficient-functions in the model corresponding to local constant estimates, i.e. $\hat{\boldsymbol{\phi}}_0(\mathbf{u}^{(i)})$, must be chosen. Also, the matrices $\mathbf{R}_{u^{(i)},0}$ must be chosen. One possibility is $\mathrm{diag}(\epsilon, \ldots, \epsilon)$, where ϵ is a small positive number.

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

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

- Construct the explanatory variables corresponding to local constant estimates using (2):
  $\mathbf{z}_t^T = [x_{1,t}\mathbf{p}_{d(1)}^T(\mathbf{u}_t) \ldots x_{p,t}\mathbf{p}_{d(p)}^T(\mathbf{u}_t)]$.
- Calculate the weight using (6) and (7):
  $w_{u^{(i)}}(\mathbf{u}_t) = (1 - (||\mathbf{u}_t - \mathbf{u}^{(i)}||/\hbar^{(i)})^3)^3$, if $||\mathbf{u}_t - \mathbf{u}^{(i)}|| < \hbar^{(i)}$ and zero

otherwise.

- Find the effective forgetting factor using (16):
  $\lambda_{eff}^{(i)}(t) = 1 - (1-\lambda) w_{u^{(i)}}(\mathbf{u}_t)$.
- Update $\mathbf{R}_{u^{(i)},t-1}$ using (15):
  $\mathbf{R}_{u^{(i)},t} = \lambda_{eff}^{(i)}(t)\mathbf{R}_{u^{(i)},t-1} + w_{u^{(i)}}(\mathbf{u}_t)\mathbf{z}_t\mathbf{z}_t^T$.
- Update $\hat{\boldsymbol{\phi}}_{t-1}(\mathbf{u}^{(i)})$ using (12):
  $\hat{\boldsymbol{\phi}}_t(\mathbf{u}^{(i)}) = \hat{\boldsymbol{\phi}}_{t-1}(\mathbf{u}^{(i)}) + w_{u^{(i)}}(\mathbf{u}_t)\mathbf{R}^{-1}_{u^{(i)},t}\mathbf{z}_t \left[y_t - \mathbf{z}_t^T \hat{\boldsymbol{\phi}}_{t-1}(\mathbf{u}^{(i)}) \right]$.
- Calculate the updated local polynomial estimates of the coefficient-functions using (8):
  $\hat{\theta}_{jt}(\mathbf{u}^{(i)}) = \mathbf{p}_{d(j)}^T(\mathbf{u}^{(i)})\, \hat{\boldsymbol{\phi}}_{j,t}(\mathbf{u}^{(i)}); \quad j = 1, \ldots p$

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

4 Simulations

Aspects of the proposed method are illustrated in this section. When the modified updating formula (15) is used the general behaviour of the method for different bandwidths is illustrated in Section 4.1. In Section 4.2 results obtained using the two updating formulas (13) and (15) are compared.

The simulations are performed using the non-linear model

$$y_t = a(t, u_{t-1}) y_{t-1} + b(t, u_{t-1}) x_t + e_t, \tag{19}$$

where $\{x_t\}$ is the input process, $\{u_t\}$ is the process controlling the coefficients, $\{y_t\}$ is the output process, and $\{e_t\}$ is a white noise standard Gaussian process. The coefficient-functions are simulated as

$$a(t,u) = 0.3 + (0.6 - \frac{1.5}{N}t)\, \exp\left(-\frac{(u - \frac{0.8}{N}t)^2}{2(0.6 - \frac{0.1}{N}t)^2} \right)$$

and

$$b(t,u) = 2 - \exp\left(-\frac{(u+1-\frac{2}{N}t)^2}{0.32},\right)$$

where $t = 1, \ldots, N$ and $N = 5000$, i.e. $a(t,u)$ ranges from -0.6 to 0.9 and $b(t,u)$ ranges from 1 to 2. The functions are displayed in Figure 2. As indicated by the figure both coefficient-functions are based on a Gaussian density in which the mean and variance varies linearly with time.

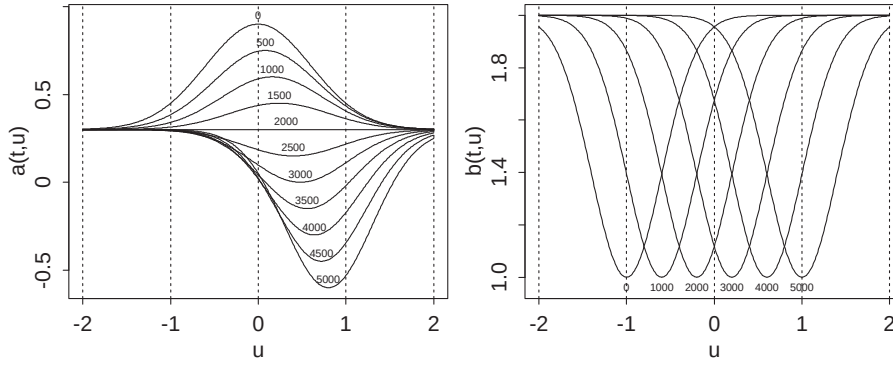


Figure 2: The time-varying coefficient-functions plotted for equidistant points in time as indicated on the plots.

Local linear adaptive estimates of the functions $a()$ and $b()$ are then found using the proposed procedure with the model

$$y_t = a(u_{t-1})y_{t-1} + b(u_{t-1})x_t + e_t. \tag{20}$$

In all cases initial estimates of the coefficient-functions are set to zero and during the initialization the estimates are not updated, for the fitting point considered, until ten observations have received a weight of 0.5 or larger.

4.1 Highly correlated input processes

In the simulation presented in this section a strongly correlated $\{\mathbf{u}_t\}$ process is used and also the $\{\mathbf{x}_t\}$ process is quite strongly correlated. This allows us to illustrate various aspects of the method. For less correlated series the performance is much improved. The data are generated using (19) where $\{x_t\}$ and $\{u_t\}$ are zero mean $AR(1)$-processes with poles in

0.9 and 0.98, respectively. The variance for both series is one and the series are mutually independent. In Figure 3 the data are displayed. Based on these data adaptive estimation in (20) are performed using nearest neighbour bandwidths, calculated assuming a standard Gaussian distribution for u_t.

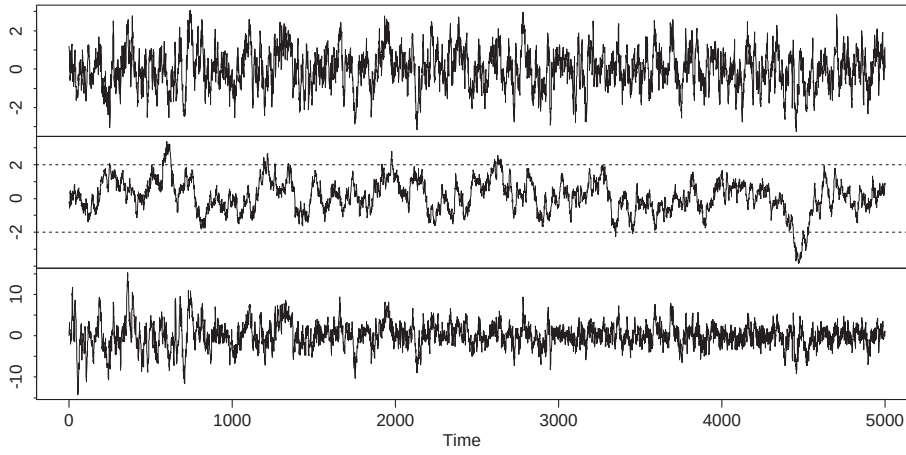


Figure 3: Simulated output (bottom) when x_t (top) and u_t (middle) are $AR(1)$-processes.

The results obtained using the modified updating formula (15) are displayed for fitting points $u = -2, -1, 0, 1, 2$ in Figures 4 and 5. For the first 2/3 of the period the estimates at $u = -2$, i.e. $\hat{a}(-2)$ and $\hat{b}(-2)$, only gets updated occasionally. This is due to the correlation structure of $\{u_t\}$ as illustrated by the realization displayed in Figure 3.

For both estimates the bias is most pronounced during periods in which the true coefficient-function changes quickly for values of u_t near the fitting point considered. This is further illustrated by the true functions in Figure 2 and it is, for instance clear that adaption to $a(t, 1)$ is difficult for $t > 3000$. Furthermore, $u = 1$ is rarely visited by $\{u_t\}$ for $t > 3000$, see Figure 3. In general, the low bandwidth ($\alpha = 0.3$) seems to result in large bias, presumably because the effective forgetting factor is increased on average, cf. Section 3.5. Similarly, the high bandwidth ($\alpha = 0.7$) result in large bias for $u = 2$ and $t > 4000$. A nearest neighbour bandwidth of 0.7 corresponds to an actual bandwidth of approximately 2.5 at $u = 2$ and since most values of u_t are below one, it is clear that the estimates at $u = 2$ will be highly influenced by the actual function values for u

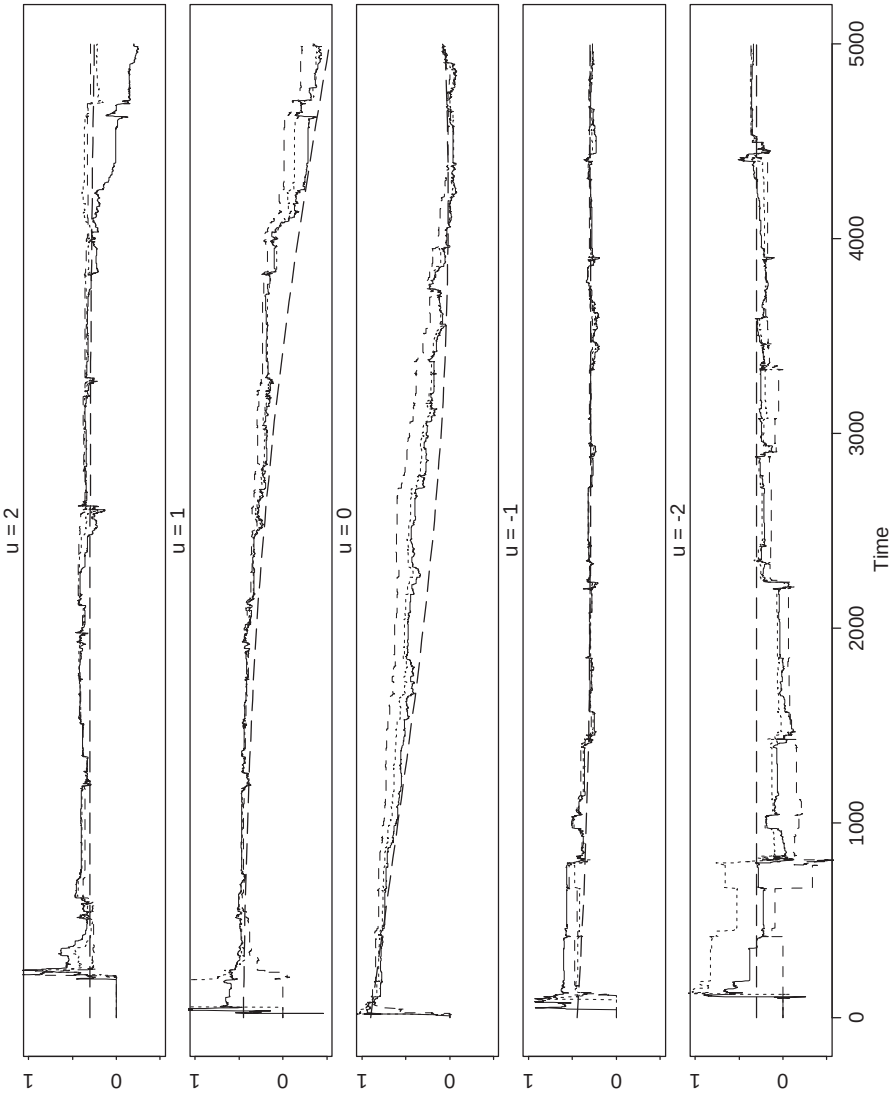


Figure 4: Adaptive estimates of $a(u)$ using local linear approximations and nearest neighbour bandwidths 0.3 (dashed), 0.5 (dotted), and 0.7 (solid). True values are indicated by smooth dashed lines.

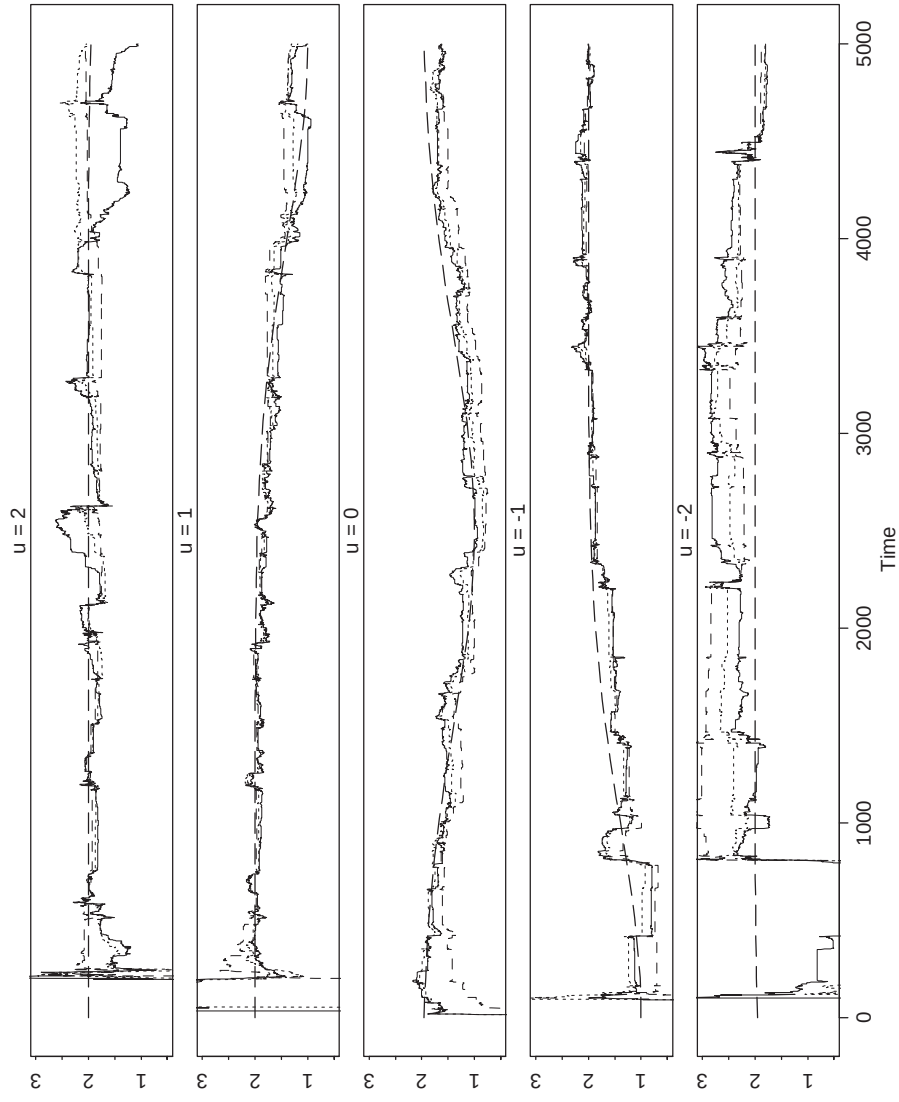


Figure 5: Adaptive estimates of $b(u)$ using local linear approximations and nearest neighbour bandwidths 0.3 (dashed), 0.5 (dotted), and 0.7 (solid). True values are indicated by smooth dashed lines.

near one. From Figure 2 it is seen that for $t > 4000$ the true values at $u = 1$ is markedly lower that the true values at $u = 2$. Together with the fact that $u = 2$ is not visited by $\{u_t\}$ for $t > 4000$ this explains the observed bias at $u = 2$, see Figure 6.

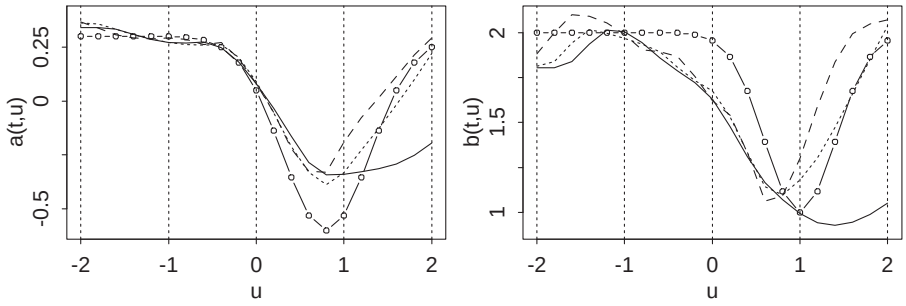


Figure 6: Adaptive estimates for the example considered in Section 4.1 at $t = 5000$ for $\alpha = 0.3$ (dashed), 0.5 (dotted), 0.7 (solid). True values are indicated by circles and fitting points ranging from -2 to 2 in steps of 0.2 are used.

4.2 Abrupt changes in input signals

One of the main advantages of the modified updating formula (15) over the normal updating formula (13) is that it does not allow fast changes in the estimates at fitting points which has not been visited by the process $\{u_t\}$ for a longer period. If, for instance, we wish to adaptively estimate the stationary relation between the heat consumption of a town and the ambient air temperature then $\{u_t\}$ contains an annual fluctuation and at some geographical locations the transition from, say, warm to cold periods may be quite fast. In such a situation the normal updating formula (13) will, essentially, forget the preceding winter during the summer, allowing for large changes in the estimate at low temperatures during some initial period of the following winter. Actually, it is possible that, using the normal updating formula will result in a nearly singular $\mathbf{R}_t$.

To illustrate this aspect 5000 observations are simulated using the model (19). The sequence $\{x_t\}$ is simulated as a standard Gaussian $AR(1)$-process with a pole in 0.9. Furthermore, $\{u_t\}$ is simulated as an iid

process where

$$u_t \sim \begin{cases} N(0,1), & t = 1,\ldots,1000 \\ N(3/2, 1/6^2), & t = 1001,\ldots,4000 \\ N(-3/2, 1/6^2), & t = 4001,\ldots,5000 \end{cases}$$

To compare the two methods of updating, i.e. (13) and (15), a fixed λ is used in (15) across the fitting points and the effective forgetting factors are designed to be equal. If $\tilde{\lambda}$ is the forgetting factor corresponding to (13) it can be varied with u as

$$\tilde{\lambda}(u) = E[\lambda^u_{eff}(t)] = 1 - (1-\lambda)E[w_u(u_t)],$$

where $E[w_u(u_t)]$ is calculated assuming that u_t is standard Gaussian, i.e. corresponding to $1 \leq t \leq 1000$. A nearest neighbour bandwidth of 0.5 and $\lambda = 0.99$ are used, which results in $\tilde{\lambda}(0) = 0.997$ and $\tilde{\lambda}(\pm 2) = 0.9978$.

The corresponding adaptive estimates obtained for the fitting point $u = -1$ are shown in Figure 7. The figure illustrates that for both methods the updating of the estimates stops as $\{u_t\}$ leaves the fitting point $u = -1$. Using the normal updating (13) of $\mathbf{R}_t$ its value is multiplied by $\tilde{\lambda}(-1)^{3000} \approx 0.00015$ as $\{u_t\}$ returns to the vicinity of the fitting point. This results in large fluctuations of the estimates, starting at $t = 4001$. As opposed to this, the modified updating (15) does not lead to such fluctuations after $t = 4000$.

5 Further topics

Optimal bandwidth and forgetting factor: So far in this paper it has been assumed that the bandwidths used over the range of $\mathbf{u}_t$ is derived from the nearest neighbour bandwidth α and it has been indicated how it can be ensured that the average forgetting factor is large enough.

However, the adaptive and recursive method is well suited for forward validation (Hjorth 1994) and hence tuning parameters can be selected by minimizing, e.g. the root mean square of the one-step prediction error (using observed $\mathbf{u}_t$ and $\mathbf{x}_t$ to predict y_t, together with interpolation between fitting points to obtain $\hat{\boldsymbol{\theta}}_{t-1}(\mathbf{u}_t)$).

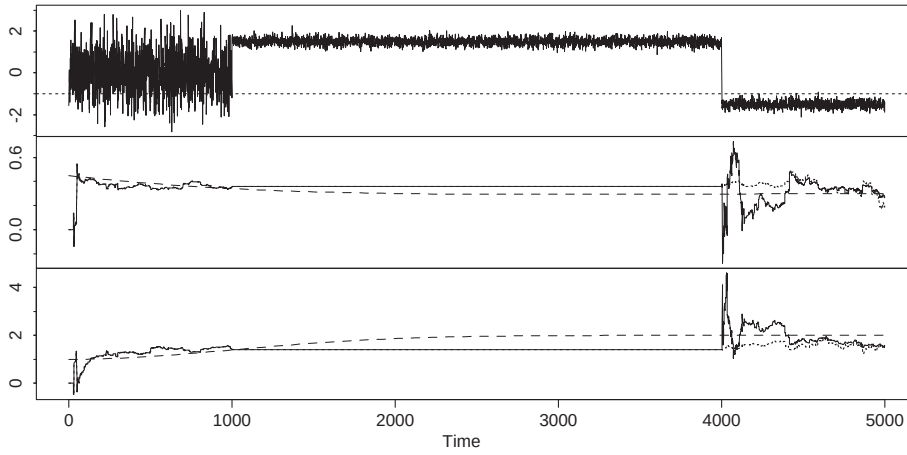


Figure 7: Realization of $\{u_t\}$ (top) and adaptive estimates of $a(-1)$ (middle) and $b(-1)$ (bottom), using the normal updating formula (solid) and the modified updating formula (dotted). True values are indicated by dashed lines.

There are numerous ways to define the tuning parameters. A simple approach is to use (λ, α), cf. (15) and (17). A more ambiguous approach is to use both λ and $\hbar$ for each fitting point $\mathbf{u}$. Furthermore, tuning parameters controlling scaling and rotation of $\mathbf{u}_s$ and the degree of the local polynomial approximations may also be considered.

If n fitting points are used this amounts to $2n$, or more, tuning parameters. To make the dimension of the (global) optimization problem independent of n and to have $\lambda(\mathbf{u})$ and $\hbar(\mathbf{u})$ vary smoothly with $\mathbf{u}$ we may choose to restrict $\lambda(\mathbf{u})$ and $\hbar(\mathbf{u})$, or appropriate transformations of these (logit for λ and log for $\hbar$), to follow a spline basis (de Boor 1978, Lancaster & Salkauskas 1986). This is similar to the smoothing of spans described by Friedman (1984).

Local time-polynomials: In this paper local polynomial approximations in the direction of time is not considered. Such a method is proposed for usual ARX-models by Joensen, Nielsen, Nielsen & Madsen (2000). This method can be combined with the method described here and will result in local polynomial approximations where cross-products between time and the conditioning variables ($\mathbf{u}_t$) are excluded.

6 Conclusion and discussion

In this paper methods for adaptive and recursive estimation in a class of non-linear autoregressive models with external input are proposed. The model class considered is conditionally parametric ARX-models (CPARX-model), which is a conventional ARX-model in which the parameters are replaced by smooth, but otherwise unknown, functions of a low-dimensional input process. These functions are estimated adaptively and recursively without specifying a global parametric form. One possible application of CPARX-models is the modelling of varying time delays, cf. (Nielsen et al. 1997).

The methods can be seen as generalizations or combinations of recursive least squares with exponential forgetting (Ljung & Söderström 1983), local polynomial regression (Cleveland & Devlin 1988), and conditional parametric fits (Anderson et al. 1994). Hence, the methods constitutes an extension to the notion of local polynomial estimation. The so called modified method is suggested for cases where the process controlling the coefficients are highly correlated or exhibit seasonal behaviour. The estimates at each time step can be seen as solutions to a range of weighted least squares regressions and therefore the solution is unique for well behaved input processes. A particular feature of the modified method is that the effective number of observations behind the estimates will be almost independent of the actual bandwidth. This is accomplished by varying the effective forgetting factor with the bandwidth. The bandwidth mainly controls the rate at which the weights corresponding to exponential forgetting goes to zero relatively to the rate at which the remaining weights goes to zero.

For some applications it may be possible to specify global polynomial approximations to the coefficient-functions of a CPARX-model. In this situation the adaptive recursive least squares method can be applied for tracking the parameters defining the coefficient-functions for all values of the input process. However, if the argument(s) of the coefficient-functions only stays in parts of the space corresponding to the possible values of the argument(s) for longer periods this may seriously affect the estimates of the coefficient-functions for other values of the argument(s), as it corresponds to extrapolation using a fitted polynomial. This problem is effectively solved using the conditional parametric model in com-

bination with the modified updating formula.

A Effective number of observations

Using the modified updating formula, as described in Section 3.3, the estimates at time t can be written as

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

where

$$\beta(t,t) = 1,$$

and, for $s < t$

$$\beta(t,s) = \prod_{j=s+1}^{t} \lambda_{eff}^u(j) = \lambda_{eff}^u(t)\beta(t-1,s),$$

where $\lambda_{eff}^u(t)$ is given by (16). It is then obvious to define the effective number of observations (in the direction of time) as

$$\eta_u(t) = \sum_{i=0}^{\infty} \beta(t,t-i) = 1 + \lambda_{eff}^u(t) + \lambda_{eff}^u(t)\lambda_{eff}^u(t-1) + \ldots \quad \text{(A.1)}$$

Suppose that the fitting point $\mathbf{u}$ is chosen so that $E[\eta_u(t)]$ exists. Consequently, when $\{\lambda_{eff}^u(t)\}$ is i.i.d. and when $\bar{\lambda}_u \in [0,1)$ denotes $E[\lambda_{eff}^u(t)]$, the average effective number of observations is

$$\bar{\eta}_u = 1 + \bar{\lambda}_u + \bar{\lambda}_u^2 + \ldots = \frac{1}{1-\bar{\lambda}_u}.$$

When $\{\lambda_{eff}^u(t)\}$ is not i.i.d., it is noted that since the expectation operator is linear, $E[\eta_u(t)]$ is the sum of the expected values of each summand in (A.1). Hence, $E[\eta_u(t)]$ is independent of t if $\{\lambda_{eff}^u(t)\}$ is strongly stationary, i.e. if $\{\mathbf{u}_t\}$ is strongly stationary. From (A.1)

$$\eta_u(t) = 1 + \lambda_{eff}^u(t)\eta_u(t-1) \quad \text{(A.2)}$$

is obtained, and from the definition of covariance it then follows, that

$$\bar{\eta}_u = \frac{1 + Cov[\lambda_{eff}^u(t), \eta_u(t-1)]}{1-\bar{\lambda}_u} \geq \frac{1}{1-\bar{\lambda}_u}, \quad \text{(A.3)}$$

since $0 < \lambda < 1$ and assuming, that the covariance between $\lambda^u_{eff}(t)$ and $\eta_u(t-1)$ is positive. Note that, if the process $\{\mathbf{u}_t\}$ behaves such that if it has been near $\mathbf{u}$ for a longer period up to time $t-1$ it will tend to be near $\mathbf{u}$ at time t also, a positive covariance is obtained. It is the experience of the authors that such a behaviour of a stochastic process is often encountered in practice.

