

Online prediction and control in nonlinear stochastic systems

Torben Skov Nielsen

**Department of Mathematical Modelling
Technical University of Denmark
Ph.D. Thesis No. 84
Kgs. Lyngby 2002**

IMM

Preface

This thesis has been prepared at Informatics and Mathematical Modelling, the Technical University of Denmark, in partial fulfillment of the requirements for earning the degree of Ph.D. in Engineering.

Various aspects concerning on-line modelling and control of non-stationary and non-linear systems are considered. The thesis is centered around two applications of the considered methods:

- Heat load prediction and control of supply temperature in district heating systems.
- Prediction of power production from the wind turbines located in a larger area.

The thesis consists of a summary report and a collection of ten research papers written and submitted for publication during a period from 1997 to 2002. The report commences with a summary of the presented papers. Hereafter follows an introduction to the considered models and estimation methods and their application to prediction and control within district heating systems and for prediction of wind power. Here also the various research papers are brought into context and the results obtained are linked together.

Acknowledgements

I am grateful towards all the people, who have contributed to this thesis in one way or another.

First of all I would like to address my gratitude to my supervisors Prof. Lic. Techn. Henrik Madsen (IMM) and Prof., Techn. Dr. Jan Holst (Division of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden) for their help and guidance.

I am also grateful for the support – both financially and otherwise – I have received from the Energy Flexible Thermal Systems Program under the Nordic Energy Research. Within this research program many interesting and fruitful discussions have taken place at various meetings and workshops.

Furthermore I wish to thank Henrik Aalborg Nielsen and Alfred Karsten Joensen for pleasant and beneficial co-operation and many fruitful discussions. I will also thank my coworkers and the administrative staff at IMM for providing an inspiring research environment.

Last but not least I am grateful to my wife and family for their patience and support during the preparation of this thesis.

Kgs. Lyngby, July 2002

Torben Skov Nielsen

Papers summarized in the thesis

Paper A

Nielsen, H. Aa., Nielsen, T. S. & Madsen, H. (1998), 'Conditional parametric arx-models', *Journal of Time Series Analysis* . Submitted.

Paper B

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

Paper C

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

Paper D

Nielsen, T. S., Madsen, H., Holst, J. & Søgaaard, H. (2002), 'Predictive control of supply temperature in district heating systems', Technical Report IMM-REP-2002-23, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

Paper E

Nielsen, T. S. & Madsen, H. (2002), 'Control of Supply Temperature in District Heating Systems', in 'Proceedings of the 8th International Symposium on District heating and Cooling', Trondheim, Norway.

Paper F

Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L. & Giebel G. (1999), 'A New Reference for Wind Power Forecasting', *Wind energy* **1**, 29–34.

Paper G

Nielsen, T. S. & Madsen, H. (1997), 'Statistical methods for predicting wind power', *in* 'Proceedings of the European Wind Energy Conference', Irish Wind Energy Association, Dublin, Eire, pp. 755–758.

Paper H

Joensen, A., Madsen, H. & Nielsen, T. S. (1997), 'Non-parametric statistical methods for wind power prediction', *in* 'Proceedings of the European Wind Energy Conference', Irish Wind Energy Association, Dublin, Eire, pp. 788 – 792.

Paper I

Nielsen, T. S., Madsen, H. & Tøfting, J. (1999), 'Experiences with statistical methods for wind power prediction', *in* 'Proceedings of the European Wind Energy Conference', James & James (Science Publishers), Nice, France, pp. 1066 – 1069.

Paper J

Nielsen, T. S., Madsen, H. & Nielsen, H. Aa., (2002), 'Prediction of wind power using time-varying coefficient-functions', *in* 'Proceedings of the 15'th IFAC World Congress on Automatic Control', Barcelona, Spain.

Summary

The present thesis consists of a summary report and ten research papers. The subject of the thesis is on-line prediction and control of non-linear and non-stationary systems based on stochastic modelling.

The thesis consists of three parts where the first part deals with on-line estimation in linear as well as non-linear models and advances a class of non-linear models which are particularly useful in the context of on-line estimation. The second part considers various aspects of using predictive controllers in connection with control of supply temperature in district heating systems – a class of systems which are inherently non-stationary. The third part concerns the issue of predicting the power production from wind turbines in the presence of Numerical Weather Predictions (NWP) of selected climatical variables. Here the transformation through the wind turbines from (primarily) wind speed to power production give rise to non-linearities. Also the non-stationary characteristics of the weather systems are considered.

The summary report commences by considering some aspects of on-line estimation of linear as well as non-linear models for time-varying and/or non-stationary system. In the following chapters the presented papers are brought into their corresponding context with respect to optimal control of supply temperature in district heating systems and prediction of power production from wind turbines located in a given geographical area.

The papers A to C focus primarily on issues regarding modelling and estimation. Paper A considers off-line estimation of conditional parametric models and a new model class – Conditional Parametric Auto-Regressive-eXtraneous (CPARX) – is suggested. The models are estimated using local polynomial

regression – an estimation method with close resemblance to that of ordinary linear Least Squares (LS) regression. Furthermore it is shown how the relationship between supply temperature and network temperature in a district heating system can be described using CPARX models. In paper B a method for on-line or adaptive estimation of time-varying CPARX models is proposed. Essentially the method is a combination of Recursive Least Squares (RLS) with exponential forgetting and local polynomial regression. The paper also suggests a method for varying the forgetting factor in order to avoid flushing information from the model in seldomly visited regions of data. These methods are a prerequisite for employing the various conditional parametric models considered in the papers in on-line applications. Paper C considers on-line estimation of linear time-varying models with a partly known seasonal variation of the model parameters. An estimation method based on local polynomial regression in the dimension of time is suggested and it is indicated that the new method is superior to ordinary RLS, if the parameter variations are smooth.

Paper D presents two predictive controllers – eXtended Generalized Predictive Controller (XGPC) and a predictive controller derived using a physical relation and considers the various issues arising when the two controllers are applied in district heating systems with the purpose of controlling the supply temperature. The proposed controllers are implemented in a software system – PRESS – and installed at the district heating system of Høje Tåstrup in the Copenhagen area, where it is demonstrated that the system can indeed lower the supply temperature without sacrificing the safe operation of the system or consumer satisfaction. The PRESS control system is also the subject of paper E. Here the results obtained for a PRESS installation at the district heating utility of Roskilde is evaluated with respect to energy and monetary savings as well as security of supply.

The papers F to J consider prediction of wind power. Paper F proposes a new reference predictor as a supplement or replacement for the often used persistence predictor. It is shown in the paper, that it is not reasonable to use the persistence predictor for prediction horizons exceeding 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.

The papers G , H and J investigate models and methods for predicting wind power from a wind farm on basis of observations and numerical weather predictions. All three papers consider multi-step prediction models, but uses different estimation methods as well as different models for the diurnal variation of wind speed and the relationship between (primarily) wind speed and

wind power (the power curve). In paper G the model parameters are estimated using a RLS algorithm and any systematic time-variation of the model parameters is disregarded. Two different parameterizations of the power curve is considered – by a double exponential Gompertz model or by a Hammerstein model – and the diurnal variation of wind speed is explained directly in the prediction models using a first order Fourier expansion. In paper H the model parameters are assumed to exhibit a systematic time-variation and the model parameters are estimated using the algorithm proposed in paper C. The power curve and the diurnal variation of wind speed is estimated separately using the local polynomial regression procedure described in paper A . In paper J the parameters of the prediction model is assumed to be smooth functions of wind direction (and prediction horizon) and the functions are estimated recursively and adaptively using the algorithm proposed in paper B. As in paper G the diurnal variation of wind speed is taken into account directly in the prediction model using a first order Fourier expansion, whereas the power curve is estimated separately.

One of the prediction models considered in paper G – the model based on a Hammerstein parametrization of the power curve – is implemented in a software system – WPPT – and installed at the control centres of Elsam and Eltra, the power production and transmission utilities in the Jutland/Funen area, respectively. Predictions of wind power for the Jutland/Funen area are calculated by upscaling predictions from 14 wind farms in the area to cover the total production. Paper I describes WPPT as used by Eltra and Elsam and evaluates the predictions of wind power for the total area. Three cases are analyzed in order to illustrate, how the operators use the predictions and with which consequences. It is concluded that WPPT generally produces reliable predictions, which are used directly in the economic load dispatch and the day to day power trade.

Contents

Preface	iii
Acknowledgements	v
Papers summarized in the thesis	vii
Summary	ix
1 Introduction	1
2 Optimization and predictive control in district heating systems	3
2.1 On optimal operation of district heating systems	4
2.2 District heating systems with one supply point	11
2.3 An Application (PRESS)	14
3 Prediction of available wind power in a larger area	25

3.1	Motivation	25
3.2	Prediction of power production from wind farms	28
3.3	An Application (WPPT)	39
4	Conclusion	45
	Bibliography	47
	Papers	
A	Conditional parametric ARX-models	53
1	Introduction	55
2	Model and estimation	56
3	Simulation study	59
4	Application to a real system	65
5	Conclusions	68
6	Discussion	71
	Bibliography	71
B	Tracking time-varying coefficient-functions	75
1	Introduction	77
2	Conditional parametric models and local polynomial estimates	79
3	Adaptive estimation	80
4	Simulations	86

5	Further topics	92
6	Conclusion and discussion	93
A	Effective number of observations	94
	References	95
C	Tracking time-varying parameters with local regression	97
1	Introduction	99
2	The Varying-coefficient approach	100
3	Recursive Least Squares with Forgetting Factor	103
4	Simulation Study	105
5	Summary	108
	References	108
D	Predictive control of supply temperature in district heating systems	111
1	Introduction	113
2	Optimal operation of district heating systems	115
3	Extended generalized predictive control	118
4	Reference curves in predictive control	132
5	The implemented models	136
6	The Høje Tåstrup district heating utility	137
7	Control of the flow rate	138
8	Control of net-point temperature	146

9	Results and discussion	166
	References	170
E	Control of supply temperature in district heating systems	173
1	Introduction	175
2	Control problem	176
3	Controller implementation	178
4	Results obtained in Roskilde	186
5	Conclusion	191
	References	192
F	A New Reference for Wind Power Forecasting	193
1	Introduction	195
2	The new reference forecast model	197
3	Examples	197
4	Summary	201
A	The Mean Square Error (<i>MSE</i>)	201
	References	202
G	Statistical Methods for Predicting Wind Power	205
1	Introduction	207
2	Data	208
3	Model estimation	209

4	Models for Predicting Wind Power	212
5	Conclusion	216
6	Acknowledgements	216
	References	216
H Non-parametric Statistical Methods for Wind Power Prediction		219
1	Introduction	221
2	Data Analysis	222
3	The Model	228
	References	230
I Experiences With Statistical Methods for Predicting Wind Power		231
1	Introduction	233
2	Implementation	234
3	The prediction model	239
4	Utility experiences	240
5	Conclusion	243
6	Acknowledgements	244
	References	244
J Prediction of wind power using time-varying coefficient-functions		245
1	Introduction	247

2	Model and estimation method	249
3	Adaptive estimation	251
4	Wind power prediction models	254
5	The prediction performance	255
6	Summary	257
	References	260

Introduction

The subject of the present thesis is model-based on-line prediction and control of non-linear systems considered in a stochastic framework. On-line estimation in linear as well as non-linear models is considered and a particular useful class of models – the conditional parametric models – is put forward. Two applications are presented: control of supply temperature in district heating system and prediction of power production from wind turbines in a given geographical area and it is demonstrated how the proposed models and estimation methods can be successfully applied within these areas.

The thesis consists of a summary report and ten research papers. In the summary report as well as the remainder of the thesis the papers are referred to according to the naming convention introduced on page vii.

The three first papers focus primarily on issues regarding modelling and estimation. Paper A suggests a new model class – Conditional Parametric Auto-Regressive-eXtraneous (CPARX) – and it is shown how the relationship between supply temperature and network temperature in a district heating system can be described using CPARX models. In paper B methods for on-line estimation of time-varying CPARX models are proposed and paper C considers on-line estimation of linear time-varying models with a partly known variation of the model parameters. Paper D presents two predictive controllers – the eXtended Generalized Predictive Controller (XGPC) and a predictive controller derived from a physical relation and considers the various

issues arising when the two controllers are applied in district heating systems with the purpose of controlling the supply temperature. The supply temperature controller has been implemented in an on-line application – PRESS – and paper E considers the results obtained for an installation of PRESS. The final five papers consider prediction of wind power. Paper F proposes a new reference predictor as a supplement or replacement for the often used persistent predictor. Paper G , H and J investigate models and methods for predicting wind power from a wind farm and finally paper I considers the prediction of wind power for a larger area and presents results obtained by an on-line application – WPPT.

The subjects of optimal control of supply temperature in district heating systems and prediction of power production from wind turbines located in a larger area are discussed in Chapter 2 and 3, respectively. The chapters are completed by presenting two on-line applications of the proposed methods. Finally the thesis is concluded in Chapter 4.

Optimization and predictive control in district heating systems

Traditionally district heating plays an important role in covering the heat demand in the Nordic countries. To illustrate the importance of district heating it can be mentioned that in Denmark more than half of the domestic heating installations are supplied by district heating. Optimal operation of district heating systems has thus been of increasing interest to researchers and practitioners alike over the last decade or so. This subject is by no means trivial though, as district heating systems are inherently non-linear and non-stationary, and the issue is further complicated by the fact, that district heating systems are very diverse with respect to production facilities, operational requirements and so forth. Section 2.1 presents the issue of optimal operation of district heating systems focusing on Danish conditions and a solution to a general setup with more than one production unit and a complex optimization criterion is outlined. Section 2.2 describes a solution tailored to a specific scenario – namely district heating systems consisting of one production unit for which optimal operation is achieved by minimizing the supply temperature to the distribution network. Here also the related papers of this thesis are presented. Finally a software package – PRESS – developed at Informatics and Mathematical Modelling (IMM) to on-line optimization of the aforementioned specific class of district heating systems is

described in Section 2.3.

2.1 On optimal operation of district heating systems

This section gives a brief introduction to optimization of the operation of district heating systems. The description will focus on how district heating is applied in Denmark, and it should be noted, that the typical operational setup for a district heating system varies considerably between countries. In (Andersen & Brydov 1987) district heating is defined as follows:

District heating may be defined as space and water heating for a number of buildings from a central plant. The heat produced in this plant is delivered to the consumers as hot water through an insulated double pipeline system. The heated water is carried in the forward pipe distribution system and having given up its heat, the cooler water returns to the plant in the other pipe for re-heating.

Thus a district heating system can be seen as consisting of three primary parts: one or more central heat producing units, a distribution network and finally the consumer installations for space heating and hot tap water production.

2.1.1 System restrictions

The objectives of the present section is to identify the main conditions under which an optimization of a district heating system is carried out. Optimal operation of the district heating system is here assumed to be achieved by minimizing the productions cost to the extent, that it can be achieved without compromising the safe operation of the system, adversely affecting the maintenance cost of the system, or sacrificing consumer satisfaction.

In most district heating systems the distribution network and consumer requirements will impose some or all of the following restrictions on the optimization :

- *A maximum allowable flow rate in the system.* The restrictions in the flow rate are due to the (always) limited pumping capacity, the risk of

cavitation in heat exchangers and difficulties maintaining a sufficiently high differential pressure in the remote parts of the network during periods with high flow rates.

- *A minimum guaranteed inlet temperature at the consumers.* This restriction is due to limitations in the consumer installations as well as minimum hot tap water temperature requirements imposed by hygienic concerns.
- *A maximum allowable supply temperature.* This restriction is put on the systems in order not to damage pipelines and consumer installations.
- *Short term variations in supply temperature.* The stresses inflicted on the network by large frequent fluctuations in the supply temperature dictate, that the short term variations in supply temperature should be limited.
- *Maximum allowable diurnal variations of the supply temperature.* In some systems the size of the expansion tanks may impose limitations on the allowable diurnal variation of the supply temperature.

The vast majority of the heat production (more than 75% in 1998) is generated from three sources: Central Combined Heat and Power (CHP) plants fueled by coal or natural gas, decentralized CHP plants using natural gas as fuel and finally pure heating plants burning waste or to a minor extent renewables such as straw and wood chipings. The remaining part of the heat production is produced by peak load and stand-by boilers in the district heating systems as well as by private sources such as process heat from the industry.

Depending on the type of heating plant different factors influence the production costs per unit energy just as there are differences in the restrictions imposed on the operation/optimization of the district heating system by the plants. Among others the following factors affecting the economics of the heat producing units has to be considered:

- *Supply temperature.* Central CHP plants will normally use steam turbines to power the electric generators. The district heating water is heated by acting as a coolant in the condenser and by steam extracted from the turbine. An increase in supply temperature (by extracting steam at a higher temperature) implies a decrease in the output of electric power (per unit of fuel). Hence the supply temperature should be kept low as electric power is more valuable than heat. Decentralized CHP plants typically use gas engines (combustion engines or turbines)

to power the electric generators. Here the district heating water is heated by acting as an engine coolant and through heat exchangers in the exhaust system of the engines. Thus similar to a traditional boiler plant the running costs by and large are independent of the supply temperature.

- *Demand for electric power.* The demand for electrical power displays a diurnal variation. The revenue from the electricity produced from the decentralized CHP units follows a similar variation in order to encourage power production during periods with peak load. For working days the revenue is low during the night and high (two levels) during the day. The demand for heat and power does not necessarily coincide and thus many CHP plants have heat storage tanks where surplus heat can be stored during periods where the heat production exceeds the heat load.

Similarly a number of factors affecting the operational costs for the distribution network is readily identified:

- *Heat loss from the network.* The heat loss in the network is a (complex) function of the supply temperature. Decreasing the supply temperature implies a lower temperature in the network in general and consequently a decrease in the heat loss from the network.
- *Pumping costs.* For most district heating utilities in Denmark the pumping costs are an order of magnitude less than the energy costs associated with the heat loss in the distribution network – hence pumping costs are left out of the optimization.
- *Maintenance costs for the network.* The operation of a district heating utility has a direct impact on the maintenance costs for the network. Large variations in supply temperature (and pressure) will increase the maintenance costs compared to a more steady operation.
- *Return temperature and peak load.* For district heating systems supplied through a common transmission system as eg. the VEKS system in the eastern part of Zealand the energy cost may be affected by return temperatures and peak loads. High peak loads and return temperatures taxes the energy transfer capacity of the common transmission system and are thus penalized by higher energy costs.

2.1.2 Optimization at plant level

The problem now consists of determining how the future heat load is allocated between the different heating plants in order to minimize the operational costs for the entire district heating system consisting of several heating plants and a distribution network. The objective is to minimize the expected operation costs within the planning horizon considered given an (uncertain) predicted heat load. The planning horizon will depend on the configuration of heating plants for the district heating system in question but for system with large CHP plants and/or heat accumulators the necessary planning horizon will be in the magnitude of days.

The task of identifying a cost function for the operation of an entire district heating system with multiple heating plants and following that finding a feasible solution to the posed minimization problem will in most cases be very difficult due to the size of the problem which is implied by the discrete nature of a start/stop schedule, the long planning horizon, the number of restrictions imposed on a solution for the entire system and finally the complexity of the models describing the distribution network. The problem is subject to continued research though. In (Arvastson 2001) an optimization scheme based on a fairly detailed physical models of the heat production units and the distribution network is proposed, where the unit commitment problem is solved by a fuzzy logic approach. Under certain simplifications a pseudo optimal control strategy is derived for the district heating system of Malmö, but as the simplifications include disregarding power production, heat accumulators and flow restrictions the control strategy is not considered operational.

In order to make the solution of the optimization problem feasible it is suggested to separate the optimization of the entire system into a scheduling between the different heat (and power) producing units including possible heat accumulators (long planning horizon) followed by a control problem for the distribution network (considerably shorter control horizon). The potential gains by optimal scheduling between several production units will typically eclipse the potential gains by optimal operation of the distribution network by a considerable margin. Hence it makes sense to let the operation of the distribution network be subordinate to the scheduling even at the cost of a (slightly) sup-optimal solution compared to an optimization which encompass the entire district heating system.

The purpose of the scheduling is to derive a plan for each heating plant stating when the plant should be running as well as the heat and power production levels. Depending on the configuration of production facilities the scheduling

horizon will typically be 1 to 3 days ahead. The scheduling between the different production units is done on basis of the following input:

- Predictions of heat load including estimated uncertainty covering the horizon considered in the scheduling.
- Predictions including estimated uncertainty covering the scheduling horizon of the necessary minimum supply temperature in order to fulfill the consumer requirements.
- The future sales price for power. At present these prices are known in advance for the decentralized CHP plants, whereas part of the power production from the centralized CHP plants are traded on the NordPool (see Section 3.1) at market rate.
- The heat and power production costs for the different productions units. These may vary with time, production level and fuel. Also the start/stop costs has to be considered.
- Limitations in the available heat and power production capacity. The heating plants may be subject to contractual obligations, which may restrict the minimum or maximum heat production. These restrictions may be time-varying. Also hydraulic limitations at the supply points as well as limitations in the allowable rate of change for heat and power production has to be considered.
- The possibilities for redistribution of heat load using the distribution network as heat storage. An approximate measure can be derived from the water volume in the forward pipe distribution system and the maximum allowable rate of change for the supply temperature.
- The charge level and temperature of any heat accumulators present in the system.

In situations where the scheduling horizon is less than one day ahead heat load can be predicted with reasonable accuracy using models based purely on observations as eg. the multi-step ARX (Auto-Regressive-eXtraneous) models described in paper D or more thoroughly in (Madsen, Palsson, Sejling & Søggaard 1990), but for longer horizons the future weather situation has to be taken into account by including numerical weather predictions (NWP) from a weather service explicitly in the model. A 39 hour heat load prediction model utilizing forecasts of ambient air temperature are currently used in PRESS. The model structure for the implemented model is derived from the

observation based model presented in (Madsen et al. 1990) and a model study in (Nielsen & Madsen 2000). The subject is further investigated in (Arvastson 2001) using a more physical orientated grey-box approach.

The required future supply temperature has to be determined for each supply point in the distribution network. One possible solution is to use meteorological forecasts of ambient temperature in combination with a control curve similar to the one used in the central control method, cf. Section 2.2. A further refinement is to establish the required consumer inlet temperature at the most critical areas of the distribution network and then use models describing the relationship between network temperature and supply temperature. This approach is closely related to the net-point temperature control problem described in Section 2.2 and in paper D.

The stochastic nature of the heat load and supply temperature predictions should be taken into account by the scheduling algorithm. Hence the scheduling should be formulated as a stochastic optimization problem, where the correlations structure of the prediction errors is included in the formulation. Possible methods to solve such a problem include stochastic dynamic programming, fuzzy logic (Arvastson 2001) and bootstrapping (Nielsen, Nielsen & Madsen 2001).

2.1.3 Optimization at distribution level

The outcome of the scheduling is a plan for the various heating plants covering the scheduling horizon as described on page 7. Only the first part of the plan corresponding to the horizon considered by the supply temperature controller for the distribution network is used. For each of the supply points the schedule is converted to a set of (time-varying) constraints and reference values used as input to the distribution network controller:

- Maximum values for the permissible supply temperature. The maximum restriction corresponds to the minimum supply temperature constraint used in the scheduling.
- The desired flow rate.
- The desired redistribution of heat load.

The optimization at the distribution level is subject to the restrictions described at page 4 – mainly restrictions related to the maximum allowable flow

rate at the supply points and consumer requirements to inlet temperature. The consumer requirements to inlet temperature are satisfied for the entire distribution system if they are satisfied in the areas which suffer the largest heat loss – hereafter referred to as “critical net-points”. Depending on the layout of the distribution network the low level pressure control in the network requires, that the flow rate is allowed to vary freely at one or more of the supply points. Only the flow rate from the supply points, which is used to control the pressure in the distribution network needs to be monitored by the supply temperature controller.

The sub-optimality of the proposed optimization scheme is caused by the supply temperature controller shifting heat load from supply points with fixed flow rate to supply points with freely varying flow rate by lowering the supply temperature, hence the realized distribution of heat load will differ from the distribution expected by the plant scheduling. The proposed solution will be close to optimum as long as the deviation between effectuated supply temperature and the minimum requirements used by the plant scheduling is reasonable, though.

The supply temperature controller requires that models describing the dynamic relationship between supply temperature(s) and respectively flow rate(s) and critical net-points temperatures are identified. One possible solution could be to extend the ARX and CPARX models presented in the papers A and D to handle multiple supply points. For more complex systems this type of models will be difficult to identify reliably as many of the input variables will exhibit little variation – eg. flow rates at supply points with fixed flow rate – or be closely correlated – eg. the supply temperatures. A different approach is to use detailed deterministic models to describe the distribution network. The computational requirements for this type of models prevents direct use in numerical optimization algorithms, but deterministic network models can serve as a mean for deriving a linearization of the system response around the current operation point by imposing the model an impulse input signal for each input variable, cf. paper D. The optimization is then carried out using the linearized model. The output error for the linearized model must be expected to be biased and coloured. Hence it will be beneficial to add an output error model to the setup, see eg. (Arvastson 2001).

One approach to implementing the supply temperature controller is to pose the problem as a constrained minimization problem with receding horizon. This formulation has some resemblance to Model Predictive Control (MPC), cf. (Mayne, Rawlings, Rao & Scokaert 2000), but with a non-quadratic cost function which makes it a more difficult problem to optimize.

A second approach is possible under the assumptions that the cost function decreases monotonously with decreasing levels of supply temperature, cf. paper E, and that each of the constrained output variables is controlled by only one supply temperature for a given time period. The optimization problem is here reduced to a set of independent minimization problems, one for each supply point. The individual problems are minimized by finding the minimum future supply temperature which observes the constraints imposed on the optimization, ie. the restrictions in flow rates and net-point temperatures. This problem is closely related to optimization problem for distribution networks with a single supply point described later in Section 2.2 and in the papers D and E.

In either case the distribution network can be used as a heat storage by varying a non-negative¹ signal which is superimposed upon the required supply temperature as determined by the controller.

2.2 District heating systems with one supply point

Traditionally the supply temperature in a district heating systems is controlled either manually by operators guided by experience or by the central control method (see (Oliker 1980)), where the supply temperature to the distribution network most frequently is determined as a function of the current ambient air temperature, possibly corrected for the current wind speed. This means, that the supply temperature control is in fact a open loop control without any feedback from the distribution network and consequently, the control curve has to be determined conservatively to ensure a sufficiently high temperature in the district heating network at all times.

The objectives of the present section is to derive a control scheme for optimal operation of a certain class of district heating systems; namely district heating systems which primarily are supplied by a single district heating station – thus scheduling problems between heat producing plant with different production costs or heat accumulators will not be considered here. Any rescheduling of the heat load in such a system will have to rely on heat storage in the distribution network. Under the assumption that the diurnal peak load and the return temperature are not adversely affected by the optimization, the production cost can then be minimized by minimizing the supply temperature, to the extent it can be achieved without compromising the safe operation

¹Non-negative in order not to compromise the restrictions.

of the system, adversely affecting the maintenance cost of the system, or sacrificing consumer goods.

2.2.1 Control strategy

The proposed control scheme for the supply temperature has two primary objectives. First of all it optimizes the operation of the district heating utility with respect to production costs. Secondly it brings the supply temperature control into a closed loop context thereby making the supply temperature control a more objective matter compared to the traditional ad hoc approach.

As described in Section 2.1.1 safe operation of the (distribution) system is tantamount to keeping the flow rate ex heating station below a certain critical limit, whereas the consumer requirements are satisfied by maintaining a sufficiently high (overall) temperature and differential pressure in the distribution network. A sufficient differential pressure in the network is ensured by keeping the flow rate below a certain limit. Paper D describes an approach to minimizing the supply temperature under the limitations, that restrictions in flow rate ex heating station and consumer inlet temperature are observed.

The optimization strategy is implemented as a set of controllers, which operates the system as close to the minimum supply temperature as possible without actually violating the restrictions. At a given time the supply temperature recommended is then selected as the maximum of the recommended supply temperatures from the individual controllers. The flow rate is monitored by a single controller whereas the consumer inlet temperature is monitored by introducing a set of *critical points* in the distribution network. The critical points are selected so that if the temperature requirements for the critical points are satisfied then the temperature requirements for all consumers are satisfied. The locations of the critical points can be determined using a deterministic network modelling tool to identify the regions in the network with the largest temperature losses under different operational conditions. Also the experience gained by the operators of the system can often be used to identify regions in the network with large temperature losses.

2.2.2 Flow rate and net-point temperature sub-controllers

A district heating system is difficult to control (optimally) as the dynamic relationship between supply temperature (the control variable) and key pa-

rameters such as network temperature and flow rate are time-varying and difficult to establish, which in both cases can be attributed to the time-varying heat load in the system. Thus the problem of controlling a district heating system calls for new control methods, which will operate reliably under these circumstances. In paper D the energy relation at the supply point

$$p_t = c_w q_t (T_{s,t} - T_{r,t}) \quad , \quad (2.1)$$

is used to convert predictions of heat load to a set point for the supply temperature. In (2.1) c_w is the specific heat of water, p_t is the heat load at time t , q_t is the flow rate and $T_{s,t}$, $T_{r,t}$ are supply and return temperature, respectively.

The temperatures at each of the critical net-points are monitored by individual controllers and thus a model describing the dynamic relationship between supply temperature and the individual net-point temperatures is needed. In paper D this relationship is described by estimating a number of models in parallel covering the range of possible time delays between supply point and net-point using an ARX model structure

$$\begin{aligned} T_t^{np} &= a_1 T_{t-1}^{np} + \sum_{i=0}^2 b_i(t - \tau - i) T_{t-\tau-i}^s + e_t \\ b_i(t) &= b_{i,1} + b_{i,2} \sin \frac{2\pi t}{24} + b_{i,3} \cos \frac{2\pi t}{24} \quad , \end{aligned} \quad (2.2)$$

where the $b_{0..2}(t)$ parameters have an embedded diurnal time variation in order to incorporate the diurnal variation of the heat load. In (2.2) T_t^{np} is the net-point temperature, T_t^s is the supply temperature, e_t is a noise sequence and τ is the time-delay of the system. The model parameters a_1 , $b_{0..2,1..3}$ are estimated adaptively using the Recursive Least Squares with exponential forgetting algorithm by Ljung & Söderström (1983).

The time delay, τ , in (2.2) can not be measured directly, but has to be estimated separately. Paper D proposes a method for identifying the time delay by using an estimate for the correlation between supply and net-point temperatures. A number of alternative procedures are considered in (Madsen, Palsson, Sejling & Sogaard 1992), but experience has shown that the method advanced in Paper D performs well.

In paper A the Conditionally Parametric ARX (CPARX) model class is proposed and it is shown that CPARX models can be used to model the temperatures in district heating network. The following model is suggested

$$T_t^{np} = a(q_t) T_{t-1}^{np} + \sum_{i=\tau_{min}}^{\tau_{max}} b_i(q_t) T_{t-i}^s + e_t \quad , \quad (2.3)$$

where q_t is the flow rate, τ_{min} , τ_{max} are minimum and maximum values for the time delay, respectively, and $a(q_t)$, $b_{\tau_{min}.. \tau_{max}}(q_t)$ are coefficient functions, which have to be estimated. Note that the time-varying time delay is incorporated in the model by making the coefficients a function of the flow rate. In paper A the coefficient functions are estimated using an off-line algorithm closely related to locally weighted regression as proposed by (Cleveland & Devlin 1988). In on-line applications it is advantageous to allow the function estimates to be updated as data becomes available and in paper B a method for recursive and adaptive estimation of the coefficient functions is proposed.

When considering the use of (2.3) for predicting the future net-point temperatures it should be noted that, given predictions of the future flow rates, the model translates into a (simple) ARX model. Hence model predictive control algorithms employing a model as (2.2) can equally well be based on a model given by (2.3).

The model (2.2) gives raise to a number of requirements on the net-point temperature controller. It must be robust toward non-minimum phase system (due to the possibility of wrongly specified time delays in the model) as well as being capable of handling time-varying systems. The controller should also be reasonably easy and robust to derive since the controller parameters are likely to change hourly as the model parameters are updated. The net-point temperature controller used in the paper D and is based on the Extended Generalized Predictive Controller (XGPC) proposed in (Palsson, Madsen & Sogaard 1994), which is a further development of the Generalized Predictive Controller (GPC) presented by Clarke, Mohtadi & Tuffs (1987). The main difference between the XGPC and GPC algorithms is found in the derivation of the control law. The XGPC uses conditional expectation to separate model output into a term with a linear dependency on future input values (control values) and a term depending on past input and output values, where a similar separation for the GPC is achieved by recursively solving a Diophantine equation. Furthermore the formulation of the GPC depends on a specific model structure (ARIMAX) whereas the only requirement on the model structure posed by the XGPC is that the future model output is separable as described above.

2.3 An Application (PRESS)

This section describes PRESS (*Danish: PRognose og EnergiStyringsSystem*) – an on-line system for prediction, control and optimization in district heating

systems.

Previous incarnations of PRESS have been installed at the district heating utilities of Esbjerg/Varde, Sønderborg and Høje Tåstrup. Currently PRESS is installed at Sønderborg Fjernvarme and Frederiksberg Varmeværk in a version without numerical weather predictions as input and without the supply temperature controller and at Roskilde Varmeforsyning in the full version with supply temperature control and numerical weather predictions. PRESS has been used operational at Roskilde Varmeforsyning since the beginning of January 2001 and the results are evaluated in Paper E with respect to energy and monetary savings as well as security of supply. The supply temperature controller were also part of the PRESS installation at Høje Tåstrup, where PRESS were used operationally during a period from autumn 1995 to spring 1997. The experiences gained from the Høje Tåstrup installation of PRESS are described in Paper D.

The PRESS system is implemented as two fairly independent parts – a numerical part and a presentation part – in the following denoted PRESS-N and PRESS-P, respectively. Data exchange between the two subsystems is implemented via a set of files. A file based interface between PRESS-N and PRESS-P has been chosen for portability reasons. Currently PRESS have been successfully tested on HP Unix systems as well as PC systems running Linux. For each installation one instance of PRESS-N is meant to be running continuously, whereas a number of presentations can run simultaneously, ie. PRESS can be used by more than one user at a time.

The calculation and estimation part of PRESS consists of several numerical modules. An overview of PRESS-N as well as a brief description of the models and numerical methods applied is found in Section 2.3.1. Where appropriate this section contains further references for a more thorough discussion of the various models and methods applied in PRESS-N.

The user interface provides an overview of the current state of the district heating system as well as access to more detailed informations through a wide range of plots. PRESS-P furthermore enables the operator to interact with the control part of PRESS by providing access to a limited set of the control parameters. PRESS-P is described in more detail in Section 2.3.2.

2.3.1 The Numerical Part of PRESS (PRESS-N)

PRESS-N is implemented in ANSI C and is designed to be straight forward to port to a new platform for which an ANSI C compliant compiler is available.

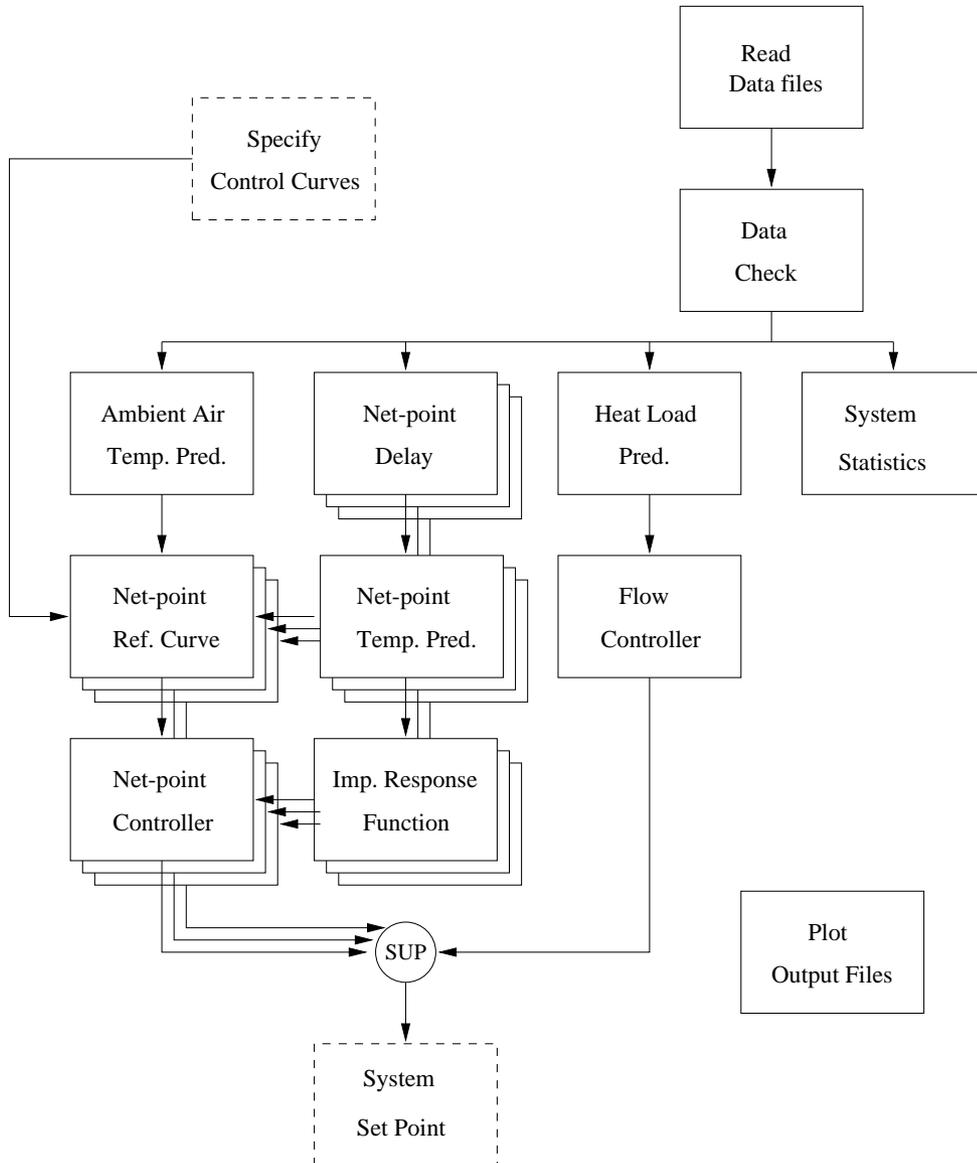


Figure 2.1: Overview of the PRESS system (numerical part).

The complete PRESS system is, if all tasks and data handling routines are considered, quite complex. From an overall point of view the system consists

of four primary parts. First of all one part is dedicated to providing the rest of the system with screened data, where errors have been detected and, if possible, corrected. The calculation of a heat load predictions and system statistics are handled by two additional packages, and finally we have the control part of PRESS-N. An overview of the different tasks and primary data flows maintained by PRESS-N is shown in Figure 2.1. From the figure it is seen that some of the above mention primary parts of the PRESS system have been divided into smaller modules. This is in particular true for the control part, which by far is the largest and most complex part of PRESS-N. In the following each module is described and where appropriate further references are made.