As an alternative to the calculations above $\lambda^u_{eff}(t)\eta_u(t-1)$ may be linearized around $\bar{\lambda}_u$ and $\bar{\eta}_u$. From this it follows, that if the variances of $\lambda^u_{eff}(t)$ and $\eta_u(t-1)$ are small then

$$\bar{\eta}_u \approx \frac{1}{1-\bar{\lambda}_u}.$$

Therefore we may use $1/(1-\bar{\lambda}_u)$ as an approximation to the effective number of observations, and in many practical applications it will be an lower bound, c.f. (A.3). By assuming a stochastic process for $\{\mathbf{u}_t\}$ the process $\{\eta_u(t)\}$ can be simulated using (A.2) whereby the validity of the approximation can be addressed.

References

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

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

Chen, R. & Tsay, R. S. (1993*a*), 'Functional-coefficient autoregressive models', *Journal of the American Statistical Association* **88**, 298–308.

Chen, R. & Tsay, R. S. (1993*b*), 'Nonlinear additive ARX models', *Journal of the American Statistical Association* **88**, 955–967.

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

de Boor, C. (1978), *A Practical Guide to Splines*, Springer Verlag, Berlin.

Friedman, J. H. (1984), A variable span smoother, Technical Report 5, Laboratory for Computational Statistics, Department of Statistics, Stanford University, California.

Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, UK.

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

Hastie, T. & Tibshirani, R. (1993), 'Varying-coefficient models', *Journal of the Royal Statistical Society, Series B, Methodological* **55**, 757–796.

Hjorth, J. S. U. (1994), *Computer Intensive Statistical Methods: Validation Model Selection and Bootstrap*, Chapman & Hall, London/New York.

Joensen, A. K., Nielsen, H. A., Nielsen, T. S. & Madsen, H. (2000), 'Tracking time-varying parameters with local regression', *Automatica* **36**, 1199–1204.

Lancaster, P. & Salkauskas, K. (1986), *Curve and Surface Fitting: An Introduction*, Academic Press, New York/London.

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

Nielsen, H. A., Nielsen, T. S. & Madsen, H. (1997), ARX-models with parameter variations estimated by local fitting, *in* Y. Sawaragi & S. Sagara, eds, '11th IFAC Symposium on System Identification', Vol. 2, pp. 475–480.

Thuvesholmen, M. (1997), 'An on-line crossvalidation bandwidth selector for recursive kernel regression', Lic thesis, Department of Mathematical Statistics, Lund University, Sweden.

Vilar-Fernández, J. A. & Vilar-Fernández, J. M. (1998), 'Recursive estimation of regression functions by local polynomial fitting', *Annals of the Institute of Statistical Mathematics* **50**, 729–754.

Paper C

C

A new reference for wind power forecasting

Originally published as

T. S. Nielsen, A. K. Joensen, H. Madsen, L. Landberg and G. Giebel. A new reference for wind power forecasting. *Wind Energy*, Vol **1**, pages 29–34, 1999.

A new reference for wind power forecasting

Torben S. Nielsen[1], Alfred K. Joensen[1], Henrik Madsen[1], Lars Landberg[2] and Gregor Giebel[2]

Abstract

In recent years some research towards developing forecasting models for wind power or energy has been carried out. In order to evaluate the prediction ability of these models, the forecasts are usually compared to those of the persistence forecast model. As shown in this paper, it is, however, not reasonable to use the persistence model when the forecast length is more than a few hours. Instead, a new statistical reference for predicting wind power, which basically is a weighting between the persistence and the mean of the power, is proposed. This reference forecast model is adequate for all forecast lengths, and like the persistence model, it requires only measured time series as input.

Keywords: Persistence, correlation, wind power, reference forecast model

1 Introduction

In this paper we propose a new reference model, which should be used instead of the persistence model (1), when short term, say up to 48 hours, forecasting models for wind power or energy are evaluated.

There are two types of wind power forecasting models, physical models as in (Landberg 1999, Landberg 1994, Landberg & Watson 1994) and statistical models as in (Joensen 1997, Nielsen & Madsen 1996, Madsen 1996). Up to now the reference for these models, and many other meteorological forecasting models, has been the persistence model given by

$$p_{t+k} = p_t + \varepsilon_{t+k} \tag{1}$$

[1]Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

[2]Department of Wind Energy and Atmospheric Physics, Risoe National Laboratory, DK-4000 Roskilde, Denmark

where t is a time index, k is the look ahead time, p is e.g. wind power or energy, and ε denotes the residual. The forecast, $\hat{p}$, obtained using this model is

$$\hat{p}_{t+k} = p_t \tag{2}$$

which states that the expected value k time steps ahead is equal to the most recent value. In statistics this is called the persistence or näive predictor. In this paper we shall denote (1) the persistence forecast model.

The model (2) is a simple description, but yet very powerful. This is because the atmosphere can be considered quasi-stationary, i.e. changing very slowly. A characteristic time scale in the atmosphere is f^{-1}, where f is the Coriolis parameter. Using $10^{-4}s^{-1}$ for f gives that this time scale is approximately 3 hours, see (Landberg 1994).

To compare the forecasts to the observations, the root mean square error (RMS) or the mean square error (MSE) is usually used. The MSE for the persistence forecast model is given by

$$MSE_p = \frac{1}{N-k}\sum_{t=1}^{N-k}(p_{t+k} - \hat{p}_{t+k})^2 = \frac{1}{N-k}\sum_{t=1}^{N-k}(p_{t+k} - p_t)^2 \tag{3}$$

where N is the number of observations. The RMS is given by

$$RMS_p = \sqrt{MSE_p} \tag{4}$$

Due to the quasi-stationarity of the atmosphere, p_{t+k} will be rather close to p_t when the time step k is less than a few hours, which means that the MSE will be small compared to the MSE for large k.

As k gets larger, $k \gg f^{-1}$, or say above 36 hours, the flow in the atmosphere will no longer remain constant, and the correlation between p_{t+k} and p_t will tend to zero. This means that the present flow provides no information about the future flow, and the model (1) which correlates the future flow to the present is no longer reasonable.

Instead the mean of the flow could be used as a simple reference when the correlation is zero. In Appendix A it is shown that the MSE for the

persistence actually is twice the *MSE* of the mean predictor, when the correlation is zero.

It is thus quite obvious to suggest a new reference forecast model as a weighting between the persistence and the mean where the weighting for different forecast lengths is determined by the correlation between p_t and p_{t+k}. In this paper such a reference is proposed. Wind power is considered, but the proposed reference can be used for many other meteorological quantities, e.g. wind speed or energy.

2 The new reference forecast model

As outlined in the introduction, the proposed reference forecast model is a weighting between the persistence and the mean, i.e. the k step forecast is written

$$\hat{p}_{t+k} = a_k p_t + (1 - a_k)\bar{p} \tag{5}$$

where p_t is the most recent measurement of the wind power, and $\bar{p}$ the estimated mean of the power given by

$$\bar{p} = \frac{1}{N}\sum_{t=1}^{N} p_t \tag{6}$$

When k is small a_k should be approximately one and the reference thus corresponds to persistence, but when k is large and the correlation is zero, a_k should be zero and the forecast is simply the mean. It is thus reasonable to define a_k as the correlation coefficient between p_t and p_{t+k}

$$a_k = \frac{\frac{1}{N}\sum_{t=1}^{N-k} \tilde{p}_t \tilde{p}_{t+k0}}{\frac{1}{N}\sum_{t=1}^{N-k} \tilde{p}_t^2} \tag{7}$$

where

$$\tilde{p}_t = p_t - \bar{p} \tag{8}$$

This actually corresponds to the value of a_k which minimizes the *MSE* for the new reference.

3 Examples

In this section measured wind power is used to calculate the correlation, and the *RMS* for the new reference is compared to the *RMS* for the mean and persistence.

3.1 Correlation

Measurements of half hourly mean values of wind power from a wind farm, located in Hollandsbjerg, Denmark, have been used to calculate an estimate of the correlation as a function of the forecast length. Two datasets are considered, namely measurements from a summer and a winter period. Each dataset contains 4380 measurements. The estimated correlation as a function of the forecast length from the summer period is shown in Figure 1 and the winter period in Figure 2.

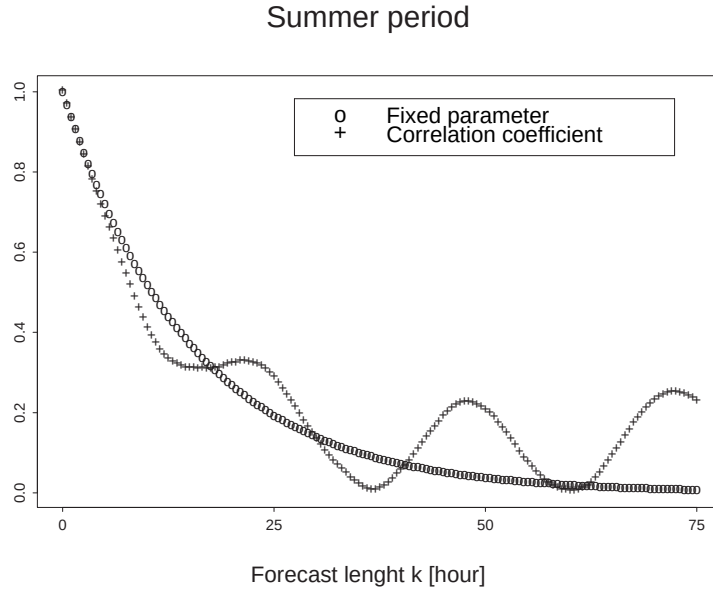


Figure 1: Estimated correlation as a function of the forecast length for 4380 half hourly mean values of observed wind power in a summer period, and the values of the fixed parameter function.

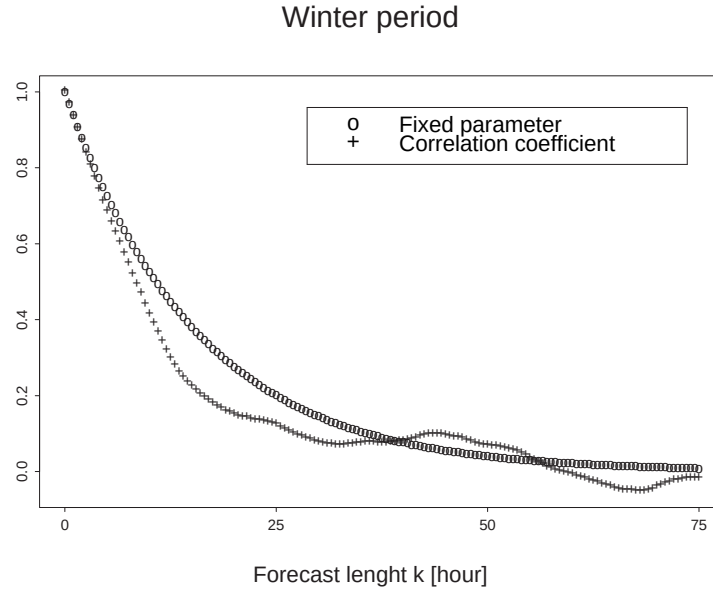


Figure 2: Estimated correlation as a function of the forecast length for 4380 half hourly mean values of observed wind power in a winter period, and the values of the fixed parameter function.

From both figures it is seen that the correlation seems to exponentially decrease as a function of the forecast length. Therefore the figures also show the values of the function

$$f(k) = \phi^k \tag{9}$$

where the values used for ϕ are the estimated correlation coefficients for $k = 1$.

The correlation for the half hour forecast ($k = 1$) is 0.968 for both periods, and the agreement between $f(k)$ and the correlation is good for both periods, as long as the forecast length is small. But for the summer period the correlation is seen to be highly periodic, which is due to the diurnal variation in the wind speed, and like latitudes like Denmark's, this diurnal variation is most significant during the summer period.

Thus, the correlation is not independent of the location of the wind farm or the time of year. Therefore it is not possible to use a simple expression like (9), or to assume global values for the correlation. It is thus recommended that the correlation is calculated for each forecast length using (7) and (8), and that the correlation which is calculated using measurements from a given location, should not be used for any other locations.

3.2 Performance

In this section the measurements from Hollandsbjerg are used to show how the *RMS* of the forecast error depends on the forecast length. One year of half hourly mean values of the power are used, and the *RMS* is calculated using: the new reference, the persistence and the mean of the power. The result is shown in Figure 3.

The figure clearly demonstrates the need for a new reference forecast model, since the *RMS* for the persistence model for large horizons is larger than the *RMS* obtains using the mean value as a forecast. For small forecast lengths, $k \leq f^{-1} \approx 3$ hours, the *RMS* for the new reference is almost identical to the *RMS* for the persistence forecast model, and for larger horizons, say k above 24 hours, the *RMS* for the new reference approximates the *RMS* of the mean. For the intermediate horizons it

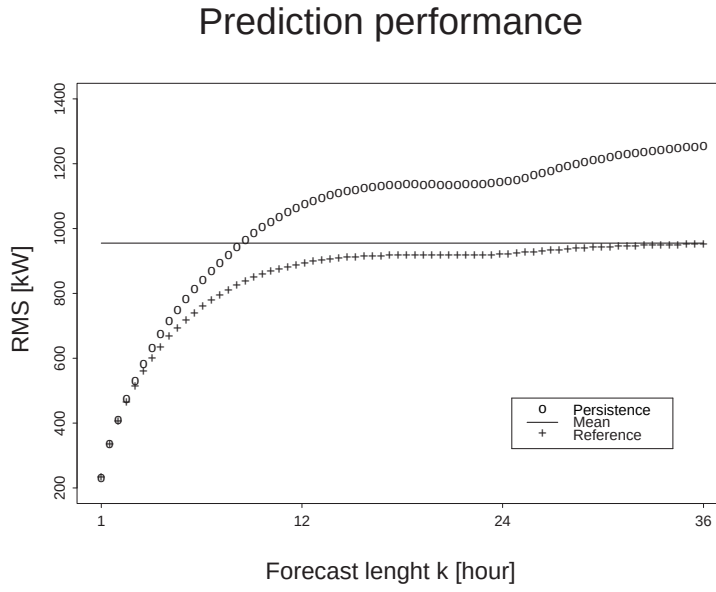


Figure 3: The *RMS* for the three simple forecast models: the mean, the persistence and the new reference. Calculated using one year of half hourly mean values of measured wind power.

is clearly seen that the new reference combines the forecasts from the persistence and the mean in such a way, that the *RMS* is significantly below the *RMS* of these two last approaches.

4 Summary

In this paper we have proposed a new reference forecast model for predictions related to wind speed and power. This reference should be used instead of the commonly used persistence forecast model, which is shown not to be reasonable for forecast lengths above a certain limit. The algorithm for calculating predictions from the new reference model is summarized below:

- Calculate the mean $\bar{p}$ using (6).
- For each forecast length k
  - Calculate the correlation coefficient a_k using (7).
  - Calculate the predictions $\hat{p}_{t+k}$ from the reference forecast model using (5).

The main difference between this algorithm and the persistence forecast model, is that the correlation coefficient has to be calculated for each forecast length. If the correlation were the same all over the world, or in other words, not depending on the location of a wind farm, the algorithm above could be simplified by omitting the calculation of the correlation coefficient. In this case the correlation coefficients could be given in a table, which could be considered globally valid. But the results in the previous section indicate that this is not the case.

The new reference forecast model is still almost as simple as the persistence forecast model, since it only requires time series of measured wind power as input. It is clearly demonstrated that if the forecast length, k, is larger than $f^{-1} \approx 3$ hours, then the new reference should be used.

A The Mean Square Error (MSE)

In this appendix it is shown that the MSE for the persistence forecast model is twice the MSE, if the mean is used as a forecast model when the flow can be considered uncorrelated.

The MSE given by (3) can be rewritten as

$$MSE_p = \frac{1}{N-k}\left(2\sum_{i=k+1}^{N-k} p_i^2 + \sum_{i=1}^{k} p_i^2 + \sum_{i=N-k+1}^{N} p_i^2 - 2\sum_{i=1}^{N-k} p_i p_{i+k}\right) \tag{A.1}$$

As the number of observations $N \to \infty$, and $k \ll N$, it is seen that the second and third sum in (A.1) becomes neglible and hence

$$MSE_p \approx \frac{2}{N-k}\left(\sum_{i=k+1}^{N-k} p_i^2 - \sum_{i=1}^{N-k} p_i p_{i+k}\right)$$

Using that the mean of two multiplied uncorrelated random variables, X and Y, is given by $E(XY) = E(X)E(Y)$, the MSE for large k can be rewritten as

$$MSE'_p \approx \frac{2}{N-k}\left(\sum_{i=k+1}^{N-k} p_i^2 - \frac{1}{N-k}\left(\sum_{i=1}^{N-k} p_i\right)^2\right)$$

If instead the mean of the flow were used as a forecast model, i.e.

$$\hat{p}_{t+k} = \bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i, \quad 1 \le t \le N$$

we see that the MSE for this model is

$$\begin{aligned} MSE_m &= \tfrac{1}{N}\sum_{i=1}^{N}\left(p_i - \tfrac{1}{N}\sum_{j=1}^{N} p_j\right)^2 \\ &= \tfrac{1}{N}\left(\sum_{i=1}^{N} p_i^2 - \tfrac{1}{N}\left(\sum_{i=1}^{N} p_i\right)^2\right) \approx \tfrac{1}{2} MSE'_p \end{aligned}$$

which means that the MSE for the persistence model will be twice the MSE of the mean model for large k, where p_{t+k} and p_t are uncorrelated.

References

Joensen, A. (1997), 'Models and methods for predicting wind power (in danish)', Department of Mathematical Modelling, Technical University of Denmark, Lyngby.

Landberg, L. (1994), Short-term prediction of local wind conditions, Technical Report Risø - R - 702 (EN), Department of Wind Power Meteorology, Risø National Laboratory, Roskilde, Denmark.

Landberg, L. (1999), 'Short-term prediction of power production from wind farms', *Jour. of Wind Engineering and Industrial Aerodynamics* **80**, 207–220.

Landberg, L. & Watson, S. (1994), 'Short-term prediction of local wind conditions', *Boundary-Layer Meteorology* **70**, 171–195.

Madsen, H. (1996), Models and methods for predicting wind power, Technical report, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Nielsen, T. S. & Madsen, H. (1996), Using meteorological forecasts in on-line predictions of wind power, Technical report, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Paper D

Statistical Methods for Predicting Wind Power

D

Originally published as

A. K. Joensen, T. S. Nielsen and H. Madsen. Statistical methods for predicting wind power. In *Wind Energy for the Next Millenium*, European Wind Energy Conference, pages 784–788, Dublin, Ireland, October 1997.

Statistical Methods for Predicting Wind Power

Alfred Joensen, Henrik Madsen and Torben S. Nielsen

Abstract

This paper describes how non-parametric statistical methods can be applied to wind power prediction. Due to the local non-stationary nature of the weather, a prediction model must be capable of adapting to the changes in the weather conditions. The approach applied here is to use locally weighted regression, where changes in the weather conditions are captured by using time dependent weighting. This approach reduces the model complexity, and the weighted regression approach is also used to model other relations, such as the power curve and the diurnal variation in wind speed and power.

Keywords: Wind power; Prediction; polynomial approximation; weighting functions.

1 Introduction

During the last decade the world has witnessed a renewed interest in wind energy. This clean source of energy has become an important and competitive alternative to conventionally fuelled plants. To fully benefit from a large amount of wind farms connected to an electrical grid, it is necessary to know in advance the electricity production generated by the wind. This knowledge enables the utility to control the conventionally fuelled plant in such a way that fossil fuels in fact can be saved. The necessary time frame for the predictions is one to two days.

The prediction models which will be described in this paper are based on measurements of wind speed w_t, power p_t and numerical weather predictions (NWPs) of wind speed ω_t and direction ϕ_t from HIRLAM (Machenhauer 1989) run by the Danish Meteorological Institute. Data from a wind farm located in Hollandsbjerg, Denmark, has been used in the study. In Figure 1 the layout of the wind farm is shown. The farm

consists of 32 wind turbines, 30 Nordtank 130 kW and 2 Nordtank 300 kW turbines.

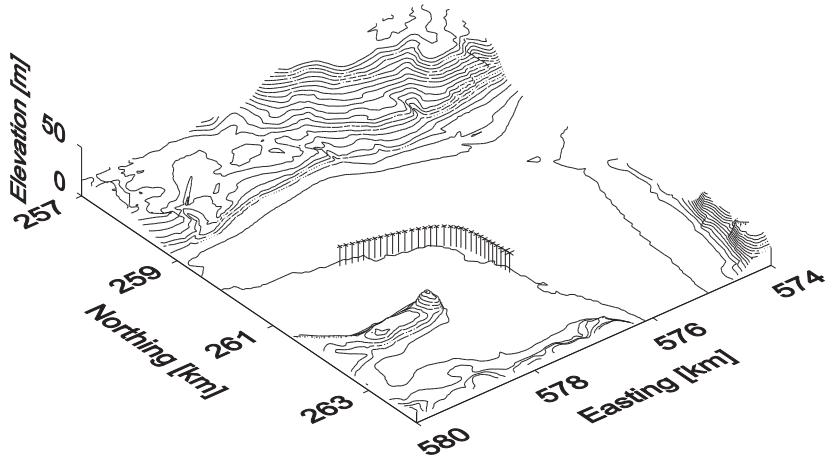


Figure 1: Layout of wind farm and the terrain surface

The data from Hollandsbjerg covers one year, and it is therefore not possible to model adequately any inter-annual dependency. Instead a local regression approach which adapts to the actual meteorological state is used.

2 Data Analysis

This section gives a short introduction to local regression, and continues with examples of this approach applied to wind power prediction. More on local regression can be found in (Hastie & Tibshirani 1993, Joensen 1997, Nielsen et al. 1997).

2.1 Local regression

Locally weighted regression is a generalisation of the kernel smoothing method (Nielsen et al. 1997), where the regression curve is approximated by local constants. If the curvature of the true regression curve is substantial then only a small fraction of the observations can be used to estimate a local constant if that estimate shall provide a reasonable non-biased (local) estimate of the regression curve. The estimated regression curve will in this case have large variance. Instead it is obvious to consider a generalisation to local regression models where e.g. polynomial approximation is used. Now it is possible to get a non-biased estimate with a larger number of observations.

The underlying model for local regression is

$$y_i = f(\boldsymbol{x}_i) + e_i, \quad i = 1, \cdots, n \tag{1}$$

where y_i are observations of a response, $\boldsymbol{x}_i$ are d- dimensional vectors of explanatory variables, e_i are independently identical distributed normal variables and n is the number of observations.

The function f is assumed to be smooth and estimated by fitting a polynomial $P(\boldsymbol{x}, \boldsymbol{x})$ model within a sliding window, and parameterized such that

$$f(\boldsymbol{x}) = P(\boldsymbol{x}, \boldsymbol{x}) \tag{2}$$

For each fitting point, and for a given parameterization $\boldsymbol{\theta}$ of the polynomial, the following least squares problem is considered

$$\hat{\boldsymbol{\theta}}(t) = \mathbf{arg}\min_{\theta} \sum_{i=1}^{n} w_i(\boldsymbol{x})(y_i - P(\boldsymbol{x}, \boldsymbol{x}))^2. \tag{3}$$

The local least squares estimate of $f(\boldsymbol{x})$ is now

$$\hat{f}(\boldsymbol{x}) = \hat{P}(\boldsymbol{x}, \boldsymbol{x}) \tag{4}$$

One common way to calculate the weights in (3) is to use a product kernel where the distance between the explanatory variables is calculated one

dimension at a time, i.e.

$$w_i(\boldsymbol{x}) = W(\frac{x_{i1} - x}{h_1}) \cdots W(\frac{x_{id} - x}{h_d}) \tag{5}$$

where the tri-cube weight function given by

$$W(u) = \begin{cases} (1 - |u|^3)^3 & u \in]-1; 1] \\ 0 & |u| \in [1; \infty[\end{cases} \tag{6}$$

can be used. The tuning parameter of the weight is called the bandwidth h, and determines how many observation are included in fitting criteria in (3).

If some of the explanatory variables are omitted in the weight calculation, the model becomes global in that variable, and is called a conditionally parametric model (Hastie & Tibshirani 1993, Nielsen et al. 1997). If the dimension d is large there will be very few observations covered by the weight function, resulting in a noisy estimate. If the true regression curve is known to be e.g. globally linear in one or more of the explanatory variables, then the conditionally paramteric model will be advantageous.

2.2 The power curve

Considering the layout of the wind farm in Figure 1 it is obvious that the power curve should depend on the wind direction. First of all, it is clear that the wind speed at each turbine is affected by the existence of other turbines since the turbines give shelter to each other. Furthermore, the wind speed also depends on the surrounding landscape, e.g. the vegetation and the topography. The power curve is therefore modelled using the measured wind speed and the NWP wind direction using the steady state relation given by

$$p_t = g(w_t, \phi_t) + e_t \tag{7}$$

The reason for using the NWP direction is that there is no measured wind direction from the site.

Using the wind direction as an explanatory variable in local regression is not straight forward. First of all the power curve at 0 deg and 360 deg should be the same as. This problem can be solved by adjusting the directions in the data set, in such a way that the direction of the fitting point becomes the midpoint of the interval used to represent the directions.

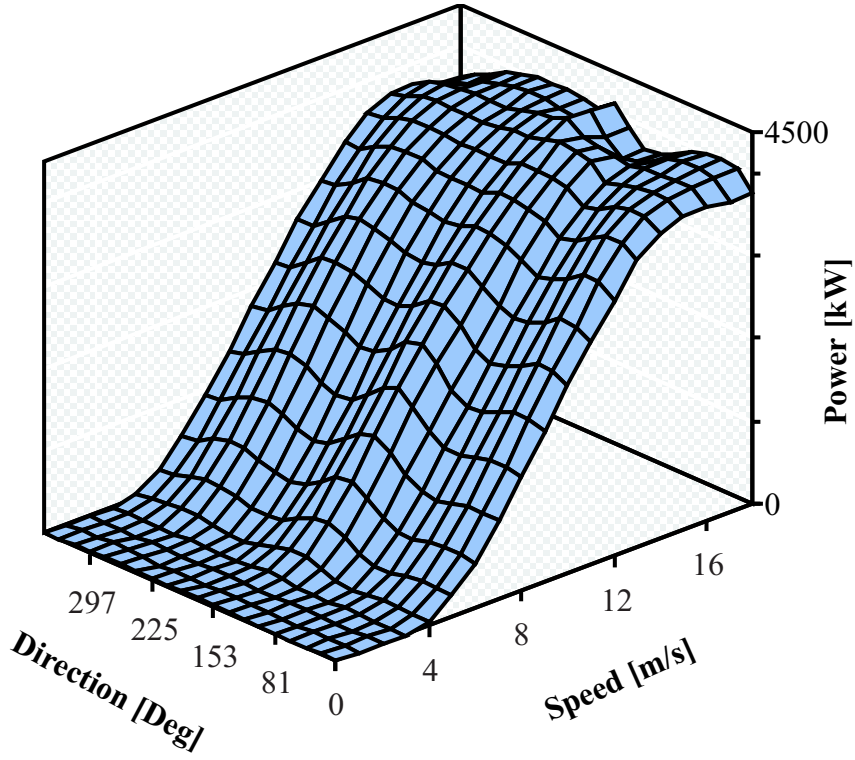


Figure 2: The power curve

The estimated power curve, using local second order polynomial approximation and fixed bandwidths of 5 ms^{-1} for the wind speed and 75 deg for the wind direction is shown in Figure 2. The surface clearly shows that the power production depends on the wind direction, but for wind speeds above 18 ms^{-1} the distribution of the data points is sparse, for some directions there are no wind speed observations above 15 ms^{-1}, resulting in a quite rugged surface for high wind speeds.

It has been assumed that the power curve does not depend on the time of year. This is however not strictly true, because the direction dependency will probably be varying from e.g. summertime where there are leafs on the trees and to wintertime where there are no leafs. But the distribution of data points would become very sparse if the power curve also should depend on the time of year.

2.3 Diurnal and annual variation

Close to the coast line it must be expected that there is a diurnal variation in the wind speed; the so called land/sea- breeze. This wind speed variation is driven by the diurnal variation in the temperature difference between land and sea. For latitudes like that of Denmark this diurnal variation is most significant during summertime. At wintertime there is usually no or only a weak diurnal variation in the wind speed. Therefore it is necessary to use both the time of day t_{day} and the time of year t as explanatory variables for the diurnal variation.

Bearing in mind that in an on-line situation only past observations are available, the diurnal variation can only be estimated using observations up to the time of the estimation. This constraint can be fulfilled by redefining the weight function in (6) in such a way that zero weight is given to future observations.

As with the wind direction, the variable used to represent the time of day is not continuos at 24 o'clock. This problem is solved in a similar manner by adjusting the time of day at each fitting point, in such a way that there are equally many observation before and after the time of day at the fitting point.

The diurnal variation in the measured and the NWP wind speed are estimated using second order polynomial approximation for the time of day and zero order for the time of year. The bandwidth used for the time of day is 5 hours and for the time of year 2 months, using only past observations. Figure 3 shows the estimated variation on a summer and winter day. First of all it is noticed that there is a clear diurnal variation at summertime, and that the mean of the wind is higher at wintertime but without any systematic variation as expected. It is also seen that

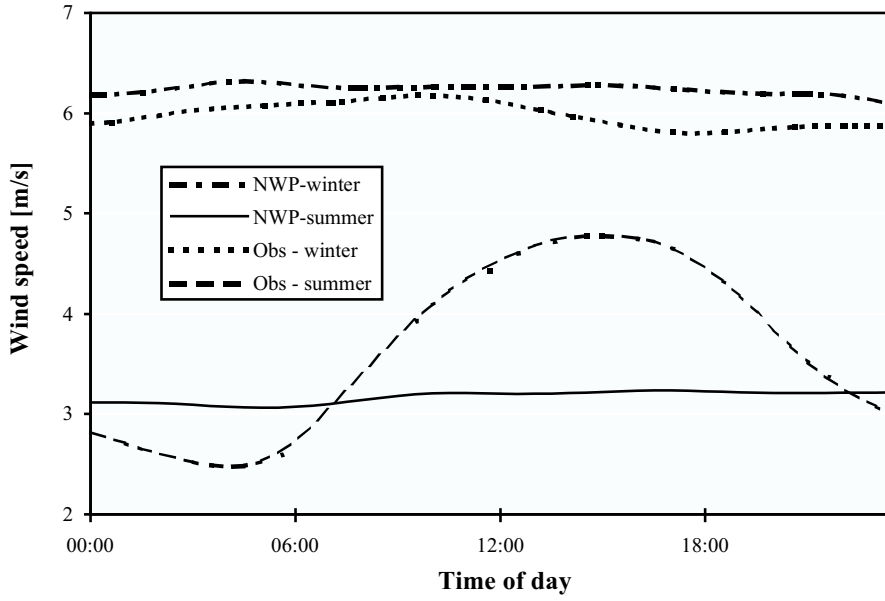


Figure 3: Diurnal variation of observed and predicted wind speed

there is no variation in the NWP wind speed in the summer period. The reason for this is that the NWP model does not take local phenomena like the land/sea-breeze properly into account.

2.4 Other relations

The purpose of Section 2 has been to emphasise the most important relations which are revealed using the data from Hollandsbjerg. But several other relations may exist which are not discussed above. For instance the NWP wind speed should depend on the wind direction, because of the spatial resolution of the NWP model. Several such relations have been examined (Joensen 1997), but without success.

The main reason for this is most likely the distribution of the wind direction. The distribution, which is depicted by a wind rose in Figure 4, shows that there is only one main wind direction, and the wind direction

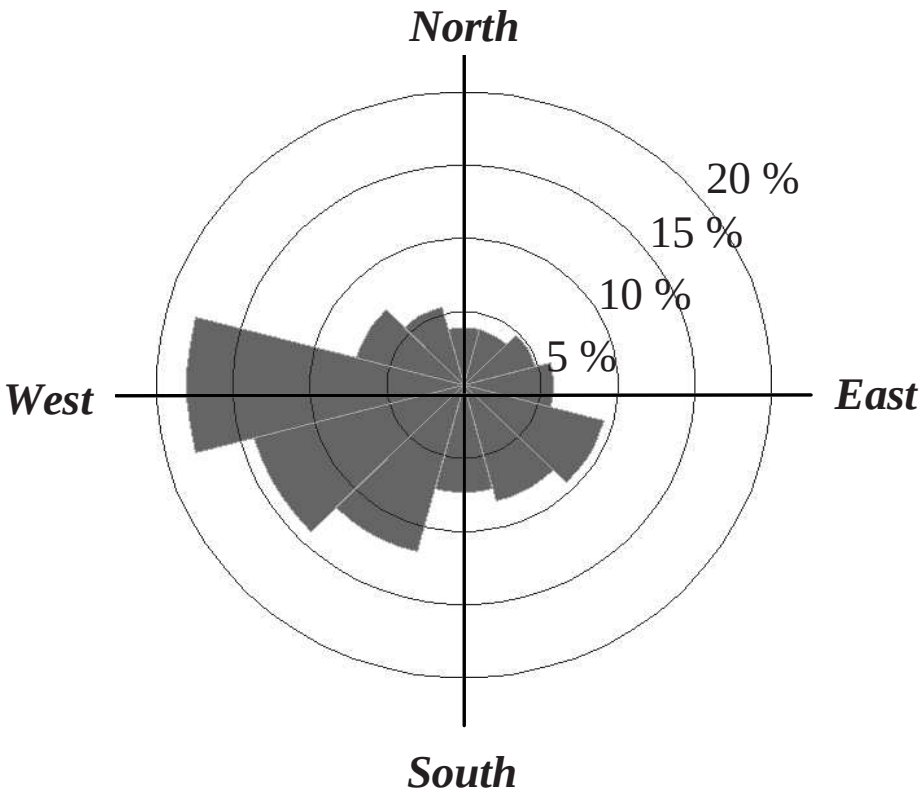


Figure 4: The distribution of the NWP wind direction

is therefore redundant.

3 The Model

There are several ways to formulate multi-step prediction models (Joensen 1997, Nielsen & Madsen 1996). The approach taken here is to formulate one model for each prediction horizon.

A number of models with several combinations of the explanatory variables have been tried out. The resulting model, which turned out to give the best overall results (Joensen 1997), includes the same explanatory

variables for each prediction horizon; but the parameters of the models are different and have to be estimated for each prediction horizon.

The overall model is

$$p_{t+k} = a(t)p_t + d(t, t_{day}) + b(t)g(\omega_{t+k}, \phi_{t+k}) + e_{t+k} \tag{8}$$

where $g(.,.)$ is the power curve described in Section 2.2, and $d(.,.)$ is the annually varying diurnal variation of the power, which is estimated similarly to the diurnal variation of the wind speed in Section 2.3. The reason for including the power observation at time t in the model, is that the power observations are auto-correlated. The result is that when k is small, say below 3 hours, then p_{t+k} is rather close to p_t. The diurnal variation of the power is included in the model because the results in Section 2.2 have shown that there were no diurnal variation in the NWPs.

The problem which arises when all the parameters of the model in (8) are to be estimated using local regression, is called the curse of dimensionality. If the parameters of all the relations in (8) are to be estimated simultaneously, there will be very few data points covered by the weight function, resulting in large parameter variance. Instead the relations have to be estimated separately, and the only parameters that have to be estimated in (8) are the coefficients of the polynomials $a(.)$, $b(.)$ and $d(.,.)$.

3.1 Results

This section gives the results from the model given in (8) compared to the so called naive predictor and a new reference model which is more adequate than the naive predictor (Nielsen, Joensen, Madsen, Landberg & Giebel 1999, Nielsen & Madsen 1996). The predictions from the new reference is a weighting between the naive predictor and the mean of the power. When the prediction horizon k is large, say above 12 hours, the correlation between the power observations p_t and p_{t+k} has almost vanished, and it can be shown that the Root Mean Square Error (RMS) of the naive predictor will be $\sqrt{2}$ times larger than the RMS of the new reference (Nielsen et al. 1999).

The models are compared using the RMS. From Figure 5 it is clearly

seen that the models which are presented in this paper outperform the reference models.

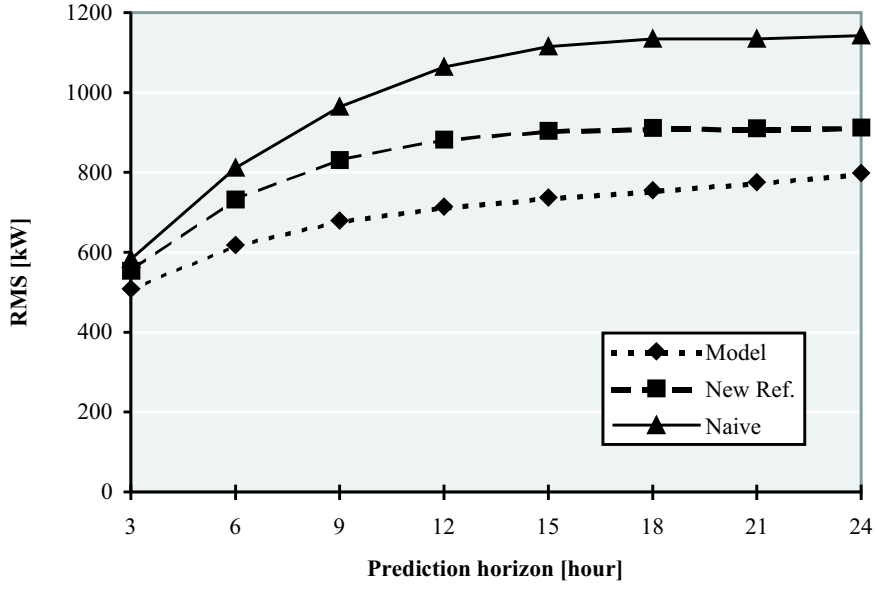


Figure 5: The RMS for different prediction horizons

3.2 Conclusion

This paper has demonstrated how locally weighted regression can be used to develop prediction models for the power production from wind turbines.

The main results in this paper are that the power production depends both on the wind speed and direction, and that there is a diurnal variation in the power production which depends on the time of year. It is described how this information can be used in an prediction model which includes NWP of wind speed and direction. Throughout all the modelling phases the locally weighted regression method has been used and demonstrated.

References

Hastie, T. & Tibshirani, R. (1993), 'Varying-coefficient models', *Journal of the Royal Statistical Society, Series B, Methodological* **55**, 757–796.

Joensen, A. (1997), 'Models and methods for predicting wind power (in danish)', Department of Mathematical Modelling, Technical University of Denmark, Lyngby.

Machenhauer, B. (1989), 'The HIRLAM Final Report', Danish Meteorological Institute, Copenhagen, Denmark.

Nielsen, H. A., Nielsen, T. S. & Madsen, H. (1997), ARX-models with parameter variations estimated by local fitting, *in* Y. Sawaragi & S. Sagara, eds, '11th IFAC Symposium on System Identification', Vol. 2, pp. 475–480.

Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L. & Giebel, G. (1999), 'A new reference for predicting wind power', *Wind Energy* **1**, 29–34.

Nielsen, T. S. & Madsen, H. (1996), Using meteorological forecasts in on-line predictions of wind power, Technical report, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Paper E

Model output statistics applied to wind power prediction

Originally published as

A. K. Joensen, G. Giebel, L. Landberg, H. Madsen, and H. Aa. Nielsen. Model output statistics applied to wind power prediction. In *Wind Energy for the Next Millenium*, European Wind Energy Conference, pages 1177–1180, Nice, France, 1–5 March 1999.

Model output statistics applied to wind power prediction

Alfred Karsten Joensen[1,2], Gregor Giebel[1], Lars Landberg[1], Henrik Madsen[2] and Henrik Aalborg Nielsen[2]

Abstract

Being able to predict the output of a wind farm online for a day or two in advance has significant advantages for utilities, such as better possibility to schedule fossil fuelled power plants and a better position on electricity spot markets.

In this paper prediction methods based on Numerical Weather Prediction (NWP) models are considered. The spatial resolution used in NWP models implies that these predictions are not valid locally at a specific wind farm, furthermore, due to the non-stationary nature and complexity of the processes in the atmosphere, and occasional changes of NWP models, the deviation between the predicted and the measured wind will be time dependent. If observational data is available, and if the deviation between the predictions and the observations exhibits systematic behaviour, this should be corrected for; if statistical methods are used, this approach is usually referred to as MOS (Model Output Statistics). The influence of atmospheric turbulence intensity, topography, prediction horizon length and auto-correlation of wind speed and power is considered, and to take the time-variations into account, adaptive estimation methods are applied.

Three estimation techniques are considered and compared, Extended Kalman Filtering, recursive least squares and a new modified recursive least squares algorithm.

Keywords: Forecasting Methods; Wind Energy; Statistical Analysis; Performance

[1]Department of Wind Energy and Atmospheric Physics, Risø National Laboratory, DK-4000 Roskilde, Denmark

[2]Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

1 Introduction

Several models for predicting the output from wind farms have already been developed, some based on observations from the wind farms (Madsen 1996), others based on numerical weather predictions (Landberg 1994), and again others on combination of both (Joensen, Madsen & Nielsen 1997).

This paper describes how statistical methods, usually referred to as model output statistics (MOS), can be used in models that combine observations and NWP model predictions, and the approach taken here is slightly different from the approach in (Joensen et al. 1997). The NWP model, HIRLAM (Machenhauer 1988), is run by the Danish Meteorological Institute (DMI). The observations, wind speed w_t and power p_t, are from four sites in Denmark: The Risø mast at Risø National Laboratory, and the Avedøre, Kappel and Østermarie wind farms.

The NWP model predicts several meteorological variables, such as temperature, surface fluxes and pressure, wind speed ω_t and direction θ_t at 31 levels/heights, see (Machenhauer 1988) for definition of the levels, and at the surface, i.e. 10 m a.g.l. The NWP model is run four times at day, at 00:00, 06:00, 12:00 and 18:00 UTC, and the predictions are given in 3 hourly steps 36 hours ahead.

2 Finding the right NWP model level

In (Landberg 1994) it was found that the NWP predicted wind from level 27 gave the best results when used as input to the neutral geostrophic drag law to determine u_*, and the neutral logarithmic profile was used to calculate the wind at hub height. It was concluded that the reason why the stability dependant relations did not improve the results, was that the heat fluxes were not predicted accurately enough. Since HIRLAM has been updated several times since the investigation in (Landberg 1994), it is reasonable to re-evaluate these results.

Based on the results in (Landberg 1994) the neutral relations are used to

transform the NWP wind down to the surface, and the prediction performance is compared to the performance of the surface wind calculated by the NWP model. HIRLAM takes the stability into account when the surface wind is calculated, and this comparison will therefore show if it is advantageous to include the stability. Furthermore, in order to make a fair comparison, the predictions should be corrected for any bias and offset, i.e. the simple MOS model

$$w_{t+k} = a_k \omega_{t+k} + b_k + \epsilon_{t+k} \tag{1}$$

where ϵ_{t+k} is assumed to be white noise and k is the prediction horizon, is fitted to the observations using the least squares method.

Observations from 44, 76 and 125 m above the surface from the Risø mast are used in the comparison, because the wind which should be used might not be the same depending on which height above the surface it is compared to.

To evaluate the performance of the predictions

$$\rho = \frac{VAR(w_{t+k}) - MSE_k}{VAR(w_{t+k})} \tag{2}$$

is used, where VAR is the estimated variance of the observations and MSE_k is the mean square error of the predictions k hours ahead. The interpretation of ρ is that it measures how much of the total variation in the observations is explained by the predictions, i.e. a value of 1 means that the predictions are perfect and 0 means that predictions are useless.

Figure 1 shows the results for each model height and each prediction horizon compared to the 44 m observations at the Risø mast. It is clearly seen that for all prediction horizons the best result is obtained using the surface wind, and the corresponding figures for the 76 and 125 m show the same results. The conclusion is therefore that it is advantageous to take the stability into account, and hence the surface wind calculated by the NWP model should be used.

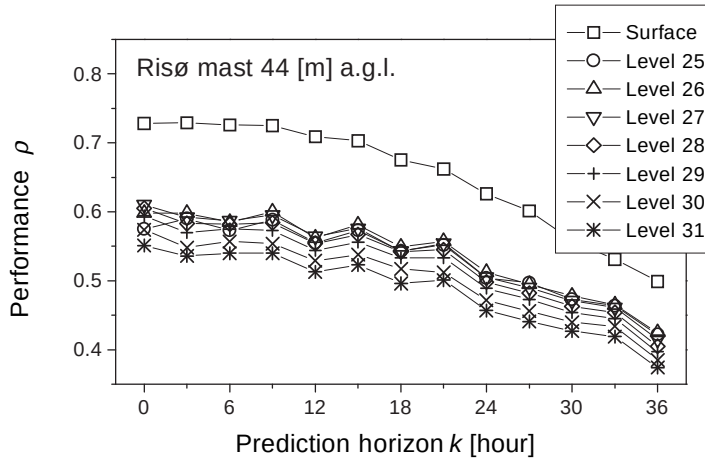


Figure 1: Performance for various NWP model levels

3 Wind direction dependency

Due to the spatial resolution of any NWP model it should be expected that some kind of wind direction dependant fine tuning to a specific site should be possible. One way to do this fine tuning is to apply a MOS model

$$w_{t+k} = a_k(\theta_{t+k})\omega_{t+k} + b_k(\theta_{t+k}) + \epsilon_{t+k} \tag{3}$$

From a physical point of view the adjustment due to the topography should be a wind direction dependant factor, but this model also includes a wind direction dependant offset. The reason for this is purely statistical, e.g. if the prediction accuracy of the NWP model is not the same for all directions the inclusion of the offset will increase the performance.

Local regression (Hastie & Tibshirani 1993) has been used to estimate the coefficient functions in (3). When using this method it has been assumed that for a given wind direction sector the coefficient functions are well approximated by second order polynomials. Using the terminology of local regression, the nearest neighbour bandwidth was chosen to include 40% of the observations at each fitting point. A physical way to take the topography into account is to perform a high resolution analysis of

the site and the surroundings, and use this analysis to correct the NWP wind for local topography effects, which obviously are not included by the NWP model resolution. To see if this is advantageous the NWP surface wind has been corrected by matrixes calculated by WAsP (Mortensen et al. 1993).

Again Risø mast data from the last half of 1997 and first half of 1998 has been used for the estimation, while validation was performed with data from the last half of 1998. In order to make a fair comparison, the MOS model (1) has been applied after the WAsP correction, and the performance has also been calculated for the MOS model applied to the raw NWP surface predictions.

Surprisingly, Figure 2 shows that the performance of the WAsP corrected forecast is worse than without, although the difference is only minor. The reason is most likely that the physical assumptions behind WASP are not satisfied when WAsP is used for predictions of wind speed and direction which contain errors. Nevertheless, it seems as if there is some dependency on the topography, because the statistical correction (3) improves the performance, but this is not the only reason. This follows from the fact that the wind speed distribution depends on the wind direction, i.e. the wind speed in Denmark is usually higher when coming from west compared to e.g. north or east. This is a feature of the overall flow, and can not be prescribed to the local topography. Because the wind speed is not perfectly predicted by the NWP model this is incorporated by the wind direction dependent offset in (3).

4 Diurnal variation

Surprisingly, as seen in Figure 2, it seems easier to predict e.g. 6 hours ahead than 3 hours ahead. The reason is that the prediction horizon k aliases with the time of day, and therefore, as shown in Figure 3, with the diurnal variation in the wind speed/atmospheric stability. This is because the NWP model is update 4 times a day, hence odd prediction horizons correspond to the following times of day: 03:00, 09:00, 15:00 and 21:00, and even horizons correspond to: 00:00, 06:00, 12:00 and 18:00.

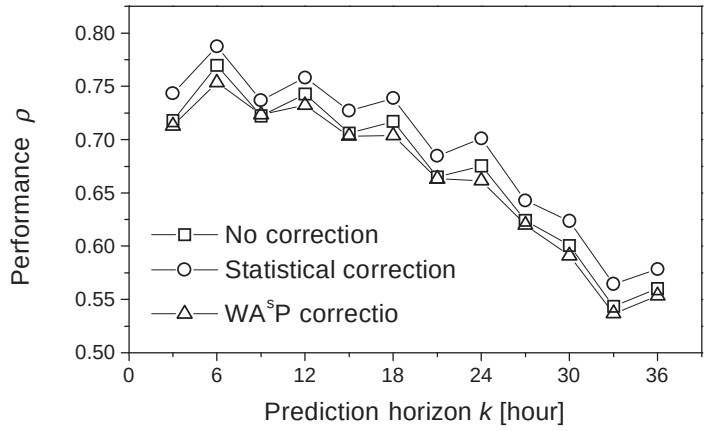


Figure 2: Performance for direction dependent models

Furthermore, Figure 3 is only based on measurements from the 1998 summer period, for latitudes like those of Denmark, there is also an annual variation in the diurnal variation (Nielsen et al. 1999). When the data used for the estimation is not from the same seasons of the year as the data used for the validation, which is the case in Section 3, the parameters of (3) become biased towards the diurnal variation in that specific period. This is the main reason for the effect seen in Figure 2.

5 Adaptive estimation

In the previous sections only wind speed models have been considered, we now turn to wind power models which are slightly more complicated due to the non-linear relation between the wind speed and the power.

Furthermore, due to the non-stationary nature of the atmosphere, it must be expected that the parameters of a MOS model will be time-varying. Hence adaptive estimation methods are considered. Two widely used methods for this purpose is Kalman filtering and recursive least squares, see (Ljung & Söderström 1983) and the references therein. The key idea behind these methods is the same, which is to discard old information as new becomes available, or to be more specific, the methods slide a time-

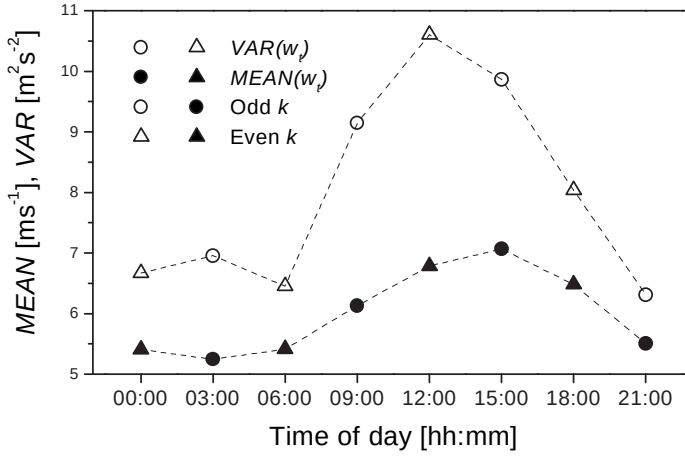


Figure 3: Diurnal variation

window of a specific width over the observations, where only the newest observations are seen. This approach has its drawbacks and advantages. If the true system is non-stationary and if this non-stationarity is not described by the model, the approach implies that the model adapts to the current state of the underlying system. But, on the other hand, because less observations are used to determine the parameters of the model, the parameters might become poorly determined, resulting in large parameter and prediction variance. The optimal model choice is therefore a model which balances simplicity and flexibility.

5.1 Extended Kalman filter

One way to simplify the model for predicting the power is, to in some way, include a known relation between wind speed and direction and power, i.e. the power curve, in the model that is to be estimated adaptively. One solution is to apply the model

$$p_{t+k} = a_k pow(b_k \omega_{t+k} + c_k, \theta_{t+k}) + d_k p_t + l_k + \epsilon_{t+k} \tag{4}$$

where $pow(\cdot, \cdot)$ is the wind farm power curve derived by the PARK application (Sanderhoff 1993). See (Landberg 1998) for a similar analysis.

The reason for the scaling of the NWP wind speed inside the power curve comes from the observation that the ratio between the measured wind speed and the NWP wind speed is different from one and time dependent. The constant inside the power curve lets the estimation determine the cut-in and cut-out wind speeds. The observed power at time t is include because for short prediction horizon the power observations are auto-correlated (Nielsen et al. 1999), this is also the reason for including the scaling of the power curve, because for short horizons more emphasis will be on the auto-correlation, i.e. p_t, and for larger horizons more weight will be put on the NWP and hence the power curve. Because this model is not linear in the parameters, the Extended Kalman Filter has been used for the estimation in this model.

5.2 Recursive least squares

A way to avoid the non-linearity is to use a polynomial approximation of the power curve, i.e.

$$p_{t+k} = a_k\omega_{t+k} + b_k\omega_{t+k}^2 + c_k\omega_{t+k}^3 + d_k p_t + l_k + \epsilon_{t+k} \tag{5}$$

This model has the same number of parameters as model (4), but it does not incorporate any knowledge about the wind farm power curve, apart from the fact that most power curves are very well approximated by a third order polynomial in the wind speed. The parameters of this model have been estimated using the standard recursive least squares algorithm.

5.3 Recursive local regression

So far we have not mentioned how the parameters in the MOS models depend on the prediction horizon. Actually we have just estimated one set of parameters for each prediction horizon. Addressing the variance problem of the Kalman Filter and the usual recursive least squares algorithm, it might be advantageous to make assumptions about how the parameters depend on k. Furthermore, in Section 3 and 4 it was shown that the NWP model predicted wind direction improved the performance for the wind speed predictions, and that there was an aliasing effect with

the time of day/atmospheric stability, caused by the update frequency of the NWP model. To take all these findings into account the following model is proposed

$$\begin{aligned} p_{t+k} = & a(k, \theta_{t+k}, t_{day})\omega_{t+k} + b(k, \theta_{t+k}, t_{day})\omega_{t+k}^2 + \\ & c(k, \theta_{t+k}, t_{day})\omega_{t+k}^3 + d(k, \theta_{t+k}, t_{day})p_t + \\ & m(k, \theta_{t+k}, t_{day}) + \epsilon_{t+k} \end{aligned} \tag{6}$$

This model is similar in structure to (5) except that the parameters/ coefficients now are assumed to be functions of the prediction horizon, the wind direction and the time of day.