- *Read Data Files.* Every 5 minutes the most recent observations are read from a data file generated by the SCADA² system in the main computer at the district heating utility. If for some reason the communication between the SCADA system and PRESS-N fails, the observations are marked as being missing in PRESS-N.
- *Data Check.* The 5 minute values are subject to some simple data consistency checks. These include high and low limit checks as well as checks to establish whether a measurement are hung up on a fixed value. If a check rejects a value the observation is marked as been unavailable. Every hour the 5 minute values are integrated into hourly values before a confidence check is carried out for the hourly values for which a prediction is available (heat load, net-point temperatures and ambient air temperature). If the check rejects the observed value it is replaced by the prediction. All subsequent tasks are performed on the hourly values.
- *Net-point Delay.* For each of the net-points the time delay between the supply point and the net-point in question is estimated. The estimate is slowly varying and reflects the diurnal average of the time delay. The estimation procedure is described in paper D.
- *Net-point Temperature Prediction.* The net-point models describe the relationship between supply temperature and net-point temperature using transfer function models with an embedded diurnal time-variation in the model parameters. For each of the net-points a number of models are running in parallel covering the range of possible time delays. At a given time only the models with a time delay within a range of ± 1 hour from the previously estimated time delay is updated. The net-point temperature models are described in more detail in paper D.

²Supervisory Control And Data Acquisition.

- *Impulse Response Function.* Based on the estimated net-point model corresponding to the actual time delay the impulse response function is calculated. As the parameters in the net-point models have a diurnal variation so do also the impulse response function.
- *Ambient Air Temperature Prediction.* The reference net-point temperature used in the net-point temperature control is depending on the ambient air temperature. The time delays in the distribution network implies, that a forecast of the ambient air temperature is necessary for calculating the future net-point temperature references. The ambient air temperature predictions are based on meteorological forecasts of ambient air temperature, past observations of the ambient air temperature and a Fourier expansion of the difference in diurnal variation between the meteorological forecasts and the observations (See paper D).
- *Net-point Reference Curve.* Using the ambient air temperature prediction and estimates of the standard deviation for the prediction errors of the ambient air temperature and net-point temperature models respectively, the net-point temperature reference is calculated. The calculated reference temperature takes into account, that the future net-point temperature only is allowed to drop below a pre-specified control curve with a given (small) probability (See paper D).
- *Net-point Controller.* Using the eXtended General Predictive Controller (XGPC) presented in paper D a set point for the supply temperature is calculated for each net-point. The primary input to the XGPC controller is the reference net-point temperature curve and the impulse response function previously calculated.
- *Heat Load Predictions.* Two predictions of heat load are made: one with a prediction horizon up to 24 based on past observations of heat load, supply temperature, ambient air temperature as well as a description of the diurnal variation of the heat load (See paper D) and a second with a prediction horizon up to 39 hours where the aforementioned inputs are extended with meteorological forecasts of ambient air temperature and wind speed. The 24 hour predictions acts as input to the flow controller, but both predictions are also made available for the operator via the graphical user interface.
- *Flow Controller.* The 24 hour heat load prediction is converted to a set point for the supply temperature using the energy relation at the supply point as described in paper D. The controller takes into account, that future values of the flow only is allowed to exceed a maximum value with a given (small) probability.

- *System Set Point.* Finally the system set point for supply temperature is selected as the maximum of the individual set points from the net-point controllers and the flow controller.
- *System Statistics.* PRESS-N makes a number of useful statistics related to the observed heat load in district heating system. The following statistics are available:
 - *Energy Signature.* The energy signature model describes the static relationship between heat load – the dependent variable – and ambient temperature, time of day and supply temperature – the explanatory variables. In PRESS the energy signature is implemented as a linear model using a fourth order polynomial expansion of ambient air temperature and a second order Fourier expansion of time of day. The model distinguishes between work and non-work days. Further details regarding the energy signature model is found in (Madsen, Nielsen & Søggaard 1996).
 - *Duration Curve.* The duration curve is one way of representing the yearly distribution of heat load. The ordinate axis covers the range from zero to the maximum heat load, and for a given value of the heat load the abscissa value on the duration curve is the number of (hourly) observations for which the heat load during last year has been above the given load.
 - *Degree Days.* Degree days is a way of representing the ambient air temperature with special emphasis on its influence on heat load and as it is seen in paper E degree days have indeed a high correlation with heat load. On a daily basis degree days are calculated as the positive difference between the diurnal average indoor temperature (per definition set to 17°C) and the diurnal average ambient temperature (see eg. (Cappelen & Jørgensen 2001)). The daily values are typically summarized to weekly, monthly, seasonal and yearly values, but only the monthly values are calculated in PRESS.
 - *Diurnal Average Heat Load.* Two diurnal average heat load profiles are calculated based on the observed heat load – one for work days and one for non-work days. The estimation of the average diurnal profiles are made adaptive by introducing an exponentially decaying weight function in the estimation. Thus the most recent observations are given the highest weight in the calculations.
- *Plot Output Files.* Various output files to the graphical user interface are generated.

The models applied in PRESS-N are estimated adaptively, they “learn” from the observed data as time goes by, thereby rendering re-calibration superfluous. On the other hand this means, that they have to run for some time before the predictions can be considered to be reliable. To overcome this drawback in applying adaptive estimation some additional features have been incorporated into PRESS-N:

- Accelerated learning allows PRESS-N to be started back in time and then use historical input files to calibrate the models before moving into real time operation.
- A backup of the current model state is generated every day at midnight. The backup ensures that the system will be able to restart quickly in case of power interrupts, system reboot or similar.

2.3.2 The Presentation Part of PRESS (PRESS-P)

PRESS-P is implemented in ANSI C++ and is based on the X11 and Motif graphical libraries. It has been tested under HP Unix and Linux and is expected to run on any platform for which X11, Motif and an ANSI C++ compliant compiler is available.

The main window provides the operators with an overview of the current system state whereas more detailed informations regarding heat load predictions, various heat load related statistics, input measurements, meteorological forecasts and the supply temperature controller is available through a number of plot windows. Figure 2.2 shows the main window of the program together with plots windows for predicted heat load and meteorological forecasts of ambient air temperature. The main window consists of a menu bar (top), a number of value fields for observations and heat load predictions (middle) and an information field (bottom).

The menu bar provides access to the various plot windows and an event listing through a number of sub-menus. All plots are initially displayed using a default setup, but through a dialogue it is possible to change the time period considered, to change the scaling of the axis as well as to compose entirely new plots.

The observation values presented in the main window are the latest available 5 minute values. In case an observation has been classified as faulty the

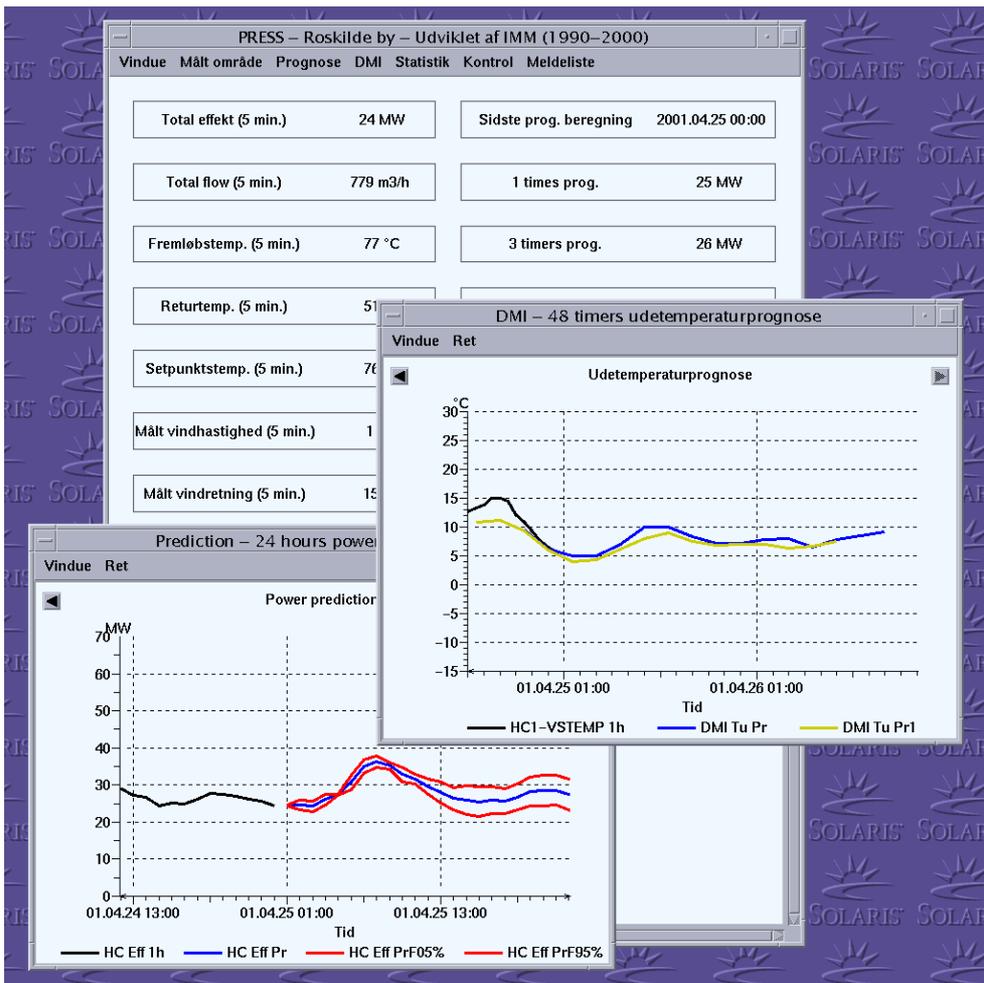


Figure 2.2: The main window in PRESS (center) shown together with plot windows for predictions of heat load (bottom left) and meteorological forecasts of ambient air temperature (center right). The heat load prediction plot shows the predicted heat load for the next 24 hours together with the observed heat load for the last 6 hours. The plot also presents some empirical uncertainty bands for the prediction. By default the 5% and the 95% quantiles corresponding to a 90% confidence interval are displayed but other quantiles are available if more appropriate. The plot of forecasted ambient air temperature shows the two most recent forecasts received from the Danish Meteorological Institute (DMI) as well as the observed value for the last 6 hours.

observation in question is marked by a red background colour in the value field.

A number of useful statistics related to heat load are presented in PRESS.

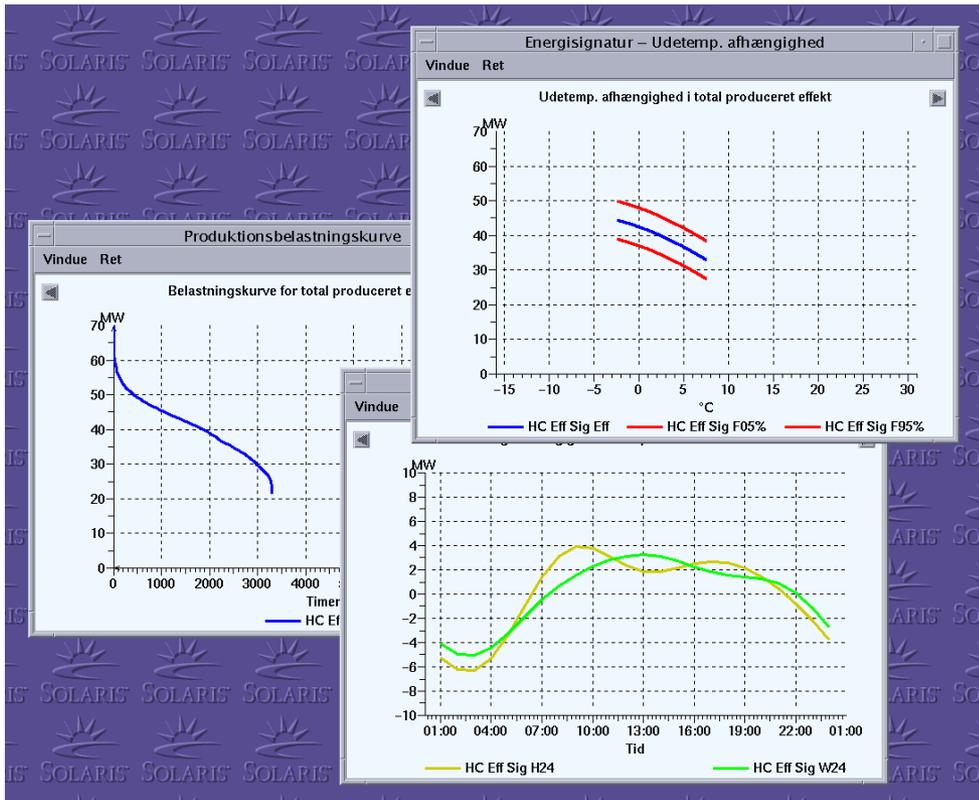


Figure 2.3: Examples of the statistical plot windows in PRESS. A plot of the duration curve (left) is shown together with plot windows presenting the ambient air temperature dependency (top right) and the diurnal dependency (bottom right) in the heat load as estimated by the energy signature model.

Figure 2.3 shows examples of the duration plot window together with two plot windows associated with the dependency of the heat load on ambient air temperature and time of day as estimated by the energy signature model described in Section 2.3.1. Also plot windows showing the average diurnal heat load for work days and non-work days as well as plots related to the local degree days are available.

Finally the PRESS system generates a number of windows related to the control function thereby enabling the operators to monitor the supply temperature controller and, to a certain extent, change the configuration of the controller on-line. Figure 2.4 gives an example of some of the control related plots windows in PRESS. Two plot windows provide an overview of the interaction between the flow and net-point temperature controllers: the first showing the observed supply temperature together with the supply temper-

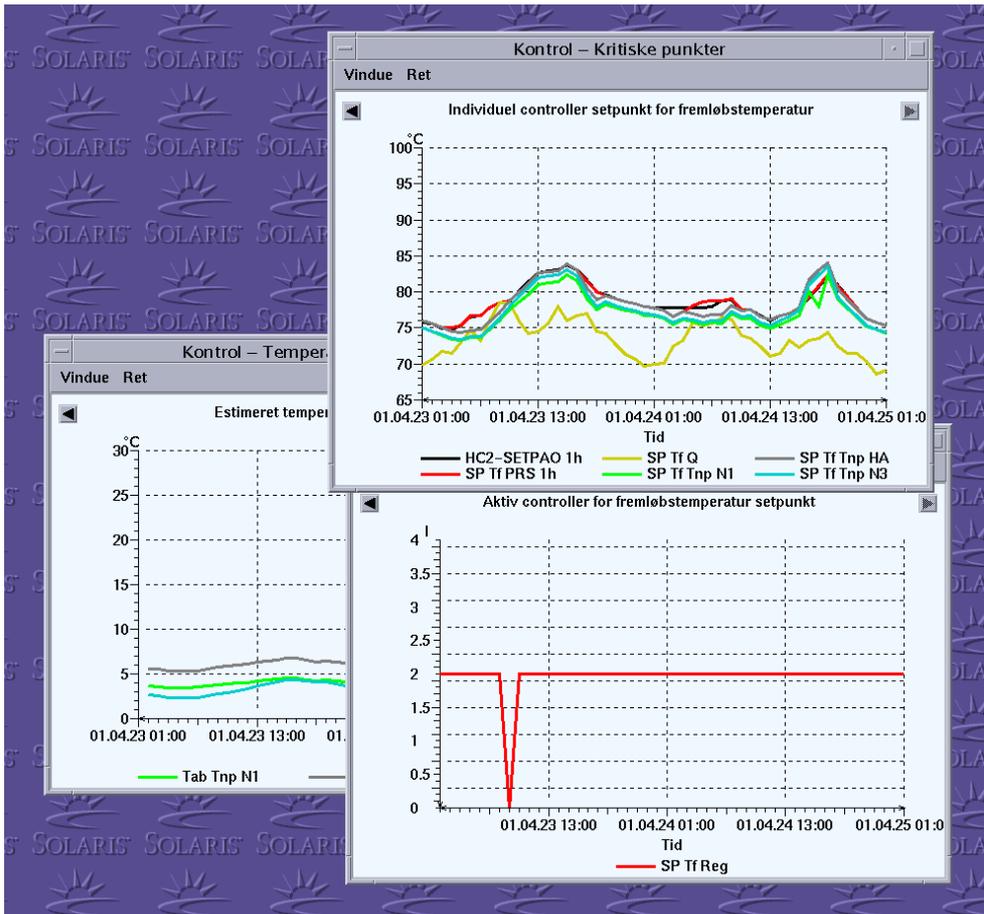


Figure 2.4: Examples of the control related plot windows in PRESS. A plot of the observed supply temperature as well as the set-points for the supply temperature as recommended by the individual net-point temperature and flow sub-controllers (top right) is shown together with plots of the active controller (bottom right) and the estimated temperature loss to the critical net-points (center left). The plots in the figure all display data for the last 48 hours.

ature set-points recommended by the individual net-point temperature and flow sub-controllers and the second displaying which of the sub-controllers that have determined the supply temperature (the active controllers). The current state of the net-point temperature models is monitored through two plot windows: the first shows the estimated (average) time delay (not shown in Figure 2.4) and the second displays the estimated temperature loss from the supply point to each of the critical net-points.

Prediction of available wind power in a larger area

In Denmark the subject of short term¹ prediction of wind power for control and surveillance purposes has been of high interest for a number of years and Informatics and Mathematical Modelling has been working within this field since 1992. Prediction of wind power is by no means a trivial matter as the underlying system – the combination of numerical weather predictions (NWP) and wind turbines – is inherently non-linear as well as non-stationary. Section 3.1 gives an account of the background and motivation for prediction of wind power. In Section 3.2 the problem is presented in more detail and the papers presented in this thesis are brought in context. Finally an application for wind power prediction – WPPT – implementing a selection of the described models and methods are presented in Section 3.3.

3.1 Motivation

In Denmark the demand for reliable wind power predictions has become more and more urgent during the recent years. This development is driven by several factors:

¹In relation to weather and wind power forecasting short-term prediction refers to predictions with a horizon from 1 hour and up to 48 – 72 hours ahead.

- The rated power of the installed wind turbines has more than quadrupled since 1994 and does now constitute a substantial fraction of production capacity for the conventional power plants. In the western part of Denmark, for instance, the wind turbines have at several occasions been close to covering the entire power demand during periods with low power load. Optimal exploitation of the transmission grid and production facilities in a system with this high a penetration of wind power will obviously require reliable predictions of the wind power production.
- A power exchange trade – NordPool – has been introduced between the Scandinavian countries. The ordinary power trading for the following day is finalized at noon, hence for utilities with a high penetration of wind power reliable wind power predictions are a prerequisite for efficient trading on the NordPool.
- As a result of the ongoing liberalization of the electricity sector a new structure is emerging. The sector is being divided into three independent types of operators:
 - The production companies which own and operate the conventional power plants and some of the wind farms.
 - The transmission companies which own and operate the high voltage transmission network. The responsibility for system reliability and endurance will typically belong to the transmission companies.
 - The distribution companies running the low voltage distribution network supplying the individual consumers.

When the liberalization is fully implemented the power trading between the various operators will be based on short term contracts typically covering the following day. Any deviations from the reported demand or production will then carry an economical penalty. Thus operators with considerably amounts of wind power will have a clear interest in precise wind power predictions.

In the western part of Denmark, where the majority of the wind turbines are located, a large fraction of the wind turbines is privately owned and situated in small groups or standing alone. As a result on-line power measurements at the utilities have up to recently only been available for a minor fraction of the wind turbines in the western part of Denmark. For most of the remaining turbines the only information regarding their production has been in form of monthly or quarterly energy readings from their accounting meters. Within the last few years this has changed, though, as the changing market conditions require that the power measurements covering all wind turbines larger than 150 kW

are made available as 15 minute average values at the utilities responsible for the accounting for the area in question. The measurements are not available on-line but is delivered in diurnal batches with a delay of a few days.

The work on prediction of wind power was initiated as a cooperation between Elsam and IMM in 1992 under the project, *Wind Power Prediction Tool in Control Dispatch Centres*, sponsored by the European Commission. During this project the first version of WPPT was developed and implemented at Elsam's control center at Fredericia. The prediction models in WPPT 1 utilized on-line measurements of wind power and wind speed. WPPT 1 went into operation in October 1994 and was subject to a three months trial period. The experience gained as well as further details regarding the models and user interface developed can be found in (Madsen, Sejling, Nielsen & Nielsen 1995) and (Madsen, Sejling, Nielsen & Nielsen 1996). In short it became apparent that WPPT 1 was capable of providing the operators with useful predictions up to 8 to 12 hours ahead, but for larger prediction horizons further model development was needed.

In (Landberg, Hansen, Vesterager & Bergstrøm 1997), (Landberg 1997a) and (Landberg 1997b) physical models describing the wind farm layout and the influence of the surroundings are used in combination with meteorological forecasts of wind speed and direction to make predictions of power production with a horizon of up to 36 hours ahead. Promising results were found for the longer prediction horizons, but the approach had poor performance on shorter horizons.

In (Nielsen & Madsen 1997) it is proposed to utilize meteorological forecasts from the national weather service as input to the previously developed statistical prediction models. Nielsen & Madsen (1997) shows, that introduction of meteorological forecasts in the prediction models results in an improved performance for all prediction horizons and especially for the larger prediction horizons very distinct improvements are found. The results from (Nielsen & Madsen 1997) are summarized in paper G. In 1997 a new project, *Implementing short-term prediction at utilities*, was initiated again with Elsam as partner and sponsored by the European Commission. The purpose of the project was to further refine the wind farm power prediction models and implement an operational wind power prediction system – WPPT 2 – at the control center of Elsam based on on-line measurements and meteorological forecasts for a number of *reference* wind farms in the western part of Denmark. Predictions of power production for the individual wind farms as well as for the entire population of wind turbines in the area are calculated by the implemented system, where the latter is accomplished by means of an

upscaling algorithm. The developed models as well as the results obtained with respect to the wind farm predictions are described in (Nielsen, Madsen, Nielsen & Tøfting 1999), whereas qualitative and quantitative assessments of the predictions for the entire area are given in paper I and (Nielsen, Madsen & Christensen 2000), respectively.

Up to 1999 most research focused on building the best possible power prediction models for wind farm, whereas the upscaling models only had received minor attention. This was due to lack of detailed information regarding the power production from wind turbines not covered by on-line measurements at the utilities. The appearance of accounting data made it possible to build more sophisticated upscaling models and in 1999 a new project *Wind farm production predictor*² with IMM, Risø and all the major Danish power utilities as partners was started. The purpose of the project was to develop a prediction system – Zephyr – for all wind turbines in a large area based on the available data – meteorological forecasts, on-line measurements for selected reference wind farms and accounting data covering almost all wind turbines in the area. The software implementation of Zephyr was a very ambitious multi-tier client server solution inspired by the Java Enterprise Beans architecture as described in (Giebel, Landberg, Joensen, Nielsen & Madsen 2000). The complicated architecture became the undoing of the first version of the Zephyr modelling system and in the end the development was abandoned. During the project new wind farm and upscaling model was developed – see (Nielsen, Madsen, Nielsen, Landberg & Giebel 2001, Marti, Nielsen, Madsen, Navarro & Barquero 2001, Nielsen, Madsen, Nielsen, Giebel & Landberg 2002) and paper J – and these are currently under implementation in a revised version of the WPPT software. This new version – WPPT 4 – will also serve as the model engine in a coming re-implementation of the Zephyr system.

3.2 Prediction of power production from wind farms

From previous investigations (Nielsen & Madsen 1997, Nielsen et al. 1999) it is known that predictions of power production from wind farms with a lead time of up to 8–12 hours depend on on-line measurements whereas predictions with a lead time larger than 12 hours mainly depend on the meteorological forecasts. For longer prediction horizons the characteristics of the meteorological forecasts are thus of major importance. These are the subject of Section 3.2.1. When employing meteorological forecasts and local observations of weather

²The project title was badly selected as it does not reflect the scope of the project.

variables, primarily wind speed, in the modelling and prediction of power production for a wind farm two conceptually different approaches for establishing the power prediction models can be taken:

1. A model of the wind speed is used to calculate predictions of wind speed as a function of forecasted and observed weather variables. The wind speed predictions are then used as input to another model describing the relationship between wind speed and power production³ in order to obtain predictions of power production. That is, the power predictions are obtained in a two stage approach, where the two models are estimated separately.
2. One model containing a description of both the correlation structure of wind speed and power production is used to obtain the power predictions. The relation between wind speed and power production can either be contained in the model or be estimated separately.

Earlier findings in (Madsen et al. 1995) and (Nielsen & Madsen 1997) suggest, that the performance of the two stage models, where the correlations structure of the power observations is disregarded, generally are inferior to models taking this structure into account. Consequently the power prediction models considered in the papers G to J all belong to the class of auto-regressive models with exogenous input - a model class where the correlation structure between current and future values of the dependent (output) variable is modelled directly. A summary of the power curve models employed in the various papers are found in Section 3.2.2 whereas the power predictions models are the subject of Section 3.2.3.

3.2.1 Characteristics of the meteorological forecasts

Wind power prediction with a lead time greater than 8 to 12 hours will to a large extent depend on the availability of reliable meteorological forecasts of wind speed and wind direction. The quality and characteristics of the meteorological forecasts from the HIRLAM model running at the Danish Meteorological Institute (DMI) have been scrutinized in several research papers eg. (Landberg, Watson, Halliday, Jørgensen & Hilden 1994, Nielsen & Madsen 1997, Nielsen et al. 1999) and paper H. The basic properties of the empiri-

³The relationship between wind speed and power production for a wind farm will be denoted the wind farm power curve. This relationship might include the influence of other weather variables, eg. wind direction is an obvious candidate.

cal distribution of the forecasts such as mean value and variance as a function of the prediction horizon are of major interest. In addition, the capability of the meteorological forecasts to describe the diurnal variation of the observed wind speed, which at many locations arise during sunny days, is of interest.

The dependency of mean value and variation of forecasted wind speed on prediction horizon has been investigated in (Nielsen et al. 1999) based on data from five locations in the Jutland/Funen area in Denmark. Here it is found that mean value and variance of the forecasted wind speed increase with increasing prediction horizon. Also the variance of the prediction errors for the forecasted wind speed and wind direction is found to increase with the prediction horizon. Hence the power curve functions described in Section 3.2.2, will have to depend on the prediction horizon.

The diurnal variation of observed and forecasted wind speeds has been analyzed in (Nielsen et al. 1999) for the same five locations as above as well as in paper H based on an older data set for one of the locations. For all but one of the locations the measured wind speed exhibits a clear diurnal variation during the summer period whereas a less distinct or no variation was found during the winter period. In (Nielsen et al. 1999) a similar pattern is found in the diurnal variation of meteorological forecasts with the reservation, that the amplitude of the profile decreases and the shape becomes distorted as the prediction horizon increases. Furthermore a distinct diurnal variation in either of the observed or the forecasted wind speed does not necessarily recur in the other. For the older meteorological forecasts investigated in paper H no diurnal variation was found regardless of time of year. It is thus concluded that the diurnal variation of wind speed can not be assumed to be described by the meteorological forecasts, but has to be described explicitly in the power prediction models.

3.2.2 Power curve models

For a single wind turbine with no obstacles in its vicinity the power production is, disregarding the influence of ambient air temperature and humidity, only a function of the wind speed “seen” by the turbine. The static relationship between power and wind speed is described by the turbine power curve specified by the manufacturer of the wind turbine. This relationship is highly non-linear as indicated in Figure 3.1, which shows scatter plots of observed power production at the Rejsby wind farm versus observed and forecasted wind speed for selected forecast horizons. The line is the estimated relationship using local regression with second order polynomial approximation, cf.

(Cleveland & Devlin 1988).

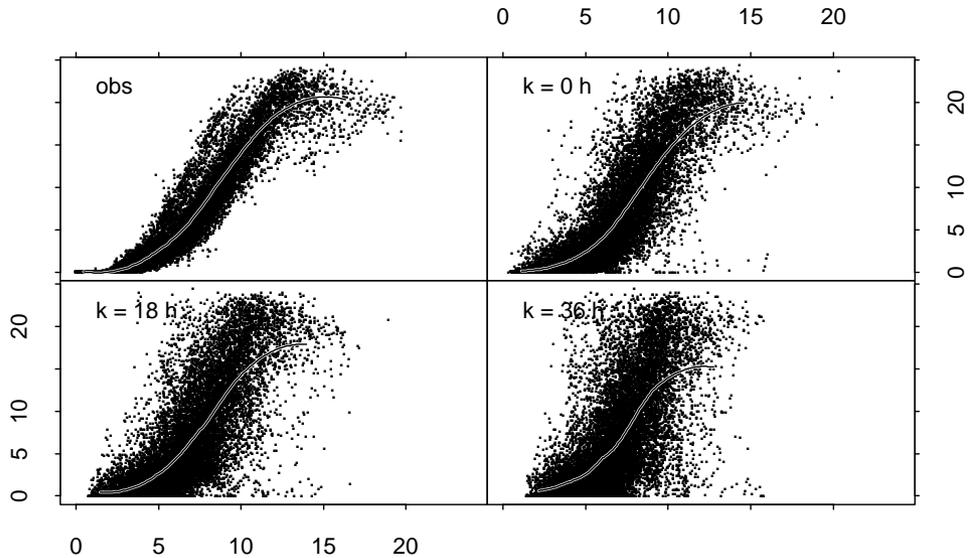


Figure 3.1: *Observed power production [MW] versus observed wind speed [m/s] (top left) and forecasted wind speed for a forecast horizon of 0 hours (top right), 18 hours (bottom left) and 36 hours (bottom right) for the Rejsby wind farm situated on the west coast of Jutland. The line is the estimated relationship using local regression with a nearest neighbour bandwidth of 40% and second order polynomial approximation.*

When the scope is changed from modelling the power output from a single free standing wind turbine to modelling the power production for an entire wind farm, it has to be taken into consideration, that the individual wind turbines now will create shadowing effects disturbing the turbines standing behind. The influence of this effect on the power production depends on the layout of the wind farm as well as the wind direction. This section considers the relationship between power production in a wind farm and local values of wind speed and wind direction – in the following denoted the wind farm power curve. Often locally measured values of wind speed (and direction) will be affected by local conditions such as wind turbines in the nearby surroundings or the location of the measurement equipment on the meteorological tower. A wind farm power curve estimated on basis of measured values will thus reflect the combined influence of wind turbine characteristics, wind farm layout and placement of the anemometer. This is confirmed in (Nielsen et al. 1999), which shows that the local conditions affecting the anemometers dominates the wind direction dependency in the estimated power curves to the extent, that local conditions affecting the entire wind farm – eg. wind farm layout, changing roughness of the surroundings and nearby obstacles – are hardly

discernible. Thus when a power curve model is used for prediction purposes with meteorological forecasts as input a dedicated power curve model for the relationship between observed power production and the corresponding meteorological forecasts should be estimated. As argued in Section 3.2.1 the power curve estimation should allow for a dependency on prediction horizon. This is also in accordance with (Jonsson 1994), who considers the estimation of linear regression models in the presence of errors in the regressors. Jonsson (1994) argues, that if errors are present in the regressors a dedicated prediction model should be estimated as use of the true system for prediction will not result in the best possible predictions.

When estimating power curve models the distribution of the local wind speed must also be taken into consideration. Most of the wind farm sitings in Denmark have a wind speed distribution with the bulk of the observations in the low to medium range and only a small fraction of the observations in the upper range with full production in the wind farms. When the wind direction dependency is taken into consideration this situation is further aggravated as illustrated in Figure 3.2, which shows a two dimensional distribution of the observed wind speed and wind direction for the Rejsby wind farm.

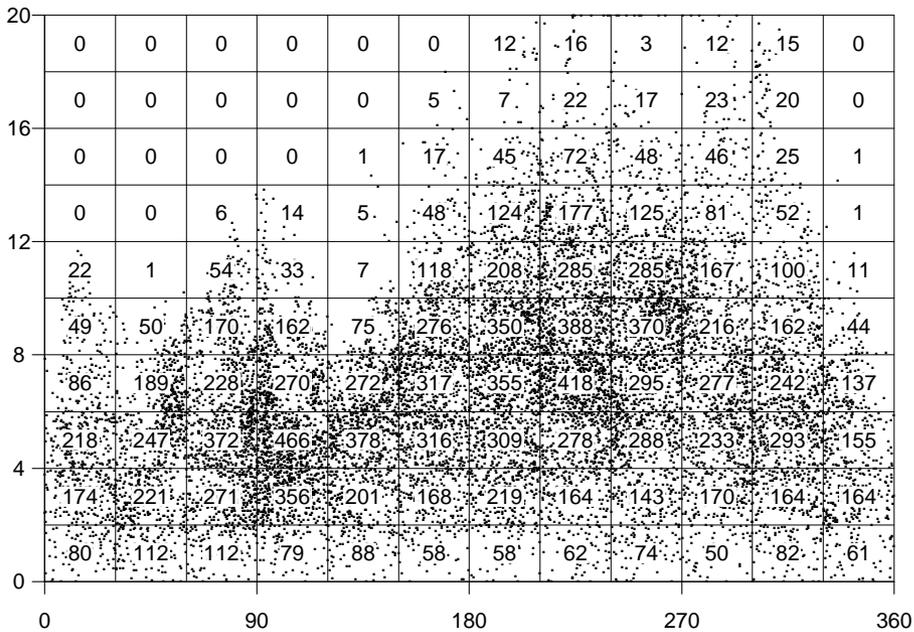


Figure 3.2: Observed wind direction [$^{\circ}$] versus observed wind speed [m/s] for the Rejsby wind farm.

Different approaches to modelling the power curve for a wind farm have been

investigated. In (Nielsen & Madsen 1997) and paper G the relationship between the observed power production, p_t , and the numerical weather predictions of wind speed and wind direction, $w_{t|t-k}^{nwp}$ and $\phi_{t|t-k}^{nwp}$, is modelled directly in the predictions models using the terms

$$\sqrt{p_{t+k}} \propto b_t^0 \sqrt{w_{t+k|t}^{nwp}} + b_t^1 w_{t+k|t}^{nwp}, \quad (3.1)$$

or by a non-linear model estimated separately

$$\begin{aligned} p_{t+k} &= m \exp \left[\beta \exp \left[\kappa (\phi_{t+k|t}^{nwp}) w_{t+k|t}^{nwp} \right] \right] + e_t \\ \kappa(\phi_t) &= \kappa_0 + \sum_{i=1}^{N_{fx}} [\kappa_{i,1} \sin i\phi_t + \kappa_{i,2} \cos i\phi_t], \end{aligned} \quad (3.2)$$

where (3.2) is a so-called Gompertz parametrization of the power curve with an additional wind direction dependency build into the parameter κ . In (3.1) the model parameters b_t^0 and b_t^1 are estimated recursively and adaptive, hence the t suffix, whereas the model parameters m , β , κ_0 and $\kappa_{1..N_{fx},1..2}$ in (3.2) are estimated by a traditional non-linear LS minimization algorithm. The use of (3.1) and (3.2) to describe the relationship between observed power production and forecasts of wind speed and for (3.2) wind direction is motivated by a model study in (Madsen et al. 1995). Here similar models were found to give a good description of the relationship between observed values of power production, wind speed and wind direction. Discussion of the modelling results is referred to Section 3.2.3.

The square root transformation of power and wind speed in (3.1) is motivated by the skew density of power and wind speed. In (Madsen et al. 1995) it is shown that the square root transformation leads to distributions of the prediction errors, which can be approximated by the Gaussian distribution. Note that the transfer function from $\sqrt{w_{t|t-k}^{nwp}}$ to $\sqrt{p_t}$ is formulated using a second order polynomial expression (a so-called Hammerstein model).

The parameters in (3.1) and (3.2) are estimated globally, that is, disregarding the time-variation in (3.1), only one set of parameters is estimated. Hence due to the skewness of the distribution of the $\{w, \phi\}$ process the low wind speeds carry a much higher weight in the parameter estimation than the high wind speeds. In paper H this problem is circumvented by defining the power curve via a non-parametric model

$$p_{t+k} = g(w_{t+k|t}^{nwp}, \phi_{t+k|t}^{nwp}) + e_{t+k} \quad (3.3)$$

where $g(\cdot)$ is a smooth function of forecasted wind speed and wind direction estimated using local polynomial regression cf. (Cleveland & Devlin 1988)

and paper A. Local polynomial regression is a very flexible estimation method, which can be tailored according to the distribution of $\{w, \phi\}$ by varying bandwidth and order of the local polynomial approximation. However, in paper H only fixed bandwidth and a second order polynomial are considered.

Motivated by the irregular distribution of wind speed and wind direction, cf. Figure 3.2, Nielsen et al. (1999) introduces upper and lower limits on the wind speeds for which it is possible to estimate a wind direction dependency. Data from five different wind farms in the Jutland/Funen area are investigated, and it is found, that the wind direction dependency between those limits has a simple form, which readily can be approximated by eg. lower order polynomials. The power curve model formulated is given as

$$\begin{aligned} p_{t+k} &= g_0(w_{t+k|t}^{nwp}) + I(w_{t+k|t}^{nwp}) \left[\sum_{j=0}^2 g_1^j(\phi_{t+k|t}^{nwp})(w_{t+k|t}^{nwp})^j \right] + e_{t+k} \\ &= g(w_{t+k|t}^{nwp}, \phi_{t+k|t}^{nwp}) + e_{t+k} \quad , \end{aligned} \quad (3.4)$$

where $I()$ is an indicator function equal to 1 for wind speeds where the wind direction dependency should be modelled with a smooth transition to zero otherwise, $g_0()$ is a smooth function of functions of wind speed and $g_1^{0..2}()$ are smooth functions of wind direction. Model (3.4) belongs to the class of *additive models* proposed in (Hastie & Tibshirani 1990). The overall model is estimated using the *back fitting* algorithm described by Hastie & Tibshirani (1990) wherein the individual terms have been estimated using local regression with a second order polynomial approximation and fixed bandwidth.

For all but one of the five wind farms investigated it is beneficial to model the wind direction dependency and the estimated wind direction dependency seems to resemble what must be expected when siting and wind turbine layout of the wind farms are considered. There seems, however, to be a reduced improvement as the prediction horizon increases.

The power curve model used in paper J is given as

$$p_{t+k} = g(w_{t+k|t}^{nwp}, \phi_{t+k|t}^{nwp}, k) + e_{t+k} \quad . \quad (3.5)$$

Model (3.5) is very similar to (3.3), but the dependency in the power curves on prediction horizon, cf. page 30, is now contained in the model. (3.5) is estimated using the method proposed in paper B, ie. the estimation of the power curve is here recursive and adaptive.