To take the stability into account it was found sufficient to estimate two sets of coefficient functions, one set for the following times of day: 00:00, 03:00, 06:00, and 21:00, i.e. mainly neutral/stable conditions, and one set for the times: 09:00, 12:00, 15:00, 18:00, i.e. mainly unstable conditions. For the wind direction and the prediction horizon, the approach described in (Nielsen, Nielsen, Joensen, Madsen & Holst 1998) have been used for the estimation of the coefficient functions. This approach is best described as recursive local regression, and it is an extension of the usual recursive least squares algorithm, where the functional shape is found by estimating the parameters locally over a grid spanning the variables, e.g. for a given wind direction θ in the grid, only observations close to this direction are used when the value of the coefficient function for this particular value of θ is estimated.

In the actual estimation the coefficient functions were estimated in a fine grid spanning the NWP model predicted wind direction, using a fixed bandwidth of 100 Deg, and for the prediction horizon an increasing bandwidth was used, i.e. for the 3 hour prediction a bandwidth spanning only the 3 hour prediction was used, increasing to a bandwidth spanning the 12 hour up to the 36 hour prediction for the 36 hour prediction, this choice reflects the fact the variation of the parameters with the prediction horizon was found to be small for large prediction horizons.

When only one set of coefficient functions were estimated for all times of day, the assumption about the variation of the coefficient functions with k failed, because in this case a 6 hourly variation is introduced in the coefficient functions with k.

Because some wind directions are rare it was found important to use a different degree of time adaptation depending on the wind direction (Nielsen et al. 1998). For frequent wind directions the optimal time window was found to be about 2-3 months, while for rare wind direction it was not to use any adaptation at all. This indicates that the variation of the coefficient functions with the wind direction is larger than the time-variation.

6 Results

Figure 4 shows the performance for the three adaptive approaches that have been described in the previous sections. It clearly seen that model (6) gives the best results, the non-linear model (4) and the linear model (5) are close in performance, neither model performs best on all prediction horizons, but overall the linear model seems to perform best. This suggests that the polynomial approximation of the power curve is adequate.

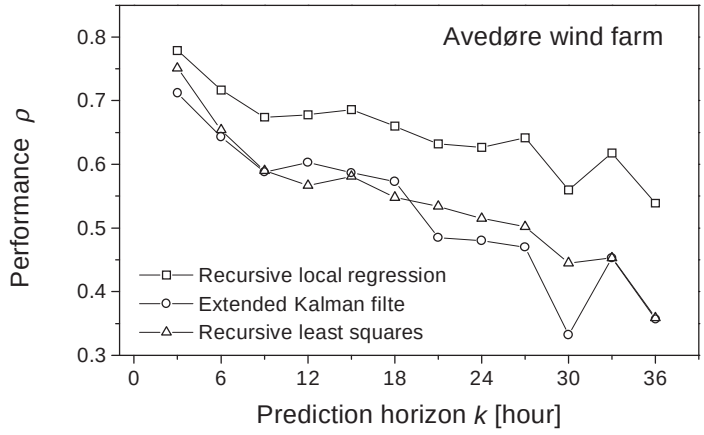


Figure 4: Performance of adaptive approaches

Figure 5 shows the prediction performance of model (6) for the three wind farms, and it is seen that there is a pronounced variation in the performance for the individual wind farms, which can be due to many

factors, e.g. the NWP model accuracy depends on the specific location, or another factor that might be of importance in this study is that the quality of the observations from the wind farms was rather poor, about 30 % of the observations were missing.

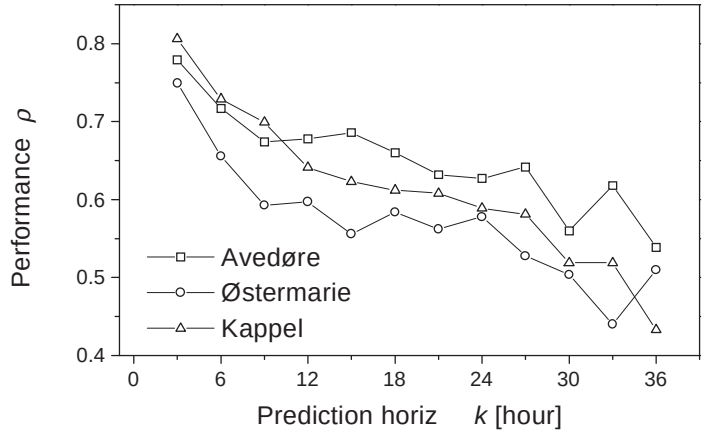


Figure 5: Performance for various wind farms

7 Summary

In this paper various MOS approaches have been proposed for wind power prediction models, which are based on numerical weather predictions and observations.

Three estimation methods have been considered: Extended Kalman filtering, recursive least squares and a new modified recursive least squares algorithm. The results indicate that the best MOS approach, is one which takes the wind direction, the time of day, the prediction horizon, and auto-correlation of the observations into account when using a wind speed polynomial approximation of the power curve to predict the future power from a wind farm. Furthermore it was found that the surface wind from the NWP model, which in this case is HIRLAM (Machenhauer 1988), should be used.

8 Acknowledgements

This work is partially funded by the European Commission (EC) under JOULE (JOR3-CT95-0008). A. Joensen is partly funded by the Danish Research Academy. G. Giebel is funded by EC Training through Research (JOR3-CT97-5004).

References

Hastie, T. & Tibshirani, R. (1993), 'Varying-coefficient models', *Journal of the Royal Statistical Society, Series B, Methodological* **55**, 757–796.

Joensen, A. K., Madsen, H. & Nielsen, T. S. (1997), Non-parametric statistical methods for wind power prediction, *in* 'Proceedings of the EWEC97', Ireland, pp. 788–792.

Landberg, L. (1994), 'Short-term prediction of local wind conditions', *Boundary-Layer Meteorology* **70**, 171–195.

Landberg, L. (1998), 'A mathematical look at a physical power prediction model', *Wind Energy* **1**, 23–30.

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

Machenhauer, B., ed. (1988), *HIRLAM Final Report*, Danish Meteorological Institute, Copenhagen, Denmark.

Madsen, H., ed. (1996), *Models and Methods for Predicting Wind Power*, Department of Mathematical Modelling, Technical University of Denmark, Denmark.

Mortensen, N. G., Landberg, L., Troen, I. & Petersen, E. L. (1993), 'Wind atlas analysis and application program (WAsP)', RisøNational Laboratory user guide: Risø-I-666(EN)(v.1) , Roskilde, Denmark.

Nielsen, H. A., Nielsen, T. S., Joensen, A. K., Madsen, H. & Holst, J. (1998), '*Tracking time-varying coefficient-functions*'. To be published.

Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L. & Giebel, G. (1999), 'A new reference for predicting wind power', *Wind Energy* **1**, 29–34.

Sanderhoff, P. (1993), 'PARK - User's Guide. A PC-program for calculation of wind turbine park performance', RisøNational Laboratory, Roskilde, Denmark. Risø-I-668(EN).

Paper F

A model to predict the power output from wind farms – an update

F

Originally published as

L. Landberg and A. K. Joensen. A model to predict the output from wind farms – an update. In *proceedings from BEWEA 20*, British Wind Energy Conference, pages 127–132, Cardiff, UK, 1998.

A model to predict the power output from wind farms – an update

Lars Landberg[2] and Alfred Joensen[1,2]

Abstract

This paper will report on the first results of the analysis of the power-production predictions made for a number of wind farms in Denmark, the UK and Greece. Because of the early stage of the analysis, focus will be on two sites in Denmark: the Risø mast at Risø National Laboratory and the Avedøre Wind Farm near Copenhagen. The predictions will be analysed to show the effect of the Model Output Statistics (MOS) modules. The results are compared to persistence, but also to a new reference model, combining persistence and the mean. Furthermore, the ability of the model to predict storms will be tested.

1 Introduction

With the increasing installed capacity of wind energy world-wide, a need to know the magnitude of this highly variable resource up to two days in advance has emerged. This need has resulted in a number of prediction models being made. This paper will focus on one such model called the HIRLAM/WAsP model developed by Risø National Laboratory in Denmark. The model has been run on-line for almost two years in a EU-JOULE III funded project.

The paper will briefly describe the method and the operational set-up, it will then quickly move on to an analysis of the results. The emphasis of the analysis will be on the wind predictions at the Risø mast and the power predictions for the Avedøre Wind Farm.

[1]Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

[2]Department of Wind Energy and Atmospheric Physics, Risø National Laboratory, DK-4000 Roskilde, Denmark

2 The Method

The method for predicting the power output of a wind farm is outlined in Figure 1. The idea is to explain (ie use physical relations) as much as possible. This is done by modelling the large-scale flow by a NWP (Numerical Weather Prediction) model. As we zoom in on the site – using the geostrophic drag law and the logarithmic wind profile – more and more detail is required, this detail is provided by the WAsP program (Mortensen, Landberg, Troen & Petersen 1993). To calculate the total production of the wind farm, taking the shadowing effects of the turbines into account, the PARK program (Sanderhoff 1993) is used. Finally, to take any effects not modelled by the physical model and general errors of the method into account two model output statistics (MOS) modules are used. For a more detailed description of the model, see (Landberg & Watson 1994, Landberg 1999*b*, Landberg, Hansen, Vesterager & Bergstrøm 1997, Landberg 1997). For an equational look at the model, see (Landberg 1999*a*). The HIRLAM model (Machenhauer 1989) of the Danish Meteorological Institute is used to generate the overall winds.

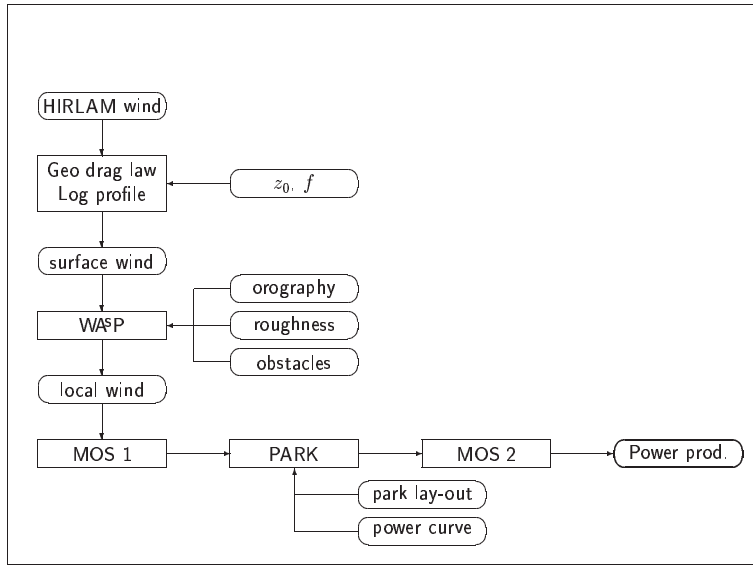


Figure 1: Flow chart of the HIRLAM/WAsP prediction model.

The emphasis in this paper will be on the effect of the two MOS modules and on the overall performance of the model.

The two MOS modules in the prediction model (MOS1 and MOS2 in Figure 1) have the following expressions

$$\text{MOS1}: \; u_{\text{MOS1}} = a(\theta)u \tag{1}$$

where u_{MOS1} is the MOS corrected wind speed, $a(\theta)$ a direction dependent (discretised in twelve sectors) scaling and u the wind speed calculated using the two physical equations and the WAsP correction (the wind labeled 'local wind' in Figure 1).

$$\text{MOS2}: \; P_{\text{MOS2}} = P + b \tag{2}$$

where P_{MOS2} is the production of the wind farm, P the production calculated by PARK, and b an off-set independent of direction.

These two MOS modules could have had a much higher degree of sofistication (e.g. dependence on forecast length, time of day etc), but to model the possible physical effects only – cf (Landberg 1999*a*) – they have been chosen as above.

3 Operational Set-Up

The model which will be evaluated in the following has been running online since the beginning of 1997 (ie for more than 18 months), predicting the production for a great number of wind farm located in Denmark, Great Britain and Greece, see Figure 3.

The HIRLAM predictions are sent from DMI via the Internet to Risø, cf Figure 2. At Risø a UNIX system is set up that runs the power prediction model every time a new HIRLAM forecast arrives. The output from this model is HTML-files (Hyper Text Mark-up Language) which are put on the World Wide Web (WWW) automatically.

The predictions from the HIRLAM model run are available around 3-4 hours after the model's verify time, the time in transit between DMI and Risø is insignificant, and the power prediction model is run in a few minutes for all the wind farms. The HIRLAM model is run twice a day, so, the WWW-site is also updated twice a day with a new 36-hour forecast. The latest development is that the HIRLAM prediction horizon has been extended to 48 hours and that the model now is run 4 times a day (ie every 6 hours).

4 Results

In the following some first results of the analysis of the performance of the prediction model will be given. The first section will focus on the prediction of the winds at the Risø mast and the second on the wind farm at Avedøre. Analysing wind-only predictions gives a clearer picture of the prediction ability of the HIRLAM/WAsP model, because the predictions are not disturbed by the no-linear power curve and is therefor a good way to begin.

4.1 The Risø mast

To evaluate the model, the predictions are compared to the actual observations for a period of 12 months (Jul 97 to Jun 98). The performance of the model is compared to three models: The persistence model, stating that the production x hours ahead is identical to the present production (cf eg (Landberg & Watson 1994)), the "new reference model" which is a combination of the persistence model and the mean of the time series (cf (Nielsen, Joensen, Madsen, Landberg & Giebel 1999)) and finally the "raw" physical version of HIRLAM/WAsP, ie the model as described in Figure 1, but without the MOS1 module corrections.

The statistical models have all had their parameters estimated using data for the last six months of 1997 and the results shown here (for all models) are predictions compared to observations from the first six months of 1998. Looking at the comparison in Figure 4 it can be seen – as expected – that the prediction model outperforms all the other models.

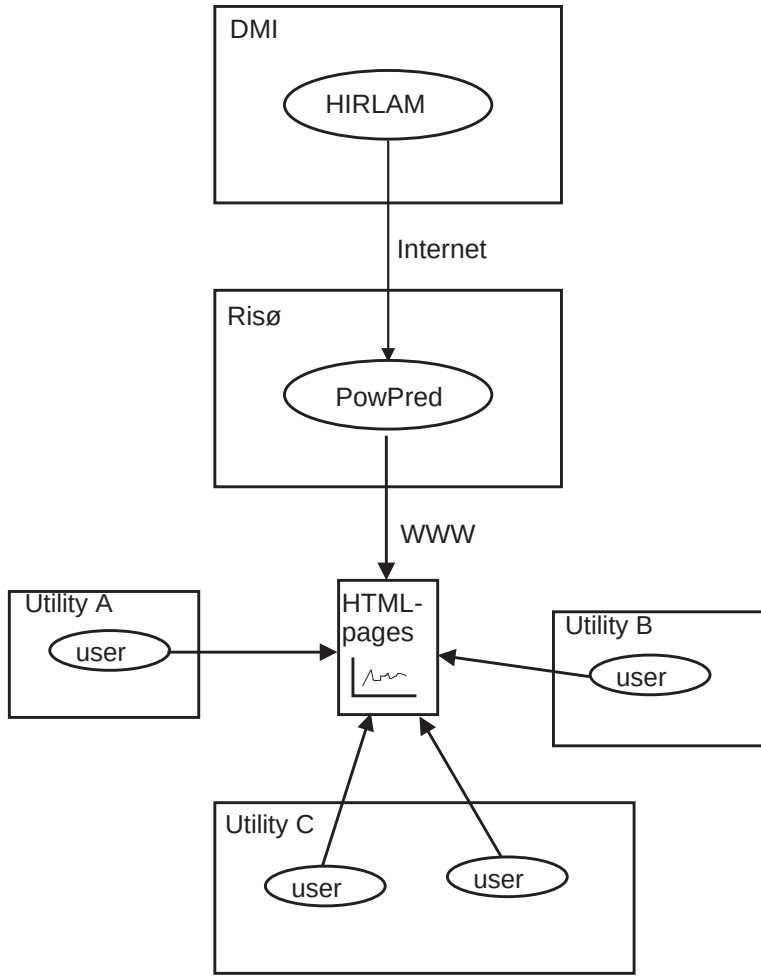


Figure 2: The operational set-up. From the Meteorological Institute where the predictions are made, via the Internet, to Risø, where the power prediction model is run. The HTML-pages can be viewed by the utility using any Web-browser, eg Netscape or Internet Explorer.

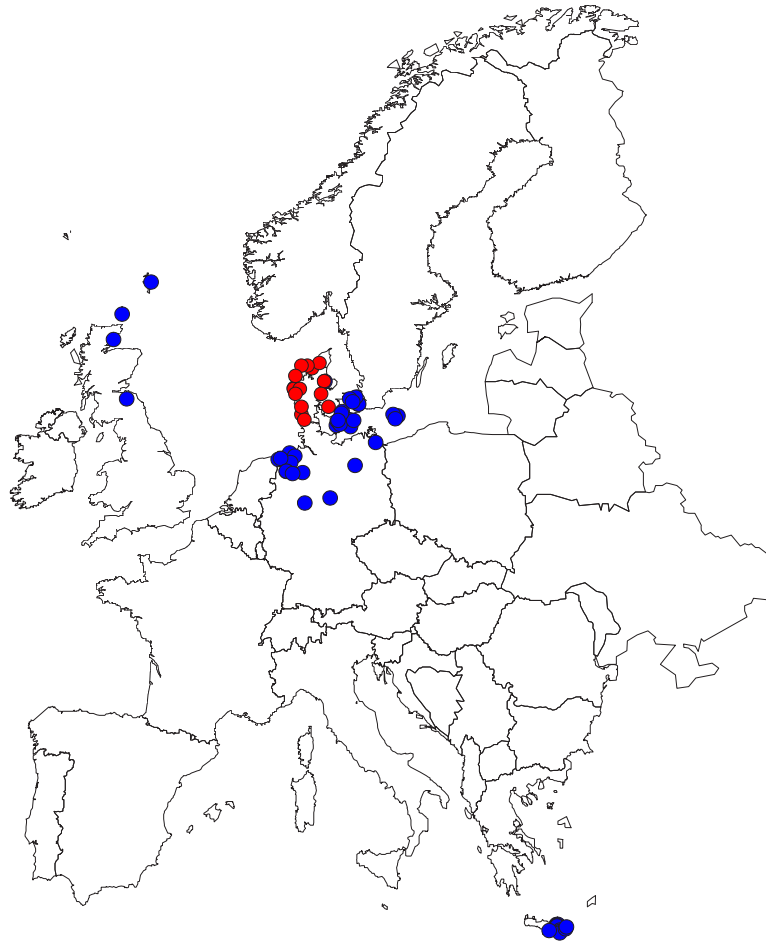


Figure 3: The location of the wind farms predicted in the on-line system.

It can also be seen, however, that excluding the MOS module leaves a model which does not perform as well; it still outperforms the various statistical models, but it is clearly seen that the MOS-module improves the prediction ability. After only 3 hours the raw physical model and the HIRLAM/WAsP model predicts better than the two reference models.

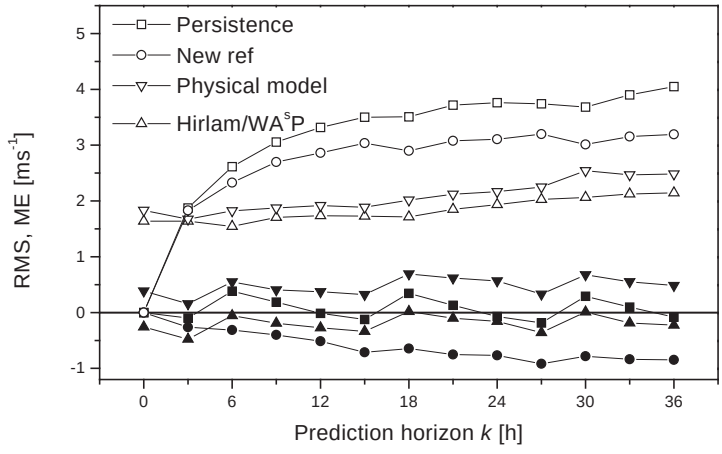


Figure 4: The performance of the prediction model for the 44.2 m wind at the Risø mast compared to persistence, the new reference (see text for explanation) and the raw physical model (see text). Full symbols refer to the mean error (ME) and open to the root mean square (RMS).

It is unexpected to see that the mean error of the "new reference model" increases with the length of the prediction. This is due to the fact that the winds in the last half year of 1997 in general were much lower than the ones in the first half year of 1998, cf Figure 5. Another very surprising fact is that looking at the *first* six months of 1997, it can be seen that the agreement between the mean wind speed as calculated by HIRLAM and the observed is quite poor. After this period the agreement is good. It is assumed that the discrepancy is due to an old version of the HIRLAM model.

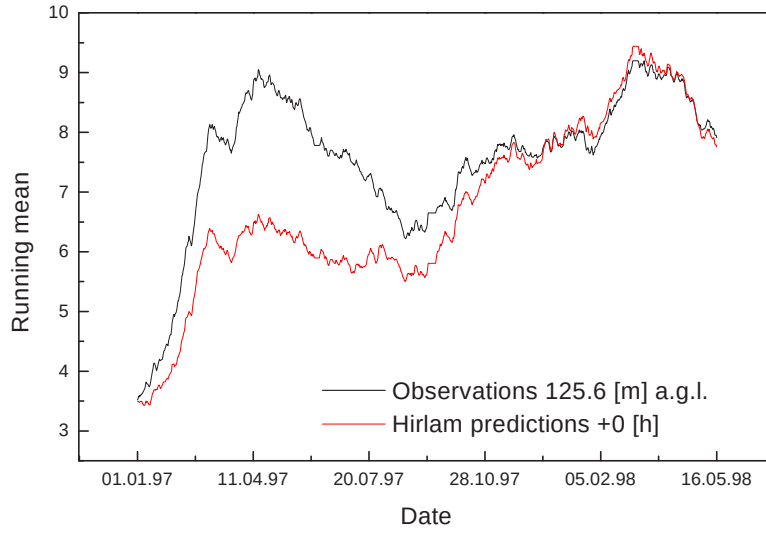


Figure 5: The running mean of the observations at the Risø mast compared to the +0 h prediction of the HIRLAM model for the 18 month period.

4.2 The Avedøre Wind Farm

In this section the performance of the prediction model for Avedøre Wind Farm will be analysed. The wind farm consists of 12 300 kW Bonus turbines and is located near Copenhagen along a dike. In Figure 6 the predictions for Avedøre are compared to the persistence predictions. As expected the prediction model outperforms the persistence model, but first after 6 hours as compared to the 3 hours for wind forecasting. It can be seen (reading the right-hand y-axis) that the mean error for the eg +12h prediction is only 13% of the total installed capacity. Comparing this result to the result of predictions done in 1996 (Landberg et al. 1997), it is found that they are in good agreement.

4.2.1 Effect of MOS

To see the effect of the two MOS modules, the change in mean error (ME) and mean absolute error (MAE) for the +24h prediction has been plotted in Figure 7. ME is the mean of the error and MAE is the mean of

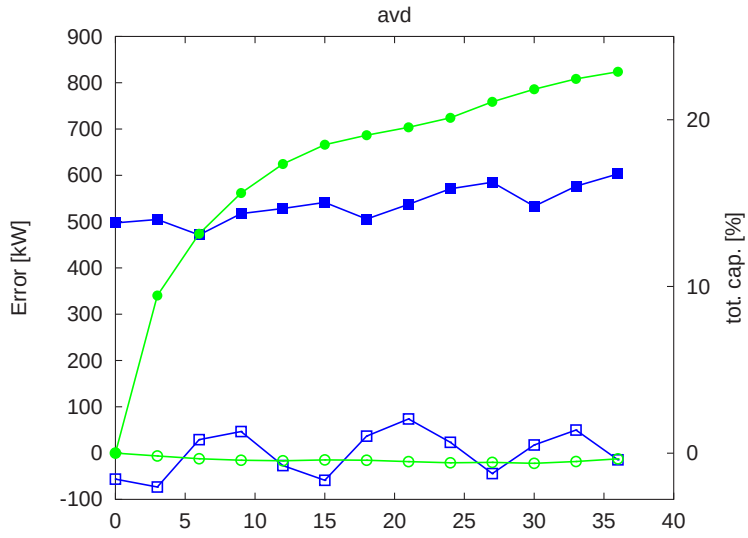


Figure 6: The performance of the prediction model for the Avedøre Wind Farm. The HIRLAM/WAsP model is marked by the squares and the persistence model by circles. Full symbols are the mean absolute error (MAE) and open the ME.

the absolute value of the error and is a measure of the scatter of the error. From the figure it can be seen that the MAE does not change significantly by using the two MOS modules, but ME is reduced considerably.

4.2.2 Storms

To try to give a feel for the prediction capability of the model all storms for the entire period (1 Jan 97 to 1 Aug 98) have been found. The definition of 'storm' is here taken to mean an event where the production rose from a very low value (ie close to 0 kW) to around 3 000 kW (approx. 80% of the total installed capacity) and then down again to a low value over a period of a few days.

Using this selection criteria 7 storms were found in 1997 and 9 in the first seven months of 1998. These 16 storms are depicted as "thumb nail images" in Figures 8, 9 and 10 and listed in Table 4.2.2.

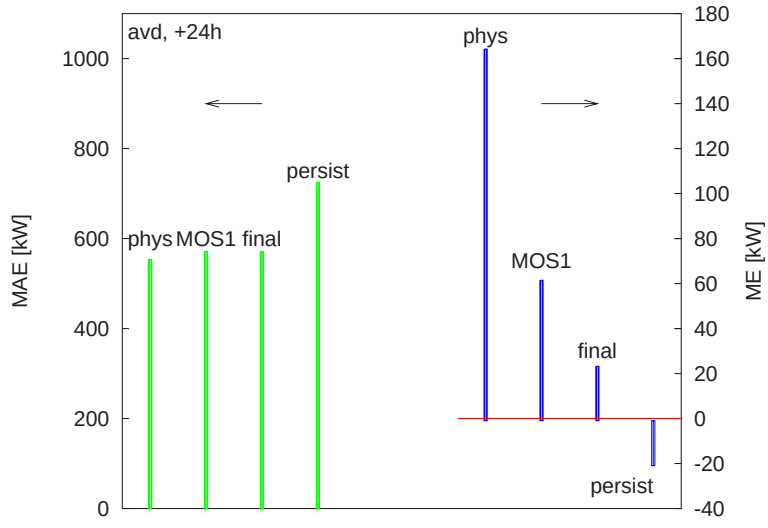


Figure 7: The mean absolute error (MAE, referring to the left-hand axis) and mean error (ME, referring to the right-hand axis) for the two MOS modules (labelled 'MOS1' and 'final') compared to the physical model ('phys') and the persistence model ('persist') for the 24 hour prediction for the Avedøre Wind Farm.

Studying the figures in detail reveals that of the seven storms in 97 one was predicted very well and two were predicted satisfactorily. In 98 four were predicted very well and three satisfactorily. This is very much in line with the results found for the wind speed at the Risø mast: in the first half of 1997 the HIRLAM model had problems, which then seems to have been fixed.

1997				
	Start	End	Quality	Notes
	7/1	9/1	–	level missed
	12/1	15/1	–	start-up missed
	18/3	20/3	=	end missed
	21/4	24/4	–	missed completely
	10/5	12/5	–	missed completely
	17/11	22/11	+	
	25/12	27/12	=	cut-out predicted
1998				
	14/1	18/1	+	
	26/1	29/1	+	
	15/2	18/2	=	
	25/2	2/3	+	cut–out predicted
	6/3	8/3	=	
	10/3	13/3	-	level missed
	19/3	20/3	=	
	15/6	19/6	+	
	14/7	16/7	=	cut-out predicted

Table 1: The storms. In the column marked 'Quality' the predictions are graded according to the following: – bad, = OK, + good.

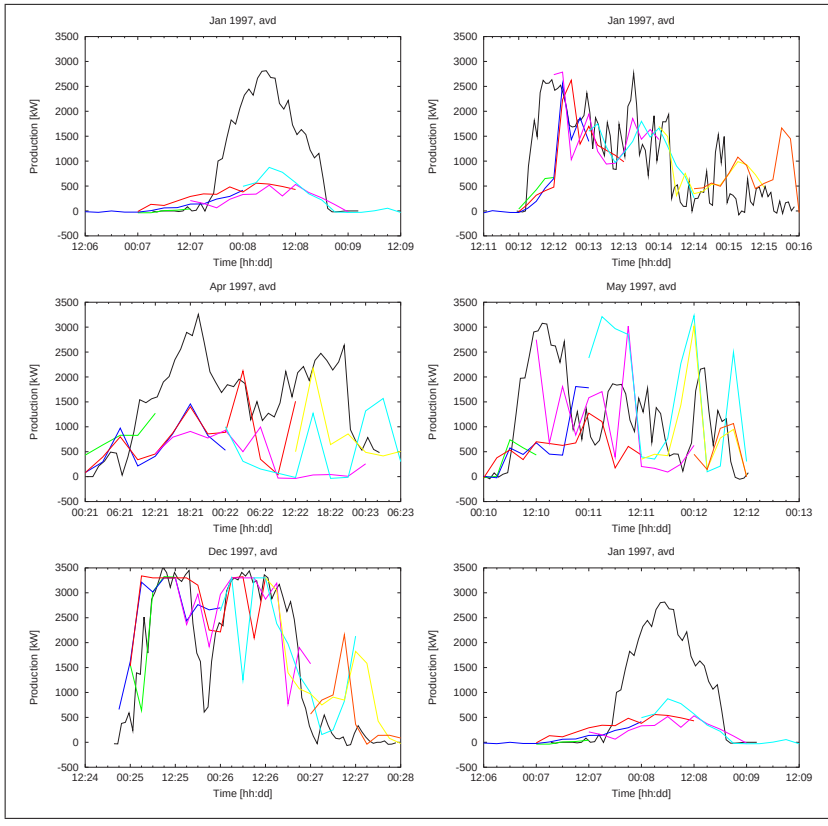


Figure 8: The storms in 1997. Solid line is the observed production, dashed lines are the predictions.

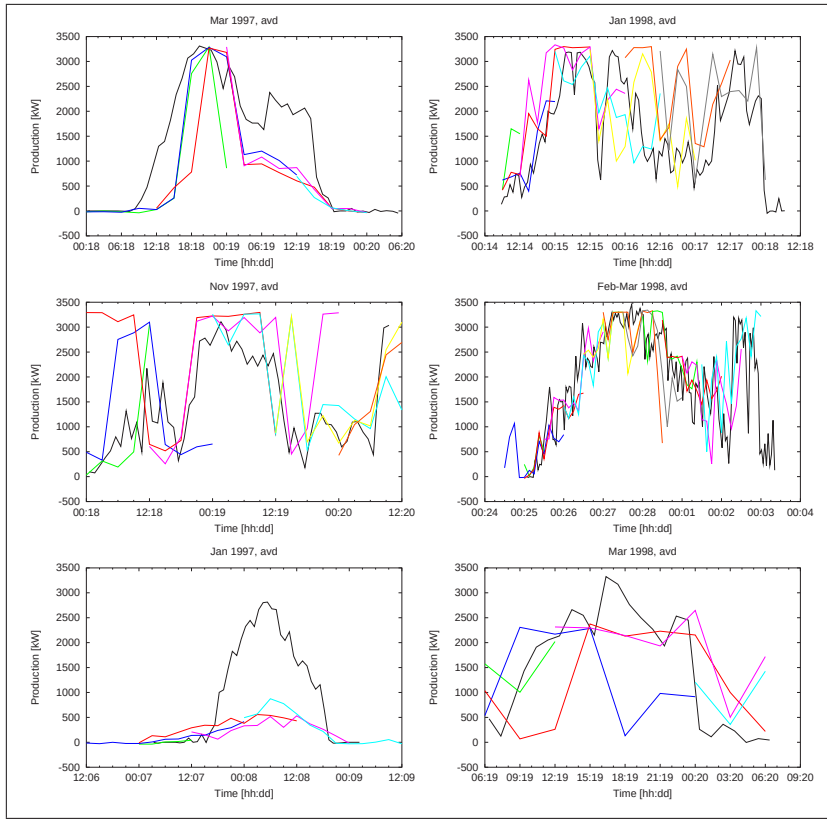


Figure 9: The storms in 1997/98. Solid line is the observed production, dashed lines are the predictions.

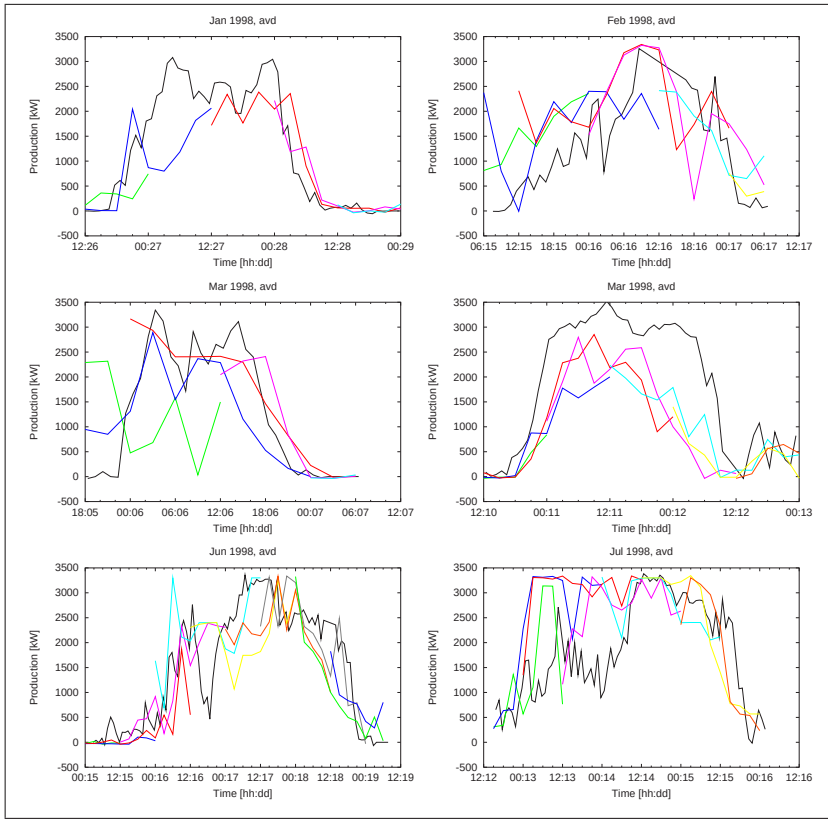


Figure 10: Storms in 1998. Solid line is the observed production, dashed lines are the predictions.

5 Summary

In this paper we have shown that the HIRLAM/WAsP model predicts wind as well as power very well. It was demonstrated that by supplementing the physical equations with MOS, the predictions were improved. Finally, it was found that strong and rapidly developing storms were predicted very well, particularly in 1998.

6 Acknowledgements

The development of the model has been funded by the EFP-programme under the Danish Ministry of Energy and the Environment, contract number 1363/94–0005 and is presently funded by the JOULE III programme of the Commission of the European Communities, contract JOR3–CT95–0008. The US part is funded by EPRI.

References

Landberg, L. (1997), 'Short-term prediction of wind farm power output – from theory to practice', Proceedings from BWEA 19, Edinburgh (UK). ed R Hunter, pp 397–402.

Landberg, L. (1999*a*), 'A mathematical look at physical power prediction model', *Wind Energy* **1**, 23–28.

Landberg, L. (1999*b*), 'Short-term prediction of the power production from wind farms', *Jour. Wind Engineering and Industrial Aerodynamics* **80**, 207–220.

Landberg, L., Hansen, M. A., Vesterager, K. & Bergstrøm, W. (1997), 'Implementing wind forecasting at a utility', RisøNational Laboratory, Roskilde, Denmark. Risø-R-929(EN).

Landberg, L. & Watson, S. J. (1994), 'Short-term prediction of local wind conditions', *Boundary-Layer Meteorol.* **70**, 171–195.

Machenhauer, B. (1989), 'The HIRLAM Final Report', Danish Meteorological Institute, Copenhagen, Denmark.

Mortensen, N. G., Landberg, L., Troen, I. & Petersen, E. L. (1993), 'Wind atlas analysis and application program (WAsP)', RisøNational Laboratory, Roskilde, Denmark. Risø-I-666(EN).

Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L. & Giebel, G. (1999), 'A new reference for predicting wind power', *Wind Energy* **1**, 29–34.

Sanderhoff, P. (1993), 'PARK - User's Guide. A PC-program for calculation of wind turbine park performance', RisøNational Laboratory, Roskilde, Denmark. Risø-I-668(EN).

Paper G

Short-term prediction towards the 21st century

G

Originally published as

L. Landberg, A. K. Joensen, G. Giebel, H. Madsen and T. S. Nielsen. Short-term Prediction towards the 21st Century. In *proceedings from BWEA 21*, British Wind Energy Conference, pages 127–136, UK, 2000.

Short-term prediction towards the 21st century

Lars Landberg[1], Alfred Joensen[1,2], Gregor Giebel[1], Henrik Madsen[2] and Torben S. Nielsen[2]

Abstract

A new chapter in the continued and exiting story of short-term prediction has begun! The paper will describe a new project funded by the Danish Ministry of Energy where all the Danish utilities (Elkraft, Elsam, Eltra, and SEAS) will participate. The goal of the project is to develop and implement on-line a prediction system combining the Risø and IMM models. This will ensure that the best forecasts are given on all prediction horizons form the short range (0-9 hours) to the long range (36-48 hours).

1 Introduction

Electrical utilities, wind farm owners, green certificate and power traders and everybody else with a commercial interest in wind energy must realise that – in a marked which expands 30% every year – the variability of the wind and thereby the power production must be dealt with as efficiently as possible.

There is no doubt that if one is a utility with more than 10% of the electrical power coming from wind energy the variability is a real problem; but recently also new players have come on the marked. This is because of the opening of the electrical market. These players include the wind farm operator who must give accurate estimates of the expected production in order to avoid severe penalties imposed by the power buyers.

One way of handling the variability is to predict the expected wind energy produced power well into the future, ie up to two days in advance. This is possible now (as first demonstrated in (Landberg & Watson 1994)) using numerical weather prediction (NWP) models.

[1]Department of Wind Energy and Atmospheric Physics, Risoe National Laboratory, DK-4000 Roskilde, Denmark

[2]Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

This paper will describe a new model which is under development in a co-operation between Risø and IMM and all the Danish electrical utilities. The paper will also begin by briefly outlining what today is considered state-of-the-art.

2 State-of-the-art

Presently there are two models to predict the power production from wind farms in operation at electrical utilities and these are both considered state-of-the-art:

1. The Risø model
2. The IMM model

The two models both use weather predictions from NWP models (here the Danish Meteorological Institute HIRLAM model) as input. The way this input is used is different for the two models:

The Risø model uses mainly physical relations to transform the predicted wind into predicted power: the geostrophic drag law, the logarithmic wind profile, WAsP corrections for local influences, PARK calculations for actual wind farm output. The results are corrected using a mathematical filter (a MOS filter). The model predicts for individual wind farms or groups of wind farms representing an area.

In *the IMM model* statistical methods are applied for predicting the expected wind power production in a larger area using on-line data covering only a subset of the total population of wind turbines in the area. The approach is to divide the area of interest into sub-areas each covered by a wind farm. Predictions of wind power with a horizon from half an hour up to 39 hours are then formed for the individual wind farms using local measurements of climatic variables as well as meteorological forecasts of wind speed and direction. The wind farm power predictions for each sub-area are subsequently up-scaled to cover all wind turbines in the sub-area before the predictions for sub-areas are summarized to form a prediction for the entire area.

3 The Project

The present paper describes a modeling system which will be developed in a project funded by the Energy Research Programme (EFP) under the Danish Ministry of the Environment and Energy. The title of the project is "Wind farm production predictor" and it has as its aim to develop a model which is a combination of the two models mentioned above and implement the system at all the Danish utilities.

The partners in the project are:

- Risø National Laboratory (coordinator)
- Institute of Mathematical Modeling at the Danish Technical University
- Elkraft, utility
- Elsam, utility
- Eltra, utility
- SEAS, utility

These partners give a good blend of research and industry.

The project started in April 1999 and will run for three years. The first version of the new model is expected to be ready in April 2000. The project involves 94 man months of work.

4 Why a New Prediction System?

The main goal is to merge the two state-of-the-art models (Risø's and IMM's), to obtain synergy between the physical and the statistical approach. This will give reliable forecasts on the short (0-9 hours) as well as the long term (24-48 hours).

Since the two "old" models were developed a lot has changed on the programming side as well, so the opportunity is taken to implement the newest programming methods: A Client/Server architecture build as Java Beans, connected to a SQL database interfaced via JDBC (see later for details).

A last advantage is that by developing a uniform model for all the Danish utilities, all efforts are focused on this model, which gives the best possible improvements and easier maintenance.

5 The New System

There are two aspects of the new system: the mathematical model and the system architecture. Both of these will be described in the following.

5.1 The Model

This section describes how a prediction model covering the total wind power production in an area can be derived. The approach is to divide the area of interest into a number of sub-areas. Predictions of wind power with a lead time from half an hour up to 48 hours are then formed for the individual sub-areas using data of the local power production as well as weather forecasts for the sub-area. The power predictions for each sub-area are then summarized to form a prediction for the entire area.

In the model setup described in the following it is assumed that data will be available from three sources:

- Weather forecasts for some selected climatic variables – primarily wind speed and wind direction.
- On-line measurements of power production in a number of (reference) wind farms in the area. Each sub-area must include at least one of the reference wind farms.

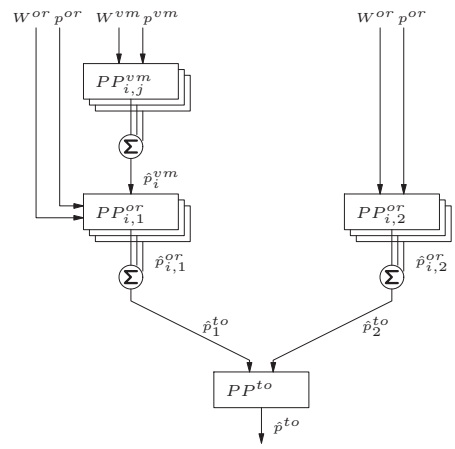


Figure 1: Model overview.

- Hourly sums of the power production for the entire population of wind turbines in each sub-area. These data will be available on a daily basis, i.e. they are not on-line measurements.

With data from these sources in hand a prediction model for the total wind power production in the area can be derived as illustrated in Figure 1. The model consists, as it is seen from the figure, of two model branches each providing a prediction of the total power production in the area but using different input data.

In the left hand model branch a dynamic wind farm prediction model ($PP^{vm}_{i,j}$) calculates predictions of the power production for the individual wind farms in a sub-area using local measurements of power production and weather forecasts. The wind farm prediction model can be written as

$$\begin{aligned} \hat{p}^{vm}_{t+k|t} &= a(k,\theta^{vm}_{t+k|t}) \cdot F(p^{vm}_{t}) + \\ &\quad b(k,\theta^{vm}_{t+k|t}) \cdot P^{vm}(W^{wm}_{t+k|t}) + \\ &\quad m(k,\theta^{vm}_{t+k|t}) \end{aligned} \tag{1}$$

where $\hat{p}^{vm}_{t+k|t}$ is the predicted power production at time $t+k$ given at time t, p^{vm}_{t} is the observed power production at time t, $F()$ is a low-pass filter function, $W^{vm}_{t+k|t}$ are the local weather forecasts for the wind farm at time $t+k$ given at time t, $P()^{vm}$ is an estimated power curve function (see

below) and finally $a()$, $b()$ and $m()$ are smooth functions of prediction horizon k and forecasted wind direction. The functions $a()$, $b()$ and $m()$ are unknown and has to be estimated (See (Nielsen 1999)). The functions must depend on k as the uncertainty of the weather forecasts increases as the prediction horizon grows.

The predictions for the individual reference wind farms in the sub-area are summarized and subsequently up-scaled by $PP^{or}_{i,1}$ to cover all wind turbines in the sub-area. Here $PP^{or}_{i,1}$ describes the static relationship between the total power production in a sub-area and the predicted power production for the related reference wind farms as a function of local weather forecasts. The upscaling model is given as

$$\begin{aligned} \hat{p}^{or}_{t+k|t} &= b(k, w^{or}_{t+k|t}, \theta^{or}_{t+k|t}) \cdot \hat{p}^{vm}_{t+k|t} \\ &\quad + m(k, w^{or}_{t+k|t}, \theta^{or}_{t+k|t}) \end{aligned} \tag{2}$$

where $\hat{p}^{or}_{t+k|t}$ is the predicted power production for the sub-area at time $t+k$ given at time t and $b()$, $m()$ are smooth functions of prediction horizon and local forecasts for the sub-area of wind speed and wind direction.

The total power prediction for the left hand model branch is subsequently found by summarizing the power predictions for the individual sub-areas.

The right hand model branch does not, as opposed to the left hand side, rely on on-line observations. Instead the power predictions for the sub-areas are calculated using the sub-area prediction model $PP^{or}_{i,2}$, which describes the static relationship between the total power production in the sub-area and local weather forecasts. The sub-area prediction model is given as

$$\hat{p}^{or}_{t+k|t} = P^{or}(W^{or}_{t+k|t}) \tag{3}$$

where $W^{or}_{t+k|t}$ are local weather forecasts for the sub-area at time $t+k$ given at time t and $P()^{or}$ is an estimated power curve function (see below). Once again the prediction of the total power production is found by summarizing the power predictions for the individual sub-areas.

The power curve for the wind farms and the sub-areas referred to above

is given as

$$\begin{aligned} \hat{P}^{xx}(W^{xx}_{t+k|t}) &= m(k, w^{xx}_{t+k|t}, \theta^{xx}_{t+k|t}) \\ &\quad + b(k, w^{xx}_{t+k|t}, \theta^{xx}_{t+k|t}) \\ &\quad \cdot S^{xx}_{t+k|t} \end{aligned} \tag{4}$$

where xx means either vm (wind farm) or or (sub-area), $S^{xx}_{t+k|t}$ is a local forecast of a stability measure (to be determined) and finally $b()$, $m()$ are smooth functions of prediction horizon as well as local forecasts of wind speed and wind direction.

Comparing the predictions from the two branches it is expected that the predictions calculated by the left hand branch will be superior for the shorter predictions horizons as they are anchored to "the real world" via the on-line power observations, whereas the opposite is expected to be true for the longer prediction horizons. The final power prediction for the total area is calculated by PP^{to} using the predictions from the two model branches as input. Using mean square error (MSE) as a measure this could be done simply by picking the best of the two predictions for the individual prediction horizons, or at bit more sophisticated by calculating the final prediction as the weighted average of the predictions from the two branches using MSE^{-1} as weighting.

5.2 The system architecture

Experience from the Danish utilities Elsam and Elkraft with the use of the existing prediction systems developed by Risø and IMM has shown that the prediction system needs to be highly flexible. Demands, such as the simultaneous use by several users, call for a Client/Server architecture, where all data assimilation and numerical calculations are performed by the server, and multiple clients can connect to the server for retrieving and viewing the predictions. Furthermore, to evaluate the performance of the models, it has to be possible to show historical predictions against the actual observations and the current status of wind farms connected to the system, e.g. in case of a wind turbine failure. Therefore the server will be responsible for maintaining a database of all relevant data collected by the system, which can be retrieved from the server and shown as e.g. tables or plots by the client.

Other features required by the Danish utilities include the possibility to add or remove wind farms to or from the system while the system is on-line, the possibility to include meteorological stations, to contain wind farms with different properties, e.g. farms where data is received hourly, and wind farms where data is only received once a day. Also, depending on the type of measurement at different wind farms, these could get different prediction models assigned. Furthermore, it should be possible to configure the system to use input from different numerical weather prediction models. The number and type of prediction models running in the system should be configurable, and each user should be able to configure personal client settings, e.g. which prediction model to use and if he wants to look at all the wind farms in the system or only a subset.

Based on all these requirements, the system will be implemented in an object oriented programming (OOP) language, which supports a Client/Server architecture and features a full set of graphical components for building GUIs (Graphical User Interfaces). The system (at least the clients, but preferably also the server) should run on various platforms.

Therefore, it has been decided to use the Java(tm) 2 platform from Sun Microsystems (`www.java.sun.com`) for the implementation of the prediction system application. The graphical user interface will be based on the Swing package in the JDK(tm) 2 release from Sun, the Client/Server architecture will be implemented using RMI (Remote Method Invocation) and EJB (Enterprise Java Beans). An SQL (Structured Query Language) database, which supports the JDBC (Java Database Connectivity) interface, will be used for the data storage.

Figure 2 illustrates the Client/Server configuration, using Java technology that complies with open standards. This implies that the system becomes platform independent, and hence the server can run on a Unix workstation as well as on Windows NT workstations, while multiple clients running under different operating systems can connect to the server.

Figure 3 focuses on a lower level of detail in the system. It illustrates how different components in the system are linked. The server holds a refer-

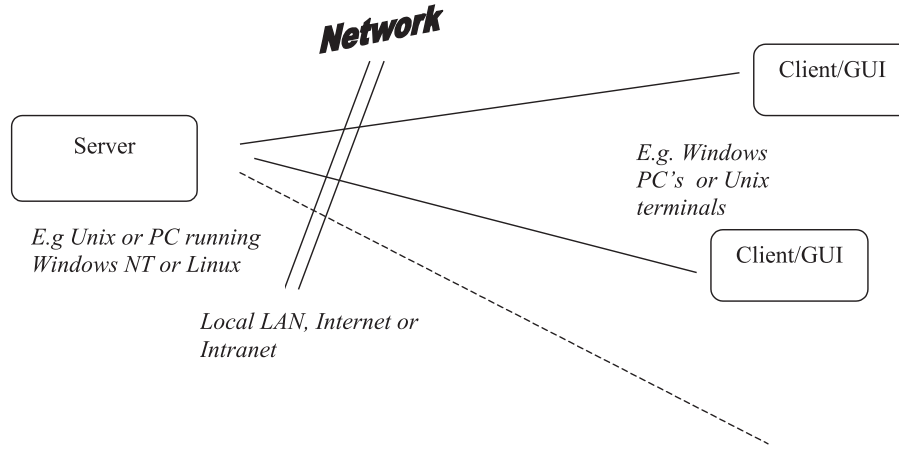


Figure 2: The Client/Server configuration using Java technology.

ence to all modules on the server side of the application, e.g. all the farm modules, which are responsible for running the farm level mathematical models, and modules running mathematical models calculating the total power predictions. Each time the model is updated, this should be reflected in the GUI components in the client, e.g. a component showing the current state of a wind farm (production, wind direction, etc.), or a plot showing the total power predictions. This is accomplished by using an *event model* over the RMI interface. All client components, which are interested in notification when a component running on the server is updated, register themselves as *listeners* to this server component. The server component maintains a list of all interested listeners, and each time the server component is updated, it runs through the list calling all listeners to notify that an update or change has occurred.

Using this concept throughout makes it easy to build a highly dynamical system. The components in the server do not need to worry about the actual implementation of a large number of components in the client, and which components will need to be notified or not, since the client components decide themselves. Figures 4 and 5 show snapshots of a preliminary GUI, which serves as an illustration of how the actual GUI will be implemented.

Figure 4 shows the map of Jutland, Denmark. Four wind turbines sym-

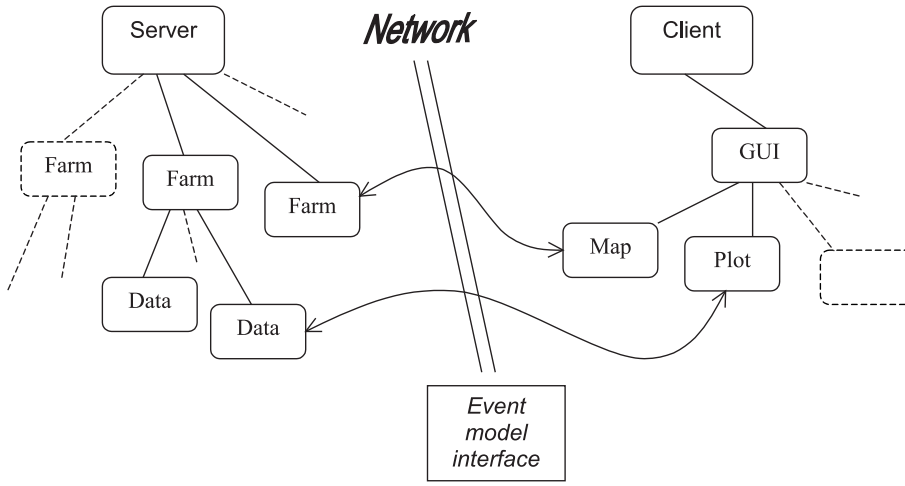


Figure 3: The linking of the different components in the system.

bolising wind farms are shown. Each time new data is received by the components, which are responsible for these wind farms in the server, the picture of the wind turbines will be updated. This information is dispatched to the "wind turbine" component on the map, using the event model interface described above. Additional information is available with a mouse-over event, by placing the mouse cursor over the turbine.

On the left hand side of the screen in Figure 5 a tree is shown, which illustrates the components running on the server, and how these components are linked. Furthermore, each node in the tree gives the user access to manipulation and/or retrieving of information about the specific component. If e.g. the wind speed node is dragged to the plot window, the wind speed measurements from the selected wind farm will be added to the plot.