3.2.3 Dynamic power predictions models for wind farms

The subject of predicting the power production from a wind farm using statistical methods is considered in the papers G, H, J and in (Nielsen et al. 1999). The models belong to the class of auto-regressive models with exogenous input and are formulated directly as k -step prediction models. The input to the prediction models are transformed values of observed wind speed as well as forecasted wind speed, where the transformation is either by a second order polynomial expansion or by estimated power curves as described in Section 3.2.2. The model parameters are estimated adaptively in order to accommodate approximations and deficiencies in the power prediction models as well as slow changes in the total system consisting of wind farm, surroundings and NWP model. These changes are caused by effects such as aging of the wind turbines, changes in the surrounding vegetation and maybe most importantly due to changes in the NWP models used by the weather service.

Models fitted to the original power series as well as models fitted to the square root transformed power series are considered. The square root transformation of power is motivated by the skew density of the power distribution (see eg. (Madsen et al. 1995) or (Joensen 1997)), but, as the model is fitted to the transformed power series, the prediction performance in the original power series must be expected to be (slightly) compromised. When a model fitted to the square root transformed power series is used to predict in the original power series, a systematic bias is introduced in the predictions, and following (Tong 1990) a second order Taylor approximation of the correction term is used in paper H and paper I to eliminate this bias (See (Nielsen et al. 1999) for further details).

The models in paper G are estimated on basis of half-hourly data and have a prediction horizon from half an hour and up to 24 hours, ie. $k \in [1, 48]$. Based on model studies by Madsen et al. (1995) and by Nielsen & Madsen (1997) the following model is proposed

$$\begin{aligned}
 \sqrt{p_{t+k}} &= a \sqrt{p_t} + b_1 \sqrt{w_t} + b_2 w_t + T_{t+k|t}^{nwp} + m_{t+k} + e_{t+k} \\
 m_{t+k} &= m + c_1 \sin \frac{2\pi h_{t+k}^{24}}{24} + c_2 \cos \frac{2\pi h_{t+k}^{24}}{24} \\
 T_{t+k|t}^{nwp} &= \begin{cases} b_3 \sqrt{w_{t+k|t}^{nwp}} + b_4 w_{t+k|t}^{nwp} & \text{or,} \\ b_3 m \exp \left[\beta \exp \left[\kappa (\phi_{t+k|t}^{nwp}) w_{t+k|t}^{nwp} \right] \right] & . \end{cases}
 \end{aligned} \tag{3.6}$$

where the transformation from meteorological forecasts of wind speed (and wind direction) to power is either by a second order polynomial or a Gompertz curve. In (3.6) h_t^{24} is the time of day and a , $b_{1..4}$, m and $c_{1..2}$ are

time-varying model parameters, which is estimated using the Recursive Least Squares algorithm with exponential forgetting by Ljung & Söderström (1983).

In (Nielsen & Madsen 1997) and paper G only minor differences between the prediction performance of (3.6) is found when using the two different parametrizations of $T_{t+k|t}$. The models employed in paper I correspond to (3.6), where the transformation of forecasted wind speed is by the second order polynomial expansion. In order to accommodate the needs of the utilities in relation to trading on the NordPool paper I considers prediction horizons up to 39 hours, though.

In paper H models with a prediction horizon up to 24 hours are identified on basis of an hourly data set, ie. $k \in [1, 24]$. The following model is put forward

$$p_{t+k} = a_t p_t + b_t g(w_{t+k|t}^{nwp}, \phi_{t+k|t}^{nwp}) + d_t(h_{t+k}^{24}) + e_{t+k}, \quad (3.7)$$

where a_t and b_t are time-varying parameters with a partly know variation, $g()$ is the power curve model given by (3.3) and $d_t()$ is a time-varying non-parametric estimate of the diurnal variation. The parameters a_t and b_t are estimated using the tracking algorithm for time-varying parameters proposed in paper C.

Nielsen et al. (1999) considers prediction models with a prediction horizon up to 39 hours. The models are identified on basis of a half hourly data set consisting of data for five wind farms situated in the western part of Denmark.

A set of models have been proposed

$$\begin{aligned}
 \sqrt{p_{t+k}} &= a_1\sqrt{p_t} + a_2\sqrt{p_{t-1}} + \\
 & b_1^o\sqrt{w_t} + b_2^o w_t + b_1^m \sqrt{w_{t+k|t}^{nwp}} + b_2^m w_{t+k|t}^{nwp} + \\
 & \sum_{i=1}^3 [c_{i,2} \cos \frac{2i\pi h_{t+k}^{24}}{24} + c_{i,1} \sin \frac{2i\pi h_{t+k}^{24}}{24}] + \\
 & m + e_{t+k} ,
 \end{aligned} \tag{3.8}$$

$$\begin{aligned}
 p_{t+k} &= a_1 p_t + a_2 p_{t-1} + \\
 & b_1^o w_t + b_2^o (w_t)^2 + b_1^m w_{t+k|t}^{nwp} + b_2^m (w_{t+k|t}^{nwp})^2 + \\
 & \sum_{i=1}^3 [c_{i,1} \cos \frac{2i\pi h_{t+k}^{24}}{24} + c_{i,2} \sin \frac{2i\pi h_{t+k}^{24}}{24}] + \\
 & m + e_{t+k} ,
 \end{aligned} \tag{3.9}$$

$$\begin{aligned}
 p_{t+k} &= a_1 p_t + a_2 p_{t-1} + \\
 & b_1^o g_{pc}^o(w_t, \theta_t) + b_1^m g_{pc}^m(w_{t+k|t}^{nwp}, \theta_{t+k|t}^{nwp}) + b_1^{24} d(h_{t+k}^{24}) + \\
 & \sum_{i=1}^3 [c_{i,1} \cos \frac{2i\pi h_{t+k}^{24}}{24} + c_{i,2} \sin \frac{2i\pi h_{t+k}^{24}}{24}] + \\
 & m + e_{t+k} ,
 \end{aligned} \tag{3.10}$$

where $a_{1..2}$, $b_{1..2}^o$, $b_{1..2}^m$, b_1^{24} , m and $c_{1..3,1..2}$ are time-varying model parameters, $g^o()$ and $g^m()$ is the power curve model given by (3.4) and as in (3.7) $d()$ is a non-parametric estimate of the diurnal variation in the power production. The model parameters are estimated using Recursive Least Squares with exponential forgetting.

Based on exhaustive search among all possible combinations of the model terms for each of (3.8), (3.8) and (3.8) a candidate for a best overall adaptive

prediction model is identified

$$p_{t+k} = \begin{cases} a_1 p_t + a_2 p_{t-1} + b_1 g^m(w_{t+k|t}^{nwp}, \theta_{t+k|t}^{nwp}) \\ \quad + m_{t+k}^1 + e_{t+k} & \text{for } k \in [1, 4] , \\ a_1 p_t + b_1 g^m(w_{t+k|t}^{nwp}, \theta_{t+k|t}^{nwp}) \\ \quad + m_{t+k}^2 + e_{t+k} & \text{for } k \in [5, 9] , \\ a_1 p_t + b_1 g^m(w_{t+k|t}^{nwp}, \theta_{t+k|t}^{nwp}) \\ \quad + m_{t+k}^3 + e_{t+k} & \text{for } k \in [10, 15] , \\ a_1 p_t + b_1 g^m(w_{t+k|t}^{nwp}, \theta_{t+k|t}^{nwp}) \\ \quad + m_{t+k}^2 + e_{t+k} & \text{for } k \in [16, 30] , \\ b_1 g^m(w_{t+k|t}^{nwp}, \theta_{t+k|t}^{nwp}) + b_2 d(t_{day}) \\ \quad + m_{t+k}^2 + e_{t+k} & \text{for } k \in [31, 54] , \\ b_1 g^m(w_{t+k|t}^{nwp}, \theta_{t+k|t}^{nwp}) + b_2 d(t_{day}) \\ \quad + m_{t+k}^3 + e_{t+k} & \text{for } k \in [55, 78] , \end{cases} \quad (3.11)$$

with

$$m_t^N = m + \sum_{i=1}^N \left[c_{i,1} \sin \frac{2\pi i h_{t+k}^{24}}{24} + c_{i,2} \cos \frac{2\pi i h_{t+k}^{24}}{24} \right] , N \in [1, 3] .$$

Two characteristics are noted for this model. First of all the transformed meteorological wind speeds are needed in the model for all prediction horizons. Secondly for the shorter prediction horizons the model relies on observed values of power, but as the prediction horizon increases the emphasis shifts to terms related to diurnal variation and for a prediction horizon larger than 15 hours no observations enter the model.

The prediction model put forward in paper J belongs to the conditional parametric auto-regressive extraneous model class proposed in paper A. The model is given as

$$\begin{aligned} p_{t+k} &= a(\theta_{t+k|t}^{nwp}, k) p_t + b(\theta_{t+k|t}^{nwp}, k) g(w_{t+k|t}^{nwp}, \theta_{t+k|t}^{nwp}, k) + \\ &\quad c_1(\theta_{t+k|t}^{nwp}, k) \cos \frac{2\pi h_{t+k}^{24}}{24} + \\ &\quad c_2(\theta_{t+k|t}^{nwp}, k) \sin \frac{2\pi h_{t+k}^{24}}{24} + e_{t+k} \end{aligned} \quad (3.12)$$

where $g()$ is the power curve model given by (3.5) and a , b , c_1 and c_2 are smooth time-varying functions estimated recursively and adaptively using the

method proposed in paper B. The coefficients in (3.12) are functions of forecasted wind direction. This is motivated by the observation that the weather situation in Denmark as a first approach can be classified by the wind direction: for westerly winds the weather is dominated by low pressure systems coming in from the North Atlantic whereas easterly winds often are associated with stable high pressure situations. Hence auto-correlation and diurnal variation of wind speed (and power production) will depend on wind direction.

3.3 An Application (WPPT)

This section describes WPPT – the Wind Power Prediction Tool – an on-line system for prediction of wind power in larger areas.

WPPT are used operationally at the control centres of Eltra, Elsam and SEAS, which are the three largest power utilities in Denmark with respect to wind power. WPPT was installed in the control centres of Eltra and Elsam in October 1997 and has been used operationally since January 1998. The third installation at the control center of SEAS went into operation in late summer 2001.

WPPT predicts the expected wind power production in a larger area using on-line data and meteorological forecasts 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 (a reference wind farm). Predictions of wind power with a horizon from 30 minute up to 39 hours are then formed 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 by a constant 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.

At present two of the wind farm models put forward in Section 3.2 are used operationally in WPPT: the installation at Eltra and Elsam uses model (3.6) where the transformation between forecasted wind speed and power production is by a second order polynomial expansion whereas the SEAS installation uses the model identified in (Nielsen et al. 1999) on basis of (3.9). The wind farm model and upscaling models proposed in paper J and (Nielsen et al. 2002), respectively, are currently under implementation in WPPT. The new upscaling models are based on account data, cf. page 26, and are expected to improve the prediction performance of WPPT significantly compared to

the constant upscaling used at present.

The experiences with the WPPT installation at Eltra and Elsam is evaluated in paper I and in (Nielsen et al. 2000). The assessment by the operators at Elsam and Eltra is, that WPPT generally produces reliable predictions, which are used directly in the economic load dispatch and the day to day electricity trade.

WPPT is implemented as two fairly independent parts – a numerical part and a presentation part – in the following denoted WPPT-N and WPPT-P, respectively. Data exchange between the two subsystems is implemented via a set of files.

WPPT-N and WPPT-P are based on the same two application frame works as PRESS-N and PRESS-P, respectively. Hence they implement similar features, accelerated learning, model backup, multiple user access, etc., and have similar system requirements as the PRESS system, cf. Section 2.3.

3.3.1 The Numerical Part of WPPT (WPPT-N)

WPPT-N can by and large be considered to consist of five major modules: data input (measurements and meteorological forecasts), data validation, model estimation and prediction, up-scaling module and finally performance logging and data output for WPPT-P. The measurements are given as 5 minute average values and the data validation is carried out on the 5 minute values before these are subsampled to form the 30 minute values used by the models. The meteorological forecasts are given as hourly values which are interpolated to form 30 minute values before being used in the models. A brief description of the functionality within each module is given in the following:

- *Data input.* The data interface for exchanging measurements and meteorological forecasts between the local SCADA system and WPPT-N is established via a set of plain ASCII files. An ASCII file interface has been selected for two reasons:
 1. The interface is simple to establish on a wide range of systems.
 2. The input files provide a historic database for the measurements. This is very helpful for fine tuning of the models as well as further model development.

The input files are checked for consistency both with respect to timing as well as the number of values.

- *Data validation.* Experience has shown that despite large efforts on-line measurements are prone to failures (errors). It is therefore essential to have some sort of automatic error classification of the measurements not only for protecting the models against the influence of erroneous measurements, but also in order to ease the surveillance tasks for the operators.
 - Range check. The measurements are checked versus predetermined minimum and maximum values.
 - Constancy check. The measurements are checked for constancy, ie. are hung up on a fixed value. Measurements of wind speed and power are allowed to become fixed around 0 for longer periods of time but otherwise fixed measurements are discarded as erroneous.
 - Confidence check. Here the output models describing the relationship between related measurements, eg. wind speed and power production, are compared with the actual measurements. If a measurement falls outside some predefined confidence bands provided by the model, it is classified as erroneous. This test is only implemented in a provisional version in the current release of WPPT.

Only the measurements are subject to the data validation methods described above, and the validation of the meteorological forecasts is left with the quality control of the national weather service.

- *Model estimation and prediction.* Each wind farm has a set of models covering the prediction horizon (30 minutes up to 39 hours) in steps of 30 minutes. Each model is a k -step prediction model for which estimation of model parameters and prediction of wind power is implemented as described in Section 3.2.3. Every 30 minutes a new 39 hour prediction is calculated for the power production of each wind farm. During periods, where model input is marked as erroneous, the model estimation is inhibited in order to protect the model from the influence of bad data and the predictions for the park are marked as being unavailable.
- *Up-scaling.* Both power production measurements and predictions for the selected wind farms are up-scaled and summarized so as to calculate an estimate of the power production in the entire Jutland-Funen supply area. For each wind farm a number of substitution wind farms have been defined and in case the values for a wind farm becomes marked as unavailable the wind farm in question is replaced by one of its predefined substitutes in the up-scaling.

- *Performance logging.* Every 30 minute all the updated predictions for the reference wind farms as well as the total prediction for the entire supply area are logged and saved in individual files.
- *Data output for WPPT-P.* Finally the interface files between WPPT-N and WPPT-P are updated (5 minute values as well as 30 minute values).

3.3.2 The Presentation Part of WPPT(WPPT-P)

In the configuration used by Elsam, Eltra and Seas WPPT-N is a rather large system taking 60-70 measurements and 14-20 meteorological forecasts as input. Thus WPPT-P has to serve several purposes:

1. Display the $\frac{1}{2}$ to 39 hour prediction of the total wind production in the supply area.
2. Provide an overview of the climatical conditions and the power production throughout the supply area.
3. Provide an overview of the current status for the measurement equipment installed in the wind farms.
4. Display detailed information for each wind farm for diagnostic purposes, eg. if a prediction seems to be unrealistic the detailed plot for the wind farms can be used to determine the reason for the bad prediction.
5. Act as interface to the up-scaling algorithm between observed and predicted power production for the local wind farms in a sub-area to the total production in that sub-area.

The need for providing both an overview as well as detailed information is reflected in the design of WPPT-P. The main window together with a number of plots directly accessible from the menu bar on the main window provides the operators with an overview of the system state whereas the system engineer has access to more detailed information through a set of sub-windows dedicated to the individual wind farms.

The main window consists of four elements - menu bar (top), map area (left), value field (right) and information field (bottom) - as shown in Figure 3.3. The menu bar provides access to overview plots of the various observations plotted together as well as plots related to observed and predicted power

production in the total supply area. As an example Figure 3.3 shows the plot of the latest prediction of power production together with the observed power production for the last 6 hours. The plot also indicates some empirical uncertainty bands for the prediction. All plots are initially displayed using a default setup but as for PRESS-P in Section 2.3.2 the plots can be excessively customized.

The map area contains a map of the supply area where the location of each reference wind farm is marked by a wind farm symbol. The symbol relays information regarding the status of the wind farm in question such as the current production as a percentage of the rated power for the wind farm as well as the current measured wind speed and direction at the wind farm. In case a measurement error has been detected in the wind farm the symbol turns red to alert the operator to the error. Furthermore more detailed information regarding the wind farm can be assessed through a wind farm window activated by clicking on the symbol with the mouse.

The value field provides some key figures regarding the current system state: Calculation time for the current power prediction, initiation time for the calculation of the last meteorological forecast received, total rated wind power in the supply area, current estimates of the total power production and finally the current power predictions for some selected prediction horizons.

An information field on the main window as well as on each of the wind farm windows provides the system engineer with a bulletin board to relay relevant information to the users of WPPT.

The wind farm window displays the most recent 5 minute values for the wind farm measurements as well as gives access to a large number of plots for the wind farms in question. In case a measurement has been classified as faulty that measurement is marked by a red background colour. Also the rated power for the wind farm as registered by WPPT-N is displayed. The plots fall into three separate categories: Plots of the various observations versus time or versus other observations, plots of the predicted power production and derived plots eg. of the historical predictions for the wind farm and of the empirical uncertainty quantiles for the predictions and finally plots of the most recent meteorological forecasts of wind speed and direction as well as plots of the historical meteorological forecasts.

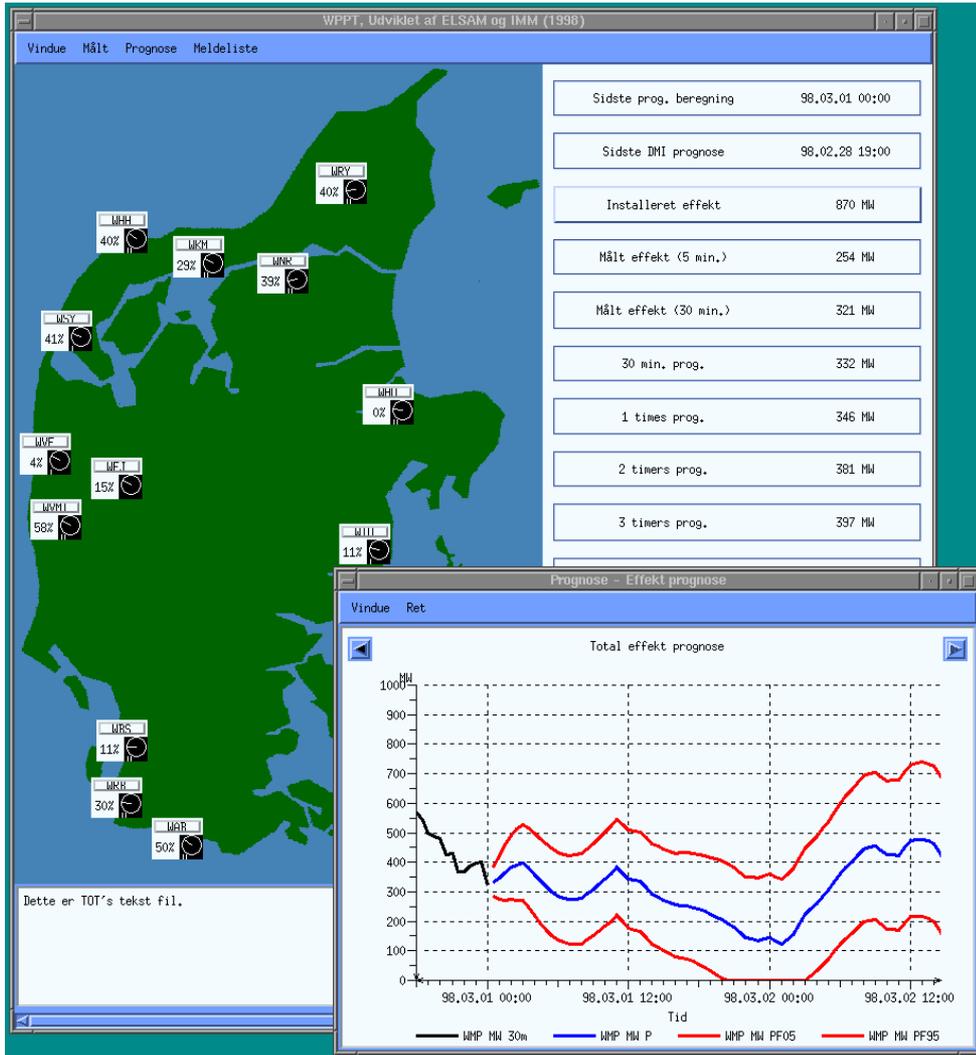


Figure 3.3: The main window in WPPT. The map area (left), where the wind farm symbols indicate position and current status of the reference wind farms, provides an overview of the current climatical situation in the supply area. The value fields (right) show the current power production together with the latest power predictions for some selected prediction horizons. The last update time for the meteorological forecasts as well as the power predictions are also displayed. The menu bar (top) provides access to overview plots of the various observations as well as plots related to the predictions of the total power production. As an example the plot overlaying the main window (bottom right) is the latest predicted power production for the total area. Dedicated windows give access to more detailed information for the individual wind farms. An information field (bottom) provides the system engineer with a bulletin board to relay relevant information to the users of WPPT.

Conclusion

The present thesis considers different aspects of modelling, prediction and control of time-varying non-linear stochastic systems.

The modelling part is concerned with non-linear systems where the non-linearities can be attributed to either a partly known time-variation of the model parameters or changes in some explanatory variables, eg. heat load and flow rate in district heating systems. The considered models have close resemblance to ordinary linear models, but with the model parameters replaced by smooth functions of either time or some explanatory variables. The latter models are named coefficient function models. Methods for recursive and adaptive estimation of the proposed models have been proposed and it has been shown that the methods can be seen as a generalization or combination of recursive least squares (Ljung & Söderström 1983) and local polynomial regression (Cleveland & Devlin 1988). Considering recursive and adaptive estimation of the class of coefficient function models it has been demonstrated, that the proposed estimation method possesses favourable stability properties in the presence of irregularly distributed observations in the space spanned by time or the explanatory variables.

The control part considers predictive control of supply temperature in district heating systems, but also the relationship between optimal operation of district heating systems and the presented supply temperature control is touched. The various predictions models necessary when applying predictive

control of supply temperature in district heating systems are considered, and a predictive control scheme – the XGPC – applicable for a wide range of linear models have been presented. The only requirement on the models is that at each time step the future model output must be additive in the term with a linear dependency on future controls, and the term depending on past control and output values. Finally it is considered how the uncertainties in the various output predictions are taken into account when determining the reference values for a predictive controller. An on-line application, PRESS, of the presented supply temperature controller have been implemented at two Danish district heating utilities. The experiences with the two implementations have been evaluated and it is shown that considerable savings can be obtained without compromising supply or comfort of the consumers.

The prediction part applies the proposed models and estimation methods to prediction of power production from wind farms one to two days a head in time. The wind farm prediction models are based on observed values of power production and wind speed as well as meteorological forecasts of wind speed and wind direction. It has been shown that the proposed methods are successful in modelling non-linear characteristics such as the time-varying diurnal variation of power production and the wind direction dependent relationship between power production and wind speed – the so-called power curve. It is also shown that models taking these characteristics into account are superior to ordinary linear models. It is found that the optimal set of parameter values depend on prediction horizon. This has been implemented either by applying a dedicated k -step prediction model for each prediction horizon or by introducing a dependency on prediction horizon in the coefficients of coefficient function types of models. The latter approach has the advantage of enforcing that the parameters must vary smoothly with the prediction horizon. The disadvantage is that the same model structure is used for all prediction horizons.

An on-line application, WPPT, implementing some of the proposed models are used at three of the major Danish power utilities with respect to wind energy. The experiences gained by the utilities have been presented, and the conclusion is that WPPT generally produces reliable predictions, which are used directly in the economic load dispatch and the day to day electricity trade.

Bibliography

- Andersen, H. N. & Brydov, P. M. (1987), District heating in denmark; research and technological development, Technical report, The Danish Ministry of Energy, Copenhagen.
- Arvastson, L. (2001), Stochastic Modelling and Operational Optimization in District Heating Systems, PhD thesis, Division of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden.
- Cappelen, J. & Jørgensen, B. V. (2001), Danmarks klima 2000, Technical Report Teknisk rapport 01-06, Danish Meteorological Institute, Danish Meteorological Institute, Copenhagen, Denmark. In Danish.
- Clarke, D. W., Mohtadi, C. & Tuffs, P. S. (1987), ‘Generalized predictive control – part I. The basic algorithm’, *Automatica* **23**, 137–148.
- 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.
- Giebel, G., Landberg, L., Joensen, A., Nielsen, T. S. & Madsen, H. (2000), The zephyr-project – the next generation prediction system, in ‘Proceedings of the Wind Power for the 21st Century Conference’, Kassel, Germany.
- Hastie, T. & Tibshirani, R. (1990), *Generalized Additive Models*, Chapman & Hall, London/New York.
- Joensen, A. (1997), Models and methods for predicting wind power, Master’s thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. In Danish.
- Jonsson, B. (1994), ‘Prediction with a linear regression model and errors in a regressor’, *International Journal of Forecasting* **10**, 549–555.
- Landberg, L. (1997a), A mathematical look at a physical power prediction model, in ‘Proceedings of the European Wind Energy Conference’, Irish Wind Energy Association, Dublin, Eire, pp. 406 – 408.
- Landberg, L. (1997b), Predicting the power output from wind farms, in ‘Proceedings of the European Wind Energy Conference’, Irish Wind Energy Association, Dublin, Eire, pp. 747 – 750.

- Landberg, L., Hansen, M. A., Vesterager, K. & Bergstrøm, W. (1997), Implementing wind forecasting at a utility, Technical Report Risø-R-929(EN), Risø, Risø National Laboratory, Roskilde, Denmark.
- Landberg, L., Watson, S. J., Halliday, J., Jørgensen, J. U. & Hilden, A. (1994), Short-term prediction of local wind conditions, Technical Report Risø-R-702(EN), Risø, Risø National Laboratory, Roskilde, Denmark.
- Ljung, L. & Söderström, T. (1983), *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA.
- Madsen, H., Nielsen, T. S. & Søgaard, H. (1996), *Control of Supply Temperature*, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Madsen, H., Palsson, O. P., Sejling, K. & Søgaard, H. (1990), *Models and Methods for Optimization of District Heating Systems*, Vol. I: Models and Identification Methods, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Madsen, H., Palsson, O. P., Sejling, K. & Søgaard, H. (1992), *Models and Methods for Optimization of District Heating Systems*, Vol. II: Models and Control Methods, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Madsen, H., Sejling, K., Nielsen, H. A. & Nielsen, T. S. (1996), *Models and Methods for Predicting Wind Power*, ELSAM, Fredericia, Denmark.
- Madsen, H., Sejling, K., Nielsen, T. S. & Nielsen, H. A. (1995), Wind power prediction tool in control dispatch centres, Technical report, ELSAM, Fredericia, Denmark.
- Marti, I., Nielsen, T. S., Madsen, H., Navarro, J. & Barquero, C. G. (2001), Prediction models in complex terrain, in 'Proceedings of the European Wind Energy Conference', Copenhagen, Denmark.
- Mayne, D. Q., Rawlings, J. B., Rao, C. V. & Scokaert, P. O. M. (2000), 'Constrained model predictive control: Stability and optimality', *Automatica* **36**, 789–814.
- Nielsen, H. A. & Madsen, H. (2000), Forecasting the heat consumption in district heating systems using meteorological forecasts, Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

- Nielsen, H. A., Nielsen, T. S. & Madsen, H. (2001), Load scheduling for decentralized CHP plants, Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Nielsen, T. S. & Madsen, H. (1997), Using meteorological forecasts in on-line predictions of wind power, Technical Report IMM-REP-1997-31, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Nielsen, T. S., Madsen, H. & Christensen, H. S. (2000), WPPT - a tool for wind power prediction, *in* 'Proceedings of the Wind Power for the 21st Century Conference', Kassel, Germany.
- Nielsen, T. S., Madsen, H., Nielsen, H. A., Giebel, G. & Landberg, L. (2002), Prediction of regional wind power, *in* 'Proceedings of the 2002 Global Windpower Conference', Paris, France.
- Nielsen, T. S., Madsen, H., Nielsen, H. A., Landberg, L. & Giebel, G. (2001), Zephyr - the prediction models, *in* 'Proceedings of the European Wind Energy Conference', Copenhagen, Denmark.
- Nielsen, T. S., Madsen, H., Nielsen, H. A. & Tøfting, J. (1999), Using meteorological forecasts in on-line predictions of wind power, Technical report, Eltra, Fredericia, Denmark.
- Oliker, I. (1980), 'Steam turbines for cogeneration power plants', *Journal of Engineering for Power* **102**, 482–485.
- Palsson, O. P., Madsen, H. & Søgaaard, H. (1994), 'Generalized predictive control for non-stationary systems', *Automatica* **30**, 1991–1997.
- Tong, H. (1990), *Non-linear Time Series, A Dynamic System Approach*, Clarendon Press, Oxford.

Papers

PAPER A

Conditional parametric ARX-models

Conditional parametric ARX-models

Henrik Aalborg Nielsen, Torben Skov Nielsen, and Henrik Madsen

Abstract

In this paper conditional parametric ARX-models are suggested and studied by simulation. These non-linear models are traditional ARX-models in which the parameters are replaced by smooth functions. The estimation method is based on the ideas of locally weighted regression. It is demonstrated that kernel estimates (local constants) are in general inferior to local quadratic estimates. For the considered application, modelling of temperatures in a district heating system, the input sequences are correlated. Simulations indicate that correlation to this extent results in unreliable kernel estimates, whereas the local quadratic estimates are quite reliable.

Keywords: Non-linear models; non-parametric methods; kernel estimates; local polynomial regression; ARX-models; time series.

1 Introduction

Linear models in which the parameters are replaced by smooth functions are denoted varying-coefficients models. Estimation in these models has been developed for the regression framework, see e.g. (Hastie & Tibshirani 1993). In this paper a special class of the models in which all coefficients are controlled by the same argument are considered, these are also denoted conditional parametric models, see e.g. (Anderson, Fang & Olkin 1994). This class of models is applied for autoregressive processes with external input and the resulting models will be denoted conditional parametric ARX-models. These models are similar to smooth threshold autoregressive models, see e.g. (Tong 1990), but more general since a transition is related to each coefficient and a non-parametric form is assumed for these transitions. The method of estimation is closely related to locally weighted regression (Stone 1977, Cleveland 1979, Cleveland & Devlin 1988, Cleveland, Devlin & Grosse 1988). Because of the autoregressive property the fitted values will not be linear combinations of the observations and results concerning bias and variance obtained for the regression framework (Cleveland & Devlin 1988, Hastie & Loader 1993) can not be used. For this reason the method is studied by simulation.

Our interest in these models originates from the modelling of temperatures in a district heating system. In such a system the energy needed for heating and hot tap-water in the individual households is supplied from a central heating utility. The energy is distributed as hot water through a system of pipelines covering the area supplied. In the system an increased energy demand is first met by increasing the flow rate in the system and, when the maximum flow rate is reached, by increasing the supply temperature at the utility. The energy demand in a district heating system typically exhibits a strong diurnal variation with the peak load occurring during the morning hours. A similar pattern can be found in the observed flow rates, although this is also influenced by variations in the supply temperature. Consequently, the time delay for an increase in the supply temperature to be observed in a household inlet also has a diurnal variation.

Models of the relationship between supply temperature and inlet temperature are of high interest from a control point of view. Previous studies have led to a library of ARX-models with different time delays and with a diurnal variation in the model parameters. Methods for on-line estimating of the varying time delay as well as a controller which takes full advantage of this model structure have previously been published (Søgaard & Madsen 1991, Palsson, Madsen & Søgaard 1994, Madsen, Nielsen & Søgaard 1996). A more direct approach is to use one ARX-model but with parameters replaced by smooth functions of the flow rate. This approach is addressed in this paper. This further has the advantage, that the need for on-line estimation of the time delay is eliminated.

In Section 2 the conditional parametric model and the estimation methods are outlined. The performance of the estimators are studied by simulation in Section 3. An application to real data from a district heating system is described in Section 4. Some of this material has previously been presented at a conference, see (Nielsen, Nielsen & Madsen 1997).

2 Model and estimation

A conditional parametric model is a linear regression model with the parameters replaced by smooth functions. The name of the model comes from observing that if the argument of the functions are fixed then the model is an ordinary linear model.

Models of this type are briefly described in the literature (Hastie & Tibshirani

1993, Anderson et al. 1994). Below a more general description of some aspects of the method is presented. The method of estimation is closely related to locally weighted regression (Stone 1977, Cleveland & Devlin 1988, Cleveland et al. 1988, Hastie & Loader 1993).

2.1 The regression case

Assume that observations of the response y_t and the explanatory variables \mathbf{x}_t and \mathbf{z}_t exist for observation numbers $t = 1, \dots, n$. An intercept is included in the model by putting the first element of \mathbf{z}_t equal to one. The conditional parametric model for this setup is

$$y_t = \mathbf{z}_t^T \boldsymbol{\theta}(\mathbf{x}_t) + e_t \quad (t = 1, \dots, n), \quad (1)$$

where e_t is i.i.d. $N(0, \sigma^2)$ and $\boldsymbol{\theta}(\cdot)$ is a vector of functions, with values in \mathcal{R} , to be estimated. The functions $\boldsymbol{\theta}(\cdot)$ is only estimated for distinct values of their argument \mathbf{x} . In this paper the approach taken is to estimate $\boldsymbol{\theta}(\cdot)$ at points sufficiently close for linear interpolation. Below \mathbf{x} denotes a single such point within the space spanned by the observations $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$.

The estimation of $\boldsymbol{\theta}(\mathbf{x})$ is accomplished by calculating the weighted least squares estimate of the parameter vector, i.e. close to \mathbf{x} the function $\boldsymbol{\theta}(\cdot)$ is approximated by a constant vector. The weight on observation t is related to the distance from \mathbf{x} to \mathbf{x}_t , such that

$$w_t(\mathbf{x}) = W(\|\mathbf{x}_t - \mathbf{x}\|/h(\mathbf{x})), \quad (2)$$

where $W : \mathcal{R}_0 \rightarrow \mathcal{R}_0$ is a nowhere increasing function. In this paper the tricube function

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

is used. $\|\mathbf{x}_t - \mathbf{x}\|$ is the Euclidean distance between \mathbf{x}_t and \mathbf{x} . The scalar $h(\mathbf{x}) > 0$ is called the bandwidth. If $h(\mathbf{x})$ is constant for all values of \mathbf{x} it is denoted a fixed bandwidth. If $h(\mathbf{x})$ is chosen so that a certain fraction (α) of the observations is within the bandwidth it is denoted a nearest neighbour bandwidth. If \mathbf{x} has dimension of two, or larger, scaling of the individual elements of \mathbf{x}_t before applying the method should be considered, see e.g. (Cleveland & Devlin 1988). A rotation of the coordinate system, in which \mathbf{x}_t is measured, could also be relevant. Note that if $\mathbf{z}_t = 1$ for all t the method of estimation reduces to determining the scalar $\hat{\theta}(\mathbf{x})$ so that $\sum_{t=1}^n w_t(\mathbf{x})(y_t - \hat{\theta}(\mathbf{x}))^2$ is minimized, i.e. the method reduces to kernel estimation, see also (Härdle 1990,

p. 30) or (Hastie & Loader 1993). For this reason the described method of estimation of $\boldsymbol{\theta}(\mathbf{x})$ in (1) is called kernel or local constant estimation.

If the bandwidth $h(\mathbf{x})$ is sufficiently small the approximation of $\boldsymbol{\theta}(\cdot)$ as a constant vector near \mathbf{x} is good. The consequence is, however, that a relatively low number of observations is used to estimate $\boldsymbol{\theta}(\mathbf{x})$, resulting in a noisy estimate, or bias if the bandwidth is increased. See also the comments on kernel estimates in (Anderson et al. 1994) or (Hastie & Loader 1993).

It is, however, well known that locally to \mathbf{x} the elements of $\boldsymbol{\theta}(\cdot)$ may be approximated by polynomials, and in many cases these will provide better approximations for larger bandwidths than those corresponding to local constants. Local polynomial approximations are easily included in the method described. Let $\theta_j(\cdot)$ be the j 'th element of $\boldsymbol{\theta}(\cdot)$ and let $\mathbf{p}_d(\mathbf{x})$ be a column vector of terms in a d -order polynomial evaluated at \mathbf{x} . If for instance $\mathbf{x} = [x_1 \ x_2]^T$ then $\mathbf{p}_2(\mathbf{x}) = [1 \ x_1 \ x_2 \ x_1^2 \ x_1x_2 \ x_2^2]^T$. Furthermore, let $\mathbf{z}_t = [z_{1t} \ \dots \ z_{pt}]^T$. With

$$\mathbf{u}_t^T = \left[z_{1t}\mathbf{p}_{d(1)}^T(\mathbf{x}_t) \dots z_{jt}\mathbf{p}_{d(j)}^T(\mathbf{x}_t) \dots z_{pt}\mathbf{p}_{d(p)}^T(\mathbf{x}_t) \right] \quad (4)$$

and

$$\hat{\boldsymbol{\phi}}^T(\mathbf{x}) = [\hat{\boldsymbol{\phi}}_1^T(\mathbf{x}) \dots \hat{\boldsymbol{\phi}}_j^T(\mathbf{x}) \dots \hat{\boldsymbol{\phi}}_p^T(\mathbf{x})], \quad (5)$$

where $\hat{\boldsymbol{\phi}}_j(\mathbf{x})$ is a column vector of local constant estimates at \mathbf{x} corresponding to $z_{jt}\mathbf{p}_{d(j)}(\mathbf{x}_t)$. The estimation is handled as described above, but fitting the linear model

$$y_t = \mathbf{u}_t^T \boldsymbol{\phi}_x + e_t \quad (t = 1, \dots, N), \quad (6)$$

locally to \mathbf{x} . Hereafter the elements of $\hat{\boldsymbol{\theta}}(\mathbf{x})$ are calculated as

$$\hat{\theta}_j(\mathbf{x}) = \mathbf{p}_{d(j)}^T(\mathbf{x}) \hat{\boldsymbol{\phi}}_j(\mathbf{x}) \quad (j = 1, \dots, p). \quad (7)$$

When $z_j = 1$ for all j this method is identical to the method by Cleveland & Devlin (1988), with the exception that Cleveland & Devlin center the elements of \mathbf{x}_i used in $\mathbf{p}_d(\mathbf{x}_t)$ around \mathbf{x} and consequently $\mathbf{p}_d(\mathbf{x}_t)$ must be recalculated for each value of \mathbf{x} considered.

Let \mathbf{U} be a matrix with rows \mathbf{u}_t^T , let $\mathbf{W}(\mathbf{x}_t)$ be a diagonal matrix containing the weights on the observations when $\mathbf{x} = \mathbf{x}_t$, and let \mathbf{y} be a column vector containing the observations. Using this notation the fitted value for observation t can be written $\hat{y}_t = \mathbf{u}_t^T [\mathbf{U}^T \mathbf{W}(\mathbf{x}_t) \mathbf{U}]^{-1} \mathbf{U}^T \mathbf{W}(\mathbf{x}_t) \mathbf{y}$. Since no element of \mathbf{y} is used in \mathbf{u}_t^T it then follows that the fitted values are linear combinations of the observations. This property is shared with, e.g., locally weighted regression and forms the basis of discussions regarding bias and variance, see e.g. (Cleveland & Devlin 1988, Hastie & Loader 1993).

The weighted least squares problem is solved by the algorithm described in (Miller 1992). The algorithm was originally written in Fortran but here a port to C by A. Shah is used (pub/C-numanal/as274.fc.tar.z from usc.edu, using anonymous ftp).

2.2 ARX-models

It is well known that a linear ARX-model can be written in the form (1), where t is the time index, $\theta(\cdot)$ is a constant parameter vector and z_t contains input and lagged values of y_t . For this reason the method described in Section 2.1 is easily extended to models of the form:

$$y_t = \sum_{i \in L_y} a_i(x_{t-m})y_{t-i} + \sum_{i \in L_u} b_i(x_{t-m})u_{t-i} + e_t, \quad (8)$$

where t is the time index, y_t is the response, x_t and u_t are inputs, $\{e_t\}$ is i.i.d. $N(0, \sigma^2)$, L_y and L_u are sets of positive integers defining the autoregressive and input lags in the model, and m is a positive integer. Finally, $a_i(\cdot)$ and $b_i(\cdot)$ are unknown but smooth functions which are to be estimated. Extensions to multivariate x_t and u_t are strait forward. Models belonging to the class defined by (8), including the multivariate extensions just mentioned, will be denoted conditional parametric ARX-models.

For this class of models the matrix U , which were introduced in the previous section, will depend on the observations of the response and hence the fitted values will not be linear combinations of the observations.

3 Simulation study

To study the estimation method described in Section 2.1 for models of the class (8) simulations using the model

$$y_t = a_1(x_{t-1})y_{t-1} + \sum_{i \in \{2,4\}} b_i(x_{t-1})u_{t-i} + e_t, \quad (9)$$

are performed. In the simulations e_t is normally distributed white noise with zero mean and variance σ_e^2 . The functions used are

$$\begin{aligned} a_1(x) &= 0.5 + 0.5/(1 + e^{4x-0.5}) \quad , \\ b_2(x) &= 0.045\sqrt{\Phi_{(0.09,0.015)}(x)} \quad , \text{ and} \\ b_4(x) &= 0.025\sqrt{\Phi_{(0.05,0.01)}(x)} \quad , \end{aligned}$$

where $\Phi_{(\mu,\sigma)}(\cdot)$ is the Gaussian probability distribution function with mean μ and standard deviation σ . Note that for all values of x used in the simulations the values of $b_2(x)$ and $b_4(x)$ are positive, and the pole ($a_1(x)$) is between 0.5 and 1. However, in the simulations the pole takes values between 0.75 and 0.79.

In the district heating case it is not possible to experiment with the system and hence the response corresponding to uncorrelated inputs can not be obtained. To study the performance of the estimation method in a relevant setting measurements of half-hourly averages of supply temperature (u_t) and flow (x_t) from a running district heating plant is used. In order to investigate certain aspects of the method simulations with white noise inputs are also presented. The distributions of the white noise inputs are chosen so that the range corresponds to the range of observations from the district heating plant.

In the simulations local constant (kernel) estimates are compared to local quadratic estimates. Local quadratic estimates are chosen rather than local linear estimates since the first type is expected to have less bias near peaks and to have similar performance in other regions.

In Section 3.1 simulated white noise input with a uniform distribution are used and the performance of local constant (kernel) estimates are compared with local quadratic estimates. Simulated white noise input with a normal distribution are used in Section 3.2 and local quadratic estimates are investigated for cases where data is sparse in some regions. Simulations with correlated input are addressed in Section 3.3 and in this section local constant and local quadratic estimates are compared when a lag not included in the simulation is included in the estimation. Histograms of the sequences of x_t are shown in Figure 1.

The length of the simulated series is approximately 40% of the length of the real data series used in Section 4. To obtain roughly the same variance of the estimates in the simulation and in the application, the simulation variance is chosen to be 40% of the variance obtained in Section 4. With a robust estimate of the variance obtained from the central 95% of the residuals, the

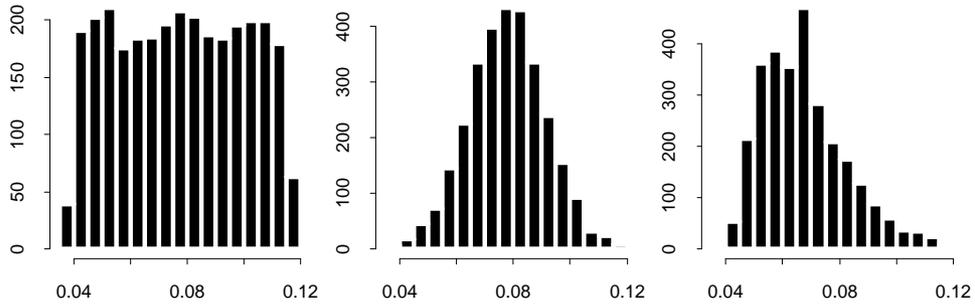


Figure 1: Histograms of x_t used in the simulations. From left to right; uniformly distributed white noise, normally distributed white noise, and measurements of flow (m^3/s) from a running plant.

simulation variance of e_t in (9) is 0.27^2 .

The function values are estimated at 50 equidistantly spaced points. Results for nearest neighbour bandwidths with $\alpha = 0.1, 0.2, \dots, 0.6$ are presented. In the figures the true values are indicated by a dotted line.

3.1 Uniformly distributed input

Minimum and maximum observed values of supply temperature and low pass filtered flow (c.f. Section 3.3) are used as limits of the uniform distribution from which the white noise input sequences, of length 3000, are generated, i.e. x_t is i.i.d. $U(0.0390, 0.116)$ and u_t is i.i.d. $U(78.2, 94.2)$.

In Figure 2 local constant (kernel) and local quadratic estimates of the functions $a_1(\cdot)$, $b_2(\cdot)$, and $b_4(\cdot)$ in (9) are shown. It is clearly seen that the local quadratic is superior to the local constant approximation. This is especially true near the border of the interval spanned by the observations and in areas with large curvature. Also note that the local quadratic approximation is quite insensitive to the choice of bandwidth.

3.2 Normally distributed input

To investigate the performance of the local quadratic estimates in cases where the data are sparse near the border of the interval spanned by the observa-

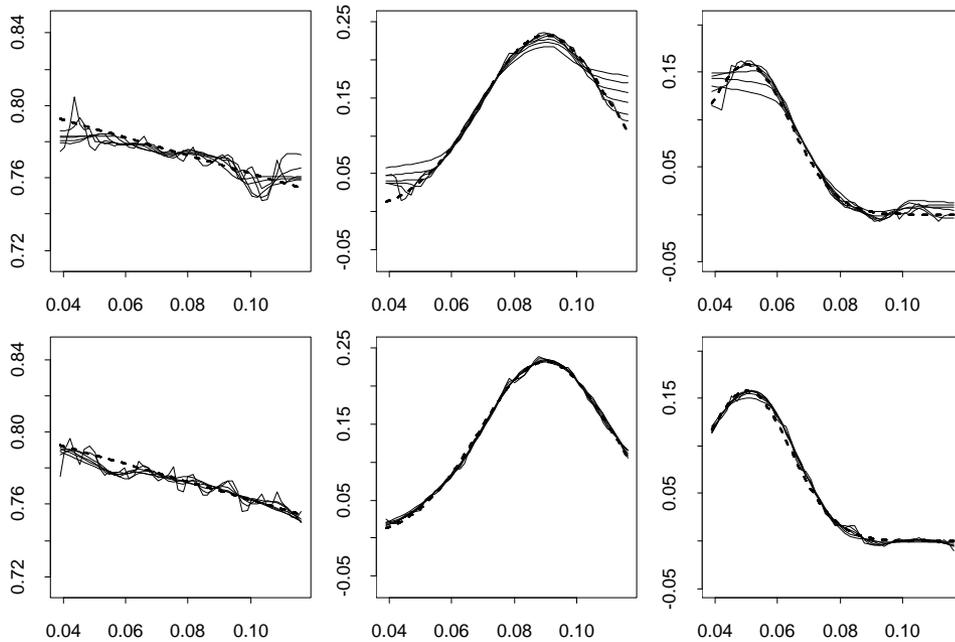


Figure 2: Uniformly distributed white noise input. Local constant (top row) and local quadratic (bottom row). From left to right; $\hat{a}_1(x)$, $\hat{b}_2(x)$, and $\hat{b}_4(x)$.

tions a simulation corresponding to Section 3.1 is performed with normally distributed white noise input.

The mean and variance of the normal distributions are chosen so that the lower and upper limits of the uniform distributions of Section 3.1 is exceeded with probability 0.25%, i.e. x_t is i.i.d. $N(0.0776, 0.0138^2)$ and u_t is i.i.d. $N(86.2, 2.86^2)$, 15 values of each of the simulated input sequences are expected to exceed the limits. Values exceeding the limits are truncated.

In Figure 3 local quadratic estimates of the functions $a_1(\cdot)$, $b_2(\cdot)$, and $b_4(\cdot)$ in (9) are shown. When compared with the distribution of x_t in Figure 1 it is seen that the estimates are relatively far from the true values in regions with sparse data. To further investigate this aspect Figure 4 indicate the pointwise distribution of the estimates, calculated for $\alpha = 0.3$ at 20 equidistantly spaced points.

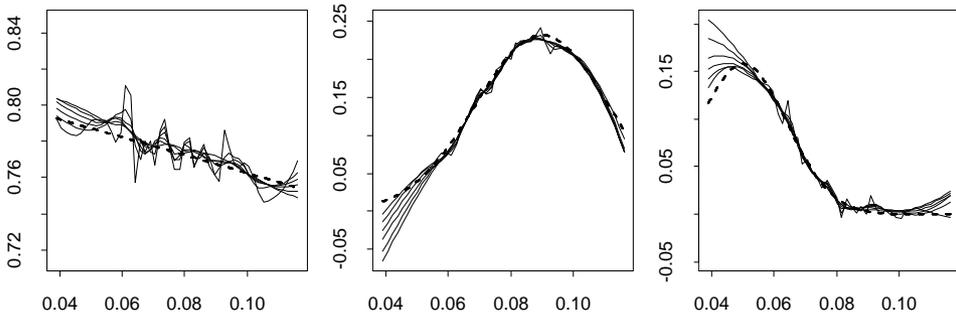


Figure 3: Normally distributed white noise input and local quadratic estimates. From left to right; $\hat{a}_1(x)$, $\hat{b}_2(x)$, and $\hat{b}_4(x)$.

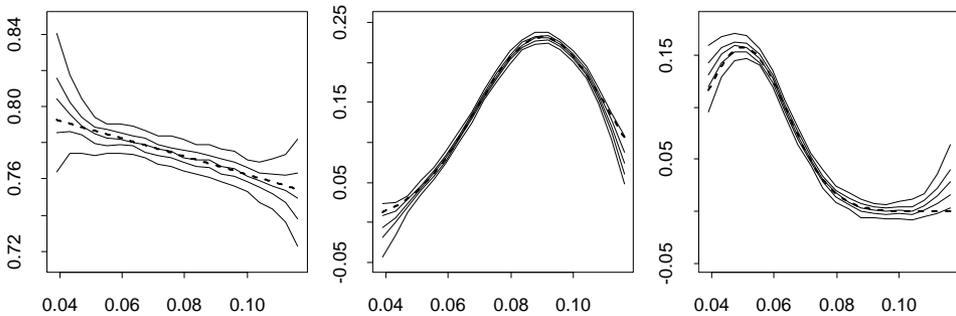


Figure 4: Quantiles (5%, 25%, 50%, 75%, and 95%) of one hundred simulations ($\alpha = 0.3$) with the same input sequences as in Figure 3. From left to right; $\hat{a}_1(x)$, $\hat{b}_2(x)$, and $\hat{b}_4(x)$.

3.3 Measurements from a running plant as input

In order to address the influence of correlated input sequences on the performance of the estimation method, measurements of 2873 half-hourly averages of flow ($x_t, m^3/s$) and supply temperature ($u_t, °C$) from a running district heating plant are used. To mimic the real situation (c.f. Section 4.2) the flow is filtered with exponential smoothing using a forgetting factor of 0.8.

Figure 5 shows the results obtained with local quadratic estimates when the structure of the true system is assumed known. Compared with the previous plots it is clear that the correlation of the input sequences deteriorates the estimates, even in regions where data is dense. The main characteristics are, however, still identifiable.

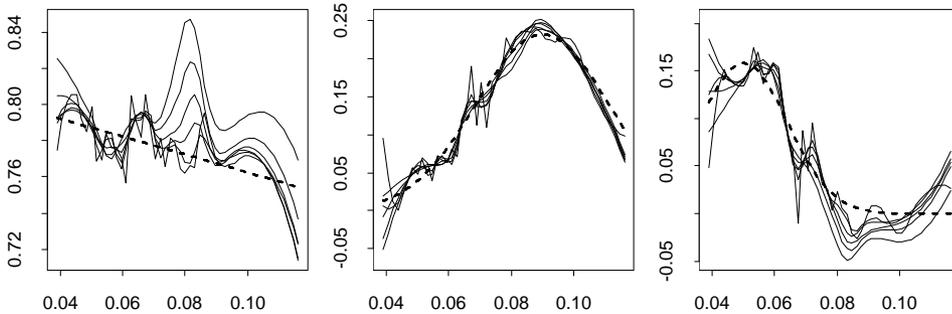


Figure 5: Measurements of flow (m^3/s) from a running plant as input and local quadratic estimates. From left to right; $\hat{a}_1(x)$, $\hat{b}_2(x)$, and $\hat{b}_4(x)$.

In Figures 6 and 7 the results when including $b_3(x_{t-1})u_{t-3}$ in the model used for estimation are shown. In this case the local constant estimates perform very poor. The local quadratic estimates perform well, except for regions where data is sparse (compare with the histogram in Figure 1).

The performance in border regions can be improved if prior knowledge about the individual functions is available. If the estimates corresponding to Figure 7 are recalculated with the local approximations $a_1(x)$ - straight line, $b_2(x)$ and $b_4(x)$ - quadratic, $b_3(x)$ - constant. This removes much of the fluctuation of $\hat{b}_3(x)$ and also increases the performance of the remaining estimates.

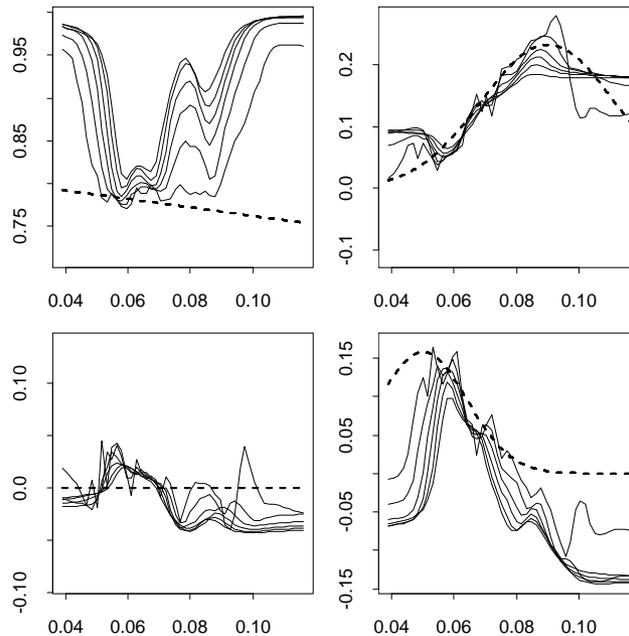


Figure 6: Measurements of flow (m^3/s) from a running plant as input and local constant estimates. From left to right; $\hat{a}_1(x)$, $\hat{b}_2(x)$, $\hat{b}_3(x)$, and $\hat{b}_4(x)$.

4 Application to a real system

In this section the method is applied to data obtained from the district heating plant “Høje Taastrup Fjernvarme” near Copenhagen in Denmark. For the periods considered the energy were supplied from one plant only.

4.1 Data

The data covers the periods from 1 April until 31 May and 1 September until 17 December, 1996. Data consists of five minute samples of supply temperature and flow at the plant together with the network temperature at a major consumer, consisting of 84 households. In 1996 that consumer used 1.2% of the produced energy.

The measurements of the flow and temperatures are occasionally erroneous. In order to find these measurements the data were investigated by visual methods; five minutes samples were excluded for 999 time values. Based on

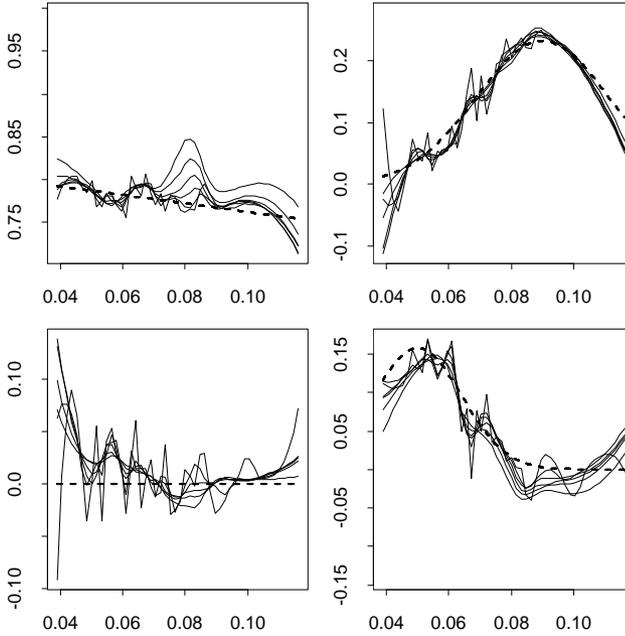


Figure 7: Measurements of flow from a running plant as input and local quadratic estimates. From left to right; $\hat{a}_1(x)$, $\hat{b}_2(x)$, $\hat{b}_3(x)$, and $\hat{b}_4(x)$.

the observations half-hour averages were calculated, and these were excluded if any of the five minute samples were excluded or missing. In total 724 half-hour averages were excluded, yielding a total of 7388 half-hour values which are assumed valid.

4.2 Models

The network temperature is modelled by models of the structure (8), where one time step corresponds to 30 minutes, y_t represents the network temperature, u_t represents the supply temperature, and x_t represents a filtered value of the flow.

Consider a simple district heating network with only one consumer and one plant. Disregarding diffusion, the time delay $\tau(t)$ of a water particle leaving the plant at time t is related to the distance d between plant and consumer by the relation $d = \int_t^{t+\tau(t)} v(s)ds$, where $v(s)$ is the velocity of the particle at time s . From this simple model it is seen that the flow should be filtered by averaging over past values, but with varying horizon. In practice this

is implemented by assuming a known volume for the district heating pipe between the plant and the major consumer. Hereafter the filtered values of the flow is found as the average of the five minute samples filling the pipe, see (Nielsen et al. 1997) for a brief description of how the volume is estimated.

Below results corresponding to two different model structures are presented, (i) the conditional parametric Finite Impulse Response (FIR) model $y_t = \sum_{i=0}^{30} b_i(x_t)u_{t-i} + e_t$, and (ii) the conditional parametric ARX-model $y_t = a(x_t)y_{t-1} + \sum_{i=3}^{15} b_i(x_t)u_{t-i} + e_t$.

4.3 Results

For both the FIR-model and the ARX-model described in the previous section local quadratic estimates and nearest neighbour bandwidths are used. The coefficient functions are estimated at 50 equally spaced points ranging from the 2% to the 98% quantile of the filtered flow.

The coefficient functions of the FIR-model are estimated for α equal to 0.1, 0.2, ..., 0.9. In Figure 8 the impulse response as a function of the flow is displayed for $\alpha = 0.4$. Equivalent plots for the remaining bandwidths revealed that for $\alpha \leq 0.2$ the fits are too noisy, whereas in all cases sufficiently smoothness is obtained for $\alpha = 0.5$. Only minor differences in the fits are observed for $\alpha \in \{0.3, 0.4, 0.5\}$.

In Figure 9 a contour plot corresponding to Figure 8 is shown. From the plot the varying time delay of the system is revealed, it seems to vary from three lags when the flow is large to approximately ten lags when the flow is near its minimum. The peak at lag zero for the lower flows is clearly an artifact. It is probably due to the correlation structure of the flow. The sample autocorrelation function shows a local peak at lag 24, and therefore the peak at lag zero can be compensated by higher lags.

The residuals of the FIR-model show a diurnal variation. The sample inverse autocorrelation function of the residuals indicates that a AR(1)-model can account for most of the correlation, this corresponds to the order of the auto regression used in (Madsen et al. 1996) which considers the same district heating system.

Based on the results obtained for the FIR-model the ARX-model described in Section 4.2 is purposed. The coefficient functions of the model are estimated for α equal to 0.3, 0.4, and 0.5. Very similar results are obtained for the three

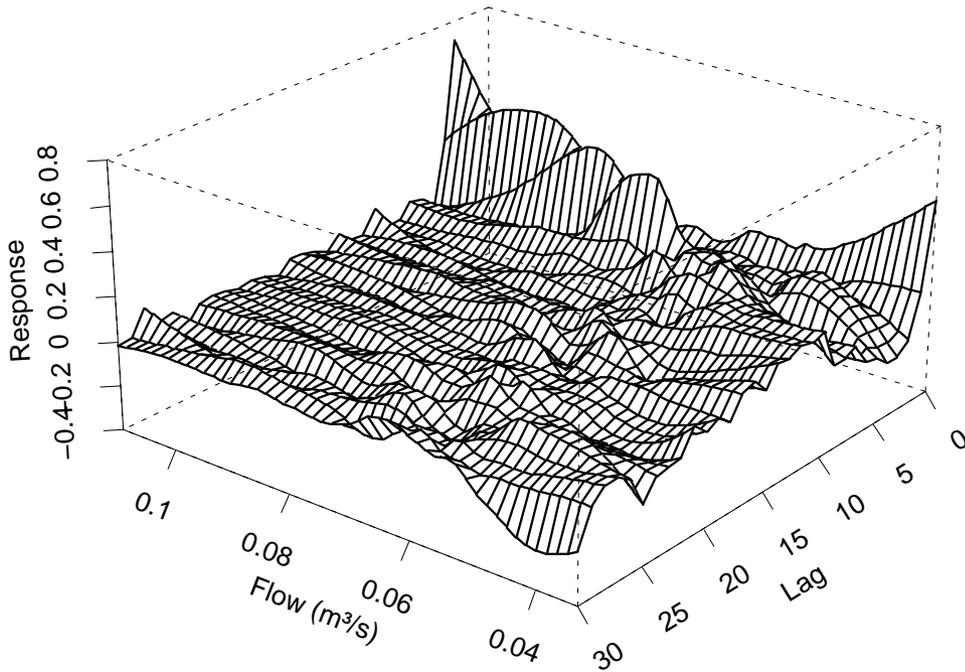


Figure 8: Impulse response function of the FIR-model.

different bandwidths. For $\alpha = 0.4$ the impulse response as a function of the flow is displayed in Figure 10. The varying time delay is clearly revealed. In Figure 11 the stationary gain of the two models and the pole of the ARX-model are shown. From the values of the stationary gain it is seen that the temperature loss changes from 6% when the flow is large to 12-15% when the flow is small.

The residuals of the ARX-model show a weak diurnal variation, only very weak autocorrelation (0.07 in lag two), and no dependence of supply temperature and flow. If the coefficients of the ARX-model are fitted as global constants the central 95% of the residuals spans 2.1 °C opposed to 1.7 °C for the conditional parametric ARX-model.

5 Conclusions

In this paper conditional parametric ARX-models are suggested and studied. These non-linear models are obtained as traditional ARX-models in which the parameters are replaced by smooth functions. Simulations show that

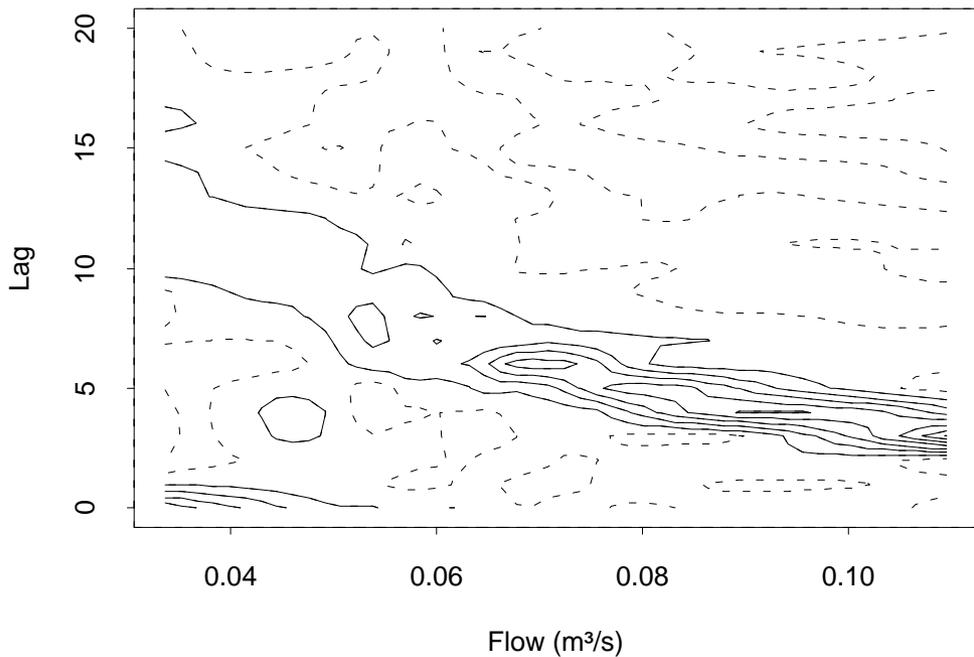


Figure 9: Contour plot of the impulse response function of the FIR-model ($\alpha = 0.4$). The contour lines is plotted from -0.1 to 0.7 in steps of 0.1 . Lines corresponding to non-positive values are dotted.

kernel estimates (local constants) in general are inferior to local quadratic estimates. In the case of correlated input sequences the kernel estimates are rather unreliable, but local quadratic estimates are in general quite reliable.

If the observations are sparse in border regions of the variable(s) defining the weight, the local quadratic estimate may have increased bias and/or variance in these regions. Consequently it is important to investigate the distribution of the variable(s) defining the weight when interpreting the estimates.

When applied to real data from a district heating system in which case numerous coefficient functions is necessary the method seem to perform well in that the results obtained are quite plausible. Furthermore, the residuals behave appropriately.

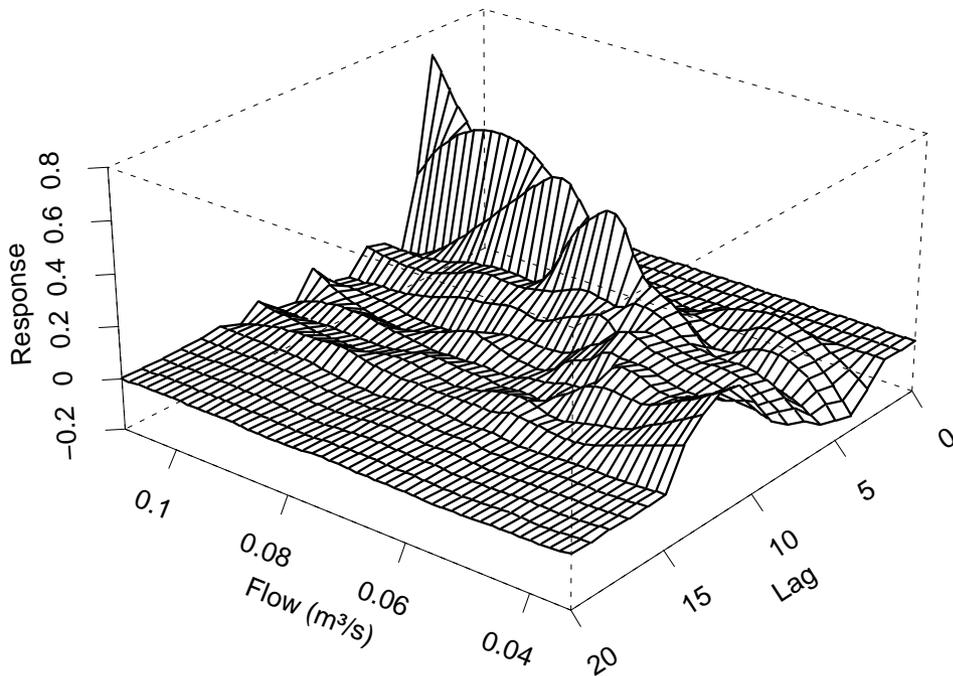


Figure 10: Impulse response function of the ARX-model.

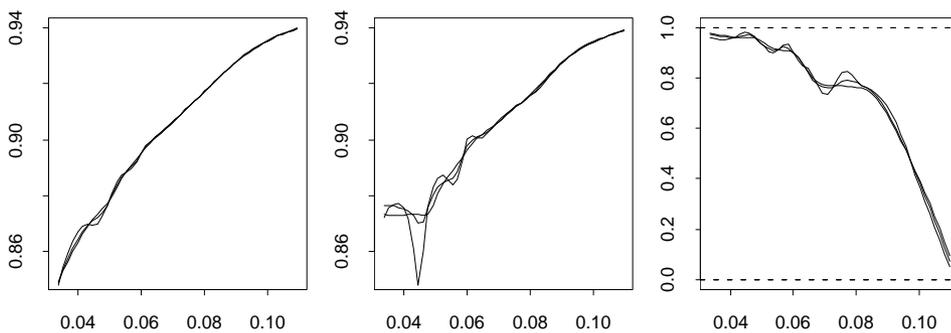


Figure 11: Stationary gain of the FIR-model (left) and ARX-model (middle) and the pole of the ARX-model (right) all plotted against the flow and for α equal to 0.3, 0.4, and 0.5.

6 Discussion

The poor performance of the kernel estimates is probably related to the inability of the approximation to fit the data locally. For the estimation of a regression function ($z_t = 1$) this has been described by Hastie & Loader (1993).

In this paper the selection of the bandwidth is not considered. If prior knowledge of the curvature of the coefficient-functions are available this may be used to select a bandwidth such that a polynomial of a certain order is a reasonable local approximation. A more informal method is to select a bandwidth for which the estimated functions attains a certain degree of smoothness. In the context of time series it seems appropriate to use forward validation (Hjorth 1994) for guidance on bandwidth selection. As for cross validation, this approach will probably yield a flat optimum (Hastie & Tibshirani 1990, p. 42), and so, some judgment is called upon.

The interpretation of the function value estimates would be greatly facilitated by results concerning the statistical properties of the estimates. It is believed that it is possible to derive such properties based on the results obtained for locally weighted regression, see e.g. (Cleveland & Devlin 1988). Without the theoretical work it is still possible to address the precision of the function estimates by for instance bootstrapping.

The conditional parametric ARX-models may be too flexible for some applications. Especially it is of interest to be able to fit some of the functions as global constants. This could probably be accomplished by applying the methods described in (Hastie & Tibshirani 1993).

For temperature control applications in district heating systems it is expected that the identified conditional parametric ARX-model is superior to more traditional adaptive controllers as considered for instance in (Madsen et al. 1996), because the non-linearity is directly modelled as opposed to the successive linearizations used in traditional adaptive estimation techniques.

Bibliography

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. (1979), ‘Robust locally weighted regression and smoothing scatterplots’, *Journal of the American Statistical Association* **74**, 829–836.
- 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.
- Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, UK.
- Hastie, T. & Loader, C. (1993), ‘Local regression: Automatic kernel carpentry’, *Statistical Science* **8**, 120–129. (Discussion: p129-143).
- Hastie, T. & Tibshirani, R. (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.
- Madsen, H., Nielsen, T. S. & Sogaard, H. T. (1996), *Control of Supply Temperature in District Heating Systems*, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Miller, A. J. (1992), ‘[Algorithm AS 274] Least squares routines to supplement those of Gentleman’, *Applied Statistics* **41**, 458–478. (Correction: 94V43 p678).
- 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.
- Palsson, O., Madsen, H. & Sogaard, H. (1994), ‘Generalized predictive control for non-stationary systems’, *Automatica* **30**, 1991–1997.
- Sogaard, H. & Madsen, H. (1991), On-line estimation of time-varying delays in district heating system, in ‘Proc. of the 1991 European Simulation Multiconference’, pp. 619–624.

Stone, C. J. (1977), 'Consistent nonparametric regression', *The Annals of Statistics* **5**, 595–620. (C/R: p620-645).

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

PAPER B

B

Tracking time-varying coefficient-functions

Tracking time-varying coefficient-functions

Henrik Aa. Nielsen¹, Torben S. Nielsen¹, Alfred K. Joensen¹,
Henrik Madsen¹, and Jan Holst²

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 (Nielsen et al. 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

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

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

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.

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

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

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

and

$$\boldsymbol{\phi}_u^T = [\phi_{u,1}^T \ \dots \ \phi_{u,j}^T \ \dots \ \phi_{u,p}^T], \quad (3)$$

where $\phi_{u,j}$ 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

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

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

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

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

$$w_u(\mathbf{u}_s) = W\left(\frac{\|\mathbf{u}_s - \mathbf{u}\|}{\tilde{h}(\mathbf{u})}\right), \quad (6)$$

where $\|\cdot\|$ denotes the Euclidean norm, $\tilde{h}(\mathbf{u})$ 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

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

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

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

where $\hat{\phi}_j(\mathbf{u})$ is the WLS estimate of $\phi_{u,j}$. 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 $\tilde{h}(\mathbf{u})$ is constant for all values of \mathbf{u} it is denoted a fixed bandwidth. If $\tilde{h}(\mathbf{u})$ is chosen so that a certain fraction α of the observations fulfill $\|\mathbf{u}_s - \mathbf{u}\| \leq \tilde{h}(\mathbf{u})$ 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 $s = 1, \dots, t$ are available, the adaptive least squares estimate of the parameters ϕ relating the explanatory variables \mathbf{z}_s to the response y_s using the linear model $y_s = \mathbf{z}_s^T \phi + e_s$ is found as

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

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

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

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

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

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 $\phi(\cdot)$ in the conditional parametric model $y_s = \mathbf{z}_s^T \phi(\mathbf{u}_s) + e_s$. 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

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

and

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

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 $\mathbf{z}_t^T \hat{\phi}_{t-1}(\mathbf{u})$ 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 $\mathbf{z}_t^T \hat{\phi}_{t-1}(\mathbf{u}_t)$ 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 $\hat{\phi}_t(\mathbf{u}) \approx \hat{\phi}_{t-1}(\mathbf{u})$ 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.

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

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

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

It is reasonable to denote

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

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 $\tilde{h}(\mathbf{u})$ corresponding to the point \mathbf{u} will be related to the nearest neighbour bandwidth α by

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

where $\mathbb{D}_u = \{\boldsymbol{\nu} \in \mathbb{R}^d \mid \|\boldsymbol{\nu} - \mathbf{u}\| \leq \tilde{h}(\mathbf{u})\}$ 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,

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

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 $\lambda_{eff}^u(t)$ 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 .

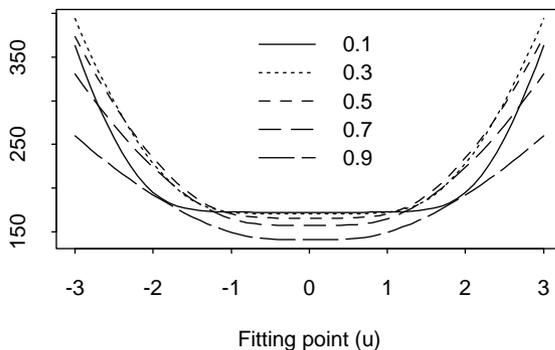


Figure 1: Effective number of observations within the bandwidth ($\alpha\tilde{\eta}_u(u)$) for $\alpha = 0.1, \dots, 0.9$ and $\lambda = 0.99$.

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

In the following description of the algorithm it will be assumed that $\mathbf{R}_{\mathbf{u}^{(i)},t}$ 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 $\mathbf{u}^{(i)}$; $i = 1, \dots, n_{fp}$ and for each fitting point:

- Construct the explanatory variables corresponding to local constant estimates using (2):

$$\mathbf{z}_t^T = [x_{1,t}\mathbf{P}_{d(1)}^T(\mathbf{u}_t) \dots x_{p,t}\mathbf{P}_{d(p)}^T(\mathbf{u}_t)].$$
- Calculate the weight using (6) and (7):

$$w_{\mathbf{u}^{(i)}}(\mathbf{u}_t) = (1 - (\|\mathbf{u}_t - \mathbf{u}^{(i)}\|/\hbar^{(i)})^3)^3, \text{ if } \|\mathbf{u}_t - \mathbf{u}^{(i)}\| < \hbar^{(i)} \text{ and zero otherwise.}$$
- Find the effective forgetting factor using (16):

$$\lambda_{eff}^{(i)}(t) = 1 - (1 - \lambda)w_{\mathbf{u}^{(i)}}(\mathbf{u}_t).$$

- 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):

$$\hat{\boldsymbol{\phi}}_t(\mathbf{u}^{(i)}) = \hat{\boldsymbol{\phi}}_{t-1}(\mathbf{u}^{(i)}) + w_{u^{(i)}}(\mathbf{u}_t)\mathbf{R}_{u^{(i)},t}^{-1}\mathbf{z}_t \left[y_t - \mathbf{z}_t^T \hat{\boldsymbol{\phi}}_{t-1}(\mathbf{u}^{(i)}) \right].$$
- Calculate the updated local polynomial estimates of the coefficient-functions using (8):

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

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

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

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

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

and

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

where $t = 1, \dots, N$ 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.

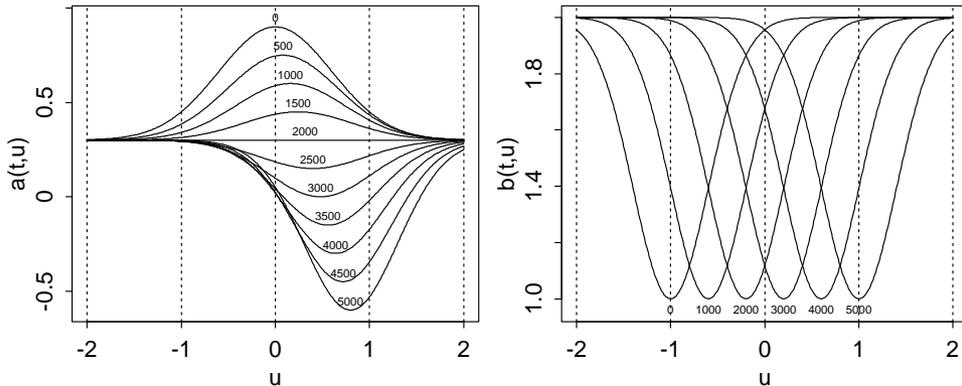


Figure 2: The time-varying coefficient-functions plotted for equidistant points in time as indicated on the plots.