It should be noted that the description of the system in this section has only been meant to illustrate the type of features, which will be available in the system. Several other features not described here will be implemented in the actual system, as well as what has been described here might change slightly, as the applications is still in the design phase. The modular design of the application, where e.g. wind farms with different characteristics can be plugged in to one or more total modules,

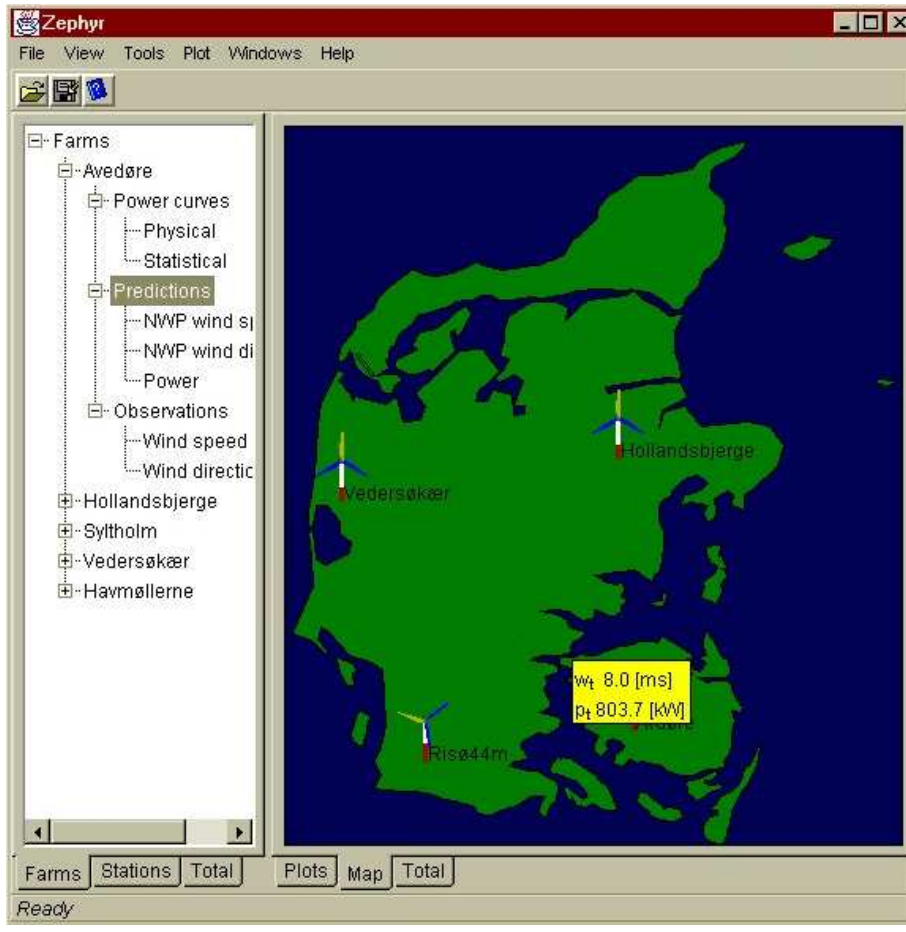


Figure 4: Snapshot of the preliminary GUI. The snapshot is of Jutland (DK) and the three wind turbines symbolise three wind farms.

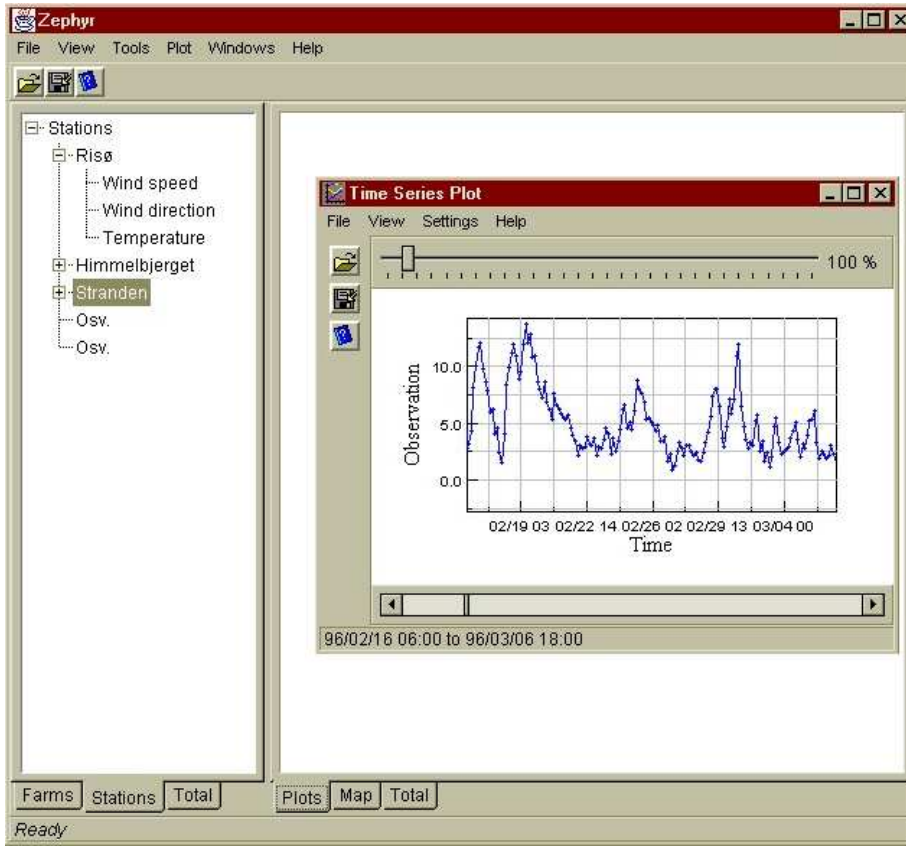


Figure 5: Snapshot of the preliminary GUI, showing a tree illustrating the components running on the server.

which calculate the total power predictions, makes it easy to customise the application to different utilities.

6 Summary

This paper has described a new system to predict the power output from wind farms. The system is being developed in a EFP-funded project which has Risø and IMM as the modeling team and all the Danish utilities as partners and users.

The mathematical model in the system is dual stringed, combining a model which uses predictions from a weather prediction model and on-line observations on the one side and weather predictions only on the other side. This is to give as accurate as possible predictions on both the short- and the long-term.

The system is designed as a Client/Server architecture, using the newest programming techniques which include: Java, Java Beans, RMI (Remote Method Invocation) and a SQL database supporting the JDBC interface. This design gives a very high degree of flexibility allowing the users to customise the system exactly to meet his or her requirements.

The final system will be ready in 2002, and the first version is expected on-line and operational in 2000.

7 Acknowledgements

The development of the prediction system is funded by the EFP-programme under the Danish Ministry of Energy and the Environment, contract number 1363/99-0017. Previous version of the models have been supported by the Danish EFP and European Union JOULE programmes.

References

Landberg, L. & Watson, S. J. (1994), 'Short-term prediction of local wind conditions', *Boundary-Layer Meteorol.* **70**, 171–195.

Nielsen, T. S. (1999), 'Using meteorological forecasts in on-line predictions of wind power', ELSAM, Fredericia, Denmark.

Paper H

Implementation of Short-term Prediction

Originally published as

L. Landberg, A. K. Joensen, G. Giebel, . Watson, H. Madsen, T. S. Nielsen, L. Laursen, J. U. Jørgensen, D. Lalas, M. Trombou, S. Pesmajoglou, J. Tøfting, H. Ravn, E. MacCarty, E. Davis and J. Chapman. Implementation of Short-term Prediction. In *Wind Energy for the Next Millenium*, European Wind Energy Conference, pages 52–57, Nice, France, March 1999.

Implementation of Short-term Prediction

Lars Landberg[1], Alfred Joensen[1], Gregor Giebel[1], Henrik Madsen[2], Torben S. Nielsen[2], SJ Watson[3], L Laursen[4], JU Jørgensen[4], DP Lalas[5], M Trombou[5], S Pesmajoglou[5], J Tøfting[6], H Ravn[7], E MacCarty[8], E Davis[9] and J Chapman[3]

Abstract

This paper will give a general overview of the results from a EU JOULE funded project ("Implementing short-term prediction at utilities", JOR3–CT95–0008). References will be given to specialised papers where applicable. The goal of the project was to implement wind farm power output prediction systems in operational environments at a number of utilities in Europe. Two models were developed, one by Risø and one by the Technical University of Denmark (DTU). Both prediction models used HIRLAM predictions from the Danish Meteorological Institute (DMI).

Keywords: Short-term prediction; wind farm power output.

1 Introduction

In many places around the world, but in Europe in particular, the number of wind farms is now so large that the electricity production from these wind farms have (sometimes critical) effect on the running and control of the overall electrical grid.

[1]Department of Wind Energy and Atmospheric Physics, Risoe National Laboratory, DK-4000 Roskilde, Denmark
[2]Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark
[3]Rutherford Appleton Laboratory (UK)
[4]Danish Meteorological Institute (DK)
[5]Observatory of Athens (GR)
[6]ELSAM (DK)
[7]ELKRAFT (DK)
[8]WECTEC (US)
[9]E Davis Consult (US)

To fully benefit from these large amounts of wind energy it is therefore necessary to have some kind of idea of the expected production in the next few days.

This will enable the electrical utilities to control the conventionally fueled plants in such a way that fossil fuels will be saved.

2 The Project

The project consisted of a group of people with skills in many areas: model development and evaluation, utility practices, implementation of models and so on.

2.1 Goal

The goal of the project was to carry out on-line implementation of prediction models developed in earlier projects. The evaluation of the models should be done in economic terms as well as the more traditional ways.

By putting most of the weight on the implementation side and not so much on the actual model development, the project aimed at (and reached) demonstrating that these very sophisticated models can run reliably in a real life operational situation.

2.2 Partners and roles

• Risø National Laboratory (DK): prediction model development, implementation and evaluation, co-ordination
• Rutherford Appleton Laboratory (UK): model evaluation using the National Grid Model (NGM) (Halliday 1988)
• Technical University of Denmark, IMM (DK): prediction model development, implementation and evaluation
• Observatory of Athens (GR): model development and evaluation
• ELSAM (DK): model implementation and implementation evaluation,

wind farm measurements
• ELKRAFT (DK): model implementation and implementation evaluation, wind farm measurements
• WECTEC (US), E Davis Consult (US), OEM (US): model evaluation for the US cases.

3 Outcome

Two prediction models were developed: the Risø and the IMM models. These models were implemented and evaluated in a number of ways described in the following. The two models both use weather predictions from Numerical Weather Prediction (NWP) (here the Danish Meteorological Institute HIRLAM model, cf (Machenhauer 1989)) models as input. The way this input is used is different for the two models and the differences and the implementation will be described in the following.

The main difference between the two models is that the Risø model was developed with a utility in mind with no on-line access to the wind farm productions, whereas the IMM model was developed with a utility with on-line wind farm productions available. Predictions were made for a number of stations shown in Figure 1. Furthermore, the Risø model was run in an off-line mode for three sites in the US.

The total sum of findings and results form the project can be found in the final report written to the European Commission (Landberg, Joensen, Giebel, Watson, Madsen, Nielsen, Laursen, Jørgensen, Lalas, Tøfting, Ravn, MacCarty, Davis & Chapman 1999).

3.1 Risø model

The Risø model uses mainly physical relations to transform the predicted wind into predicted power: the overall HIRLAM-predicted wind is transformed to the surface using the geostrophic drag law and the logarithmic wind profile, the surface wind is corrected for local influences using the WAsP model (Mortensen, Landberg, Troen & Petersen 1993, Troen & Petersen 1989), and the PARK program (Sanderhoff 1993) is used for

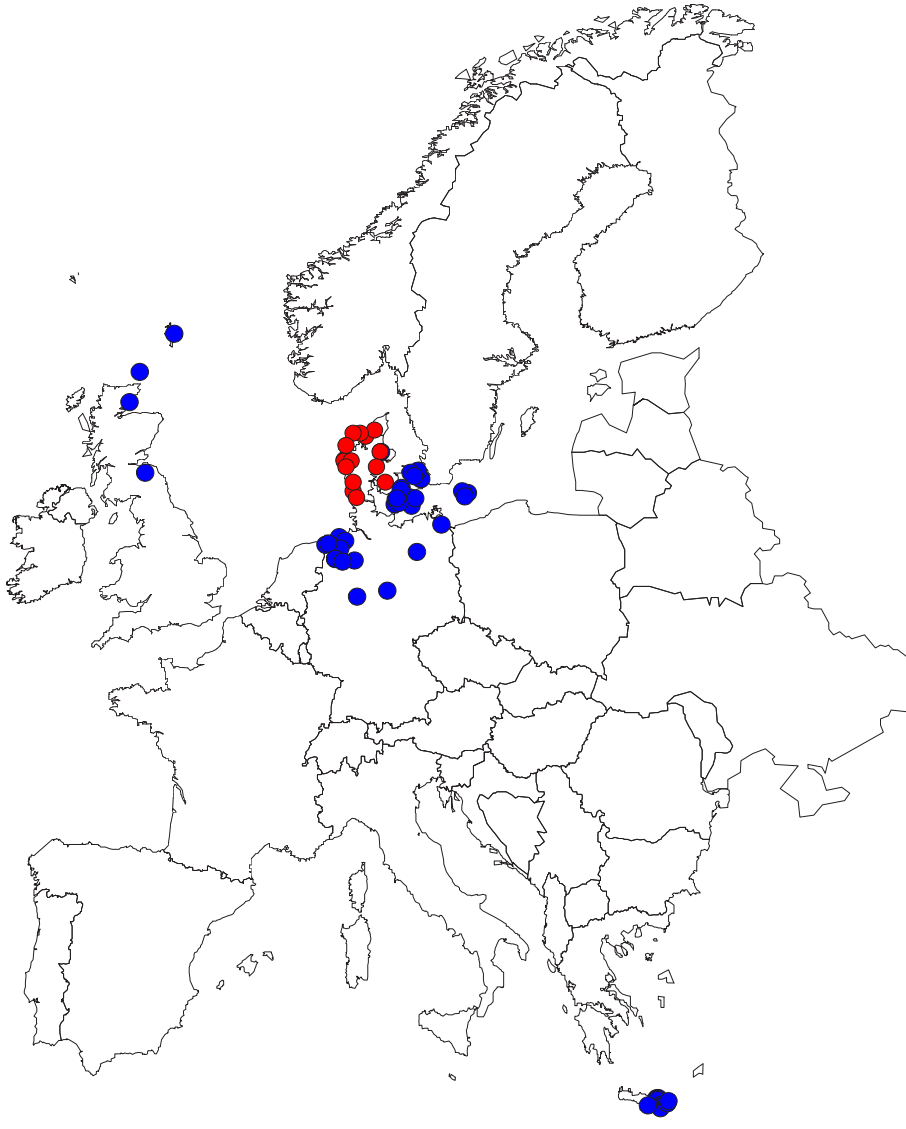


Figure 1: The location of the wind farms. Black dots are the wind farms for which Risø predicts, and gray are the ones IMM predicts for.

calculations of actual wind farm output. The results are corrected using a mathematical filter (a MOS filter). For detailed description and analysis of the model see (Landberg 1999*b*, Landberg 1998, Landberg 1997, Landberg 1999*a*).

The on-line implementation of the Risø model is shown in Figure 2. An example of the predictions as seen of the WWW is shown in Figure 3.

To give an example of the Risø model's ability to predict storms Figure 4 shows the development of a storm and how well the predictions agreed with the observations.

3.2 IMM model

In the IMM model statistical methods are applied for predicting the expected wind power production in a larger area using on-line data covering only a subset of the total population of wind turbines in the area. The approach is to divide the area of interest into sub-areas each covered by a wind farm. Predictions of wind power with a horizon from half an hour up to 39 hours are then formed for the individual wind farms using local measurements of climatic variables as well as meteorological forecasts of wind speed and direction. The wind farm power predictions for each sub-area are subsequently up-scaled to cover all wind turbines in the sub-area before the predictions for sub-areas are summarized to form a prediction for the entire area. The model is described in great detail in (Nielse & Madsen 1997).

The idea behind the implementation is shown in Figure 5. The over-view screen of the prediction system at Elsam is shown in Figure 6

3.3 Elkraft implementation

Elkraft Power Company coordinates energy cooperation in the eastern part of Denmark.

In the Zealand area there are now installed wind turbines with a total ca-

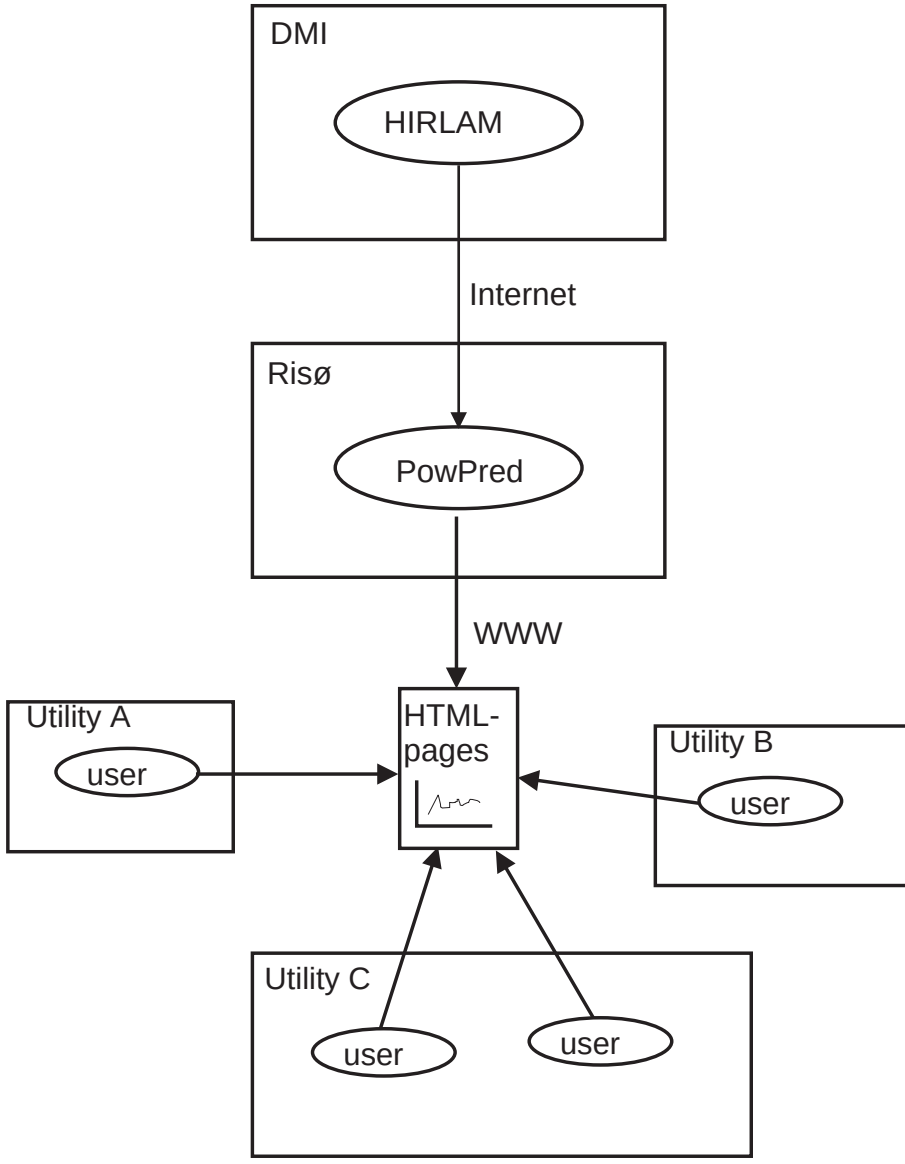
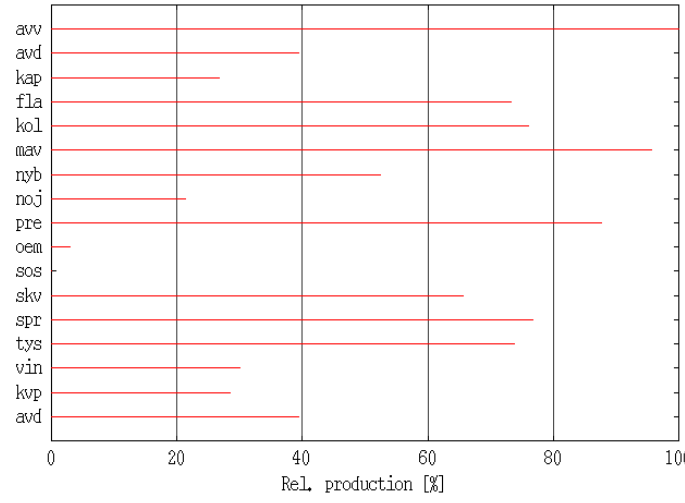


Figure 2: The idea behind the on-line implementation of the Risø model.

Predictions Page 1 of 1

Predictions +12 hours ahead

0 3 6 9 12 15 18 21 24 27 30 33 36



Absolute numbers [kW]

Click on wind farm for full predictions

Click on prediction length to get that prediction

12hrel.htm

Figure 3: The page viewable on the WWW showing the Risø predictions.

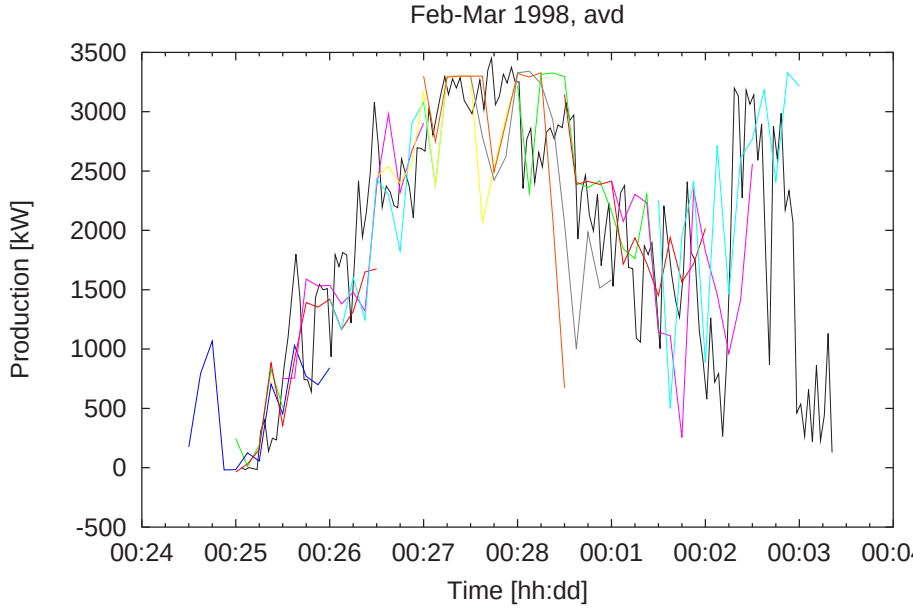


Figure 4: The storm on the 27th February as seen from the Avedøre Wind Farm. Solid line is the observed production and dashed lines are the predictions using the Risø model.

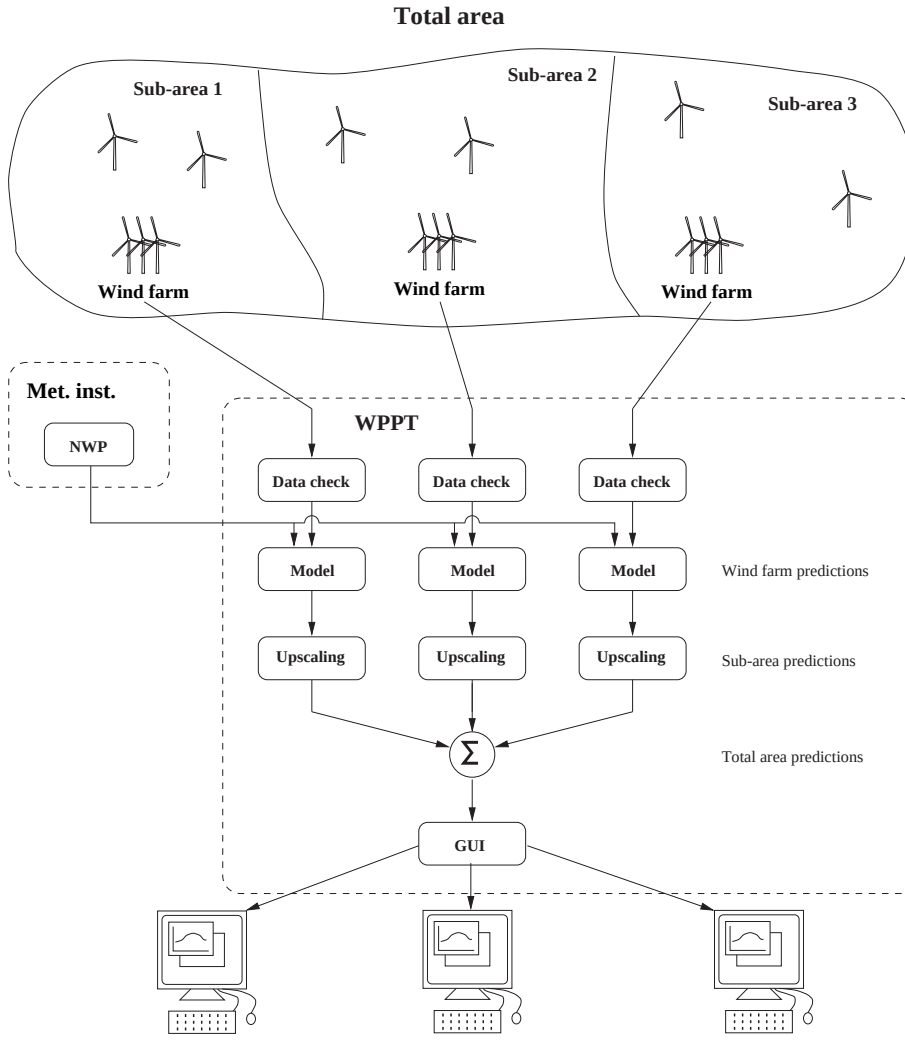


Figure 5: The idea behind the on-line implementation of the IMM model.

Figure 6: The over-view screen of the IMM model as seen at the Elsam utility

pacity of around 300 MW. This figure will double or triple within the next decade. The influence of the fluctuations of the wind power is already being felt in the daily control and operation of the system. Therefore efforts are being undertaken to predict the wind power production.

In the present version of the wind power prediction system the data flows may be sketched as follows. The Danish Meteorological Institute produces predictions of wind speeds for a number of specified locations, 15 in total, where major wind farms are located. The prognoses are represented as values for every third hour, with a time horizon of 36 hours.

The prognoses are transmitted to Risø National Laboratory, where the predictions of wind speed are automatically transformed to predictions of power production, based on WAsP analyses of the specific wind farms. These predictions may be seen at the homepage at Risø National Laboratory.

Elkraft Power Company takes the predictions from Risø National Laboratory via the Internet. The predictions are then combined with the available knowledge to produce a prognosis for the whole are of interest. In particular, the wind turbines, for which individual prognoses are not made, are included by using an up-scaling factor. Further, tuning of the prognoses is made, for instance to account for major wind farms under construction or revision. Longer time biases in the prognoses are detected by comparison with the available measurements, given as hourly or monthly production values.

The prognoses and online measurements are distributed via the local area network to the relevant persons, in particular to those in the control room and to those that trade power on short-term basis.

The system was introduced mid-1998, and has been functional during the last quarter of 1998. An example of the user interface at Elkraft is shown in Figure 7.