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

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

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 .

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.

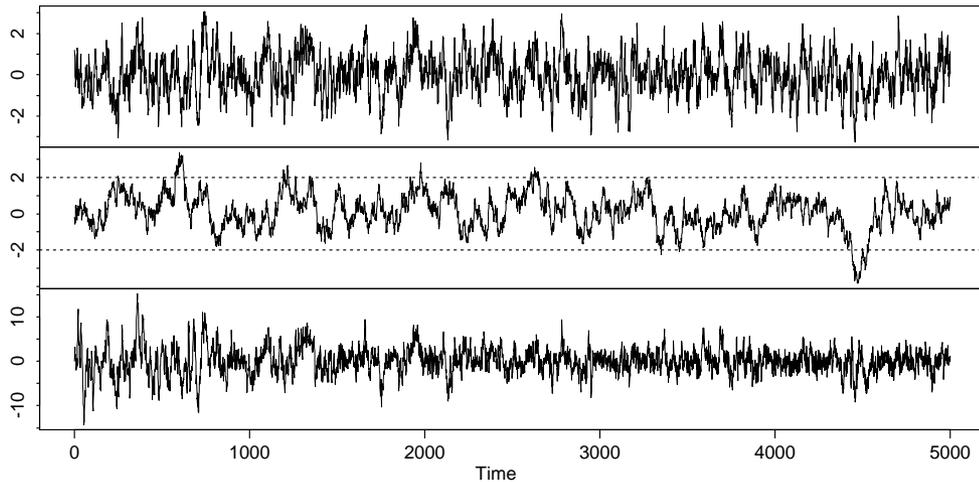


Figure 3: Simulated output (bottom) when x_t (top) and u_t (middle) are $AR(1)$ -processes.

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 near one. From Figure 2 it is seen that for $t > 4000$ the true values at $u = 1$ is markedly lower than 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.

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 tem-

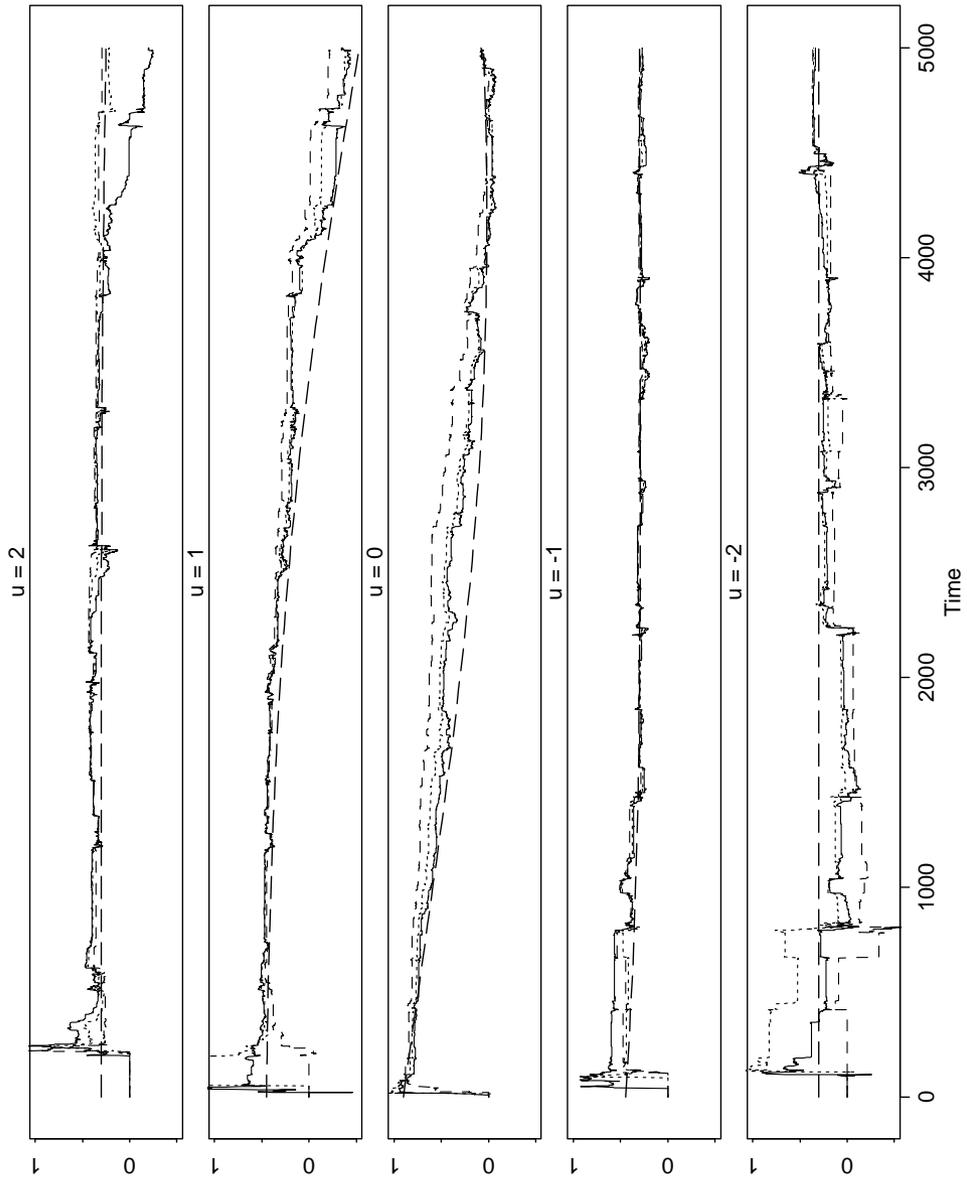


Figure 4: Adaptive estimates of $a(u)$ using local linear approximations and nearest neighbour bandwidths 0.3 (dashed), 0.5 (dotted), and 0.7 (solid). True values are indicated by smooth dashed lines.

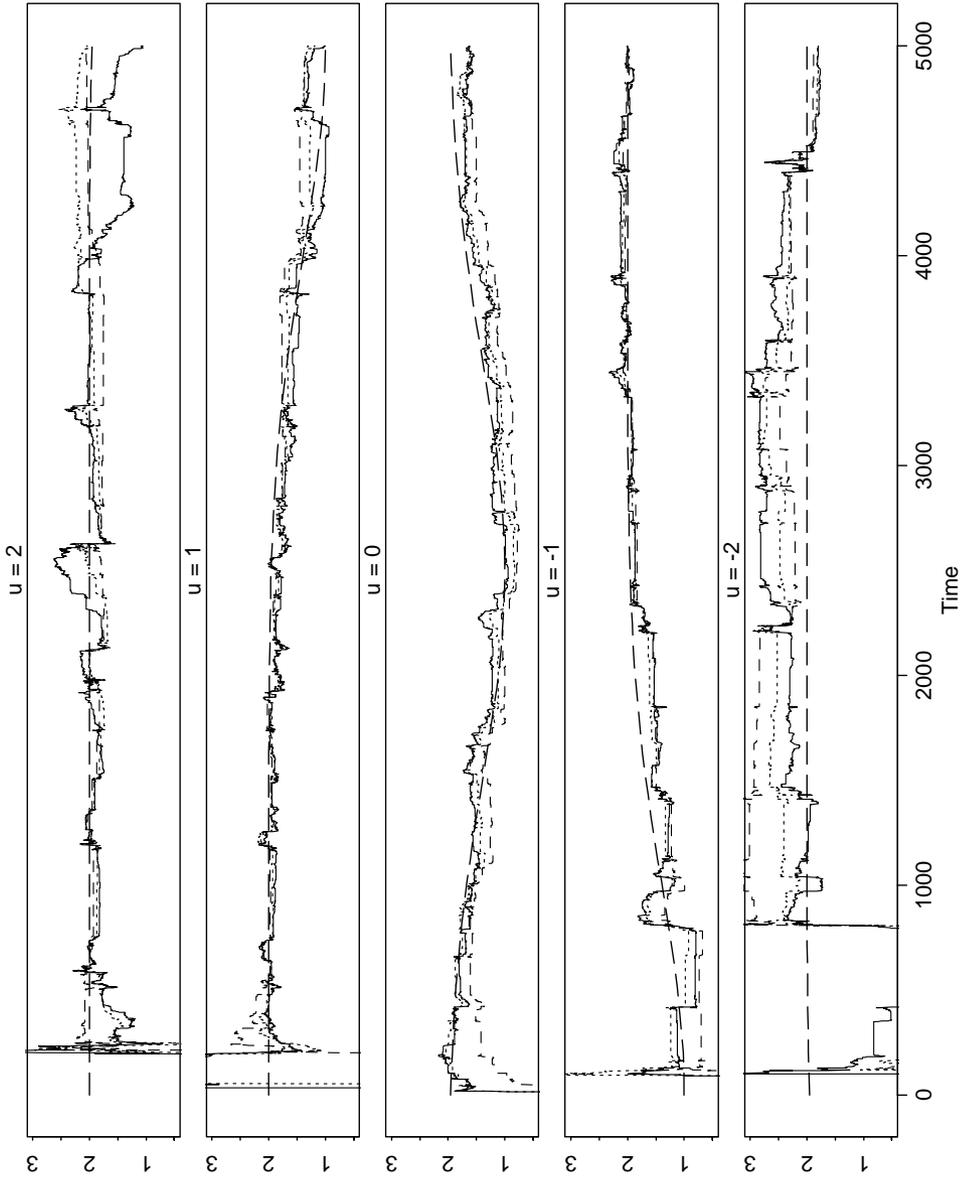


Figure 5: Adaptive estimates of $b(u)$ using local linear approximations and nearest neighbour bandwidths 0.3 (dashed), 0.5 (dotted), and 0.7 (solid). True values are indicated by smooth dashed lines.

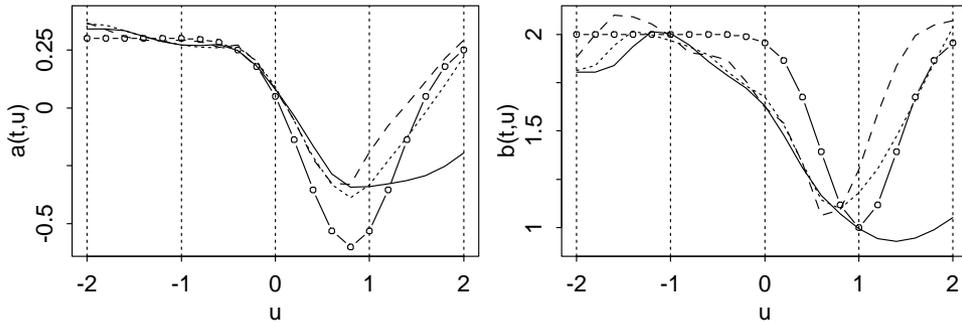


Figure 6: Adaptive estimates for the example considered in Section 4.1 at $t = 5000$ for $\alpha = 0.3$ (dashed), 0.5 (dotted), 0.7 (solid). True values are indicated by circles and fitting points ranging from -2 to 2 in steps of 0.2 are used.

perature 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

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

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

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

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 $t = 4000$.

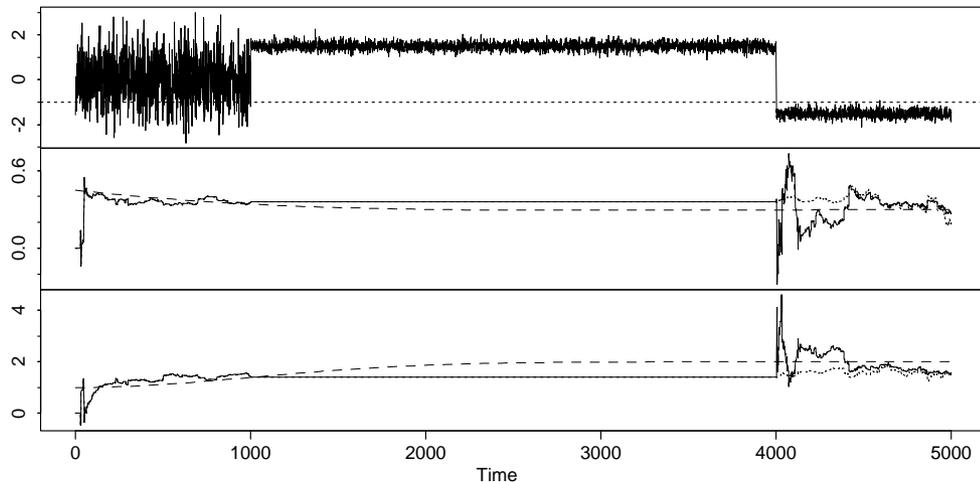


Figure 7: Realization of $\{u_t\}$ (top) and adaptive estimates of $a(-1)$ (middle) and $b(-1)$ (bottom), using the normal updating formula (solid) and the modified updating formula (dotted). True values are indicated by dashed lines.

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 $\hat{\boldsymbol{\theta}}_{t-1}(\mathbf{u}_t)$).

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 \tilde{h} 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 $\tilde{h}(\mathbf{u})$ vary smoothly with \mathbf{u} we may choose to restrict $\lambda(\mathbf{u})$ and $\tilde{h}(\mathbf{u})$, or appropriate transformations of these (logit for λ and log for \tilde{h}), 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 (1999). 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 combination 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

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

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

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

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

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

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 $\lambda_{eff}^u(t)$ 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 $\lambda_{eff}^u(t)\eta_u(t-1)$ may be linearized around $\bar{\lambda}_u$ and $\bar{\eta}_u$. From this it follows, that if the variances of $\lambda_{eff}^u(t)$ 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. (1993a), ‘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, Dept. of Statistics, Stanford Univ., 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. (1999), '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

Tracking time-varying parameters with local regression

Tracking time-varying parameters with local regression

Alfred Joensen, Henrik Madsen,
Henrik Aa. Nielsen, and Torben S. Nielsen

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 \bar{h} , of the time-window determines how fast the model adapts to the variations, and the most adequate value of \bar{h} depends on how fast the parameters actually vary in time. If the time variations are fast, \bar{h} 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 \bar{h} can be seen as a bias/variance trade off.

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, Joensen, Madsen & Holst 2000) 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, Hurdle & 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 individual 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

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

where y_i is a response, \mathbf{x}_i and \mathbf{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 \mathbf{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 \mathbf{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 $\mathbf{x}_i = i$ 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.

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

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

$$w_i(t) = W \left(\frac{|i - t|}{\bar{h}(t)} \right). \quad (3)$$

The scalar $\bar{h}(t)$ is called the bandwidth, and determines the size of the neighbourhood that is spanned by the weight function. If e.g. $\bar{h}(t)$ 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 $\mathbf{x}_i = i$, 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 $\bar{h}(t)$ 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 j th element of $\boldsymbol{\theta}(t)$ and let $\mathbf{p}_d(t)$ be a column vector of terms in a d -order polynomial evaluated at t , i.e. $\mathbf{p}_d(t) = [t^d \ t^{d-1} \ \dots \ 1]$. Furthermore, introduce $\mathbf{z}_i = [z_{1i} \ \dots \ z_{pi}]^T$,

$$\mathbf{u}_{i,t}^T = [z_{1i}\mathbf{p}_{d_1}^T(t-i) \ \dots \ z_{ji}\mathbf{p}_{d_j}^T(t-i) \ \dots \ z_{pi}\mathbf{p}_{d_p}^T(t-i)], \quad (4)$$

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

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

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

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

$$y_i = \mathbf{u}_{i,t}^T \boldsymbol{\phi} + e_i; \quad i = 1, \dots, t, \quad (7)$$

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

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

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

$$y_i = \mathbf{z}_i^T \boldsymbol{\theta} + e_i \quad (9)$$

which corresponds to Eq. 1 when $\boldsymbol{\theta}(\mathbf{x}_i)$ 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

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

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

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

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 $\hat{h}(t) = \hat{h} = -(\ln \lambda)^{-1}$, 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)

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

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

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 \mathbf{z}_t by $\mathbf{u}_{i,t}$. However, the explanatory variable $\mathbf{u}_{i,t}$ is a function of t , which means that as we step forward in time,

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

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

$$(s+1)^d = s^d + ds^{d-1} + \frac{d(d-1)}{2!} s^{d-2} + \dots + 1. \quad (16)$$

it follows that

$$\begin{aligned} \mathbf{p}_{d_j}(s+1) &= \begin{bmatrix} 1 & d_j & \frac{d_j(d_j-1)}{2!} & \frac{(d_j-1)(d_j-2)}{3!} & \dots & 1 \\ 0 & 1 & d_j-1 & \frac{(d_j-1)(d_j-2)}{2!} & \dots & 1 \\ & & 1 & d_j-2 & \dots & 1 \\ & & & 1 & & \\ \vdots & & & & \ddots & \vdots \\ 0 & \dots & & & \dots & 1 \end{bmatrix} \begin{bmatrix} s^{d_j} \\ s^{d_j-1} \\ \vdots \\ 1 \end{bmatrix} \\ &= \mathbf{L}_j \mathbf{p}_{d_j}(s) \end{aligned} \quad (17)$$

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

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{L}_2 & 0 & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & \dots & & \dots & \mathbf{L}_p \end{bmatrix} \quad (18)$$

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

$$\mathbf{R}(t) = \lambda \mathbf{L} \mathbf{R}(t-1) \mathbf{L}^T + \mathbf{u}_t \mathbf{u}_t^T, \quad (19)$$

and the updating formula for the parameters Eq. 13 is left unchanged¹. 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 $\mathbf{u}_{i,t} = \mathbf{u}_{i,t-1}$, 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 $t \rightarrow \infty$.

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

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

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

¹Clarification of the original article: Naturally the parameter estimate must be shifted forward in time, hence Eq. 13 becomes

$$\hat{\boldsymbol{\theta}}(t) = \mathbf{L}^{-1} \hat{\boldsymbol{\theta}}(t-1) + \mathbf{R}^{-1}(t) \mathbf{u}_t \left[y_t - \mathbf{u}_t^T \mathbf{L}^{-1} \hat{\boldsymbol{\theta}}(t-1) \right].$$

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

$$MSE_p = \frac{1}{N-1} \sum_{j=2}^N \left\{ \frac{1}{n-s+1} \sum_{i=s}^n (y_i - \hat{a}(i-1)y_{i-1} - \hat{b}(i-1)z_i)^2 \right\}. \quad (23)$$

Only observations for which $i \geq s = 350 > \max(\hat{h}_{opt})$, where \hat{h}_{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. 23, has not been used for the estimation of the parameters, therefore the optimal bandwidth, \hat{h}_{opt} , can be found by minimizing Eq. 23 with respect to the bandwidth \hat{h} , 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 es-

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

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

pecially when 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

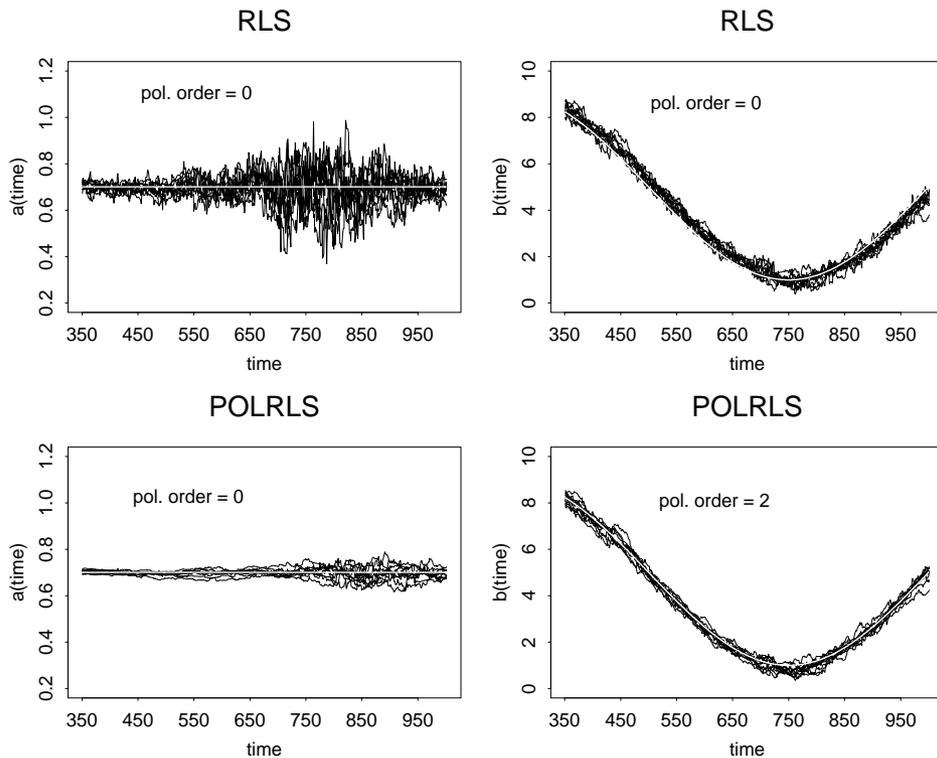


Figure 1: Estimated parameter trajectories. The first row shows the trajectories from the *RLS* algorithm, the second row shows the result from the *POLRLS* algorithm where a has been approximated by a zero order polynomial, and b by a second order polynomial.

be used in the new algorithm, means that the variance can be significantly 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 \bar{h} 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 a significant 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 seasons 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., Joensen, A., Madsen, H. & Holst, J. (2000), ‘Tracking time-varying coefficient-functions’, *International Journal of Adaptive Control and Signal Processing* **14**, 813–828.
- 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 D

**Predictive control of supply
temperature
in district heating systems**

D

Predictive control of supply temperature in district heating systems

Torben Skov Nielsen, Henrik Madsen, Jan Holst, and Henning T. Søgaard

Abstract

This report considers a new concept for controlling the supply temperature in district heating systems using stochastic modelling, prediction and control. A district heating systems is a difficult system to control (optimally) as the dynamic relationship between supply temperature (the control variable) and key parameters such as network temperature and flow rate are time-varying and difficult to establish, which in both cases can be attributed to the time-varying heat load in the system. Thus the problem of controlling a district heating system calls for new control methods, which will operate reliably under these circumstances. Here two new controllers – the Extended Generalized Predictive Controller and a predictive controller derived from a physical relation – are presented. The controllers has been implemented in a software package, PRESS, which has been installed in the Høje Tåstrup district heating utility – a major district heating system in the south-west part of the greater Copenhagen area. The results obtained for the Høje Tåstrup district heating utility are evaluated with respect to security of supply as well as the savings obtained when compared to the previously used control strategy.

1 Introduction

Traditionally the supply temperature in a district heating system is controlled by the central control method (see (Oliker 1980)), where the supply temperature to the distribution network most frequently is determined as a function of the current ambient air temperature, possibly corrected for the current wind speed. This means, that the supply temperature control is in fact an open loop control without any feedback from the distribution network¹ and consequently, the control curve has to be determined conservatively to ensure a sufficiently high temperature in the district heating network at all times.

In Section 2 it is argued that the production costs for a district heating utility with one primary heating station can be minimized by minimizing the supply

¹Return temperature and in some cases a few selected net-point temperatures are normally measured, but are not used by an automatic control algorithm.

temperature assuming that some restrictions on flow rate ex heating station and temperature at selected network points – hereafter referred to as *critical net-points* – are observed. The proposed control scheme for the supply temperature has two primary objectives. First of all it optimizes the operation of the district heating utility with respect to production costs. Secondly it brings the supply temperature control into a closed loop context thereby making the supply temperature control a subject for general control actions compared to the traditional ad hoc approach.

The flow rate in the system is monitored by a flow controller. In Section 7 it is shown how the energy relation at the supply point (cf. (37)) can be used to convert predictions of heat load to a set point for the supply temperature in order to control the flow rate ex the district heating station.

The temperature at each of the critical net-points is monitored by individual controllers and thus models describing the dynamic relationship between supply temperature and each of the individual net-point temperatures are needed. Here a relatively simple linear stochastic model describing the relationship using an ARX (Auto-Regressive-eXtraneous) model structure has been employed. In this context the dynamic relationship is time-varying with both an diurnal variation and a long term trend. Hence both a relatively fast diurnal variation and a more slow long term variation have to be considered. This is accomplished in the proposed approach by explicitly describing the diurnal variation by time-varying parameters whereas the long term variations are taken into account by recursive and adaptive estimation. The above considerations raise the following requirements to a model based predictive controller:

1. It must be robust towards non-minimum phase system (due to the possibility of wrongly specified time delays in the model).
2. It must be reasonably easy and robust to derive since the controller parameters are likely to change hourly as the model parameters are updated.
3. It must be capable of handling time-varying models since it has to cope with both the significant diurnal variations and the long term variations.

The General Predictive Control (GPC) scheme proposed in (Clarke et al. 1987) fulfills the first and second requirement but can not handle models with time-varying parameters. In Section 3 an extended version of the GPC – the Extended Generalized Predictive Controller (XGPC) – is presented, which

enables the use of a deterministic time-variation in the model parameters as well as extends the model class handled by the GPC scheme, and the application of the XGPC, with the purpose of controlling the temperature at the critical net-points, is described in Section 8.

The proposed control scheme for supply temperature in district heating systems has been implemented in a software system – PRESS – and the usefulness of the system has been demonstrated during an installation for a feasibility study at the Høje Tåstrup district heating utility. The settings in Høje Tåstrup are described in Section 6 and the results obtained here are presented in Section 9.

2 Optimal operation of district heating systems

The objectives of the present section is to derive a control scheme for optimal operation of a certain class of district heating systems; namely district heating systems which primarily are supplied by a single district heating station – thus scheduling problems between heat producing plants with different production costs will not be considered here. Optimal operation of the district heating system is assumed to be achieved by minimizing the production costs to the extent, that it can be achieved without compromising the safe operation of the system, adversely affecting the maintenance cost of the system, or sacrificing consumer satisfaction. This requires, that an optimization criterion is defined and the operational restrictions under which the optimization is carried out are identified.

The following conditions, which have potential impact on the operational costs for district heating utility, should be considered:

- *The structure of the account rate between the heat producing units (the supplier) and the district heating utility (the distributor).* For CHP units the cost of production per unit of energy will increase with increasing supply temperature. Often also factors such as the return temperature and the diurnal peak loads are considered in the accounting so that high return temperatures and peak loads will increase the price payed per unit energy.
- *The heat loss from the network.* The heat loss in the network is a (complex) function of the supply temperature. A lowering of the supply

temperature implies a lower temperature in the network in general and consequently a decrease in the heat loss from the network.

- *The pumping costs.* In this context the pumping costs are negligible and consequently left out of the optimization.
- *The maintenance costs for the network.* The operation of a district heating utility has a direct impact on the maintenance costs for the network. Large variations in supply temperature (and pressure) will increase the maintenance costs compared to a more steady operation.

In most district heating systems some or all of the the following restrictions will be imposed on the operation of the system:

- *A maximum allowable flow rate in the system.* The restrictions in the flow rate are due to the (always) limited pumping capacity, the risk of cavitation in heat exchangers and difficulties in maintaining a sufficiently high differential pressure in the remote parts of the network during periods with high flow rates. An increased supply temperature implies a lower flow rate.
- *A minimum guaranteed inlet temperature at the consumers.* This restriction is mainly due to limitations in the consumer installations as well as minimum hot tap water temperature requirements imposed by hygienic concerns. An increase in supply temperature naturally implies a higher inlet temperature at the consumers.
- *A maximum allowable supply temperature.* This restriction is imposed on the systems in order not to damage pipelines and consumer installations.
- *Short term variations in supply temperature.* The stresses implied in the network by large and frequent fluctuations in the supply temperature dictates, that the short term variations in supply temperature should be limited.
- *Maximum allowable diurnal variations of the supply temperature.* In some systems the size of the expansion tanks may impose limitations on the allowable diurnal variation of the supply temperature.

Under the assumption that diurnal peak load and return temperature are not adversely affected by the optimization, the production cost can be minimized by minimizing the supply temperature, but at the same time some of the

system restrictions are activated by lowering the supply temperature. Based on the above considerations the goal of the optimization can be formulated as:

The supply temperature should be minimized under the restriction, that flow rate, consumer inlet temperatures and variations of supply temperature are kept within acceptable bounds.

The optimization strategy is implemented as a set of controllers, which operate the system as close to the minimum supply temperature as possible without actually violating the restrictions. The flow rate is monitored by a single controller whereas the consumer inlet temperature is monitored by introducing a set of *critical points* in the distribution network. The critical points are selected so that if the temperature requirements for the critical points are satisfied then the temperature requirements for all consumers are satisfied. The critical net-points should be located in the “weak” areas of the distribution network, i.e. in areas with a high temperature loss. The critical net-points must be placed in the distribution network and not in the inlet pipelines to individual consumers in order to avoid that the temperature response at the critical net-points is dominated by the behavior of individual consumers. The weak areas in a distribution network are identified either by using a deterministic network simulation program as e.g. “SYSTEM RØRNET” by Rambøll (cf. (Hansen & Højlund 1997)) or by using the experience of the operators.

The PRESS control system can then be seen as consisting of a flow controller and a set of net-point temperature controllers, namely one for each critical point in the network. At a given time the supply temperature recommended by PRESS is then selected as the maximum of the recommended supply temperatures from the individual controllers as illustrated in Figure 1.

The flow rate and net-point temperature restrictions are directly monitored by controllers, thus making them *hard* restrictions, whereas the restrictions in the variations of supply temperature are fulfilled by tuning of the controller design parameters in the flow rate and net-point temperature controllers, thus making this restriction *soft* restrictions.

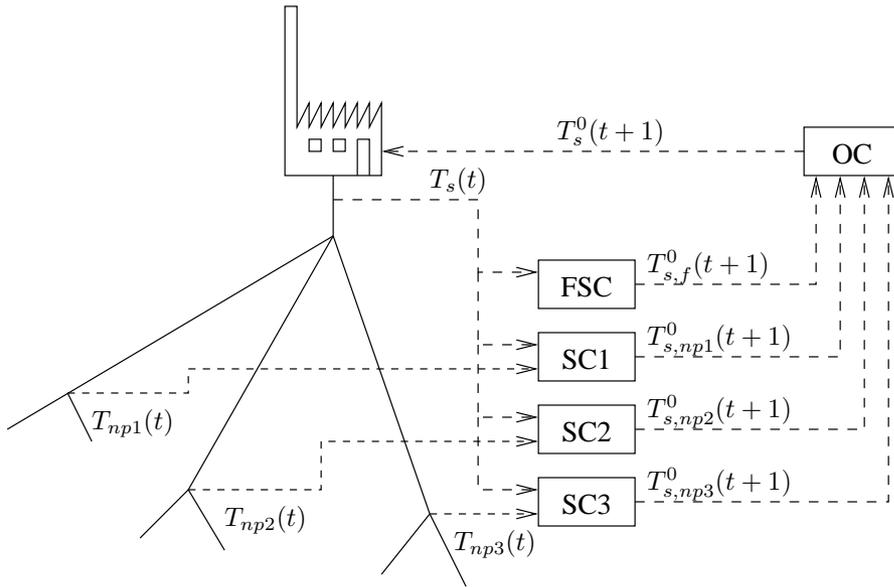


Figure 1: *Schematic overview of a district heating network, with three critical net-points and the corresponding controllers. OC is the overall controller, FSC is the flow sub-controller, SC# are the supply temperature sub-controllers, $T_{np\#}$ are the supply temperatures in the network, T_s is the supply temperature from the plant, and $T_{s,\#}^0$ are the supply temperatures required by the sub-controllers.*

3 Extended generalized predictive control

This section describes how the GPC scheme proposed in (Clarke et al. 1987) can be extended to handle systems governed by models with time-varying parameters. Compared to the original GPC method a more general point of view is used in the development of the eXtended Generalized Predictive Control (XGPC), and it is shown, that the GPC method is a special case of the XGPC. The main difference is, that time-variation of the model parameters is permitted, but a generalization of the cost function and the associated constraints is also introduced.

Under the GPC the future controls are calculated by minimizing a criterion function of the predicted deviations from an output reference as well as the necessary control effort. The very foundation in the GPC scheme is the separation of the future system response into two separate terms: the term only depending on past input and output values, and the term only depending on future input values. In the GPC algorithm proposed by Clarke et al. (1987) this separation is achieved by recursively solving a Diophantine equa-

tion, which implies that only models with time-invariant parameters can be employed. Clarke et al. (1987) base the formulation of the GPC on a specified ARIMAX model

$$A(q^{-1})y(t) = B(q^{-1})u(t) + \frac{C(q^{-1})e(t)}{\nabla}, \quad (1)$$

where $\{e(t)\}$ is a white noise sequence with zero mean, ∇ is the difference operator, $1 - q^{-1}$, and A , B and C are time-invariant polynomials in the backward shift operator, q^{-1} . In the actual derivation of the GPC control law C is set to 1, though.

In (Clarke et al. 1987) it is argued that the integral part in (1) should be part of the model as it ensures an offset free control, but it also facilitates the derivation of the GPC algorithm. From a model builders point of view, the model class imposed by the original GPC formulation is somewhat problematic as the introduction of an unjustified integration in the noise term may lead to poorly estimated system dynamics (See e.g. (Madsen et al. 1992)).

In (Palsson et al. 1994) an alternative or extended formulation of the GPC is proposed – the XGPC. Under XGPC predictions of the output are calculated as the conditional expectation of the future output conditioned on past output values as well as on past and future input values (controls). The use of conditional expectation in computing the output predictions makes it possible to employ the XGPC for a wide range of linear models, where the only requirement is that the future model output is additive in the term with a linear dependency on future controls, and the term depending on past control and output values. In the following the XGPC will be formulated for an ARMAX model, but most of the formulation only rely on the assumption that the (linear) model is additive as specified above.

3.1 The ARMAX model

Consider a ARMAX model with time varying

$$A_t(q^{-1})y(t) = B_t(q^{-1})u(t) + C_t(q^{-1})e(t), \quad (2)$$

where $\{e(t)\}$ is a white noise sequence with mean zero, and A , B and C are time-varying polynomials in q^{-1} :

$$\begin{aligned} A_t(q^{-1}) &= 1 + a_{1,t}q^{-1} + \cdots + a_{n,t}q^{-n} \\ B_t(q^{-1}) &= b_{1,t}q^{-1} + \cdots + b_{m,t}q^{-m} \\ C_t(q^{-1}) &= 1 + c_{1,t}q^{-1} + \cdots + c_{r,t}q^{-r}. \end{aligned}$$

The time-variation of the parameters in these polynomials is assumed to be described by known² functions of time. Note that if both A and B have a root equal to 1, C is 1, and the model parameters are time-invariant, the ARIX model used for the formulation of the ordinary GPC in (Clarke et al. 1987) is obtained.

3.1.1 The j -step-ahead predictor

The optimal j -step-ahead predictor, $\hat{y}(t+j|t)$, corresponding to (2) is found as the conditional expectation of $y(t+j)$ conditioned on observations of the output and external signals up to and including time t . Due to the time-variation of the parameters the predictor *cannot* be obtained by means of a Diophantine equation. Instead the predictor can be obtained using conditional expectation as described in e.g. (Abraham & Ledolter 1983, Madsen 1989)). After the influence of the initial values has subsided the predictor is given as

$$\begin{aligned} \hat{y}(t+j|t) = & - \sum_{i=1}^n a_{i,t+j} \hat{y}(t+j-i|t) + \sum_{i=1}^m b_{i,t+j} u(t+j-i) \\ & + \sum_{i=1}^r c_{i,t+j} \hat{e}(t+j-i|t), \quad j \geq 1 \end{aligned} \quad (3)$$

where

$$\hat{y}(t+j|t) = y(t+j), \quad j < 1, \quad (4)$$

$$\hat{e}(s|t) = \begin{cases} e(s) = y(s) - \hat{y}(s|s-1) & \text{if } s \leq t \\ 0 & \text{if } s > t \end{cases} \quad (5)$$

and the parameters $a_{i,t+j}$, $b_{i,t+j}$ and $c_{i,t+j}$ are assumed known.

In general, a j -step-ahead prediction can be computed as an one-step-ahead prediction from time origin $t+j-1$ where unknown output values are replaced by their predicted counterparts and unknown noise components are fixed to 0.

3.1.2 The predictor as a linear function of future controls

From (3) it is found that $\hat{y}(t+j|t)$ is linear in the input values up to time $t+j-1$ and output values up to time t . This is seen by eliminating the \hat{y} 's

²In practice the time dependency of the A , B and C polynomials has to be estimated, but at the time of control the time dependency is considered known.

and the \hat{e} 's in (3) recursively using (4) and (5). If $v_j(t)$ denotes the sum of all terms on the right-hand side of (3) involving known data ($y(t), y(t-1), \dots$ and $u(t-1), u(t-2), \dots$), the j -step-ahead predictor can be expressed as

$$\hat{y}(t+j|t) = \sum_{i=1}^j \bar{h}_{j,i,t} u(t+j-i) + v_j(t), \quad j = 1, 2, 3, \dots, \quad (6)$$

where $\bar{h}_{j,i,t}$ ($i = 1, 2, 3, \dots, j$) are the coefficients describing the linear dependence on future controls. These coefficients and $v_j(t)$ can, of course, be found by eliminating the \hat{y} 's and the \hat{e} 's in (3) recursively, but there is an easier way.

Write the model (2) as

$$y(t) = H_t(q^{-1})u(t) + G_t(q^{-1})e(t), \quad (7)$$

where the time-varying transfer functions

$$H_t(q^{-1}) = \sum_{i=0}^{\infty} h_{i,t} q^{-i}$$

and

$$G_t(q^{-1}) = \sum_{i=0}^{\infty} g_{i,t} q^{-i}$$

correspond to $B(q^{-1})/A(q^{-1})$ and $C(q^{-1})/A(q^{-1})$ in the time-invariant case. The coefficients $h_{0,t}, h_{1,t}, h_{2,t}, \dots$ are clearly the weights of the time-varying impulse response function describing the dynamic relationship between input and output – i.e. $h_{i,t}$ is the extra contribution to $y(t)$ if $u(t-i)$ is increased by one. The coefficients of $H_t(q^{-1})$ are difficult to computed by polynomial division of $B_t(q^{-1})$ by $A_t(q^{-1})$ since the q^{-1} operator affects the t index of the coefficients of A and B , but the coefficients are readily derived by passing impulses through the system $A_t(q^{-1})y(t) = B_t(q^{-1})u(t)$. From (7) the j -step ahead output prediction is found by conditional expectation:

$$\hat{y}(t+j|t) = \sum_{i=0}^{\infty} h_{i,t+j} u(t+j-i) + \sum_{i=j}^{\infty} g_{i,t+j} e(t+j-i), \quad j = 1, 2, 3, \dots$$

Comparison with (6) shows that

$$\begin{aligned} \bar{h}_{j,i,t} &= h_{i,t+j}, \quad i = 1, \dots, j \\ v_j(t) &= \sum_{i=j+1}^{\infty} h_{i,t+j} u(t+j-i) + \sum_{i=j}^{\infty} g_{i,t+j} e(t+j-i), \end{aligned}$$

where $j = 1, 2, 3, \dots$. Thus

$$\hat{y}(t+j|t) = \sum_{i=1}^j h_{i,t+j} u(t+j-i) + v_j(t). \quad (8)$$

Equation (8) shows that the $h_{i,t+j}$ -values and $v_j(t)$ can be calculated as follows:

1. To find $v_j(t)$ ($j = 1, 2, 3, \dots$): Let $u(t+j-i) = 0$ for $i = 1, 2, \dots, j$ and compute $\hat{y}(t+j|t)$ using (3). This “prediction” becomes $v_j(t)$.
2. To find $h_{i,t+j}$ ($j = 1, 2, \dots$ and $i = 1, \dots, j$): Let

$$u(t+j-l) = \begin{cases} 1 & \text{if } l = i \\ 0 & \text{otherwise} \end{cases},$$

and $y(t-l) = 0$, $e(t-l) = 0$ for $l = 0, 1, 2, \dots$ (corresponding to $v_j(t) = 0$) and compute $\hat{y}(t+j|t)$ using (3). This “prediction” becomes $h_{i,t+j}$.