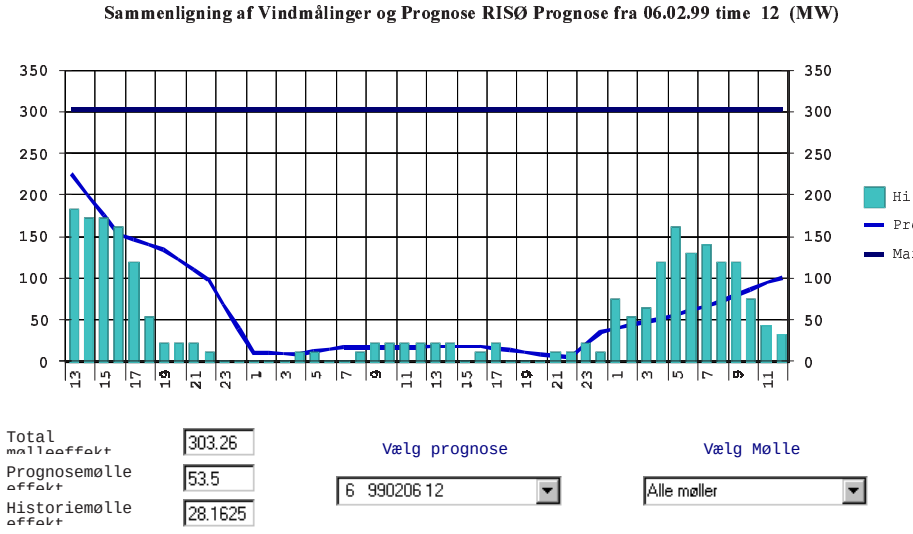


Figure 7: The predictions as displayed in the Elkraft dispatch and control centre. The columns are hourly measured wind power production, the straight horizontal line corresponds to maximum production (i.e installed capacity) and the curved line is the prediction. It is possible for the dispatcher to choose between current and previous predictions and productions, and which turbines to have displayed.

3.4 Elsam/Eltra implementation

In the Western part of Denmark Elsam is responsible for the economical load dispatch of the production from the primary power stations, whereas Eltra controls the transmission grid and has the system responsibility. The power production set-up consists of 6 primary power stations equipped with 4250 MW of CHP (Combined Heat and Power) units, a large number of local CHP units with a total installed power of 1400 MW and finally wind turbines with a total rated power of approximately 1000 MW. The production from the local CHP units and the wind turbines is treated as priority production, which implies, that the available power from these sources has to be accepted by the system responsible operator. On a yearly basis the load in the Elsam/Eltra area ranges between 1200 MW and 3700 MW. It is obvious, that the management of 1000 MW of wind power in such a setup will have to rely on the availability of dependable wind power predictions.

The IMM model is implemented in a software package called WPPT (Wind Power Prediction Tool). WPPT was installed in the control centres of Elsam and Eltra in October 1997 and has been used operationally since January 1998. The assessment by the operators is that WPPT generally produces reliable predictions, which are used directly in the economic load dispatch and the day-to-day electricity trade. In periods with unstable weather the operators may choose to modify the predictions (typically smooth the pattern of the prediction) before further usage though. The economical value of the wind power predictions is difficult to evaluate directly mainly due to the problem of assessing the course of action had the predictions not been available. Instead two cases have been analysed in order to illustrate how the predictions are used and with which consequences:

- *Case 1.* On October 17th 1998 the wind power production varied from 600 MW during the morning hours down to 300 MW at 6 pm before increasing to 800 MW at midnight. At 10:30 am the day before WPPT had predicted a wind production around 600 MW during the first half of the day rising to maximum production (930 MW at that time) from 8 pm and onwards. The next wind power prediction based on a new set of meteorological forecasts was

available at 4:30 pm October 16th and predicted a different course for the last part of the following day - from 700 MW just before noon down to 350 MW at 5 pm increasing to maximum production at midnight. The first prediction was so different from the actual production, that the running reserve would not have been capable of covering the missing production from 5 PM to 7:30 pm. The missing power would have had to be purchased from NordPool at a total extra price of approximately DKK 16,000. The second prediction was so much better, that the deviation could be countered by the normal means of regulation without any additional costs compared to a perfect forecast. See Figure 8 for details.

- *Case 2.* On November 9th the wind power production varied from 350 MW at midnight 8th increasing up to 800 MW at noon before decreasing down to 100 MW at midnight 9th. This course of the wind power production was accurately predicted the day before and consequently did not imply any costs for the operation.

As indicated by the examples above the operators rely on the wind power production from WPPT in the daily planning since the predictions are markedly better than what can be derived from other sources. This is not to say, that there is no room for improvement, and thus WPPT is subject to continues improvements based on the experiences of the operators.

3.5 RAL calculations

The National Grid Model was run for the England & Wales, Crete and Iowa grids and the results were:

- The forecasts give improved fossil fuel savings over persistence for the England & Wales grid - at least 13% better at 40% penetration (cf Figure 9).
- The results for Crete are poor because the site forecasts are poor.
- Crete has a lot of fast response plant and so forecasting is not so beneficial anyway unless models can significantly improve upon

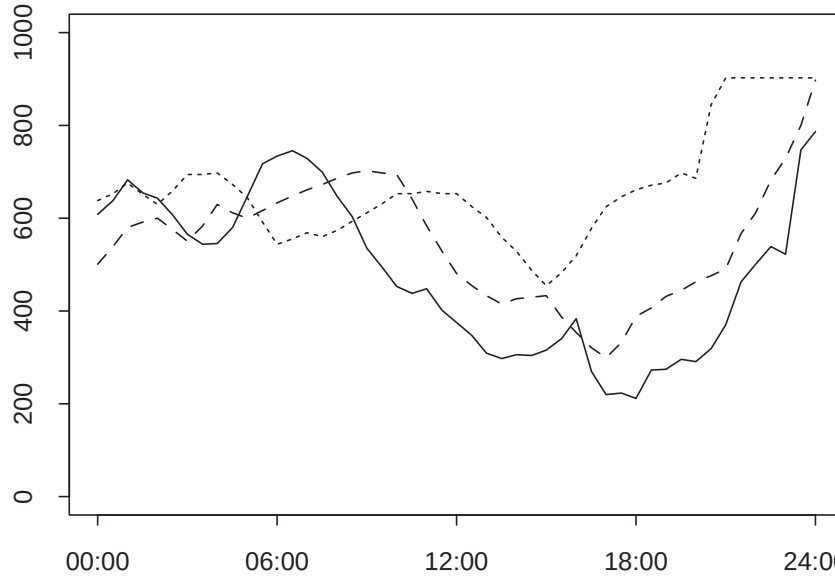


Figure 8: Case 1 (October 17th 1998) predicted by the IMM model. Solid line is the observed production, dashed lines are the predictions.

persistence at up to 4 hours ahead. Also a study of Crete would benefit from a hybrid of the RAL NGM and the RAL islands model (which can simulate diesel start-ups on a minute by minute basis).

- The Iowa results are disappointing, this is because of the crude temporal resolution of the forecasts (six-hours).

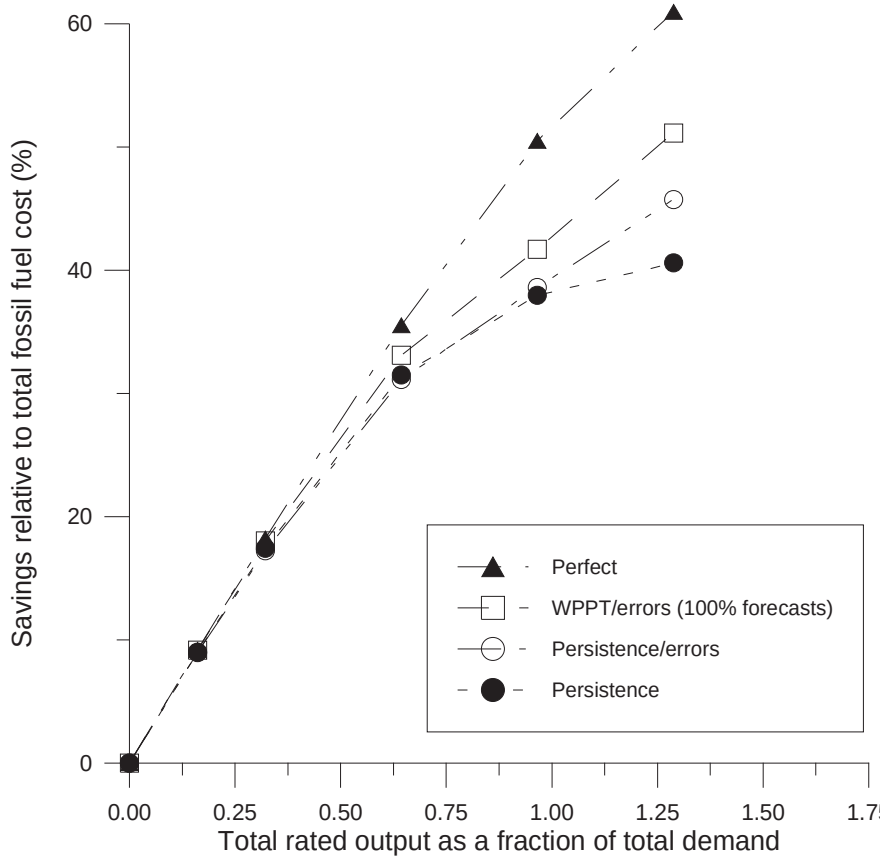


Figure 9: Fossil fuel savings during the calendar year 1994 for different installed wind power capacities into the England and Wales grid using different forecasting methods.

3.6 US results

The Risø forecasting model was applied to selected sites in the US. The general goal of this a application was to understand whether the approach used successfully in Europe could be transferred to facilities in the US. When this EPRI program was begun in 1996, it was thought that data either from the EPRI managed eight station North Dakota Wind Resource Assessment Program or the first large EPRI/DOE TVP (Turbine Verification Programme) project at Ft. Davis, Texas would be used. The availability of data from these two projects was limited, forcing EPRI to look elsewhere.

Datasets from regions where wind energy projects were either operational or being considered was a significant consideration in the choice of sites. In addition, areas where the terrain was not too complex, that is, hilly or mountainous, and areas where numerical weather prediction models might have sufficient valid data to perform successfully. The Great Plains was the prime area as the terrain is principally flat or rolling farm- and grasslands. In addition, there is a sufficient observational data base upwind of the Great Plains which should allow for good performance of numerical weather prediction models. The projected development of large wind electric generating facilities in the upper Great Plains of Minnesota and Iowa also focused EPRI on this region.

EPRI obtained data from the first operational wind plant on the Buffalo Ridge in Southwestern Minnesota for use in the forecasting application. This 25MW wind plant came on-line in mid-1994 and power data was made available by the wind plant owner for slightly over a 2-year period. Wind speed and wind direction data was also available for wind resource assessment programs conducted in Iowa during the same time period. These meteorological data were also obtained.

Application of the Risø model requires the use of numerical weather prediction data. The availability of historical data for the concurrent time period, mid 1994 to mid 1996, was researched. The National Center for Atmospheric Research (NCAR) was contacted and there archives were reviewed. The only complete prediction data set available for the US at that time was for the Nested Grid Model, the operational weather prediction model used by the National Center for Environmental Prediction

(NCEP). These data sets were assembled and provided to Risø National Laboratory for testing their modeling approach on US sites.

The Risø model was applied to the Buffalo Ridge Wind Plant, the meteorological site at Alta, and the meteorological site at Sibley. For the wind plant, preparation for application of the model included:

- Creation of a digitized terrain file. This file included all terrain contours within a 10 km radius of the wind plant.
- Creation of a roughness data file. This file included an estimate of the terrain roughness for each of twelve 30 degree sectors.
- Creation of the PARK data files. These files include the power curve for the KVS33, a thrust curve for the KVS33, a meteorological data file consisting of shape, scale and frequency of occurrence of the wind speed in twelve 30 degree sectors, and the location of each (73) KVS33 turbine.

For the two meteorological tower sites, a roughness file is the only required input since these are single sites, wind speed is the predictand, and the files for WAsP and PARK are not required.

From the numerical weather prediction data file, the 10 meter, 950 millibar (mb), 850 mb and 700 mb wind speed were extracted. The data for a matrix of four sites, coordinate point 42, 43, 51, and 52 are extrapolated separately to the Buffalo Ridge, Alta, and Sibley sites. This forms the basis for the predicted wind speed. For example, the predictions from the model run of 12159412 (December 15, 1994 at 1200GMT) would consist of wind speed and wind direction values for 8 time periods at six hour intervals, out to 12179412 (December 17, 1994 at 1200GMT).

The model was first applied to the meteorological data at the Alta site. Predictions of average wind speed at eight different time periods in the future twice each day are made using the model. These predictions are based on the numerical weather prediction model. These predictions are then compared to the observed values from the Alta tower. A matrix is then created comparing the predicted and observed values. This matrix

is plotted in Figure 10. The poor correlation and pronounced lack of linearity between the predicted and actual values is disappointing

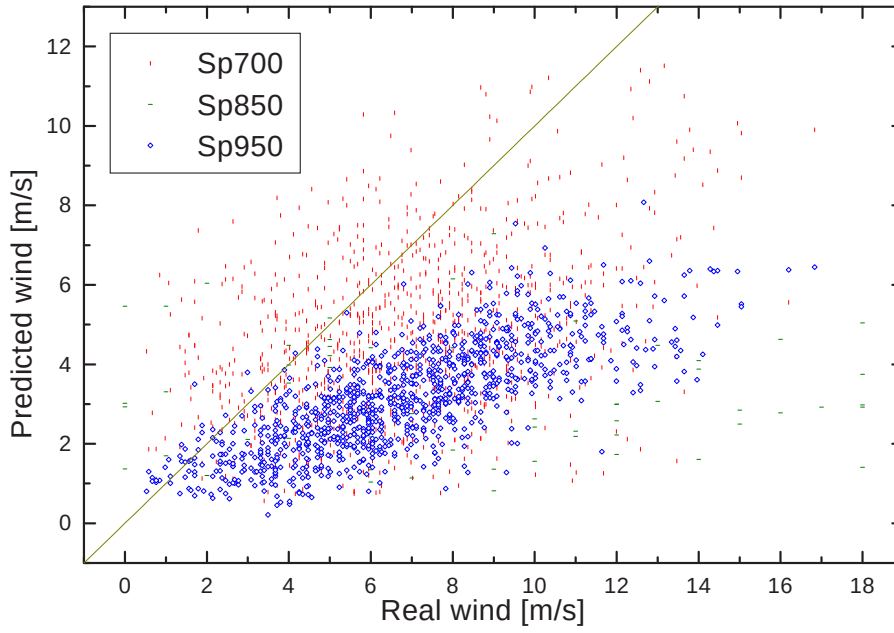


Figure 10: Scatter Plot of the Forecast Wind Speed Versus Actual Wind Speed at the Alta Site.

The reason for this is not yet clear. The data set, consisting of predicted values and actual values, either wind speed (Alta and Sibley) or wind power (Buffalo Ridge) is still being analyzed by staff at Risø. It is possible that the large grid spacing in the NGM model, compared to the smaller grid spacing in the HIRLAM model, could be one of the causative factors.

4 The Future

This section will describe the different ways the results of this project are planned to be used in future projects.

In Denmark the Energy Agency via the EFP99 research programme has funded a project where the Danish utilities will be provided with a predic-

tion system which is a combination of the Risø and the IMM prediction systems described in this paper. It is planned that at the end of the project all Danish utilities with a high penetration of wind energy will have this prediction system integrated in the daily dispatch and scheduling.

In the US it is hoped that a EPRI/DOE-funded project will start in 1999. The main goal is to implement the Risø prediction system in the US, predicting for a number of wind farms in the US.

5 Summary

This paper has described the results of a now finished JOULE project, which had as its goal to implement and evaluate in both traditional and economic term models for predicting the power output from wind farms. It was shown that two such models were implemented successfully and that the two major Danish utilities used the predictions in the daily dispatch and planning.

6 Acknowledgements

The project has been funded by the Commission of the European Communities under the JOULE programme, contract JOR3–CT95–0008. The development of the Risø model was as funded by the EFP-programme under the Danish Ministry of Energy and the Environment, contract 1363/94–0005 and JOULE contract JOUR–0091–MB(C). The IMM model was developed in a JOULE II project contract JOU2-CT92-0083.

References

Halliday, J. A. (1988), Wind meteorology and the integration of electricity generated by wind turbines, PhD thesis, University of Strathclyde, England. RAL internal report RAL T 075.

Landberg, L. (1997), 'Short-term prediction of local wind conditions', *Boundary-Layer Meteorology* **70**, 171–195.

Landberg, L. (1998), 'A mathematical look at a physical power prediction model', *Wind Energy* **1**, 23–30.

Landberg, L. (1999*a*), 'Operational results from a physical power prediction model', Proceedings from EWEC99, Nice (FR).

Landberg, L. (1999*b*), 'Short-term prediction of the power production from wind farms', *Jour. Wind Engineering and Industrial Aerodynamics* **80**, 207–220.

Landberg, L., Joensen, A., Giebel, G., Watson, S., Madsen, H., Nielsen, T., Laursen, L., Jørgensen, J., Lalas, D., Tøfting, J., Ravn, H., MacCarty, E., Davis, E. & Chapman, J. (1999), 'Implementing short-term prediction at utilities', Published for the CEC by Risø National Laboratory, Roskilde,Denmark. Final report to the European Commission, JOULE project JOR3–CT95–0008.

Machenhauer, B. (1989), 'The HIRLAM Final Report', Danish Meteorological Institute, Copenhagen, Denmark.

Mortensen, N. G., Landberg, L., Troen, I. & Petersen, E. L. (1993), 'Wind atlas analysis and application program (WAsP)', RisøNational Laboratory, Roskilde, Denmark. Risø-I-666(EN).

Nielse, T. S. & Madsen, H. (1997), Experience with statistical methods for predicting wind power, *in* Y. Sawaragi & S. Sagara, eds, 'In proceedings from EWEC97', Vol. 2, pp. 475–480.

Sanderhoff, P. (1993), 'PARK - User's Guide. A PC-program for calculation of wind turbine park performance', RisøNational Laboratory, Roskilde, Denmark. Risø-I-668(EN).

Troen, I. & Petersen, E. L. (1989), 'The european wind atlas', Published for the CEC by Risø National Laboratory, Roskilde,Denmark. 656 pp.

Paper I

HIRLAM - Analysis of vertical model levels

Submitted for publication.

HIRLAM - Analysis of vertical model levels

Alfred K. Joensen

Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark

Abstract

Using numerical weather predictions as input to calculations of corresponding wind power predictions from wind farms, requires an assessment of the wind at each wind turbine position. This paper examines how the wind from different vertical levels from a numerical weather prediction model, HIRLAM, corresponds with the actual measured wind speeds.

Keywords: Numerical Weather Prediction, HIRLAM, turbulence intensity, Wind Power.

1 Introduction

This paper presents an analysis of the numerical weather predictions from HIRLAM run by the Danish meteorological institute. A thorough description of HIRLAM can be found in (Sass et al. 1999) and (Machenhauer 1989). The emphasis is on applications of the predictions for wind power prediction. More specifically, wind speed prediction variables from HIRLAM at different vertical model levels are compared to measured wind speeds. Furthermore, the influence of the turbulence intensity on the vertical wind speed variation is examined.

2 Finding the correct HIRLAM model level

In (Landberg & Joensen 1998) it is argued that HIRLAM must have changed in 1997. This assumption is based on a comparison of the running mean of the measured wind speed at Risø National Laboratory and

the HIRLAM predicted wind for the same location. In the description of HIRLAM in (Sass et al. 1999) it is stated that various parameterizations schemes have been used in HIRLAM. It is therefore to be expected that it is possible to identify the change more accurately by looking at the relation between the wind at different model levels, as the wind at different model levels is linked via the turbulence parameterization.

Figure 1 verifies this assumption. It is clearly seen that the relation between the special output wind from HIRLAM, corresponding to 10 m agl, and the wind at the lowest model level changes, and the date of the change is found to be the 10th of September 1997.

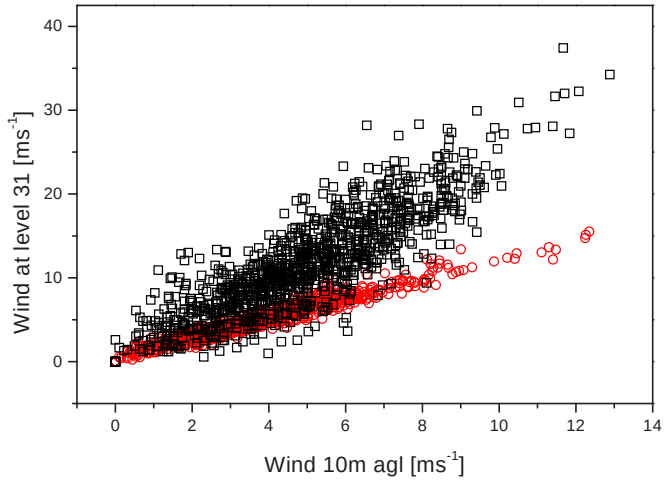


Figure 1: HIRLAM wind from the lowest model level versus the special 10m agl output wind from HIRLAM. Square symbols correspond to data before the 10th September 1997 and round symbols correspond to data received after the 10th of September

The results presented in (Joensen, Giebel, Landberg, Madsen & Nielsen 1999) are based on data received after the 10th of September. But it is interesting to see what the consequences of the change of the parameterization has for the accuracy of the predictions.

In Figure 2 and 3 the performance of the predictions is shown as a func-

tion of HIRLAM model level and prediction horizon, for each period. The performance measure is defined in (Joensen et al. 1999), and the simple model

$$w_{t+k} = a\omega_{l,t+k} + b + \varepsilon_{t+k}, \tag{1}$$

is used to link the observed wind speed w_{t+k} to the magnitude of the HIRLAM predicted wind ω_{t+k} from model level l, k is the prediction horizon. As the purpose of the analysis in this section is to identify the wind from HIRLAM which gives the highest degree of explanation, the simple model (1) should suffice.

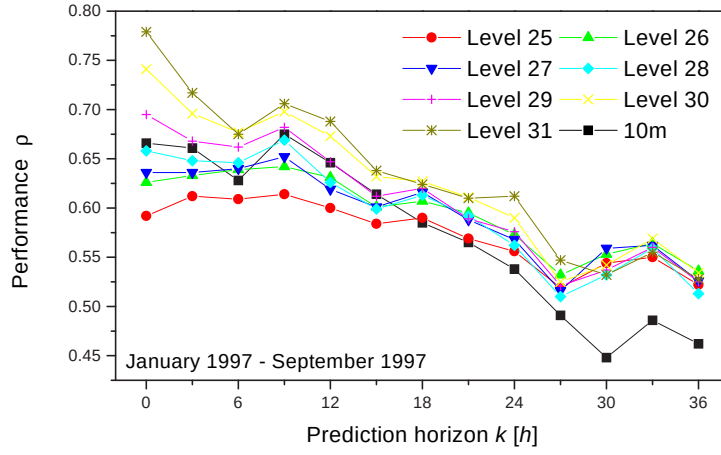


Figure 2: Wind speed prediction performance using the wind from various HIRLAM levels and for different prediction horizons. Data is from January 1997 to September 1997.

The most significant change that can be observed in these two figures is that in the second period, the model level winds give the poorest performance, while the special output wind from HIRLAM corresponding to 10 m agl clearly gives the best results. When comparing the performance in the two periods, using the HIRLAM wind which gives the best performance in each period, then it does not seem as if the performance is improved by the new parameterization scheme used in HIRLAM. In the first period there is a close run between the wind from the lowest HIRLAM levels, i.e. level 31 and 30.

In the next section these findings are further analyzed, as it is not possible to deduce from these figures what really has happened.

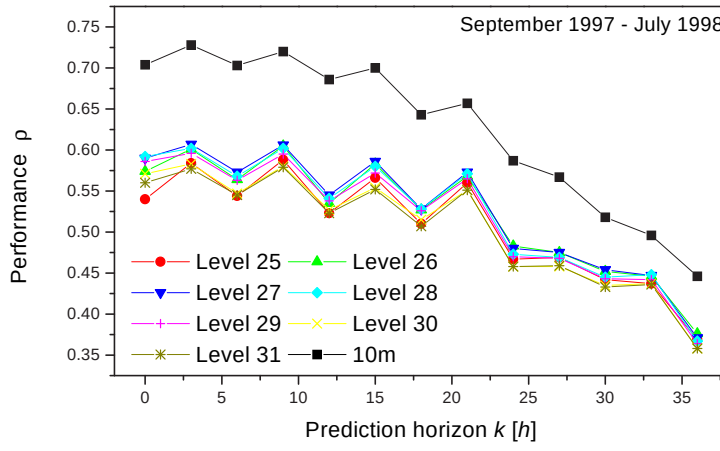


Figure 3: Wind speed prediction performance using the wind from various HIRLAM levels and for different prediction horizons. Data is from the September 1997 to July 1998

The systematic variation of the performance is described in (Joensen et al. 1999), which is related to the aliasing effect between the prediction horizon and the diurnal variation in the atmospheric turbulence intensity.

3 Influence of turbulence intensity

In (Landberg 1994) it was concluded that the turbulence intensity was not predicted accurately enough to be used in the stability dependent geostrophic drag law and the vertical wind profile. This is the reason for using the neutral versions of these relations in the Risø prediction system.

In this section we will take a closer look at these assumptions, and try to find out if some measure of the turbulence intensity can be used in some

modified relation. In (Sass et al. 1999) it is described how the stability dependency is handled by HIRLAM, and it is also noted that several stability schemes have been tested in HIRLAM. In the previous section it was found that the properties of the HIRLAM predictions changed on the 10th of September 1997, and the most likely reason for this change was that a new turbulence scheme had been introduced.

In general turbulence is generated by convective forces and wind shear (Stull 1988). The variables which are available for examining the turbulence influence from HIRLAM are therefore the sensible and latent heat fluxes and the wind speed.

The Buoyancy flux, B_s, will be used as a measure of the turbulence intensity. The Buoyancy is defined as (Stull 1988)

$$B_s = \beta \frac{H_s}{c_p \rho} + 0.608 \frac{g H_L}{L_c \rho}, \tag{2}$$

where $\beta = g/\theta$ is the buoyancy parameter, H_s is the sensible heat flux, c_p the heat capacity of air at constant pressure, ρ the density of the air, g the gravitational acceleration, H_L the latent heat flux, L_c the latent heat of vaporization, and θ is the potential temperature.

The buoyancy is negative for unstable and positive for stable conditions, due to the definition of the sensible and latent heat flux in HIRLAM. A way to verify how well the turbulence dependency used in HIRLAM fits reality is to examine how the difference between the observed wind speed and HIRLAM predicted wind speed depends on the Buoyancy flux.

The measurements which are used in this section are from the mast at Risø National Laboratory and the measurement height is 44 m agl. Figure 4 shows scatter plots of the difference between the HIRLAM predicted wind speed from various model levels and the measured wind speed from the Risø mast versus the Buoyancy flux. The 3 and 6 hour predictions have been used and the model levels which have been used are level 25, 31 and the special output wind corresponding to 10 m agl. The first column in the figure correspond to the period before HIRLAM changed and the second column corresponds to the following period. From the figure it is clearly seen that the turbulence scheme has changed, as the shape of the scatter in the two columns is very different.