In step 1 we simply compute $v_j(t)$ as the prediction of $y(t+j)$ in case that the control signal is zero from time t onwards. In step 2 we compute $h_{i,t+j}$ as the response at time $t+j$ to a unit input pulse at time $t+j-i$.

Now introduce a maximum prediction horizon N (≥ 1) which corresponds to N_2 in the cost function of the traditional GPC. Considering (8) it is found that the j -step-ahead predictions, j running from 1 up to N , can be written as a linear matrix expression:

$$\hat{\mathbf{y}}(t) = \mathbf{H}(t)\mathbf{u}(t) + \mathbf{v}(t), \quad (9)$$

where

$$\begin{aligned} \hat{\mathbf{y}}(t) &= (\hat{y}(t+1|t), \dots, \hat{y}(t+N|t))^T \\ \mathbf{u}(t) &= (u(t), \dots, u(t+N-1))^T \\ \mathbf{v}(t) &= (v_1(t), \dots, v_N(t))^T \\ \mathbf{H}(t) &= \begin{pmatrix} h_{1,t+1} & 0 & \cdots & 0 & 0 \\ h_{2,t+2} & h_{1,t+2} & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ h_{N-1,t+N-1} & h_{N-2,t+N-1} & \cdots & h_{1,t+N-1} & 0 \\ h_{N,t+N} & h_{N-1,t+N} & \cdots & h_{2,t+N} & h_{1,t+N} \end{pmatrix}. \end{aligned}$$

In (9) it has been assumed that the same model is applicable for all prediction horizons. Actually, this need not be the case. If different models are used, the j th row of $\mathbf{H}(t)$ and the j th element of $\mathbf{v}(t)$ belong to a special model designed for j -step-ahead prediction. Making use of an individual model for each horizon is often relevant if a non-linear system is approximated by linear models, i.e. if the optimal linearized predictor for the system depends on the prediction horizon.

3.2 The cost functions

The aim of the controller is to keep the system output “close to” some pre-determined reference signal, $\{y^0(t)\}$ which can be constant (set-point) or time-varying. The control error, i.e. the difference between the output and its reference, should obviously be small but at the same time the control effort must be kept within reasonable bounds.

For the GPC these requirements are met by the following time dependent optimization

$$\begin{aligned} & \min_{\nabla u(t+j-1)_{j=1,\dots,N_2}} J[N_1, N_2, N_u, \lambda_1, \dots, \lambda_{N_u}; t, \nabla u(t+j-1)_{j=1,\dots,N_2}] \\ & = E_t \left[\sum_{j=N_1}^{N_2} [y(t+j) - y^0(t+j)]^2 + \sum_{j=1}^{N_2} \lambda_j [\nabla u(t+j-1)]^2 \right] \end{aligned} \quad (10)$$

subject to the model in (1) and

$$\nabla u(t+j-1) = 0, \quad j > N_u, \quad (11)$$

where

- N_1 is the minimum cost (output) horizon,
- N_2 is the maximum cost (output) horizon,
- N_u ($< N_2$) is the control horizon,
- λ_j (≥ 0) is a control-weighting sequence,

and $E_t[\cdot]$ denotes the conditional expectation of its argument conditioned on data up to time t . Using matrix notation (10) may be rewritten as

$$\begin{aligned} & \min_{\nabla \mathbf{u}(t)} J(\mathbf{\Lambda}; t, \nabla \mathbf{u}(t)) = \\ & E_t [(\mathbf{y}(t) - \mathbf{y}^0(t))^T (\mathbf{y}(t) - \mathbf{y}^0(t)) + \nabla \mathbf{u}(t)^T \mathbf{\Lambda} \nabla \mathbf{u}(t)], \end{aligned} \quad (12)$$

where $\mathbf{y}(t)$ is a vector of future output values, $\mathbf{y}^0(t)$ is a vector of future set-points, $\nabla \mathbf{u}(t)$ is a vector of future differenced control values and $\mathbf{\Lambda}$ is a diagonal matrix weighting future control effort:

$$\begin{aligned}\mathbf{y}(t) &= (y(t + N_1), \dots, y(t + N_2))^T \\ \mathbf{y}^0(t) &= (y^0(t + N_1), \dots, y^0(t + N_2))^T \\ \nabla \mathbf{u}(t) &= (\nabla u(t + 1), \dots, \nabla u(t + N_u))^T \\ \mathbf{\Lambda} &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_{N_u} \end{pmatrix}.\end{aligned}$$

For the XGPC the control strategy is based on minimization of a more general cost function than the ordinary GPC cost function in (10):

$$\begin{aligned}\min_{\mathbf{u}(t)} J(\mathbf{\Gamma}(t), \mathbf{\Lambda}(t), \boldsymbol{\omega}(t); t, \mathbf{u}(t)) \\ = E_t[(\mathbf{y}(t) - \mathbf{y}^0(t))^T \mathbf{\Gamma}(t)(\mathbf{y}(t) - \mathbf{y}^0(t)) + \\ \mathbf{u}(t)^T \mathbf{\Lambda}(t)\mathbf{u}(t) + 2\boldsymbol{\omega}(t)^T \mathbf{u}(t)],\end{aligned}\quad (13)$$

where

$$\begin{aligned}\mathbf{y}(t) &= (y(t + 1), \dots, y(t + N))^T \\ \mathbf{y}^0(t) &= (y^0(t + 1), \dots, y^0(t + N))^T \\ \mathbf{u}(t) &= (u(t), \dots, u(t + N - 1))^T\end{aligned}$$

and

N is the maximum prediction horizon, which corresponds to N_2 in (10), $\mathbf{\Gamma}(t)$ is a positive semidefinite³ and symmetric matrix which weights the control errors, $\mathbf{\Lambda}(t)$ is a positive semidefinite³ and symmetric matrix which weights the squared control values, and $\boldsymbol{\omega}(t)$ is a vector weighting the control values linearly.

Note that by selecting appropriate values for the weights in (13) the cost function of the GPC controller may be obtained, cf. page 130.

³ The matrix $\mathbf{H}(t)^T \mathbf{\Gamma}(t) \mathbf{H}(t) + \mathbf{\Lambda}(t)$ must be non-singular.

3.3 Choice of the design parameters for the XGPC controller

Apart from the fact that the $\mathbf{\Gamma}(t)$ -matrix should be symmetric and positive semidefinite, no further restrictions apply to it, and it may, for instance, be the identity matrix as in the ordinary GPC presented by (Clarke et al. 1987). Another possibility is to choose the inverse covariance matrix of the prediction errors, i.e. $\mathbf{\Gamma}(t) = \text{Var}[\mathbf{e}(t)]^{-1}$, where $\mathbf{e}(t) = \mathbf{y}(t) - \hat{\mathbf{y}}(t)$. This choice means that less attention is paid to the control error if the corresponding prediction is very uncertain. For open-loop stable systems this seems reasonable since the current control cannot do very much about a prediction error, which is primarily a result of stochastic disturbances.

If the control strategy should aim at keeping the control values close to zero, $\mathbf{\Lambda}(t) = \lambda \mathbf{I}$ ($\lambda > 0$) and $\boldsymbol{\omega} = \mathbf{0}$ would be appropriate. If, on the contrary, the control value $u(t)$ should be kept close to some pre-specified value $u^0(t)$ and the control increments should be simultaneously damped, $\mathbf{\Lambda}(t)$ and $\boldsymbol{\omega}(t)$ may be chosen such that

$$\mathbf{u}(t)^T \mathbf{\Lambda}(t) \mathbf{u}(t) + 2\boldsymbol{\omega}(t)^T \mathbf{u}(t) + \text{const.} = \sum_{j=1}^N \left(\lambda_{1,j} [u(t+j-1) - u^0(t+j-1)]^2 + \lambda_{2,j} [\nabla u(t+j-1)]^2 \right)$$

where $\lambda_{1,j}$ and $\lambda_{2,j}$ ($1 \leq j \leq N$) are non-negative real numbers and “const.” is a term which is independent of $\mathbf{u}(t)$ but dependent on $u^0(t)$. This leads to

$$\mathbf{\Lambda}(t) = \begin{pmatrix} \lambda_{1,1} + \lambda_{2,1} + \lambda_{2,2} & -\lambda_{2,2} & \cdots & 0 & 0 \\ -\lambda_{2,2} & \lambda_{1,2} + \lambda_{2,2} + \lambda_{2,3} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & \lambda_{1,N-1} + \lambda_{2,N-1} + \lambda_{2,N} & -\lambda_{2,N} \\ 0 & 0 & \cdots & -\lambda_{2,N} & \lambda_{1,N} + \lambda_{2,N} \end{pmatrix} \quad (14)$$

and

$$\boldsymbol{\omega}(t) = - \begin{pmatrix} \lambda_{1,1}u^0(t) + \lambda_{2,1}u(t-1) \\ \lambda_{1,2}u^0(t+1) \\ \lambda_{1,3}u^0(t+2) \\ \vdots \\ \lambda_{1,N}u^0(t+N-1) \end{pmatrix}. \quad (15)$$

Particularly if $\lambda_{1,j} = 0$ and $\lambda_{2,j} = \lambda_j$ ($j = 1, \dots, N$), the controls are weighted as in the ordinary GPC cost function (10).

It is also possible to choose $\mathbf{\Lambda}(t)$ and $\boldsymbol{\omega}(t)$ so that filtered controls, $u^f(t)$, are weighted in (13). Let

$$u^f(t) = F_t(q^{-1})u(t),$$

where $F_t(q^{-1})$ is a FIR filter given by

$$F_t(q^{-1}) = 1 + f_{1,t}q^{-1} + f_{2,t}q^{-2} + \cdots + f_{p,t}q^{-p}.$$

If, for instance, $u^f(t)$ should be weighted as

$$\sum_{j=1}^N \lambda_j \left[u^f(t+j-1) \right]^2,$$

the corresponding structures of $\mathbf{\Lambda}(t)$ and $\boldsymbol{\omega}(t)$ would turn out to be

$$\mathbf{\Lambda}(t) = \begin{pmatrix} \overbrace{x \cdots x}^{p+1 \text{ columns}} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x & \cdots & x & \cdots & x & 0 \\ & \ddots & \vdots & \vdots & \ddots & \\ 0 & & x & \cdots & x & \cdots & x \\ \vdots & \ddots & & \ddots & \vdots & & \vdots \\ 0 & \cdots & 0 & & x & \cdots & x \end{pmatrix}$$

$$\boldsymbol{\omega}(t) = \left(\overbrace{x \cdots x}^{p \text{ elements}} \ 0 \cdots 0 \right)^T,$$

where the x 's represent non-zero elements.

From the above discussion one can conclude that the cost function in (13) offers a very flexible criterion for defining the optimal control.

The cost horizon, N , which implicitly is included in the control criterion, should be chosen so that all future output values significantly affected by the current control are included in the cost function. For open-loop stable systems the horizon, which includes all impulse responses in the impulse response function for the system significantly different from zero, provides an appropriate value of N .

3.4 The resulting controller

In order to minimize the cost function in (13) the expectation of the cost function is evaluated and the resulting function is minimized by equating its derivative with respect to future controls with zero. If a vector of prediction errors ($j = 1, \dots, N$),

$$\mathbf{e}(t) = (e(t+1|t), e(t+2|t), \dots, e(t+N|t))^T ,$$

is introduced, the expectation of the first term of the cost function can be written as

$$\begin{aligned} & E_t [(\hat{\mathbf{y}}(t) + \mathbf{e}(t) - \mathbf{y}^0(t))^T \mathbf{\Gamma}(t)(\hat{\mathbf{y}}(t) + \mathbf{e}(t) - \mathbf{y}^0(t))] \\ &= (\hat{\mathbf{y}}(t) - \mathbf{y}^0(t))^T \mathbf{\Gamma}(t)(\hat{\mathbf{y}}(t) - \mathbf{y}^0(t)) + E_t [\mathbf{e}(t)^T \mathbf{\Gamma}(t)\mathbf{e}(t)] , \end{aligned}$$

where the projection theorem by (Brockwell & Davis 1987), which states that the predictions and the prediction errors are independent and thus uncorrelated, has been used. If we substitute for $\hat{\mathbf{y}}(t)$ from (9) the cost function becomes:

$$\begin{aligned} & J(\mathbf{\Gamma}(t), \mathbf{\Lambda}(t), \boldsymbol{\omega}(t); t, \mathbf{u}(t)) \\ &= (\mathbf{H}(t)\mathbf{u}(t) + \mathbf{v}(t) - \mathbf{y}^0(t))^T \mathbf{\Gamma}(t)(\mathbf{H}(t)\mathbf{u}(t) + \mathbf{v}(t) - \mathbf{y}^0(t)) \\ & \quad + \mathbf{u}(t)^T \mathbf{\Lambda}(t)\mathbf{u}(t) + 2\boldsymbol{\omega}(t)^T \mathbf{u}(t) + E_t [\mathbf{e}(t)^T \mathbf{\Gamma}(t)\mathbf{e}(t)] . \end{aligned}$$

The last term in this expression does not depend on $\mathbf{u}(t)$ so the optimal control can be found by minimizing:

$$\begin{aligned} \tilde{J}(t, \mathbf{u}(t)) &= (\mathbf{H}(t)\mathbf{u}(t) + \boldsymbol{\beta}(t))^T \mathbf{\Gamma}(t)(\mathbf{H}(t)\mathbf{u}(t) + \boldsymbol{\beta}(t)) \\ & \quad + \mathbf{u}(t)^T \mathbf{\Lambda}(t)\mathbf{u}(t) + 2\boldsymbol{\omega}(t)^T \mathbf{u}(t) , \end{aligned} \quad (16)$$

where

$$\boldsymbol{\beta}(t) = \mathbf{v}(t) - \mathbf{y}^0(t) . \quad (17)$$

Computing the derivative of this function with respect to $\mathbf{u}(t)$ gives (omitting the time arguments):

$$\frac{\partial \tilde{J}(\mathbf{u})}{\partial \mathbf{u}} = 2\mathbf{H}^T \mathbf{\Gamma} \mathbf{H} \mathbf{u} + 2\mathbf{H}^T \mathbf{\Gamma}^T \boldsymbol{\beta} + 2\mathbf{\Lambda} \mathbf{u} + 2\boldsymbol{\omega} .$$

The optimal control is obtained by equating this expression with zero and solving the resulting equation with respect to \mathbf{u} :

$$\mathbf{u}(t) = - [\mathbf{H}(t)^T \mathbf{\Gamma}(t) \mathbf{H}(t) + \mathbf{\Lambda}(t)]^{-1} [\mathbf{H}(t)^T \mathbf{\Gamma}(t)^T \boldsymbol{\beta}(t) + \boldsymbol{\omega}(t)] . \quad (18)$$

Here it is assumed that $\mathbf{\Gamma}(t)$ and $\mathbf{\Lambda}(t)$ have been chosen so that the matrix $\mathbf{H}(t)^T \mathbf{\Gamma}(t) \mathbf{H}(t) + \mathbf{\Lambda}(t)$ becomes non-singular. As only the first element of the control vector, $\mathbf{u}(t)$, is implemented, the control law may be written

$$u(t) = -\mathbf{l}(t) [\mathbf{H}(t)^T \mathbf{\Gamma}(t)^T \boldsymbol{\beta}(t) + \boldsymbol{\omega}(t)] , \quad (19)$$

where $\mathbf{l}(t)$ is the first row of $[\mathbf{H}(t)^T \mathbf{\Gamma}(t) \mathbf{H}(t) + \mathbf{\Lambda}(t)]^{-1}$.

The elements in the vector $\boldsymbol{\beta}(t)$ defined in (17) are predicted values of the control errors given future controls equal zero, i.e. the predictions if $\mathbf{u}(t) = \mathbf{0}$, cf. (9). Hence the optimal control in (19) is obtained as a linear feedback from the predicted values of the control errors given future controls equal zero plus a constant term coming from $\boldsymbol{\omega}(t)$.

3.5 Linear equality constraints

For some practical systems it might be desirable to constrain the sequence of future controls. In the GPC developed by Clarke et al. (1987), for instance, only the first N_u controls are free while the subsequent controls are considered to be constant and equal to the value of the last free control signal. This restriction makes up a set of simple linear constraints: $u(t+j-1) = u(t+j-2)$ for $j > N_u$. The advantage of introducing such constraints is obvious: the dimension of the optimization problem associated with the computation of the optimal control is reduced to N_u . Furthermore, it does not reduce the performance of the controller significantly when reducing the degrees of freedom of the control in this way as long as the value of N_u is not too low compared to the complexity of the system. Clarke et al. (1987) suggests that for complex systems N_u should be at least equal to the number of unstable or badly-damped poles.

In this section it is shown how to add general linear constraints to the minimization criterion in (13) (or the equivalent in (16)). Hence consider the following set of constraints:

$$\mathbf{S}(t)\mathbf{u}(t) = \mathbf{d}(t) , \quad (20)$$

where the matrix $\mathbf{S}(t)$ is $M \times N$ ($M \leq N$) and has full rank ($= M$). The vector $\mathbf{d}(t)$ is a column with M elements. To solve a minimization problem being subject to (20), the M equations can simply be solved with respect to a subset of the elements in $\mathbf{u}(t)$. This subset of the future control values can then be eliminated in the cost function. Since the cost function is a quadratic function

of the unknown controls, and the constraints are linear, this elimination leads to an unconstrained quadratic cost function. The procedure is shown below.

Split the \mathbf{u} -vector⁴ into two parts, \mathbf{u}_1 and \mathbf{u}_2 :

$$\mathbf{u} = [\mathbf{u}_1^T \ \mathbf{u}_2^T]^T ,$$

where \mathbf{u}_1 is a vector of the dimension $N - M$ and \mathbf{u}_2 is an M -vector. \mathbf{S} is partitioned similarly,

$$\mathbf{S} = [\mathbf{S}_1 \ \mathbf{S}_2] , \tag{21}$$

where it is assumed that the columns of \mathbf{S}_2 are linearly independent (i.e. \mathbf{S}_2 is a non-singular matrix).

Now (20) can be written as

$$\mathbf{S}_1 \mathbf{u}_1 + \mathbf{S}_2 \mathbf{u}_2 = \mathbf{d} . \tag{22}$$

Solving this equation with respect to \mathbf{u}_2 yields

$$\mathbf{u}_2 = \mathbf{S}_2^{-1}(\mathbf{d} - \mathbf{S}_1 \mathbf{u}_1) . \tag{23}$$

Matrices and vectors in (16) are partitioned so that they match the partitioning of \mathbf{u} :

$$\begin{aligned} \mathbf{H} &= [\mathbf{H}_1 \ \mathbf{H}_2] \\ \mathbf{\Lambda} &= \begin{bmatrix} \mathbf{\Lambda}_{11} & \mathbf{\Lambda}_{12} \\ \mathbf{\Lambda}_{21} & \mathbf{\Lambda}_{22} \end{bmatrix} \\ \boldsymbol{\omega} &= [\boldsymbol{\omega}_1^T \ \boldsymbol{\omega}_2^T]^T . \end{aligned}$$

Using this (16) can be rewritten as

$$\begin{aligned} \tilde{J}([\mathbf{u}_1^T \ \mathbf{u}_2^T]^T) &= (\mathbf{H}_1 \mathbf{u}_1 + \mathbf{H}_2 \mathbf{u}_2 + \boldsymbol{\beta})^T \boldsymbol{\Gamma} (\mathbf{H}_1 \mathbf{u}_1 + \mathbf{H}_2 \mathbf{u}_2 + \boldsymbol{\beta}) + \\ &\quad \mathbf{u}_1^T \mathbf{\Lambda}_{11} \mathbf{u}_1 + 2\mathbf{u}_1^T \mathbf{\Lambda}_{12} \mathbf{u}_2 + \mathbf{u}_2^T \mathbf{\Lambda}_{22} \mathbf{u}_2 + 2\boldsymbol{\omega}_1^T \mathbf{u}_1 + 2\boldsymbol{\omega}_2^T \mathbf{u}_2 , \end{aligned} \tag{24}$$

where the fact that $\mathbf{\Lambda}$ is symmetric has been utilized (i.e. $\mathbf{\Lambda}_{12} = \mathbf{\Lambda}_{21}^T$). Substituting \mathbf{u}_2 in (24) by the right-hand side of (23), it is found that

$$\begin{aligned} \tilde{J}([\mathbf{u}_1^T \ \mathbf{u}_2^T]^T) \Big|_{\mathbf{u}_2 = \mathbf{S}_2^{-1}(\mathbf{d} - \mathbf{S}_1 \mathbf{u}_1)} &= \\ &= (\mathbf{H}^* \mathbf{u}_1 + \boldsymbol{\beta}^*)^T \boldsymbol{\Gamma} (\mathbf{H}^* \mathbf{u}_1 + \boldsymbol{\beta}^*) + \mathbf{u}_1^T \mathbf{\Lambda}^* \mathbf{u}_1 + 2\boldsymbol{\omega}^{*T} \mathbf{u}_1 + \text{const.} , \end{aligned}$$

⁴The t -arguments are omitted in the derivation of the constrained cost function.

where

$$\begin{aligned}
\mathbf{H}^* &= \mathbf{H}_1 - \mathbf{H}_2 \mathbf{S}_2^{-1} \mathbf{S}_1 \\
\boldsymbol{\beta}^* &= \mathbf{H}_2 \mathbf{S}_2^{-1} \mathbf{d} + \boldsymbol{\beta} \\
\boldsymbol{\Lambda}^* &= \boldsymbol{\Lambda}_{11} + (\mathbf{S}_1^T (\mathbf{S}_2^{-1})^T \boldsymbol{\Lambda}_{22} - 2\boldsymbol{\Lambda}_{12}) \mathbf{S}_2^{-1} \mathbf{S}_1 \\
\boldsymbol{\omega}^* &= (\boldsymbol{\Lambda}_{12} - \mathbf{S}_1^T (\mathbf{S}_2^{-1})^T \boldsymbol{\Lambda}_{22}) \mathbf{S}_2^{-1} \mathbf{d} + \\
&\quad \boldsymbol{\omega}_1 - \mathbf{S}_1^T (\mathbf{S}_2^{-1})^T \boldsymbol{\omega}_2 .
\end{aligned} \tag{25}$$

Thus the optimal control is found by minimizing

$$\tilde{J}^*(\mathbf{u}_1) = (\mathbf{H}^* \mathbf{u}_1 + \boldsymbol{\beta}^*)^T \boldsymbol{\Gamma} (\mathbf{H}^* \mathbf{u}_1 + \boldsymbol{\beta}^*) + \mathbf{u}_1^T \boldsymbol{\Lambda}^* \mathbf{u}_1 + 2\boldsymbol{\omega}^{*T} \mathbf{u}_1 ,$$

with respect to \mathbf{u}_1 . The structure of this cost function is similar to the structure of the cost function in (16). Therefore, the solution to the constrained optimization problem is obtained by replacing \mathbf{H} , $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\omega}$ in (18) by \mathbf{H}^* , $\boldsymbol{\beta}^*$, $\boldsymbol{\Lambda}^*$ and $\boldsymbol{\omega}^*$, respectively.

3.5.1 The constraints used by Clarke *et al.*

Below it is shown how to choose \mathbf{H}^* , $\boldsymbol{\beta}^*$, $\boldsymbol{\Lambda}^*$ and $\boldsymbol{\omega}^*$ in order to obtain the GPC method presented by (Clarke et al. 1987). According to (11), this controller is subject to

$$u(t+j-1) = u(t+j-2) , \quad j > N_u > 0 .$$

These constraints can easily be expressed by means of (20). If, for instance, $N = 5$ and $N_u = 2$, then $M = N - N_u = 3$ and

$$\mathbf{S}(t) = \underbrace{\begin{pmatrix} 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}}_{\mathbf{S}_1(t)} , \quad \mathbf{d}(t) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

If $N_1 = 1$ and $N_2 = N = 5$ in the cost function in (10) it is found that (cf.

(14) and (15)):

$$\begin{aligned} \mathbf{\Lambda} &= \left(\begin{array}{cc|ccc} \lambda_1 + \lambda_2 & -\lambda_2 & 0 & 0 & 0 \\ -\lambda_2 & \lambda_2 + \lambda_3 & -\lambda_3 & 0 & 0 \\ \hline 0 & -\lambda_3 & \lambda_3 + \lambda_4 & -\lambda_4 & 0 \\ 0 & 0 & -\lambda_4 & \lambda_4 + \lambda_5 & -\lambda_5 \\ 0 & 0 & 0 & -\lambda_5 & \lambda_5 \end{array} \right) \\ \boldsymbol{\omega}(t) &= \underbrace{(-\lambda_1 u(t-1) \ 0)}_{\boldsymbol{\omega}_1(t)^T} \underbrace{0 \ 0 \ 0}_{\boldsymbol{\omega}_2(t)^T} \\ \mathbf{\Gamma} &= \mathbf{I} \text{ (5 \times 5 identity matrix).} \end{aligned}$$

For the time-invariant ARIMAX model used by (Clarke et al. 1987) the “impulse response matrix”, \mathbf{H} , introduced in (9) becomes time-invariant

$$\mathbf{H} = \left(\begin{array}{cc|ccc} h_1 & 0 & 0 & 0 & 0 \\ h_2 & h_1 & 0 & 0 & 0 \\ h_3 & h_2 & h_1 & 0 & 0 \\ h_4 & h_3 & h_2 & h_1 & 0 \\ h_5 & h_4 & h_3 & h_2 & h_1 \end{array} \right).$$

Observing that

$$\mathbf{S}_2^{-1} \mathbf{S}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 0 & -1 \\ 0 & -1 \end{pmatrix}, \quad (26)$$

it is readily found from (25) that

$$\mathbf{H}^* = \mathbf{H}_1 + \mathbf{H}_2 \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} h_1 & 0 \\ h_2 & h_1 \\ h_3 & h_1 + h_2 \\ h_4 & h_1 + h_2 + h_3 \\ h_5 & h_1 + h_2 + h_3 + h_4 \end{pmatrix} \quad (27)$$

$$\begin{aligned} \boldsymbol{\beta}^*(t) &= \boldsymbol{\beta}(t) \\ \mathbf{\Lambda}^* &= \begin{pmatrix} \lambda_1 + \lambda_2 & -\lambda_2 \\ -\lambda_2 & \lambda_2 \end{pmatrix} \\ \boldsymbol{\omega}^*(t) &= \boldsymbol{\omega}_1(t) = (-\lambda_1 u(t-1) \ 0)^T, \end{aligned}$$

where the vector $\boldsymbol{\beta}(t)$ contain the predicted values of the input-free control errors (defined in (17)).

Note that the first column of \mathbf{H}^* is equal to the first column of \mathbf{H} , and the second column is the sum of columns 2 to 5 of \mathbf{H} . As a general rule it can be established that the first $N_u - 1$ columns of \mathbf{H}^* are copied from \mathbf{H} while the N_u th column is the sum of the remaining columns of \mathbf{H} . This result was also heuristically proposed by (Bjerre 1992). Note also that $\mathbf{\Lambda}^*$ and $\boldsymbol{\omega}^*$ depend on λ_1 and λ_2 , but not on λ_3 , λ_4 and λ_5 . This is quite natural since the latter λ 's have no influence on the optimal solution as the corresponding control increments have been fixed by means of the equality constraints. Moreover, this is in conformity with the results of (Clarke et al. 1987).

4 Reference curves in predictive control

All predictive controllers require that an output reference covering the control horizon is specified. In some applications the reference may be constant over time or a fixed (time dependent) function, whereas in other applications the reference is given as a function of one or more explanatory variables. In the latter case predictions of the explanatory variable(s) will be needed in order to calculate the future reference values.

The XGPC cost function (13) minimized in Section 3 in order to obtain the optimal controller (19) contains terms penalizing the control errors as well as the control effort. Thus, disregarding the influence of the control effort in the control criterion, positive and negative control errors are weighted equally in the control criterion. This implies, that the controller will aim at minimizing the control errors, but not discriminate between realizations above or below the reference signal. In many situations this is as intended, but in some applications the reference signal acts as an output restriction and must be treated as such. Thus the inherent uncertainty in the predictions of system output and explanatory variable(s) must be taken into account, when the output reference values are determined.

A possible solution to this problem could be simply to determine the output reference as the restriction (reference) plus an additional safety margin sufficiently large to avoid violations of the restrictions where the latter is established by experience. Such a margin will often be determined conservatively thus resulting in the system operating too far away from the restrictions (and thus the optimum) compared to a more systematic approach.

Alternatively the reference signal is determined, so that the probability of the controlled variable observing the restrictions are equal to a specified (and

large) value.

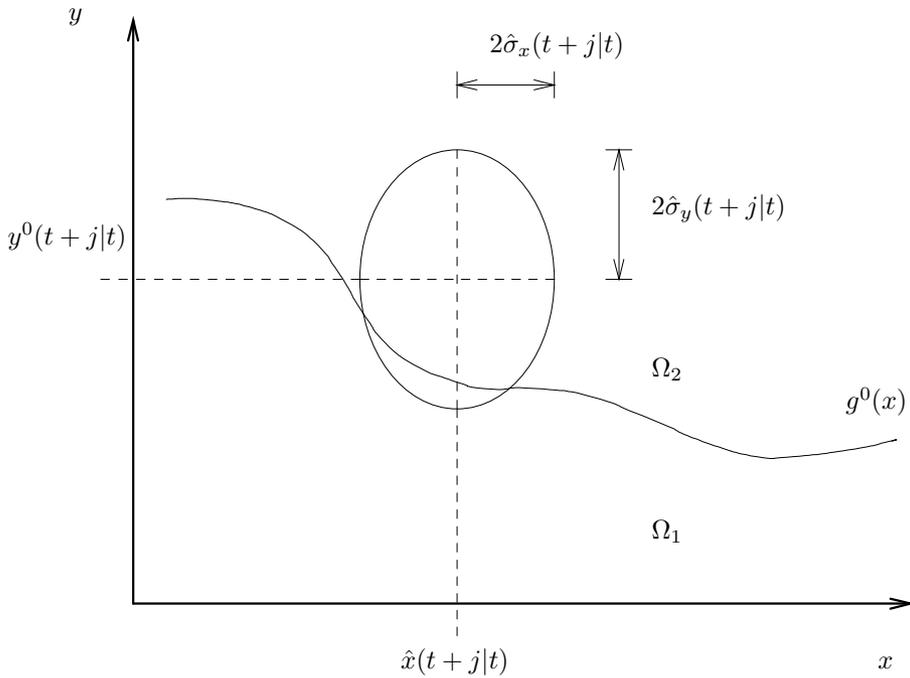


Figure 2: *Inference of an output reference based on model uncertainty for the case where the reference or restriction curve, $g^0(x)$, is a function of one explanatory variable, x . The model uncertainty is specified by estimated standard deviation $\hat{\sigma}_y()$ and $\hat{\sigma}_x()$ for the system model and the predictions of the explanatory variable, respectively. In the figure the “legal” Ω_2 is the area above $g^0(x)$ and the “illegal” Ω_1 is the area below $g^0(x)$.*

Consider a situation where the restrictions $g^0(x)$ are given as a function of one explanatory variable x with the restrictions being violated in the area Ω_1 below $g^0(x)$ and being observed in the area Ω_2 above $g^0(x)$. As stated above the problem is to determine the future reference, $y^0(t+j)$, so that the probability of future controlled and explanatory variables, $y(t+j)$ and $x(t+j)$, respectively, will belong to the “legal” Ω_2 area, is equal to π – see Figure 2. In other words $y^0(t+j)$ should be determined so that

$$P\{(x(t+j), y(t+j)) \in \Omega_2 | I_t\} = \pi, \quad 0 < j < N, \quad (28)$$

where I_t is the information set at time t , N is the cost horizon, and

$$\Omega_2 = \{(x(t), y(t)) \in \mathcal{R}^2 | y(t) \geq g^0(x(t))\}.$$

The future values of x and y can be written as a sum of predictions and

prediction errors, i.e.

$$\begin{aligned} x(t+j) &= \hat{x}(t+j|t) + e_x(t+j|t) \\ y(t+j) &= \hat{y}(t+j|t) + e_y(t+j|t), \end{aligned} \quad (29)$$

where $\hat{x}(t+j|t)$ and $\hat{y}(t+j|t)$ are j -step predictions given at time t of the explanatory and the output variables, respectively, and $\{e_x(t+j|t)\}$ and $\{e_y(t+j|t)\}$ are mutually independent Gaussian distributed but possibly coloured noise sequences with zero mean and standard deviation σ_x^j and σ_y^j , respectively.

Assuming that the system is controlled according to the model used to obtain the predictor $\hat{y}(t+j|t)$ the uncertainty in achieving $y^0(t+j)$ is equal to the uncertainty of $\hat{y}(t+j|t)$. Now $y^0(t+j)$ is found by substituting $\hat{y}(t+j|t)$ in (29) by $y^0(t+j)$ before inserting (29) into (28) thereby obtaining

$$\begin{aligned} P\{(\hat{x}(t+j|t) + e_x(t+j|t), y^0(t+j) + e_y(t+j|t)) \\ \in \Omega_2\} = \pi, \quad j > 0. \end{aligned} \quad (30)$$

Applying $P\{(x, y) \in \Omega_2\} + P\{(x, y) \in \Omega_1\} = 1$ and $P\{(x, y) \in \Omega_1\} = \int_{\Omega_1} f(x, y) dx dy$, where f is a (time dependent) probability distribution function, (30) is rewritten as

$$1 - \pi = \int_{-\infty}^{\infty} \int_{-\infty}^{g^0(x)} f_{t+j}(x(t+j), y(t+j)) dx dy \quad (31)$$

Using (29) and the fact that the noise terms $\{e_x(t+j|t)\}$ and $\{e_y(t+j|t)\}$ are assumed to be mutually independent and Gaussian distributed with zero

mean and standard deviation σ_x^j and σ_y^j , (31) is reformulated as

$$\begin{aligned}
 1 - \pi &= \int_{-\infty}^{\infty} \int_{-\infty}^{g^0(x)} \phi \left(\frac{x - \hat{x}(t+j|t)}{\sigma_x^j(t+j)} \right) \\
 &\quad \phi \left(\frac{y - y^0(t+j)}{\sigma_y^j(t+j)} \right) dy dx \\
 &= \int_{-\infty}^{\infty} \phi \left(\frac{x - \hat{x}(t+j|t)}{\sigma_x^j(t+j)} \right) \\
 &\quad \left[\int_{-\infty}^{g^0(x)} \phi \left(\frac{y - y^0(t+j)}{\sigma_y^j(t+j)} \right) dy \right] dx \\
 &= \int_{-\infty}^{\infty} \phi \left(\frac{x - \hat{x}(t+j|t)}{\sigma_x^j(t+j)} \right) \\
 &\quad \Phi \left(\frac{g^0(x) - y^0(t+j)}{\sigma_y^j(t+j)} \right) dx, \tag{32}
 \end{aligned}$$

where ϕ and Φ are the density, respectively, the probability distribution function for the standardized Gaussian distribution.

Since (32) has no analytical solution $y^0(t+j)$ must be found by numerical methods. In the application presented in Section 8, $y^0(t+j)$ is found using an algorithm based on a combination of Simpson integration and Newton Raphson zero searching.

The proposed method for determining $y^0(t+j)$ does not restrict the function g^0 defining the restriction or control curve to functions of one explanatory variable. If for instance the restrictions are specified as a function of p explanatory variables, $g^0(x_1, \dots, x_p)$, then the output reference, $y^0(t+j)$, is found as the solution to

$$\begin{aligned}
 1 - \pi &= \\
 &\int_{-\infty}^{\infty} \phi \left(\frac{x_1 - \hat{x}_1(t+j|t)}{\sigma_{x_1}^j(t+j)} \right) \cdots \int_{-\infty}^{\infty} \phi \left(\frac{x_p - \hat{x}_p(t+j|t)}{\sigma_{x_p}^j(t+j)} \right) \\
 &\quad \Phi \left(\frac{g^0(x_1, \dots, x_p) - y^0(t+j)}{\sigma_y^j(t+j)} \right) dx. \tag{33}
 \end{aligned}$$

where it is assumed that the prediction errors for the output and the p explanatory variables are mutually independent and Gaussian distributed with zero mean standard deviation $\sigma_y^j(t+j), \sigma_{x_1}^j(t+j), \dots, \sigma_{x_p}^j(t+j)$. The solution

to (33) is found as for (32). However, it should be noted, that the computational burden of the proposed scheme will increase rapidly with the number of explanatory variables in g^0 .

5 The implemented models

The control scheme implemented in the software package PRESS developed for control of district heating systems contains, as described in the following, several different sub-problems, for which model based solutions are needed. The models implemented in PRESS are all linear parametric prediction models dedicated to either 1-step or k -step prediction, where a step corresponds to one hour. The model parameters are in all cases estimated adaptively using recursive least squares estimation with exponential forgetting as proposed by Ljung & Söderström (1983) thereby allowing for time-varying models in the estimation. Where k -step models are employed each prediction horizon is covered by an individual model – that is a separate set of model parameters is estimated for each prediction horizon.

The performance of the implemented models have been evaluated on basis of data collected at the Høje Tåstrup district heating utility during a period from mid April 1996 to end of May 1996 – a period in which the PRESS system was controlling the supply temperature continuously. In all cases the model parameters in the (adaptive) recursive least squares algorithm have been initialized using data from the period prior to the trial period.

The criterion used in the evaluation of the various prediction models is the estimated standard deviation for the prediction error, $\hat{\sigma}_e$. The prediction horizon selected in the evaluation is the 1-step predictions for the 1-step models and 1, 6, 12, 18 and 24 step predictions for the k -step models. The estimated standard deviation for the prediction errors is calculated as

$$\begin{aligned}\hat{\sigma}_e(k) &= \sqrt{\frac{1}{N_{obs} - 1} \sum_{t=1}^{N_{obs}} (e(t|t - k) - \hat{\mu}_e(k))^2} & (34) \\ \hat{\mu}_e(k) &= \frac{1}{N_{obs}} \sum_{t=1}^{N_{obs}} e(t|t - k) \\ e(t|t - k) &= Y(t) - \hat{Y}(t|t - k)\end{aligned}$$

where N_{obs} is the number of observations used in the evaluation, $Y(t)$ is the observed value a time t and $\hat{Y}(t|t - k)$ is the predicted value at time t given

observations up to time $t - k$.

The performance of the selected models has been evaluated by comparison to three simple prediction models: The mean value predictor, the persistent predictor and the persistent diurnal predictor. A verbal description of the three predictors is given in Table 1.

Model	Predictions
$y(t) = \mu + e_{\mu}(t)$	The predictions are equal to the average, $\hat{\mu}$, of the entire series and independent of time and the prediction horizon: $\hat{y}(t+k) = \hat{\mu}$. This predictor will be referred to as the <i>mean value predictor</i> . Note that this predictor is actually non-causal.
$\nabla y(t) = e_{\nabla}(t)$	The predictions are equal to the latest observation and independent of the prediction horizon: $\hat{y}(t+k t) = y(t)$. This predictor is commonly known as the <i>persistent or naive predictor</i> .
$\nabla_{24}y(t) = e_{\nabla_{24}}(t)$	In order to predict $y(t)$ at a future hour, the observation at the same hour of the preceding day is used: $\hat{y}(t+k t) = y(t+k-24)$. This predictor is called the <i>persistent diurnal predictor</i> .

Table 1: *Description of the predictors resulting from some simple models.*

6 The Høje Tåstrup district heating utility

The PRESS control system has been used operationally at the Høje Tåstrup district heating utility during a period from autumn 1995 to spring 1997. The Høje Tåstrup district heating utility is a large system supplying a great part of the Høje Tåstrup suburb⁵ with heat and hot tap water. The system is divided into several sub-networks, which are operated individually. The results described in the present report have been obtained by employing the PRESS system on the sub-network known as the Gasværksvej network, which distributes approximately 30% of the energy in the Høje Tåstrup system. Figure 3 shows an overview over the Gasværksvej sub-network including the branches supplying three net-points, which have been selected as critical points (cf. page 117).

⁵The supplied floor area corresponds to some 2.000.000 m^2 .

The main source for heat supply in the Gasværksvej sub-network is the Gasværksvej heat exchanger unit, but in prolonged periods with cold weather ($T_a < -2$ °C) a peak load boiler at Malervej is put into operation. During these periods the PRESS control is disabled.

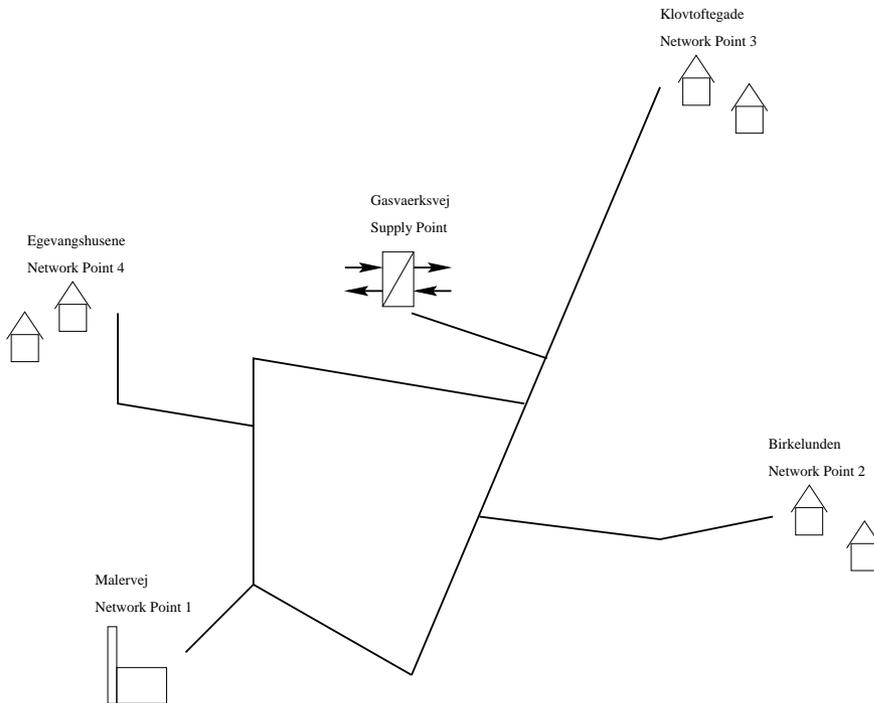


Figure 3: *An overview of the Gasværksvej network in the Høje Tåstrup district heating utility.*

7 Control of the flow rate

Due to limitations in pumping capacity and allowable pressure differences in various parts of the system the flow rate ex district heating station has to be kept below a certain limit – the maximum flow rate. This section describes how a predictive control scheme can be derived for controlling the flow rate ex district heating station. In this context the system output is the flow rate and the input is the supply temperature. A prediction model relating the future flow rates to past and future supply temperatures will have to take the future heat load into account, i.e. it will have to depend on heat load predictions. Rather than identifying a (complex) flow model depending on a heat load prediction model the presented control scheme employs the heat load

predictions directly to calculate the minimum necessary supply temperature imposed by the flow rate restrictions.

Section 7.1 describes the heat load model, and in Section 7.2 it is shown how a control law relating supply temperature and flow rate can be derived from the energy balance equation ex district heating station and the heat load predictions.

7.1 The heat load model

As outlined in the previous section the flow controller depends on predictions of the heat load. Previous model studies carried out in (Madsen et al. 1990), (Madsen et al. 1992) and (Sejling 1993)) have suggested, that a good description of the relationship between past climatic and operational observations, and the future heat load in a district heating system can be obtained by a transfer function model of the following form for the 1-step predictions

$$\begin{aligned} \hat{p}(t+1|t) = & a_1 p(t) + a_2 p(t-1) + a_{24} p(t-23) + a_{25} p(t-24) + a_{26} p(t-25) + \\ & b_{1,0} \nabla T_s(t+1) + b_{1,1} \nabla T_s(t) + b_{1,24} \nabla T_s(t-23) + b_{1,25} \nabla T_s(t-24) + \\ & b_{2,1} T_a(t) + b_{2,2} T_a(t-1) + b_{2,24} T_a(t-23) + b_{2,25} T_a(t-24) + \\ & \sum_{i=1}^3 \left[c_{1,i} \sin \frac{2i\pi(t+1)}{24} + d_{1,i} \cos \frac{2i\pi(t+1)}{24} \right] + \\ & I_{\{A\}}(t) \sum_{i=1}^3 \left[c_{2,i} \sin \frac{2i\pi(t+1)}{24} + d_{2,i} \cos \frac{2i\pi(t+1)}{24} \right] + l, \end{aligned} \quad (35)$$

where

- t is the time,
- $\hat{p}(t+1|t)$ is the one-step-ahead prediction of the heat load at time $t+1$,
- $p(t)$ is the heat load,
- $T_s(t)$ is the supply temperature ($\nabla T_s(t) = T_s(t) - T_s(t-1)$),
- $T_a(t)$ is the ambient air temperature,
- $I_{\{A\}}(t)$ is an indicator which is 0 on workdays and 1 otherwise, and
- l is an adjusting parameter accounting for the variables having mean values different from zero.

This heat load prediction model takes into account the energy storage in the distribution system and the supplied buildings as well as the influence from air

temperature. In addition to the transfer function components (35) contains a third order Fourier expansion of the diurnal load profile for work days and non work days, respectively. Note that all variables are recorded as hourly averages and not as instantaneous values. This is why the prediction at time $t + 1$ is a function of the supply temperature up to time $t + 1$ and not only up to time t . Initially the model employed in Høje Tåstrup also contained terms describing the dependency of wind speed, but the model performance was not improved by including wind speed. This is in accordance with (Sejling 1993), who came to a similar conclusion based on data from Esbjerg.

The corresponding predictors for prediction horizons, j , between 2 and 22 are

$$\begin{aligned}
 \hat{p}(t + j|t) = & \\
 & a_j p(t) + a_{j+1} p(t - 1) + \\
 & a_{24} p(t + j - 24) + a_{25} p(t + j - 25) + a_{26} p(t + j - 26) + \\
 & b_{1,0} \nabla T_s(t + j) + b_{1,1} \nabla T_s(t + j - 1) + b_{1,2} \nabla T_s(t + j - 2) + \\
 & b_{1,3} \nabla T_s(t + j - 3) + b_{1,24} \nabla T_s(t + j - 24) + b_{1,25} \nabla T_s(t + j - 25) + \\
 & b_{2,j} T_a(t) + b_{2,j+1} T_a(t - 1) + b_{2,j+2} T_a(t - 2) + b_{2,j+3} T_a(t - 3) + \\
 & b_{2,24} T_a(t + j - 24) + b_{2,25} T_a(t + j - 25) + \\
 & \sum_{i=1}^3 \left[c_{1,i} \sin \frac{2i\pi(t+j)}{24} + d_{1,i} \cos \frac{2i\pi(t+j)}{24} \right] + \\
 & I_{\{A\}}(t) \sum_{i=1}^3 \left[c_{2,i} \sin \frac{2i\pi(t+j)}{24} + d_{2,i} \cos \frac{2i\pi(t+j)}{24} \right] + l \quad (36)
 \end{aligned}$$

For $j = 21, 24$ some parameters appear twice. In that case additional components following the diurnal components of the dependency on autoregressive heat load and air temperature are introduced to end up with the same number of parameters in the model.

Each prediction horizon is covered by a separate model for which the model parameters are estimated adaptively using a Recursive Least-Squares algorithm as mentioned in Section 5.

The results obtained at the Høje Tåstrup district heating utility by the heat load prediction model (35, 36) are shown in Table 2 for j equal 1, 6, 12, 18 and 24 hours together with the results obtained by the three simple predictors from Section 5. It is seen, that for all horizons but 24 hours the proposed model performs significantly better than any of the three simple predictors. For $k = 24$ the observed results are comparable to the results obtained by the

k	$\hat{\sigma}_e$	$\hat{\sigma}_{e_{\nabla}}$	$\hat{\sigma}_{e_{\nabla_{24}}}$	$\hat{\sigma}_{e_{\mu}}$
1	0.590	0.742	1.501	1.901
6	1.187	1.959	1.501	1.901
12	1.350	1.884	1.501	1.901
18	1.407	2.074	1.501	1.901
24	1.503	1.501	1.501	1.901

Table 2: *The estimated standard deviation of the predictions errors the heat load model (35, 36) ($\hat{\sigma}_e$) and by way of comparison, persistence ($\hat{\sigma}_{e_{\nabla}}$), diurnal persistence ($\hat{\sigma}_{e_{\nabla_{24}}}$) and the mean ($\hat{\sigma}_{e_{\mu}}$) for the prediction horizons 1, 6, 12, 18 and 24 hours. The standard deviations of the prediction errors are estimated on basis of data from the period April 15 1996 to May 31 1996. A forgetting factor, $\lambda = 0.99851$, has been used for all prediction horizons.*

persistent diurnal predictor, which indicates, that the 24-step model is in need of further tuning. Here it should be noted that the same forgetting factor, λ , is used for all prediction horizons and it is most likely that improvements of the model performance can be achieved by optimizing λ for each prediction horizon.

Figure 4 and 5 show plots of observed and predicted heat load as well as the observed prediction errors for k equal 12 and 24 hours.

7.2 The control law for flow rate

The heat load predictions are used as a basis for predictive control of the supply temperature. The purpose of the controller is to perform an on-line control of the supply temperature in order to ensure that the flow rate, $q(t)$, is kept below a maximum value, q_{\max} . Hence it would be more appropriate to have prediction models of the flow rate rather than models of the heat load. It will, however, be shown that heat load models can be used instead of flow models. In order to do that the following relationship between flow rate and heat load is used:

$$p(t) = c_w q(t)(T_s(t) - T_r(t)), \quad (37)$$

where c_w is the heat capacity of water and $T_r(t)$ is the return temperature. Suppose that the future values of the flow rate at time $t + j$ should be kept below q_{\max} with probability π . If the actual time is t this means that

$$P\{q(t + j) \leq q_{\max}\} = \pi, \quad 0 < N_1 \leq j \leq N_2, \quad (38)$$

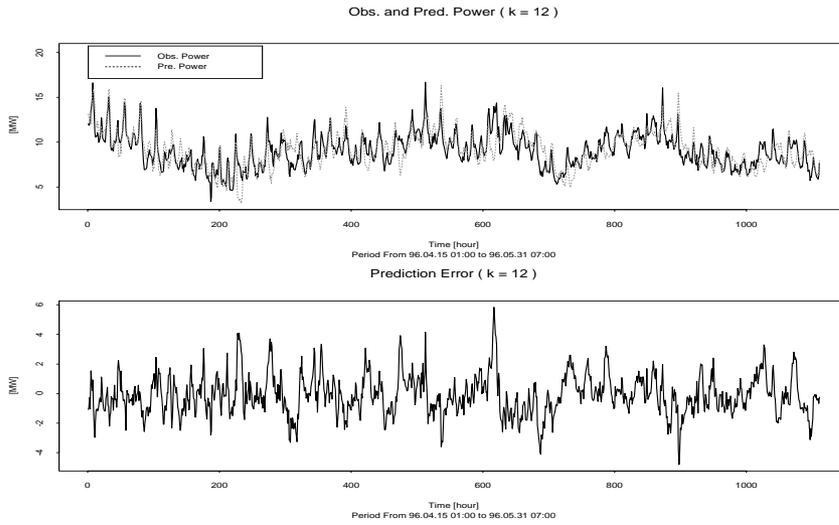


Figure 4: *Top)* Observed heat load (full line) and the 12 hour predictions of heat load (dotted line) at the Høje Tåstrup district heating utility in the period from April 15 1996 to May 31 1996. *Bottom)* The 12 hour prediction error in the same period.

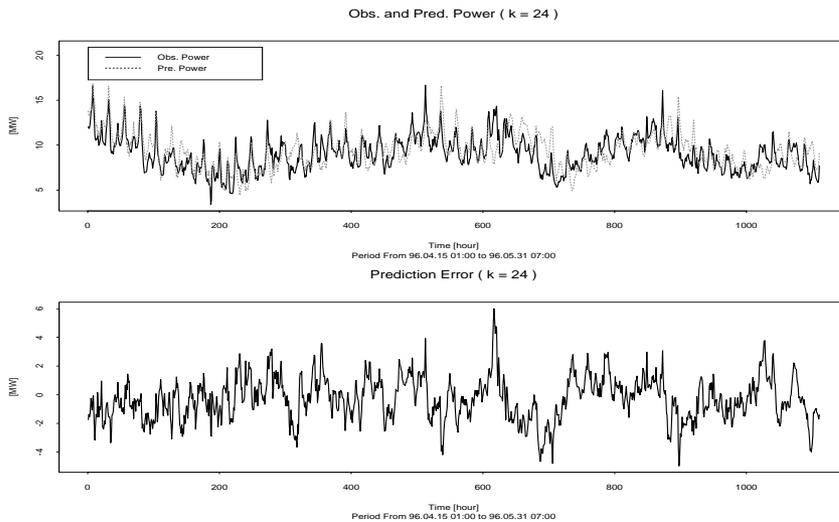


Figure 5: *Top)* Observed heat load (full line) and the 24 hour predictions of heat load (dotted line) at the Høje Tåstrup district heating utility in the period from April 15 1996 to May 31 1996. *Bottom)* The 24 hour prediction error in the same period.

where N_1 and N_2 are chosen so that $q(t+N_1), \dots, q(t+N_2)$ encompass future flow rate values significantly affected by the next control, $T_s(t+1)$. Later on

it will be shown, that these equations provide the reference values, $q^0(t+j)$, of the flow rate implicitly. If (37) is used to substitute for $q(t+j)$ in (38) we obtain

$$P\{p(t+j) \leq c_w q_{\max}(T_s(t+j) - T_r(t+j))\} = \pi. \quad (39)$$

Since the return temperature j steps ahead is unknown, a predictor has to be introduced. As the variation of the return temperature is very slow and with a variance which is much less than the variance of the supply temperature it is sufficient in the present context to use the following random walk model

$$T_r(t) = T_r(t-1) + e(t),$$

where $\{e(t)\}$ is assumed to be white noise with mean zero. The corresponding j -step-ahead predictor is

$$\hat{T}_r(t+j|t) = T_r(t), \quad j = 1, 2, 3, \dots \quad (40)$$

The heat load and the return temperature j steps ahead can be written as a sum of the prediction and the prediction error:

$$\begin{aligned} p(t+j) &= \hat{p}(t+j|t) + e_p(t+j|t) \\ T_r(t+j) &= \hat{T}_r(t+j|t) + e_{T_r}(t+j|t), \end{aligned} \quad (41)$$

where the prediction errors, $e_p(t+j|t)$ and $e_{T_r}(t+j|t)$, are assumed to be mutually independent and to have mean zero and variances $\sigma_p^2(j)$ and $\sigma_{T_r}^2(j)$. Inserting (41) into (39) gives (after a few manipulations)

$$\begin{aligned} &P\{e_p(t+j|t) + c_w q_{\max} e_{T_r}(t+j|t) \\ &\leq c_w q_{\max}(T_s(t+j) - \hat{T}_r(t+j|t)) - \hat{p}(t+j|t)\} = \pi. \end{aligned}$$

Assuming that the prediction errors are Gaussian distributed this can be rewritten as

$$\frac{c_w q_{\max}(T_s(t+j) - \hat{T}_r(t+j|t)) - \hat{p}(t+j|t)}{\sqrt{\sigma_p^2(j) + c_w^2 q_{\max}^2 \sigma_{T_r}^2(j)}} = u_\pi, \quad (42)$$

where u_π is the 100π % quantile in the standardized normal distribution. The next step is to substitute for $\hat{T}_r(t+j|t)$ and $\hat{p}(t+j|t)$ in this equation. But it is necessary to rewrite (35, 36) first to fit the control strategy.

Simulations of the district heating system in Høje Tåstrup at typical flow rates during winter times shows (See (Hansen & Højlund 1997)), that the

heat produced at the Gasværksvej heat exchanger reaches the first consumers within 1 hour and almost all consumers within 3 hours. Therefore the control is based on predictions of the heat load 1, 2 and 3 steps ahead. This means that $j = 1, 2, 3$ are the relevant prediction horizons in (35, 36).

Now introduce a control horizon $N_u = 1$. This means that the choice of the next control, $T_s(t + 1)$, is subject to

$$T_s(t + j) = T_s(t + 1), \quad j = 2, 3 \quad (43)$$

($\Rightarrow \nabla T_s(t + j) = 0$ for $j = 2, 3$). Under this restriction, (35, 36) can be expressed as

$$\begin{aligned} \hat{p}(t + j|t) &= \beta_j \nabla T_s(t + 1) + \gamma_j(t) \\ &= \beta_j (T_s(t + 1) - T_s(t)) + \gamma_j(t), \end{aligned} \quad (44)$$

where

$$\beta_j = \begin{cases} b_{1,0} & \text{if } j = 1 \\ 0 & \text{if } j = 2, 3 \end{cases}$$

and, respectively,

$$\begin{aligned} \gamma_j(t) &= a_1 p(t) + a_2 p(t - 1) + \\ & a_{24} p(t - 23) + a_{25} p(t - 24) + a_{26} p(t - 25) + \\ & b_{1,24} \nabla T_s(t - 23) + b_{1,25} \nabla T_s(t - 24) + \\ & b_{2,1} T_a(t) + b_{2,2} T_a(t - 1) + b_{2,3} T_a(t - 2) + b_{2,4} T_a(t - 3) + \\ & b_{2,24} T_a(t - 23) + b_{2,25} T_a(t - 24) + \\ & \mu_1(t + 1) + I_{\{A\}}(t + 1) \mu_2(t + 1) + l \quad , \quad j = 1, \end{aligned}$$

and

$$\begin{aligned} \gamma_j(t) &= a_j p(t) + a_{j+1} p(t - 1) + \\ & a_{24} p(t + j - 24) + a_{25} p(t + j - 25) + a_{26} p(t + j - 26) + \\ & b_{1,24} \nabla T_s(t + j - 24) + b_{1,25} \nabla T_s(t + j - 25) + \\ & b_{2,j} T_a(t) + b_{2,j+1} T_a(t - 1) + b_{2,j+2} T_a(t - 2) + b_{2,j+3} T_a(t - 3) + \\ & b_{2,24} T_a(t + j - 24) + b_{2,25} T_a(t + j - 25) + \\ & \mu_1(t + j) + I_{\{A\}}(t + j) \mu_2(t + j) + l \quad , \quad j = 2, 3. \end{aligned}$$

Here the diurnal profiles in (35, 36) of the heat load during working and non-working days have been abbreviated μ_1 and μ_2 , respectively.

Inserting (40), (43) and (44) into (42) it is readily seen that

$$\frac{c_w q_{\max}(T_s(t+1) - T_r(t)) - \beta_j(T_s(t+1) - T_s(t)) - \gamma_j(t)}{\sqrt{\sigma_p^2(j) + c_w^2 q_{\max}^2 \sigma_{T_r}^2(j)}} = u_\pi ,$$

or if the equation is solved with respect to $T_s(t+1)$,

$$T_s(t+1) = \frac{S_j + K T_r(t) - \beta_j T_s(t) + \gamma_j(t)}{K - \beta_j} , \quad (45)$$

where

$$\begin{aligned} K &= c_w q_{\max} \\ S_j &= u_\pi \sqrt{\sigma_p^2(j) + K^2 \sigma_{T_r}^2(j)} . \end{aligned}$$

For each of the prediction horizons (1, 2 and 3 hours), (45) results in a value, $T_{s,j}(t+1)$, of $T_s(t+1)$, which is a solution to (38), but the final controller should, of course, provide a unique value of $T_s(t+1)$. Such a value is obtained by constructing a weighted average of $T_{s,1}(t+1)$, $T_{s,2}(t+1)$ and $T_{s,3}(t+1)$:

$$T_s(t+1) = \boldsymbol{\chi}^T \mathbf{T}_s(t+1) , \quad (46)$$

where

$$\boldsymbol{\chi} = (\chi_1 \ \chi_2 \ \chi_3)^T , \quad \sum_{j=1}^3 \chi_j = 1 ,$$

and

$$\begin{aligned} \mathbf{T}_s(t+1) &= (T_{s,1}(t+1) \ T_{s,2}(t+1) \ T_{s,3}(t+1))^T \\ &= \begin{pmatrix} \frac{S_1 + K T_r(t) - b_{1,0} T_s(t) + \gamma_1(t)}{K - b_{1,0}} \\ \frac{S_2 + K T_r(t) + \gamma_2(t)}{K} \\ \frac{S_3 + K T_r(t) + \gamma_3(t)}{K} \end{pmatrix} . \end{aligned}$$

Equation (46) constitute the controller in its operational form. In the implementation at Høje Tåstrup a weighting $\boldsymbol{\chi}^T = (0.4 \ 0.3 \ 0.3)$ have been used in (46), where the weights χ_i roughly correspond to the fraction of the heat produced at time t , which is consumed at time $t+i$ (cf. (Hansen & Højlund 1997)).

8 Control of net-point temperature

The net-point temperature controller monitors the restriction on the operation of the district heating system imposed by the inlet temperature requirements for the consumers. The consumer inlet temperature is monitored by introducing a set of *critical net-points* in the distribution network. The net-points are selected, so that if the temperature requirements for the critical points are satisfied, then the temperature requirements for all consumers are satisfied (cf. page 117).

The net-point temperatures are monitored by a set of controllers – one controller for each net-point – based on the XGPC optimal controller (18). In order to ensure that the requirements of all consumers are satisfied the supply temperature ex the district heating station is selected as the maximum of the supply temperature requirements dictated by the individual controllers.

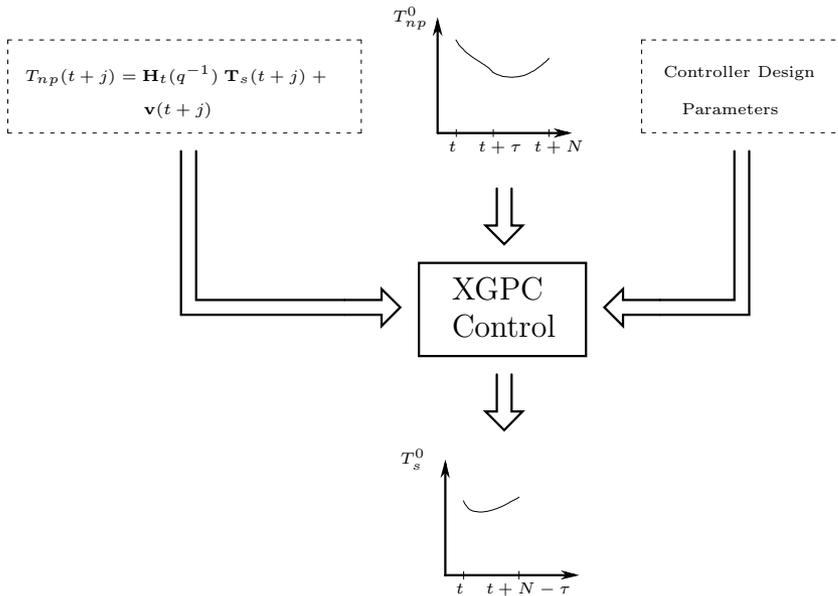


Figure 6: Overview of XGPC net-point temperature controller. From left to right the XGPC requires, that the following informations are supplied: A system model describing how past and future supply temperatures will affect the future net-point temperature, a reference for the future net-point temperature and a specification of the full set of controller design parameters.

This section presents the implemented XGPC control of the net-point temperatures and the different components in the net-point temperature controller are described with focus on aspects of implementation. General considerations regarding the application of the XGPC method in the context of controlling

the net-point temperature in a district heating system are presented and the specific settings used in the Høje Tåstrup implementation of PRESS are described. An overview of the implemented controller is found in Figure 6. As it is seen from the figure the controller requires input from three different sources:

1. *A system model.* The very foundation of the XGPC is a description of how past and future controls will affect the future output from the system. Here the network model presented later on in Section 8.1 is used to calculate the impulse response from supply temperature to net-point temperature (the influence of future controls) and the input free system response (the influence of past control actions).
2. *A reference for the future output (the net-point temperature).* The XGPC takes a sequence of future output reference values into account when calculating the next control value, and consequently a net-point temperature reference covering the output horizon is a prerequisite. The reference values are determined in such a way, that the probability of future net-point temperature observations below a value determined by the future air temperature is fixed (and small). The calculation of a net-point temperature reference is described in Section 8.2.
3. *A set of controller parameters.* Finally the full set of parameters in the XGPC has to be specified. The control parameters used at the Høje Tåstrup district heating utility is specified in Section 8.3.

Finally some additional requirements imposed on the controller by the district heating system are described in Section 8.4.

8.1 The net-point temperature model

In order to control the temperature at the critical points in the distribution network a model describing the dynamics in the network is needed. This could be a deterministic model employing well known physical relationships governing a district heating distribution network. Such a model requires, that a large amount quantities is specified, e.g. the physical layout of the network, pipe dimensions, thermal transmissions coefficients for the pipes and so forth. In older systems many of these informations are simply not available and even for newer systems many of the parameters are not known precisely primarily due to aging of the various components in the transmission

network. Furthermore it should be noted, that when a deterministic network model is used for prediction purposes, it will be depending on predictions of variables such as heat load, which by nature are inherently stochastic. Thus even a “perfect” deterministic model of the distribution network will exhibit discrepancies from the true system, which must be taken into account in the control of the system.

In a situation where on-line data is available and only the relationship between the supply temperature and a limited number of net-points is needed, it seems a better approach to use a simple model, which allows for on-line tuning of the model parameters, to describe this relationship. In PRESS the net-point temperature controller is based on a more simple stochastic model describing the relationship between hourly observations of supply and net-point temperature using an ARX (Auto-Regressive-eXtraneous) model structure. In this context a district heating system is considered as a non-stationary system, where one of the main contributors to the non-stationarities are the changing flow rates in the system as the heat load varies. These non-stationarities can be separated into variations caused by diurnal variations in the heat load and variations caused by the annual change in heat load and climatic conditions. The diurnal variations are incorporated in the model by introducing a deterministic diurnal variation in the model parameters (embedded time-variation), whereas the slow drift in the model parameters caused by the annual changes is accommodated by estimating the model parameters adaptively using a Recursive Least-Squares algorithm.

The changing flow rates in the system introduces a time-varying time delay from supply point to net-point. A number of models are estimated in parallel – one for each possible value of the time-delay. At a given time only the estimates of the ARX models corresponding to the current time delay ± 1 hour is updated. The diurnal variation in the flow rate is accommodated via the embedded diurnal parameter variation of the individual models whereas longer lasting trends in the flow rate is handled by enabling the model which encompass the relevant time delay. In practice this time delay can not be measured, but has to be estimated. The time delay estimation procedure and the ARX-model corresponding to a given time delay are described in the two following sections.

8.1.1 Time-varying time delays between supply and net-point temperature

One of the crucial tasks in a control system is to identify the time delay between input, the supply temperature T_s , and output, the net-point temperature T_{np} . Given a reasonable set of control parameters the XGPC control applied in PRESS is fairly robust towards a wrongly estimated time delay, but never the less the time delay between T_s and T_{np} remains a key parameter in the control system.

The estimation scheme employed at Høje Tåstrup for the time delay between supply point and the various net-points is a further development of an idea first applied in a prior version of PRESS installed at Vestkraft in Esbjerg.

The scheme is based on the assumption, that the time delay τ between two time series, as a first approach, can be found as the lag, where the cross-covariance, γ , between the two time series has the largest numerical value, i.e.

$$\hat{\tau}_{T_s, T_{np}}(t) = \arg \max_{\tau} \hat{\gamma}_{T_s, T_{np}}(\tau, t) \quad , \quad \tau \in [\tau_{min}; \tau_{max}] .$$

An estimate for the time delay, $\hat{\tau}_{T_s, T_{np}}(t)$, at time t is found by maximizing the estimated cross-covariance over τ

$$\hat{\gamma}_{T_s, T_{np}}(\tau, t) = \frac{1}{t - \tau} \sum_{i=\tau+1}^t \left(T_s(i) - \frac{1}{t} \sum_{j=1}^t T_s(j) \right) \left(T_{np}(i - \tau) - \frac{1}{t} \sum_{j=1}^t T_{np}(j) \right) , \quad (47)$$

In (47) it is assumed that two processes $\{T_s(t)\}$ and $\{T_{np}(t)\}$ are second order stationary. Due to the changing heat loads caused by the changing climatical conditions $\{T_s(t)\}$ and $\{T_{np}(t)\}$ are neither first nor second order stationary and thus the cross-covariance estimation algorithm has to be modified. Here the non-stationarity of the $\{T_s(t)\}$ and $\{T_{np}(t)\}$ have been handled by introducing exponentially decaying weight function in the estimation of the mean values and cross-covariance function. Thus in the estimation of the (time-

varying) time delay (47) is replaced by the recursive estimation algorithm

$$\begin{aligned}
 \hat{\gamma}_{T_s, T_{np}}(\tau, t) &= (1 - \lambda_\gamma) (T_s(t) - \bar{T}_s(t)) (T_{np}(t - \tau) - \bar{T}_{np}(t - \tau)) + \\
 &\quad \lambda_\gamma \hat{\gamma}_{T_s, T_{np}}(\tau, t - 1) \\
 \bar{T}_s(t) &= (1 - \lambda_m) T_s(t) + \lambda_m \bar{T}_s(t - 1) \\
 \bar{T}_{np}(t) &= (1 - \lambda_m) T_{np}(t) + \lambda_m \bar{T}_{np}(t - 1)
 \end{aligned} \tag{48}$$

where $\bar{T}_s(t)$ and $\bar{T}_{np}(t)$ are exponentially smoothed average values of $\{T_s(t)\}$ and $\{T_{np}(t)\}$ and λ_γ and λ_m are forgetting factors between 0 and 1.0 in the update of the cross-covariance and the average value estimations, respectively. In (48) it has been used that

$$\lim_{N \rightarrow \infty} \sum_{i=0}^N \lambda^i = \frac{1}{1 - \lambda}$$

The above mentioned scheme supplies a base value for $\hat{\tau}(t)$, which can be altered by plus or minus one time step depending on the parameters of the estimated network models. If, for instance, the net-point model corresponding to the estimated $\hat{\tau}(t)$ shows, that the influence of $T_s(t)$ on $T_{np}(t + \hat{\tau}(t))$ is negligible, the estimated time delay is shifted 1 time step up, and similar, if the net-point model corresponding to the time delay $\hat{\tau}(t) - 1$ shows, that $T_s(t)$ has a considerable impact on $T_{np}(t + \hat{\tau}(t) - 1)$, the time delay is shifted 1 time step down.

8.1.2 The net-point temperature model for a specified time delay

For control purposes a good candidate to a useful net-point temperature model is a simple model with a deterministic diurnal time variation build into the model parameters. Previous model studies (See (Søgaard 1988)) have demonstrated, that a good description of the relationship between the supply temperature and the net-point temperature in a district heating system can be obtained by using a linear transfer function model (an ARX model) with time-varying parameters in the transfer function. The network models used

in Høje Tåstrup are based on a similar model, which takes the form

$$\begin{aligned}
 T_{np}(t) &= a_1 T_{np}(t-1) + b_0(t-\tau) T_s(t-\tau) + \\
 &\quad b_1(t-\tau-1) T_s(t-\tau-1) + \\
 &\quad b_2(t-\tau-2) T_s(t-\tau-2) + e(t) \\
 b_0(t-\tau) &= b_{0,1} + b_{0,2} \sin \frac{2\pi(t-\tau)}{24} + \\
 &\quad b_{0,3} \cos \frac{2\pi(t-\tau)}{24} \\
 b_1(t-\tau-1) &= b_{1,1} + b_{1,2} \sin \frac{2\pi(t-\tau-1)}{24} + \\
 &\quad b_{1,3} \cos \frac{2\pi(t-\tau-1)}{24} \\
 b_2(t-\tau-2) &= b_{2,1} + b_{2,2} \sin \frac{2\pi(t-\tau-2)}{24} + \\
 &\quad b_{2,3} \cos \frac{2\pi(t-\tau-2)}{24}
 \end{aligned} \tag{49}$$

where $T_{np}(t)$ is the observed net-point temperature at time t , $T_s(t)$ is the observed supply temperature at time t , $e(t)$ is a noise sequence, τ is the time delay between net-point and supply temperature and the diurnal variation of $b_0(t)$, $b_1(t)$ and $b_2(t)$ has been implemented by a first order Fourier expansion. The model parameters are estimated adaptively using a Recursive Least Squares algorithm as described in Section 5.

The model formulated in (49) is a 1-step prediction model. The k -step predictions are then obtained by recursive use of the 1-step prediction model. A number of models corresponding to the possible values of the time delay are estimated in parallel, but at a given time only the model encompassing the current time delay as explained previously in Section 8.1.1 is updated.

The evaluation of the net-point model is based on data from the period in May 23 1996 to May 31 1996, where the weather conditions and consequently the estimated network delay have remained stable. However, as the network model is depending on a reliable estimate of the network delay, it should be noted, that the evaluation of the network model carried out in this section in fact encompass both model (49) *and* the time delay estimation scheme previously described in Section 8.1.1.

For each of the net-points the model is evaluated by comparing the standard deviation of the observed 1-step prediction error to the standard deviation of the prediction errors for the three simple prediction models described in Section 5, and the results are shown in Table 3. From the table it is seen, that net-point model (49) clearly out-performs all the simple predictors previously suggested. This is true for all three net-points, but it is in particular notable for the net-points at Birkelunden and Klovtoftegade.

Figure 7, 8 and 9 show plots of observed supply temperature, observed net-point temperature, predicted net-point temperature and prediction error for

NP	$\hat{\tau}$	$\hat{\sigma}_e$	$\hat{\sigma}_{e_{\nabla}}$	$\hat{\sigma}_{e_{\nabla_{24}}}$	$\hat{\sigma}_{e_{\mu}}$
Birkelunden (2)	4	0.383	0.576	1.286	1.183
Klovtoftegade (3)	3	0.579	0.800	2.146	2.217
Egevangshusene (4)	5	0.742	0.835	1.946	1.633

Table 3: The estimated standard deviation of the predictions errors for the net-point temperature model (49) ($\hat{\sigma}_e$) and by way of comparison, persistence ($\hat{\sigma}_{e_{\nabla}}$), diurnal persistence ($\hat{\sigma}_{e_{\nabla_{24}}}$) and the mean ($\hat{\sigma}_{e_{\mu}}$) for the net-points at Birkelunden (2), Klovtoftegade (3) and Egevangshusene (4). The standard deviations of the prediction errors are estimated on basis of data from the period May 23 1996 to May 31 1996. A forgetting factor, $\lambda = 0.99405$, has been used for all net-points.

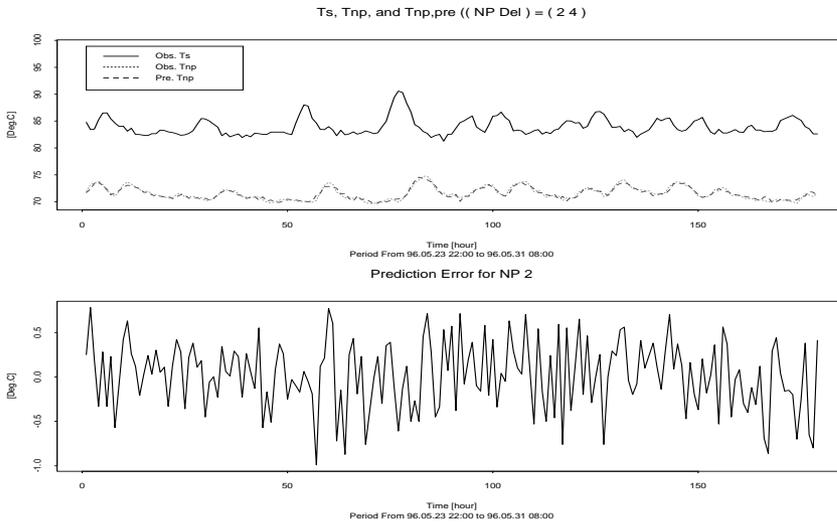


Figure 7: Top) Observed supply temperature (full line), observed net-point temperature (dashed line) and the 1 hour predictions of net-point temperature (dotted line) for the net-point at Birkelunden during the period from May 23 1996 to May 31 1996. Bottom) The 1 hour prediction error in the same period.