The scatter in the upper left corner shows the difference between the HIRLAM predicted wind speed from model level 25 and the measured wind speed, using data from the first period. If the sensible and latent heat fluxes are predicted accurately by HIRLAM, then this difference should depend on the Buoyancy flux in the following way: when the buoyancy is negative the atmosphere is unstable and the difference should be small, as the turbulence intensity decreases the buoyancy goes from negative to positive values and the difference between the wind speed should increase. From the figure it is therefore seen that the qualitative shape of this scatter is as expected.

The next scatter in the first column is similar to the first one, except that now the predicted wind from level 31 is used. Level 31 corresponds on average to approximately 30–40 m agl, therefore it should be expected that there is little of no dependency on the buoyancy in this case, as the measurements are from 44 m agl. Actually it seems as it the influence of the turbulence is slightly overemphasized. In last scatter in the first column the special output wind from HIRLAM corresponding to 10 m agl is used, compared to the scatter where level 31 is used, it is seen that now the influence of the stability is even more overemphasized. In this case this is what should be expected, as the height which the predicted wind corresponds to is well below measurement height of the measured wind speed.

In the second period the scatter using the level 25 looks as expected, but, surprisingly, after HIRLAM has changed, there is a clear dependency on the Buoyancy flux for the difference between the wind speed from level 31 and the observed wind speed. As described above, these wind speeds correspond to roughly the same height above the ground, therefore this dependency should not be expected. The shape of the scatter suggests that the influence of the turbulence intensity is not taken into account; the shape of the scatter is very similar to the shape of the scatter for level 25.

Given the shape of the scatter using level 31 in the second period, it is not a surprise to see that there is no dependency on the stability when the 10 m wind is used. As the measurements are from 44 m agl, it should be expected that this scatter should be similar to the scatter in the first period. This is not the case, and therefore it actually seems as

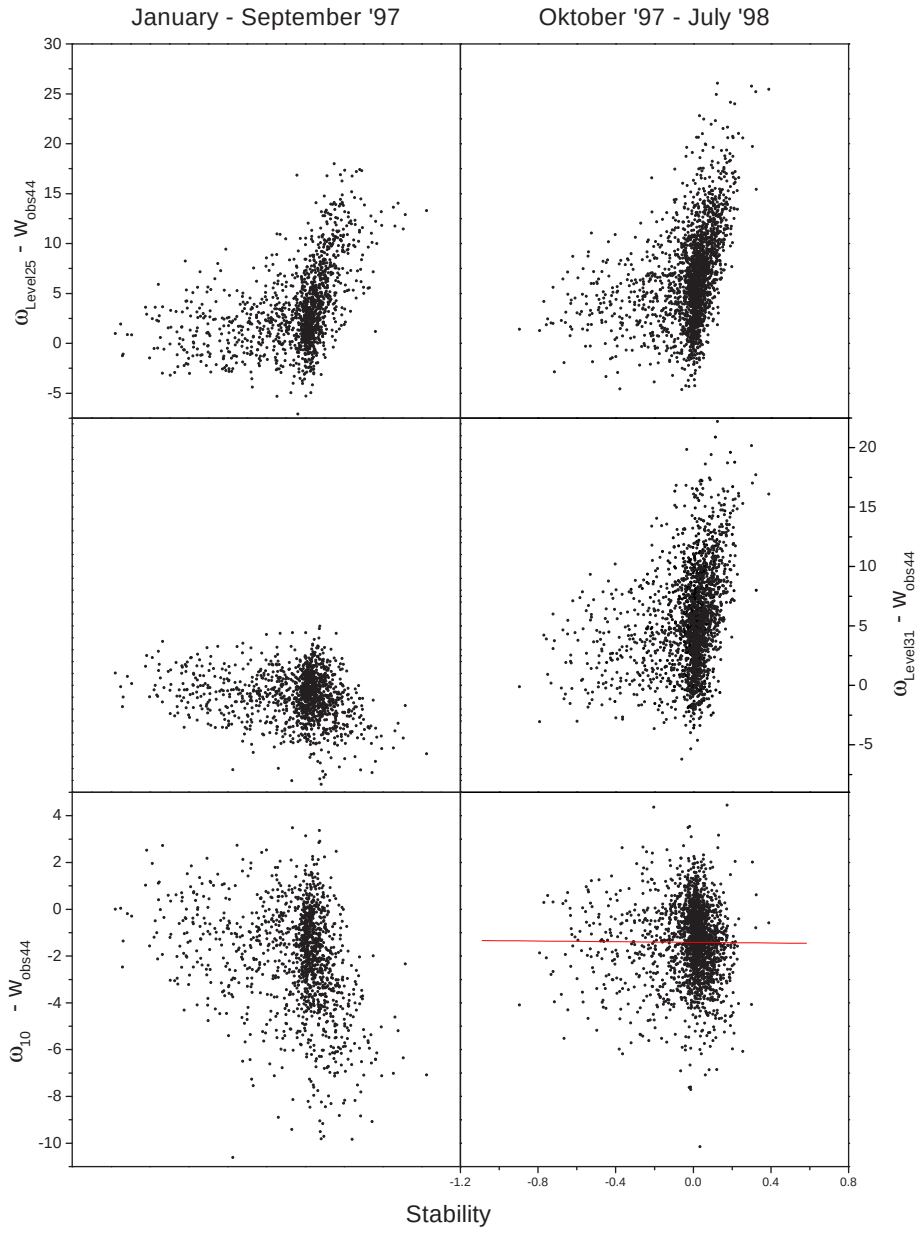


Figure 4: Scatter plots of the difference between the HIRLAM predicted wind speed and the observed wind speed at 44m agl versus the Buoyancy flux ($B_s \cdot 10^3$).

if HIRLAM performs worse in modelling the turbulence intensity after the change. Nevertheless, by using the 10 m wind from HIRLAM in the second period, the dependency on the stability seems to be taken into account.

As described in (Sass et al. 1999) special parameterizations are used in HIRLAM, which have been modified to fit the resolution of the model, and to retain numerical stability in the model integration. This means that the parameterizations used in HIRLAM might be constructed with emphasizes on the integration of the model, not necessarily on the accuracy or influence of the turbulence at a given height and specific location. In (Petersen, Mortensen, Landberg, Højstrup & Frank 1997) it is argued that numerical weather prediction models, where the flow is driven by the pressure gradient only, and the influence of turbulence and surface drag is neglected, perform remarkably well in predicting the overall atmospheric flow compared to models like e.g. HIRLAM. This is because the turbulence is difficult to predict, especially as the prediction horizon increases. The magnitude of the drag imposed on the flow from the surface is directly related to the turbulence intensity, and uncertainty in the turbulence intensity corresponds to random input to the integration of the model equations. Therefore, a more conservative guess of the turbulence intensity, like assuming neutral conditions or a mean diurnal variation, might give better performance for larger prediction horizons than a highly sophisticated turbulence scheme. This fact ought to be realized, and the physical parameterizations used in numerical weather prediction models should therefore also depend on the prediction horizon length.

The findings in this section clearly demonstrate that there is a problem related to the use of physical relations derived from idealized conditions on variables from a numerical weather prediction model, in this case HIRLAM. The reason is that the properties of the input variables to the physical relations do not fit the assumptions for which the relations have been derived, therefore these relations are simply not adequate. In (Jonsson 1994) a formal treatment of this problem is given in a regression context. These finding also support the results in (Joensen et al. 1999), where it was found that the performance of using the physical relations in the Risø system was worse than not using any of these relations at all.

4 Summary

It has been illustrated how the predicted wind speed from HIRLAM corresponds to actual measured values of wind speed at three different heights above the ground. It is found that the turbulence intensity dependency is not properly taken into account, although, the modeling of dependency has improved since the model has been updated.

References

Joensen, A. K., Giebel, G., Landberg, L., Madsen, H. & Nielsen, H. A. (1999), Model output statistics applied to wind power prediction, *in* 'In Wind Energy for the Next Millenium', Nice, France, pp. 1157–1161.

Jonsson, B. (1994), 'Prediction with a linear regression model and errors in a regressor', *International Journal of Forecasting* **10**, 549–555.

Landberg, L. (1994), Short-term prediction of local wind conditions, Technical Report Risø - R - 702 (EN), Department of Wind Power Meteorology, Risø National Laboratory, Roskilde, Denmark.

Landberg, L. & Joensen, A. (1998), 'A model to predict the output from wind farms – an update', proceedings from BWEA 20, British Wind Energy Conference. pp 127–132, Cardif, UK.

Machenhauer, B. (1989), 'The HIRLAM Final Report', Danish Meteorological Institute, Copenhagen, Denmark.

Petersen, E. L., Mortensen, N. G., Landberg, L., Højstrup, J. & Frank, H. P. (1997), Wind power meteorology, Technical Report Risø - I - 1206 (EN), Department of Wind Power Meteorology, Risø National Laboratory, Roskilde, Denmark.

Sass, B., Nielsen, N., Jørgensen, J. & Amstrup, B. (1999), The operational HIRLAM system at DMI, Technical report, Danish Meteorological Institute, Copenhagen, Denmark.

Stull, R. B. (1988), *An Introduction to boundary layer meteorology*, Kluwer Academic Publishers, Dordrecht, Boston, London.

Paper J

The economic value of acurate wind power forecasting to utilities

Originally published as

S. J. Watson, G. Giebel and A. K. Joensen. The Economic Value of Acurate Wind Power Forecasting to Utilities. In *Wind Energy for the Next Millenium*, British Wind Energy Conference, pages 1177–1180, Nice, France, March 1999.

The economic value of acurate wind power forecasting to utilities

Simon J Watson[1], Gregor Giebel[2] and Alfred Joensen[2]

Abstract

With increasing penetrations of wind power, the need for accurate forecasting is becoming ever more important. Wind power is by its very nature intermittent. For utility schedulers this presents its own problems particularly when the penetration of wind power capacity in a grid reaches a significant level (¿20%). However, using accurate forecasts of wind power at wind farm sites, schedulers are able to plan the operation of conventional power capacity to accommodate the fluctuating demands of consumers and wind farm output. The results of a study to assess the value of forecasting at several potential wind farm sites in the UK and in the US state of Iowa using the Reading University/Rutherford Appleton Laboratory National Grid Model (NGM) are presented. The results are assessed for different types of wind power forecasting, namely: persistence, optimised numerical weather prediction or perfect forecasting. In particular, it will shown how the NGM has been used to assess the value of numerical weather prediction forecasts from the Danish Meteorological Institute model, HIRLAM, and the US Nested Grid Model, which have been 'site tailored' by the use of the linearised flow model WASP and by various Model Output Statistics (MOS) and autoregressive techniques.

Keywords: Computer Programs, Forecasting Methods, Integration, Fossil Fuel Power Generation

[1]Building R63, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, UK, Tel: +44 1235 445559, Fax: +44 1235 446863, E-mail: sjwatson@easynet.co.uk

[2]Dept of Wind Energy and Atmospheric Physics, Risø National Laboratory, DK-4000 Roskilde, Denmark, Tel: +45 46 77 5095, Fax: +45 77 5970, E-mail: Gregor.Giebel@risoe.dk, Alfred.Joensen@risoe.dk

1 Introduction

When the penetration of wind power in a national electricity grid reaches a given level (¿20%) the fluctuating nature of wind farm output starts to become important (Bossanyi 1983). In certain areas of Denmark and Northern Germany, penetration levels are exceeding this figure and interest in the accurate forecasting of wind power is becoming more widespread. If it is possible to accurately forecast up to a day ahead the expected output from wind farms connected to a network, it is then possible to schedule and dispatch the conventional power plant more efficiently.

This paper shows a study of how numerical weather prediction (NWP) forecasting of wind power can aid central power unit dispatch and result in overall fossil fuel savings. The two study cases of the England and Wales national grid and the grid of the US State of Iowa are presented.

2 The National Grid Model

2.1 Outline

The Reading University/Rutherford Appleton Laboratory National Grid Model (NGM) (Bossanyi 1983) has been developed over a number of years to simulate the scheduling and dispatch of conventional power plant connected to a national electricity grid system and can also simulate the integration of renewable energy sources, namely wind and photovoltaic power plant. The NGM has been used to assess the value of forecasting at UK Meteorological Office sites (Watson, Landberg & Halliday 1994). In this paper, the model is used to assess the value of forecasting at real or potential wind farm sites.

The model works by tracking hour–by–hour the status of each power unit. At each hour, the start–up and shutdown of plant is planned up to a day ahead in order to meet the predicted hourly demand. This planning of start-up and shutdown of plant is dependent on the prediction of load and wind power (if wind power is integrated into the network) for each

hour ahead.

2.2 Power Plant

The model can simulate the operation of several different types of conventional plant namely:

• Nuclear - assumed to operate as inflexible base load

• Hydro - assumed to operate partly as base load and partly as fast response.

• Combined cycle gas turbine (CCGT) - modelled as having a three-hour start-up time from cold (idle for ¿120 hours), two hour start-up time if idle for ¿8 hours but ¡120 hours and a one hour start-up time if idle for less than eight hours.

• Coal/oil-fired steam turbine - modelled as having an eight-hour start-up from cold, but can be on a given number of hours warm standby if required in less than eight hours.

• Pump-storage - assumed to be available 'instantaneously' within the hourly time resolution of the model.

• Open cycle gas turbine (OCGT) - also assumed to be available 'instantaneously'. Plants are scheduled in a merit order, which depends on plant type and overall fuel efficiency. Nuclear plant is used first in the merit order followed by CCGT, oil/coal thermal, pump-storage and OCGT. Hydro plant is used both as base load and peaking plant in a similar manner to pump-storage.

2.3 Wind Power

If wind power is present in the system then it is accepted after nuclear plant in the merit order. This is subject to whether all the thermal plant is minimum part-loaded. If the supply is greater than the demand for

any given hour even if all the thermal plant is at minimum part-loading (50%) then excess wind power is discarded until this is no longer the case. This has to be done as the thermal plant cannot be switched off instantaneously, it has to be done within the one-hour time resolution of the model.

2.4 Spinning reserve

In general, for a national power system, there is only a limited amount of fast response plant available (generally pump-storage, hydro or OCGT). In the case of the England and Wales grid, this is mainly in the form of OCGTs, which are expensive to run. Fast response plant is therefore only used as a last resort at times of peak demand or when there is an unexpected surge in demand.

Alternatively, 'spinning reserve' can be scheduled in advance. The term spinning reserve corresponds to thermal power plant at part-load whose loading can be quickly increased to meet a surge in demand or, if there is a significant amount of wind power integrated into the system, an unexpected fall in generated wind power. The NGM therefore plans at each hour the amount of spinning reserve that is required up to a day ahead.

This spinning reserve is planned as a fixed fraction of the predicted load (called SR1) and a fixed fraction of the predicted wind power, if any wind energy is integrated into the system (called SR2). An extension to the spinning reserve algorithm can be used, whereby SR2 can also be specified as a fraction of the expected standard deviation of the forecast error divided by the predicted wind power. The expected standard deviation is assessed statistically off-line beforehand using historic data classified by key parameters such as time-of-day, season, direction, forecast lead-time, etc.

When the NGM is run it carries out a simulation for the period of interest, usually one year, with fixed values of SR1 and SR2. The model repeats each yearly run, optimising the values of SR1 and SR2 until the total fossil fuel cost is minimised subject to there being no loss-of-load events (power cuts) during the period of simulation. The load is much more

predictable than the wind power and so SR1 is generally far smaller than SR2.

2.5 Load data and load prediction

Load data is accepted by the model as an hourly or half-hourly time series. It is assumed that the standard deviation accuracy of load prediction is 1.5% for all hours up to a day ahead. The model is run such that the actual load is used as the forecast and the real load at dispatch, as far as the model is concerned, is the actual load multiplied by a gaussian uncertainty factor with mean 1 and standard deviation 0.015.

2.6 Wind power data and wind power prediction

Wind power data is accepted as one dataset of hourly wind power values. These values are a combination of the power output from whatever sites are being used as input to the grid. The wind power data can be real or 'simulated' (where wind speed at a site is converted to wind power using a suitable wind turbine power curve).

The NGM can accept wind power predictions in several formats:

• Persistence - The wind power at time x hours ahead $(t+x)$ is predicted to be same as it is at the present time, t.

• Numerical weather prediction model hybrid - Forecasts of the wind power are accepted from whatever hybrid forecast is available whether it be statistical, enhanced numerical weather prediction or a hybrid of the two. A forecast can be accepted either hourly or 12-hourly.

• Perfect - Wind power is assumed to be forecasted perfectly. The term 'numerical weather prediction model' refers to a computer model used typically by a national meteorological bureau to predict meteorological parameters. Such models output values on a regular grid. In this case, values from a regular grid are interpolated to the wind farm site and 'tailored' to the site using Model Output Statistics (MOS) (Glahn &

Lowry 1972), a wind flow model such as WASP (Mortensen et al. 1993), a time series autoregressive model (ELSAM (Ed.) 1995), or a hybrid of these, depending the data available.

3 The Grids Studied

3.1 England and Wales

The England and Wales 1994 plant mix is summarised in Table 1. There are interconnections with Scotland (2200MW) and France (1988MW). These are treated by the NGM as output from coal units. This is an approximation but allows the source of the power transmitted by the interconnections to be treated as plant in the merit order with a given start-up time. It can be seen that the grid is dominated by thermal plant with only a small fraction of fast response plant (pump-storage and OCGT).

3.2 Iowa

The US State of Iowa 1996 plant mix is also summarised in Table 1. There are interconnections to neighbouring states which are neglected by NGM. This is to simulate the situation where the Iowa grid is self-sufficient in terms of electricity. The Iowa grid is also dominated by thermal plant but there is a larger proportion of fast response plant (OCGT) than for the England and Wales grid.

Table 1. England & Wales (1994) (National Grid Company plc 1996) and Iowa (1996) plant mix (Mid-Continent Area power Pool 1997).

Plant	Capacity (MW)	
	England and Wales	Iowa
Nuclear	10642	475
Coal	31370	2286
Oil	5934	3102
CCGT	8915	107
OCGT	1559	1443
Pump-storage	2088	-
Total	60508	7413

4 Wind Farm Sites

4.1 United Kingdom

Wind speed data from eleven sites monitored as potential wind farm sites were used as supplied by the UK wind farm developer Renewable Energy Systems Ltd. The data from these sites, which were used for this study, covered the calendar year 1994. Wind speed forecast data were interpolated from the nearest grid points of the Danish Meteorological Institute's HIRLAM model (Machenhauer 1989) to the potential wind farm sites.

HIRLAM forecasts are made at 00Z and 12Z up to t+36 hours with a temporal resolution of 3 hours. The high-level geostrophic wind speed from HIRLAM was then transformed to the ground using the WASP model. Orography data, from UK Ordnance Survey 50m Panorama grid point files, and roughness data, from UK Ordnance survey 1:25000 Pathfinder maps, for each site were used as input to the WASP model. In addition, the HIRLAM forecasts were combined with historic observed values using an autoregressive time series model, called the Wind Power Prediction Tool (WPPT) (ELSAM (Ed.) 1995) to produce hourly forecasts. It was found that the WPPT forecasts gave more accurate predictions than

the HIRLAM/WASP forecasts and it is the results using these forecasts, which are reported here. These forecasts are henceforth referred to as the WPPT forecasts.

Wind speed was transformed to potential wind power using the power curve of the largest commercial wind turbine on the market at the time when this project was started, namely the Vestas V66 1650kW wind turbine.

4.2 The US State of Iowa

Wind speed data for two potential wind farm sites in the State of Iowa were made available by Wectec. Data for the calendar year 1996 were used in the NGM analysis. In this case, wind speed forecast data were interpolated from the nearest grid points of the US National Weather Service Nested Grid Model to the two potential wind farm sites. The Nested Grid Model forecasts are made at 00Z and 12Z up to $t + 48$ with a temporal resolution of 6 hours. As no orography and roughness data were available, interpolated geostrophic wind speed values were 'tailored' to the two sites using historic data and Model Output Statistics (MOS) as a function of direction and time-of-day. These forecasts are referred to as the NWS/MOS forecasts.

Wind speed was transformed to potential wind power using the power curve of the Vestas V66 1650kW turbine once again.

5 Results

5.1 Overview

For both the England and Wales grid and the grid covering the US State of Iowa, fossil fuel savings for a one-year simulation period were calculated for different penetrations of wind power and different forecasting methods. Fossil fuel savings are defined as the difference between the fossil fuel cost with no wind power and the fossil fuel cost for a given

penetration of wind power. The wind power penetration is adjusted by multiplying the Vestas V66 wind turbine output as calculated in Section 4 by an appropriate number of installed turbines at each site. For simplicity, the same number of turbines was assumed to be installed at each site.

5.2 England and Wales

Figure 1 shows the fossil fuel savings resulting from different penetrations of wind power for the England and Wales grid for 1994. The x-axis displays the total rated energy output as a fraction of the total energy demand for the year (273TWh), assuming that the wind turbines would be operating at maximum output throughout the entire year. The y-axis displays the fossil fuel saving relative to the total fossil fuel cost when no wind power is integrated into the grid. What is noticeable from Figure 1 is:

• The value of forecasting becomes significant when the rated output as a fraction of total demand reaches around 0.6, which corresponds to a penetration level of 25%.

• It is beneficial to use the standard deviation of the historic forecast errors in planning the spinning reserve (see Section 2.4). This can also be seen when using persistence forecasting at high penetration, where the rated output as a fraction of total demand reaches 1.3 (penetration level of 40%).

When the rated output as a fraction of total demand reaches 1.3 (penetration level of 40%), the fossil fuel savings using the WPPT forecasts and the standard deviation of the historic forecast errors to plan the spinning reserve are 26% higher than persistence. However, it should be noted that if persistence forecasting is used in conjunction with the standard deviation of the historic persistence forecast errors to plan the spinning reserve, the savings are still 13% better than persistence alone. NB Penetration is defined as the installed wind power capacity as a fraction of the total conventional and wind power plant.

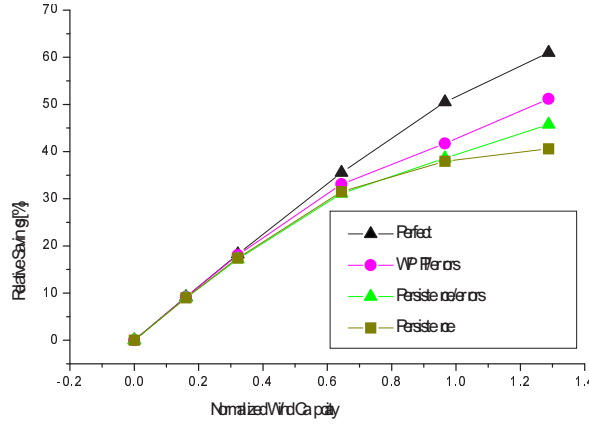


Figure 1: Fossil fuel savings for different forecasting methods for the England and Wales grid (1994)

5.3 Iowa

Figure 2 shows the equivalent fossil fuel savings for the Iowa grid comparing perfect, persistence and NWS/MOS forecasts. The total demand here was 28TWh.

It can be seen that:

• Perfect forecasting performs increasingly better than persistence as the penetration level increases.

• The NWS/MOS forecasts perform significantly worse than persistence at all penetrations.

The first point indicates that 'intelligent' forecasting is beneficial from the point of plant scheduling. However, the second point shows that the NWP forecasts produced every 12 hours at lead-times in increments of six hours are not good enough in comparison with persistence.

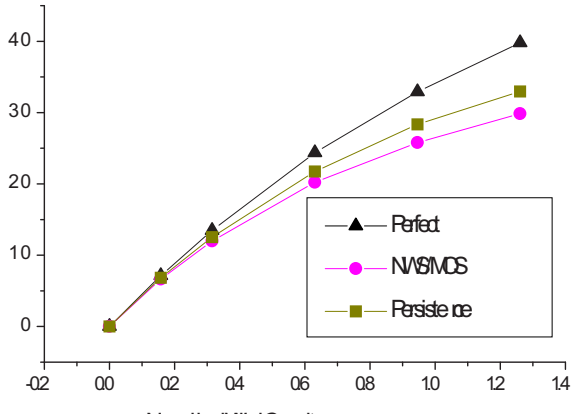


Figure 2: Fossil fuel savings for different forecasting methods for the Iowa grid (1996).

6 Discussion

The WPPT forecasts give greater fossil fuel savings than persistence in the case of the England and Wales grid. In the case of Iowa, the NWS/MOS forecasts perform worse than persistence. However, if one compares the forecasts with persistence as a function of forecast lead-time for the individual sites then it is found that the forecasts outperform persistence at lead-times of 3–4 hours and this is approximately the same for the sites in England and Wales and those in Iowa.

Although the NGM results for Iowa seem disappointing, it should be noted that the WPPT forecasts are optimised as far as possible for the UK sites giving hourly forecasts which are based on three-hourly HIRLAM forecasts, whereas the NWS/MOS are only available every six hours. This means that much of the time, quite 'old' forecasts are being used by the NGM in the case of the Iowa grid to plan plant scheduling and spinning reserve.

7 Conclusions

It has been shown that enhanced numerical weather prediction forecasts of wind power at potential wind farm sites can improve the efficiency of plant scheduling resulting in fossil fuel savings when compared with persistence. Indeed, for the England and Wales grid WPPT forecasts can result in a saving of up to 26% compared to persistence at a wind power penetration level of 40%. However, it has also been shown that the quality of the forecasts is quite critical. If thermal plant has a start-up time of eight hours then six-hourly forecasts updated every 12 hours do not show an improvement over persistence. The next step would be to optimise the NWS/MOS forecasts using a site 'tailored' WPPT model.

8 Acknowledgements

The authors would like to acknowledge the support of the Directorate General XII of the European Commission who are providing part funding for this project under the JOULE programme, contract JOR3-CT95-0008, 'Implementing Short-Term Prediction at Utilities' and also the Marie-Curie-Fellowship, JOR3-CT97-5004, 'European Wind Energy Capacity Effects'.

References

Bossanyi, E. A. (1983), 'Use of a grid simulation model for longer-term analysis of wind energy integration', *Wind Eng.* **7**, 233–246.

ELSAM (Ed.) (1995), 'Wind power prediction tool in central dispatch centres', Final report to the EC for contract JOU2-CT92-0083.

Glahn, H. R. & Lowry, D. A. (1972), 'The use of models output statistics (mos) in objective weather forecasting', *J. Appl. Met.* **11**, 1203–1211.

Machenhauer, B. (1989), 'The HIRLAM Final Report', Danish Meteorological Institute, Copenhagen, Denmark.

Mid-Continent Area power Pool (1997), '1996 coordinated bulk power supply/load and capability report', Published by the Mid-Continent Area power Pool, USA.

Mortensen, N. G., Landberg, L., Troen, I. & Petersen, E. L. (1993), 'Wind atlas analysis and application program (WAsP)', RisøNational Laboratory user guide: Risø-I-666(EN)(v.1) , Roskilde, Denmark.

National Grid Company plc (1996), '1996 seven year statment', Published by the National Grid Company plc, UK.

Watson, S. J., Landberg, L. & Halliday, J. A. (1994), 'Application of wind speed forecasting to the integration of wind energy into a large scale power system', *IEE Proc. C* **141**, 357–362.