Birkelunden, Klovtoftegade and Egevangshusene, respectively. As expected from Table 3 the plots reveal, that the net-point model is working well for Birkelunden and Klovtoftegade, whereas it seems as if the predicted net-point temperature is trailing the observed net-point temperature in most of the observed period for Egevangshusene. This could indicate, that the time delay is wrongly estimated (too large) for Egevangshusene. Changing the time delay downwards to 4 and later 3 hours did not improve the matter though, and the

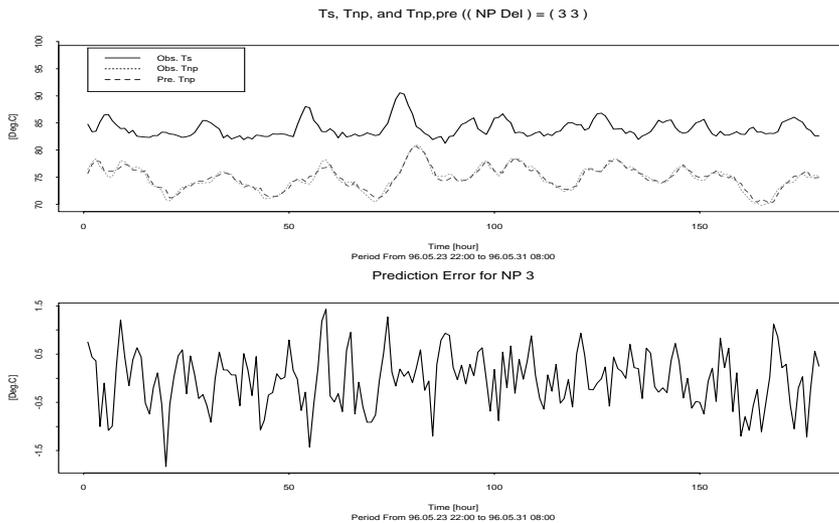


Figure 8: *Top) Observed supply temperature (full line), observed net-point temperature (dashed line) and the 1 hour predictions of net-point temperature (dotted line) for the net-point at Kløvtoftegade during the period from May 23 1996 to May 31 1996. Bottom) The 1 hour prediction error in the same period.*

problem is probably caused by a poor relationship between supply temperature and net-point temperature for Egevangshusene. The problem is most likely due to the fact that the net-point temperature gauge at Egevangshusene is placed at the end of a network branch supplying a series of strip buildings and in periods with low heat demands the temperature readings at the network point are seriously affected by the behavior of the last few consumers at the end of the branch, thus rendering the measurement inapplicable in this context. This notion is further supported by Figure 9, which shows, that the pattern in supply temperature is hardly recognizable in net-point temperature for Egevangshusene.

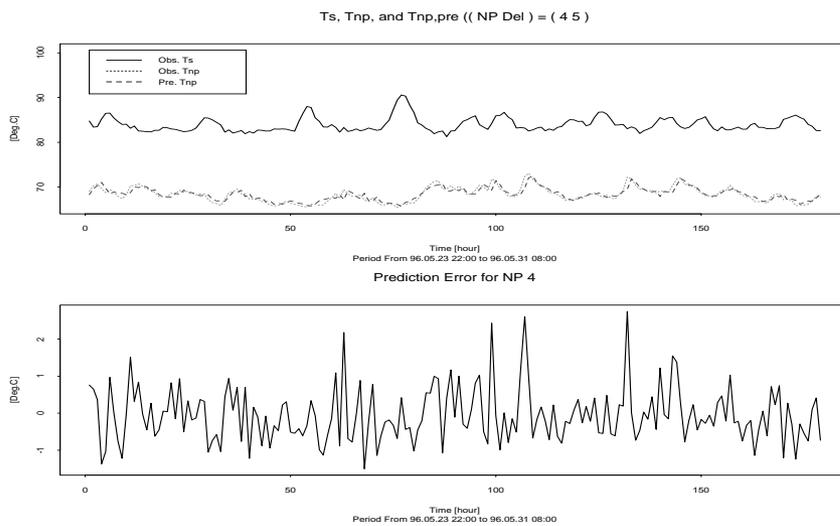


Figure 9: *Top)* Observed supply temperature (full line), observed net-point temperature (dashed line) and the 1 hour predictions of net-point temperature (dotted line) for the net-point at Egevangshusene during the period from May 23 1996 to May 31 1996. *Bottom)* The 1 hour prediction error in the same period.

8.2 The reference net-point temperature

In a traditionally controlled district heating system the supply temperature is determined as a function of the ambient air temperature and it seems reasonable to let the minimum acceptable net-point temperature for the critical net-points be governed by a similar function.

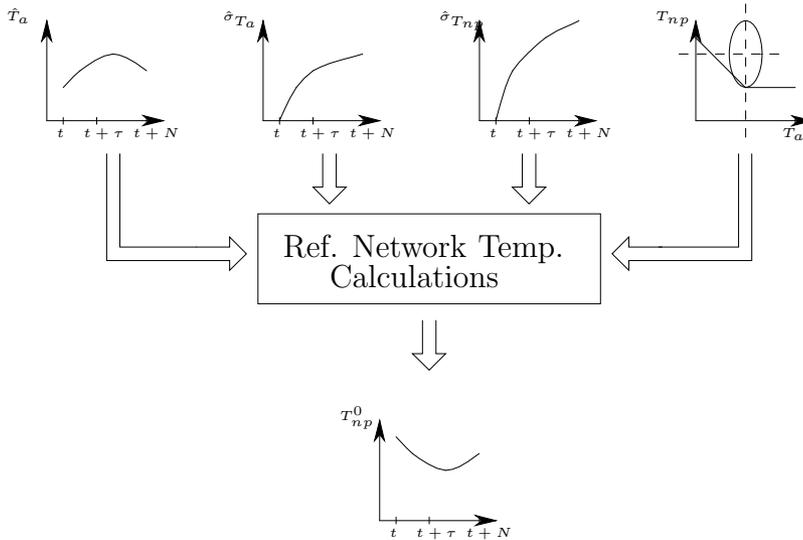


Figure 10: *Reference net-point temperature overview. From left to right the reference net-point temperature calculations require, that the following informations are supplied: A prediction of ambient temperature covering the output horizon (\hat{T}_a), estimates of the standard deviation for the ambient air ($\hat{\sigma}_{T_a}$) and the net-point temperature predictions ($\hat{\sigma}_{T_{np}}$) and finally a control curve expressing the minimum acceptable net-point temperature (T_{np}) as a function of the air temperature (T_a).*

In PRESS the reference net-point temperature is determined using the algorithm outlined in Section 4. Here this temperature is determined, so that the probability of future net-point temperatures below the minimum acceptable inlet temperature is equal to a specified value – chosen to 10% for the installation of PRESS in Høje Tåstrup. The future reference net-point temperature is, as is seen from Figure 10, calculated on basis of the following input:

- A prediction of the ambient air temperature covering the output horizon. The prediction model used in PRESS is described later on in this section.
- The estimated standard deviation of the prediction errors for the am-

bient air temperature and net-point temperature prediction models in Section 8.2.2 and 8.1. In PRESS the standard deviation is estimated using an adaptive and recursive updating scheme

$$S_k^2(t) = (1 - \lambda) (y(t) - \hat{y}(t|t-k))^2 + \lambda S_k^2(t-1), \quad (50)$$

where $S_k(t)$ is the estimated standard deviation for the k -step predictions, $y(t)$ and $\hat{y}(t|t-k)$ are the observed and predicted values of ambient air temperature and net-point temperature, respectively, and λ is a forgetting factor between 0 and 1.0. In the installation of PRESS in Høje Tåstrup $\lambda = 0.98810$ has been used.

- A control curve expressing the minimum acceptable net-point temperature as a function of the air temperature. A method to infer a control curve based on observed data is proposed later on in this section.

8.2.1 Net-point temperature control curve

The net-point reference temperature for a given net-point is, as previously described, calculated on the basis of a predefined relationship between the ambient air temperature and the minimum acceptable net-point temperature for the net-point – the net-point temperature control curve for the critical net-point. One of the problems in establishing a predictive controller for the net-point temperature is, that often very little information is initially available regarding these control curves.

One solution to this problem could be to make a survey of a number of buildings represented by the critical net-point as described in (Hansen & Bøhm 1996) in order to establish the inlet temperature requirements. Here it is found, that the minimum inlet temperature to the investigated consumer installations is governed by a function of ambient air temperature very similar to the curve shown in Figure 11. The increasing net-point temperature with decreasing air temperature reflects the limited capacity in the consumers heating installations, whereas the minimum is determined by the hot tap water installations. If the critical point in question represents an area with a great diversity in the heating installations the amount of work involved may make such an approach infeasible, though.

Another approach is to try to establish the control curve, which should be used in order to maintain status quo for the net-point when introducing the predictive controller. This curve should not be used operationally but serves as a starting point. The following describes a simple technique to determine

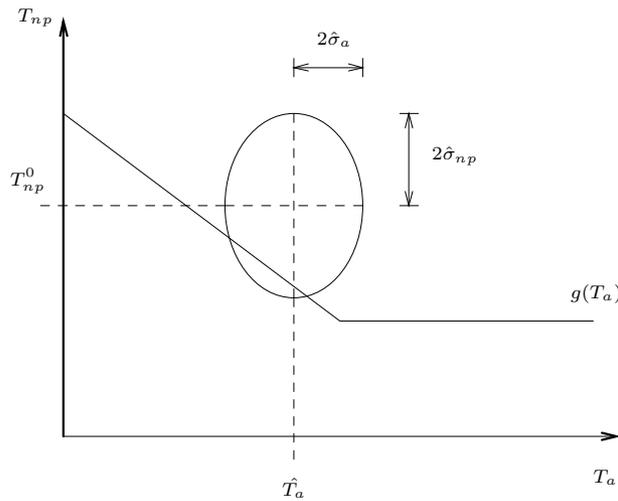


Figure 11: *Reference net-point temperature curve. The reference curve g^0 is a function of the ambient air temperature T_a . For a given (observed) ambient air temperature the corresponding net-point temperature should be in the area above g^0 .*

such a control curve as a function of one explanatory variable – in our case the air temperature. The proposed method is not limited to one variable – if the supply temperature previously has been determined as a function of more than one variable, the proposed scheme can easily be extended to cover this case as well.

The estimation scheme can be outlined as follows:

- The observed range of the observations of the explanatory variable, here the ambient air temperature, is divided into a number of sub-intervals.
- The observations of ambient air and net-point temperature are grouped into the above defined sub-intervals, and the quantile of the net-point temperature distribution corresponding to $1 - \pi$ is established for each interval. The number of observations in each interval must be sufficiently large to give a reasonable estimate of the $1 - \pi$ quantile. As previously defined π is a specified probability of the controller observing the net-point temperature restrictions.
- Finally, the selected control curve – e.g. a curve as defined by g^0 in Figure 11 – is fitted to the observed quantiles by a regression analysis.

In the Høje Tåstrup implementation of PRESS the initial control curves have

been established by the proposed method. Figure 12 illustrates the method using the net-point at Birkelunden as an example.

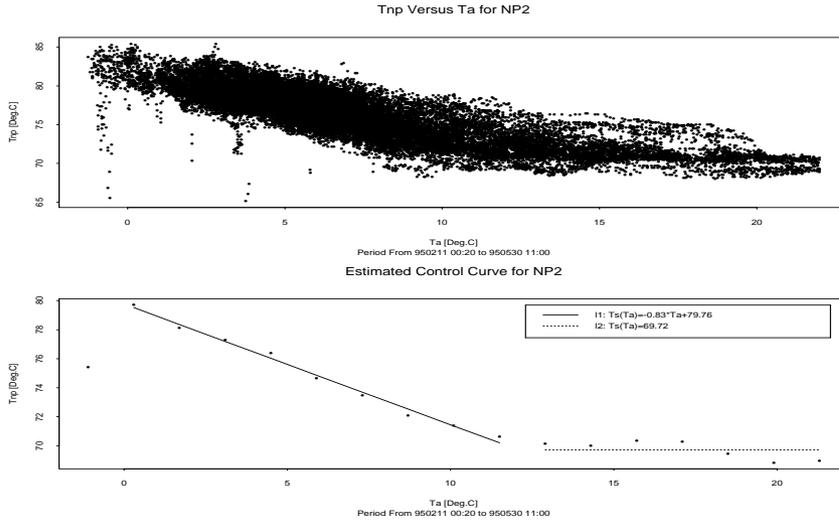


Figure 12: *Estimated Control Curve for the net-point at Birkelunden (2). Top) Observed net-point temperature versus observed ambient air temperature from a period – February 11 1995 to May 30 1995 – prior to the installation of PRESS. Bottom) The estimated control curve based on the 10 % quantile for the observed net-point temperature. The dots indicate the 10 % quantiles in each of the sub-intervals (defined by the ambient air temperature), 11 is the linear regression line for quantiles in intervals below 13 °C and 12 is the mean of quantiles in intervals above 13 °C.*

The estimated control curves are then subject to further adjustment after the trial and error concept – the control curves are simply lowered in small steps until customer complaints occur where after the last step(s) is reversed. The parameters in the control curves used in Høje Tåstrup during the final stages of the trial period are listed in Table 4.

<i>Net – point</i>	$T_{np,min}$	$T_{np,0}$	$T_{a,disc}$
Birkelunden (2)	66.0	74.0	16.0
Klovtoftegade (3)	68.0	78.0	16.0
Egevangshusene (4)	63.0	73.0	16.0

Table 4: *The final control curve parameters used at Høje Tåstrup from May 1996 and onwards. Here $T_{np,min}$ corresponds to 12 in Figure 12, $T_{np,0}$ is the net-point temperature for an ambient air temperature of 0 °C and $T_{a,disc}$ is the ambient air temperature in the intersection between 11 and 12.*

8.2.2 Ambient air temperature prediction model

The predictions of the ambient air temperature are generated by a k -step prediction model, where each of the prediction horizons from 1 to 24 hours ahead has its own dedicated model. Structurally, however, the models covering the different prediction horizons are similar. The k -step prediction model is given as

$$T_a(t) = a_k T_a(t - k) + a_{24} T_a(t - 24) + \mu \sum_{i=1}^2 \left[c_{i,1} \sin \frac{2i\pi t}{24} + c_{i,2} \cos \frac{2i\pi t}{24} \right] + e(t) \quad (51)$$

where $T_a(t)$ is the air temperature at time t , a_k , a_{24} , μ and $c_{i,1}, c_{i,2}$ $i = 1, 2$ are the model parameters to be estimated, and finally $e(t)$ is a noise sequence. The model parameters are estimated adaptively using a Recursive Least-Squares algorithm as described in Section 5. The individual k -step models are, as it is seen from (51), implemented as a weighting between the three simple prediction models from Table 1 with an additional second order Fourier expansion of the diurnal variation of the air temperature.

The selected trial period from April 15 1996 to May 31 1996 covers a wide variety of weather situations typically found during a Danish spring. Due to a change of measurement equipment in early spring 1996, the trial period has not been extended backwards into the winter 1995-1996.

k	$\hat{\sigma}_e$	$\hat{\sigma}_{e_{\nabla}}$	$\hat{\sigma}_{e_{\nabla 24}}$	$\hat{\sigma}_{e_{\mu}}$
1	0.819	1.067	3.055	4.035
6	2.370	4.507	3.055	4.035
12	2.755	6.187	3.055	4.035
18	2.815	4.883	3.055	4.035
24	2.885	3.055	3.055	4.035

Table 5: *The estimated standard deviation of the predictions errors for the ambient air temperature model (51) ($\hat{\sigma}_e$) and by way of comparison, persistence ($\hat{\sigma}_{e_{\nabla}}$), diurnal persistence ($\hat{\sigma}_{e_{\nabla 24}}$) and the mean ($\hat{\sigma}_{e_{\mu}}$) for the prediction horizons 1, 6, 12, 18 and 24 hours. The standard deviations of the prediction errors are estimated on basis of data from the period April 15 1996 to May 31 1996. A forgetting factor, $\lambda = 0.99702$, has been used for all prediction horizons.*

The obtained results in the trial period are listed in Table 5 for k equal 1, 6, 12, 18 and 24 hours together with the performance for the simple predictors.

From the table it is readily seen, that the proposed air temperature model is superior to the three simple prediction models for all the listed prediction horizons.

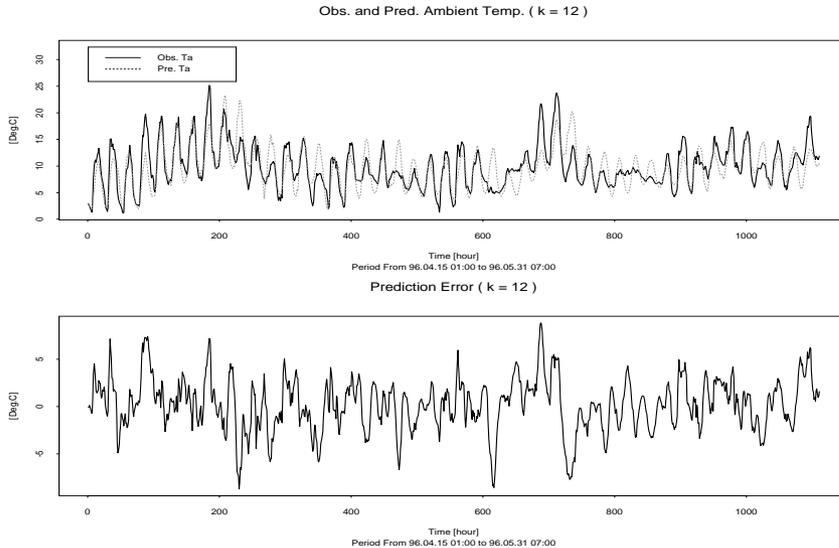


Figure 13: *Top) Observed ambient air temperature (full line) and the 12 hour predictions of ambient air temperature (dotted line) at the Høje Tåstrup district heating utility in the period from April 15 1996 to May 31 1996. Bottom) The 12 hour prediction error in the same period.*

Figure 13 and 14 show plots of observed and predicted air temperature as well as the prediction errors for k equal 12 and 24 hours, respectively. For both prediction horizons it is seen, that during stable weather conditions the predicted air temperature follows the observed air temperature closely. During periods where the weather is changing, i.e. dominated by the passage of pressure fronts, the behavior of the predictors is less favourable, but when considering that the proposed model only utilizes past observations, such a behavior is inevitable. A significant improvement of the forecasts for the local air temperature in those situations will most certainly require, that forecasts from large scale meteorological models are included in the model. The Danish Meteorological Institute (DMI) is about to initiate a new service, SAFE-Energy, which enables the public to subscribe to on-line meteorological forecasts, thus making it realistic for even minor district heating utilities to have access to on-line meteorological forecasts of critical climatical variables.

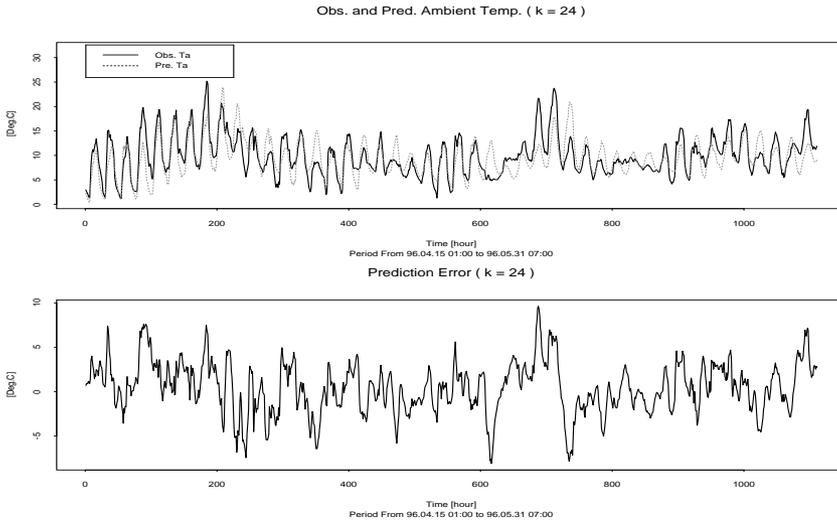


Figure 14: *Top) Observed ambient air temperature (full line) and the 24 hour predictions of ambient air temperature (dotted line) at the Høje Tåstrup district heating utility in the period from April 15 1996 to May 31 1996. Bottom) The 24 hour prediction error in the same period.*

8.3 The design parameters of net-point temperature controller

When implementing an XGPC-control a number of design parameters have to be considered. More specifically the precise formulation of the control criterion as well as the value of the various controller design parameters in the selected criterion has to be determined. In this case the XGPC-control is implemented using the linear equality constrained version of the control criterion from Section 3.5:

$$\begin{aligned}
 \tilde{J}(t, \mathbf{u}_1(t), \mathbf{u}_2(t)) &= (\mathbf{H}^*(t)\mathbf{u}_1(t) + \boldsymbol{\beta}^*(t))^T \boldsymbol{\Gamma}(t) (\mathbf{H}^*(t)\mathbf{u}_1(t) + \boldsymbol{\beta}^*(t)) \\
 &\quad + \mathbf{u}_1(t)^T \boldsymbol{\Lambda}^*(t) \mathbf{u}_1(t) + 2\boldsymbol{\omega}^*(t)\mathbf{u}_1(t) \\
 \boldsymbol{\beta}(t) &= \mathbf{v}(t) - \mathbf{y}^0(t) \\
 \mathbf{H}^*(t) &= \mathbf{H}_1(t) - \mathbf{H}_2(t)\mathbf{S}_2(t)^{-1}\mathbf{S}_1(t) \\
 \boldsymbol{\beta}^*(t) &= \mathbf{H}_2(t)\mathbf{S}_2(t)^{-1}\mathbf{d}(t) + \boldsymbol{\beta}(t) \\
 \boldsymbol{\Lambda}^*(t) &= \boldsymbol{\Lambda}_{11}(t) + \\
 &\quad (\mathbf{S}_1(t)^T(\mathbf{S}_2(t)^{-1})^T \boldsymbol{\Lambda}_{22}(t) - 2\boldsymbol{\Lambda}_{12}(t))\mathbf{S}_2(t)^{-1}\mathbf{S}_1(t) \\
 \boldsymbol{\omega}^*(t) &= (\boldsymbol{\Lambda}_{12}(t) - \mathbf{S}_1(t)^T(\mathbf{S}_2(t)^{-1})^T \boldsymbol{\Lambda}_{22}(t))\mathbf{S}_2(t)^{-1}\mathbf{d}(t) + \\
 &\quad \boldsymbol{\omega}_1(t) - \mathbf{S}_1(t)^T(\mathbf{S}_2(t)^{-1})^T \boldsymbol{\omega}_2(t).
 \end{aligned}$$

Here

is a vector of future controls (supply temperatures)

$$\mathbf{u}(t) \quad \mathbf{u}(t) = [\mathbf{u}_1(t) \ \mathbf{u}_2(t)] \ ,$$

where $\mathbf{u}_1(t)$ and $\mathbf{u}_2(t)$ are vectors of the free controls and the constrained controls, respectively.

$\mathbf{y}^0(t)$ is a vector of future output references (reference net-point temperatures),

$\mathbf{v}(t)$ is a vector containing the expected response from the input free system,

is a matrix containing the time-varying impulse response of the system

$$\mathbf{H}(t) \quad \mathbf{H}(t) = [\mathbf{H}_1(t) \ \mathbf{H}_2(t)] \ ,$$

is a vector containing the expected control errors from the input free system

$$\boldsymbol{\beta}(t) \quad \boldsymbol{\beta}(t) = [\boldsymbol{\beta}_1(t) \ \boldsymbol{\beta}_2(t)] \ ,$$

$\boldsymbol{\Gamma}(t)$ is positive semidefinite and symmetric matrix penalizing the expected control errors given $\mathbf{u}(t)$,

is a positive semidefinite and symmetric matrix penalizing the squared controls and, in combination with $\boldsymbol{\omega}(t)$, defining a (possible) filtering of $\mathbf{u}(t)$

$$\boldsymbol{\Lambda}(t) \quad \boldsymbol{\Lambda}(t) = \begin{bmatrix} \boldsymbol{\Lambda}_{11}(t) & \boldsymbol{\Lambda}_{12}(t) \\ \boldsymbol{\Lambda}_{12}(t)^T & \boldsymbol{\Lambda}_{22}(t) \end{bmatrix} \ ,$$

is a vector penalizing the controls linearly and, in combination with $\boldsymbol{\Lambda}(t)$, defining a (possible) filtering of $\mathbf{u}(t)$

$$\boldsymbol{\omega}(t) \quad \boldsymbol{\omega}(t) = [\boldsymbol{\omega}_1(t) \ \boldsymbol{\omega}_2(t)] \ ,$$

is a matrix which in conjunction with $\mathbf{d}(t)$ defines the linear equality constraints

$$\mathbf{S}(t) \quad \mathbf{S}(t) = [\mathbf{S}_1(t) \ \mathbf{S}_2(t)] ,$$

is a vector which in conjunction with $\mathbf{S}(t)$ defines the equality constraints

$$\mathbf{S}_1(t)\mathbf{u}_1(t) + \mathbf{S}_2(t)\mathbf{u}_2(t) = \mathbf{d}(t) ,$$

and the partitioning of $\mathbf{H}(t)$, $\boldsymbol{\beta}(t)$, $\boldsymbol{\Gamma}(t)$, $\boldsymbol{\omega}(t)$, $\mathbf{S}(t)$ and $\mathbf{d}(t)$ is done as described in Section 3.5. The following tuning parameters have to be selected:

- *The cost horizon N .* In the original presentation of the GPC by (Clarke et al. 1987) it is suggested, that for open-loop stable system the cost horizon of the controller is set equal to or larger than the rise time of the system. For the Høje Tåstrup district heating utility investigations have shown, that $N = 12$ is an appropriate choice. This choice is further supported by the rise times calculated for the system by “SYSTEM RØRNET” in (Hansen & Højlund 1997).
- *The control horizon N_u .* Due to computational costs it is in general suggested that N_u should be kept small. A proper value of this design parameter depends on the complexity (order) of the system. For simple systems $N_u = 1$ gives a reasonably good control. Increasing N_u leads to more active control and smaller control errors. As short term fluctuations of the supply temperature should be kept low this suggests that N_u should be kept low. Here a value, $N_u = 1$, has been chosen.
- *The control error matrix $\boldsymbol{\Gamma}(t)$.* As a first approach this matrix has been selected as the $N \times N$ identity matrix, i.e. all output horizons are weighted identically in the control criterion. Another possibility is to use the inverse covariance matrix of the predictions errors for the net-point model. Thereby control errors, where the corresponding prediction is uncertain, carries less weight in the control criterion.
- *The $\mathbf{S}_1(t)$, $\mathbf{S}_2(t)$ and $\mathbf{d}(t)$ linear equality constraint matrices and vectors.* These matrices define the linear equality constraints imposed on $\mathbf{u}(t)$. To improve the stability of the controller the constraints proposed by (Clarke et al. 1987), where $u(t+j) = u(t+N_u)$, $j > N_u$, are imposed on $\mathbf{u}(t)$. In that case $\mathbf{S}_1(t)$, $\mathbf{S}_2(t)$ and $\mathbf{d}(t)$ become time-invariant

and – as stated in Section 3.5.1 – are then given by

$$\begin{aligned} \mathbf{S}_1 &= \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \\ \mathbf{S}_2 &= \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ \mathbf{d} &= [0 \ 0 \ 0]^T . \end{aligned}$$

- *The quadratic control weight matrix $\mathbf{\Lambda}(t)$.* The only restriction in the choice of $\mathbf{\Lambda}(t)$ is that $\mathbf{\Lambda}(t)$ should be a positive semidefinite and symmetric matrix of dimension N . The matrix $\mathbf{\Lambda}(t)$ can together with $\boldsymbol{\omega}(t)$ be used to define a filtering of $\mathbf{u}(t)$ before $\mathbf{u}(t)$ enters the control criterion. For a district heating system frequent variations of the supply temperature are undesirable. Consequently a filtering of $\mathbf{u}(t)$ has been implemented, which penalizes changes in $\mathbf{u}(t)$ instead of the actual level of $\mathbf{u}(t)$. For simplicity $\nabla \mathbf{u}(t)$ has been weighted uniformly for the different control horizons. In that case $\mathbf{\Lambda}$ is given by

$$\mathbf{\Lambda} = \begin{bmatrix} 2\lambda & -\lambda & \cdots & 0 & 0 \\ -\lambda & 2\lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 2\lambda & -\lambda \\ 0 & 0 & \cdots & -\lambda & \lambda \end{bmatrix} .$$

Experience from the implementation in Høje Tåstrup has shown, that a reasonable compromise between the quality of the control (the discrepancy between expected $\mathbf{y}(t)$ and the reference $\mathbf{y}^0(t)$) and control effort (the changes in $\mathbf{u}(t)$) can be obtained by selecting $\lambda = 0.2$.

- *The linear control weight vector $\boldsymbol{\omega}(t)$.* As argued above it is the changes in $\mathbf{u}(t)$ and *not* the actual level of $\mathbf{u}(t)$ which should be penalized, and together with the choice for $\mathbf{\Gamma}(t)$ this means that $\boldsymbol{\omega}(t)$ is given by

$$\boldsymbol{\omega}(t) = [-\lambda u_{t-1}, 0, \dots, 0]^T ,$$

where $\boldsymbol{\omega}(t)$ is of dimension N .

8.4 Additional requirements in the case of Høje Tåstrup

Provided that the net-point control curve and the critical net-points are reasonably selected, the reference net-point temperature determined in Section 8.2 acts as a reference curve for the net-points, which guarantees, that

the consumer requirements are satisfied. In the Høje Tåstrup case a sensible operation of the system imposes some additional requirements.

The ratio between peak heat load and diurnal average heat load has a significant impact on the price payed by the Høje Tåstrup district heating utility for a given amount of energy – ratios above a certain level are heavily penalized. Hence a heat load displacement has been implemented by superimposed a non-negative⁶ signal upon the required supply temperature as determined by the controller during the morning hours thereby using the distribution network as a heat reservoir. The additional diurnal supply temperature signal used in Høje Tåstrup is found in Table 6.

<i>Hour of day</i>	00:00 – 05:00	05:00 – 06:00	07:00 – 09:00	09:00 – 10:00	10:00 – 23:00
∇T_s^0	0.0	1.0	2.0	1.0	0.0

Table 6: *Additional diurnal supply temperature signal.*

In the implementation of PRESS at the Høje Tåstrup district heating utility the minimum allowable net-point temperature was initially determined as a function of the hourly predictions of ambient air temperature as described in Section 8.2 – an approach previously applied in Esbjerg. However, it soon became apparent, that in Høje Tåstrup this approach leads to an unacceptable large diurnal variation in the supply temperature in periods with large diurnal variations of the ambient air temperature. During spring and fall it is not uncommon with diurnal variation of the ambient air temperature in the region of 10°C to 15°C which, using the control curves defined in Table 3, translates into a diurnal variation of the net-point reference temperature of up to 9°C and – taking the temperature loss into account – more than 10°C in supply temperature. This variation is difficult to reduce significantly by tuning of the controller design parameters (increasing λ) without seriously compromising the quality of control, as it is caused by a badly specified reference value.

An obvious solution is to introduce a low pass filtering of the air temperature before it enters the calculation of the net-point reference curve. Introducing such a low pass filtering is not only advantageous or even necessary from a control point of view, but as the heat capacity of the total thermal mass of the buildings acts as a low pass filter on the influence of air temperature on heat demand, it seems reasonable that a low pass filter should be introduced

⁶Non-negative in order not to compromise the restrictions.

before the air temperature is used to determine the reference net-point temperature. The filter used in Høje Tåstrup is implemented as a rectangular weight function of the air temperature for the past 24 hours

$$f(T_a, t) = \frac{1}{24} \sum_{i=0}^{23} T_{a,t-i} \quad . \quad (52)$$

The use of (52) reduces the diurnal variation of the net-point temperature significantly and reflects the expected filtering of the buildings reasonably well. If future values of filtered air temperature are needed the future values of $T_{a,t-i}$ in (52) are replaced by predictions, cf. page 159. A more physical motivated filter is proposed in (Nielsen & Madsen 2000), where model studies have indicated, that the ambient air temperature should be filtered through a simple first-order filter

$$f(T_a, t) = \frac{0.06}{1 - 0.94q^{-1}} T_a(t) \quad , \quad (53)$$

before the relationship between heat load and ambient air temperature is estimated, and it seems reasonable to expect the required net-point temperature to exhibit a similar dependency on ambient air temperature. (53 has recently been implemented in PRESS.

9 Results and discussion

A control strategy consists, as previously described in Section 2, of an optimization criterion and a set of restrictions imposed by the system. This is reflected in the evaluation of the implemented control, which focuses not only on the controllers capabilities with respect to the optimization criterion – minimizing the supply temperature – but also on the controllers ability to observe the imposed flow and net-point temperature restrictions.

To examine how the implementation of PRESS at the Høje Tåstrup district heating utility has influenced the supply temperature a comparison between data from a three month period during spring 1997 and data from a three month period during spring 1995 prior to the installation of the PRESS controller has been carried out. The two periods are comparable with respect to the observed range and distribution of the ambient air temperature. A quantitative comparison of the supply temperature control in the two periods have been obtained in Figure 15 by estimating a non-parametric regression line for

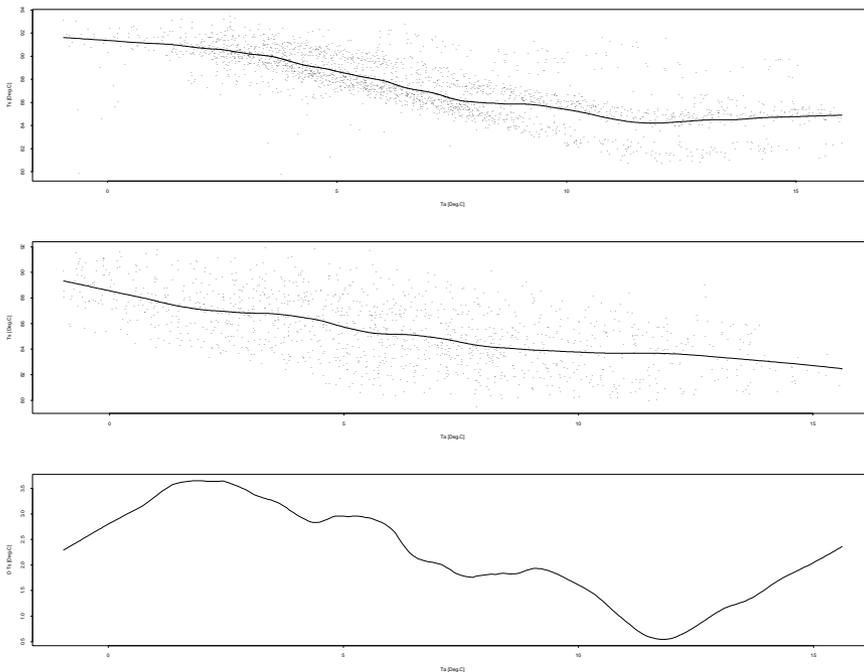


Figure 15: *Top) Supply temperature versus air temperature at Høje Tåstrup for a three month period during spring '95. Middle) Supply temperature versus air temperature at Høje Tåstrup for a three month period during spring '97. Bottom) The savings in supply temperature obtained by using PRESS versus air temperature.*

the relationship between air temperature and supply temperature in each of the two periods. By comparing the two curves it may be concluded that the implementation of the PRESS controller have implied a lowering of the supply temperature ranging from 3.5°C for an air temperature of 2°C to 0.5°C for an air temperature of 12°C . Here it should be noted, that the low-level valve controller at the Gasværksvej heat exchanger station in periods have been quite unwilling to let the supply temperature drop below 82°C . Without this problem the use of PRESS could probably have lowered the supply temperature even more for the higher air temperatures. An important assumption in the above conclusion is, that the implementation of the PRESS controller is the most influential change in the Gasværksvej network from May 1995 to May 1997 – an assumption which to the best of our knowledge is valid.

Due to the operation constraints under which the PRESS controller has operated, it has not been possible to perform a formal evaluation of the implemented flow controller. In the Høje Tåstrup installation the flow restriction normally gets active – i.e. determines the set point for the supply tempera-

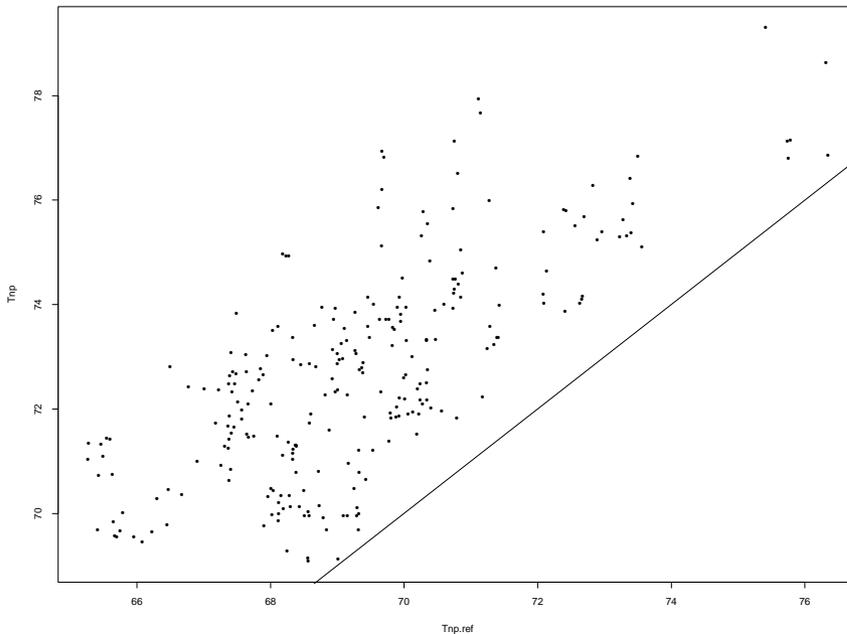


Figure 16: *Observed net-point temperature (T_{np}) versus net-point reference temperature (T_{np}^{ref}) for the net-point at Birkelunden (2) during periods, where the supply temperature has been governed by the corresponding net-point temperature controller. For observations above the full line the controller has observed the net-point temperature restrictions.*

ture – when the air temperature is just above 0°C and even then only around the peak load in the morning hours. As the peak load boilers are in operation for air temperatures below -2°C and the PRESS control consequently is disabled in those periods, it has been impossible to find sufficiently long periods, where the flow controller has been active, to evaluate the controller. It can be said though, that, after some initial problems in determining the maximum allowable flow rate in the system, PRESS has operated the system within the limits imposed by the flow rate.

The second set of restrictions imposed on the control is, that the net-point temperature at the selected critical net-points should be above a minimum value given by the control curve for the net-point with a specified probability. The supply temperature is determined as the present maximum of the supply temperatures recommended by the individual sub-controllers, and during the three month period in spring 1997 the supply temperature has mainly been determined by net-point 4 interrupted by a few periods where net-point 2

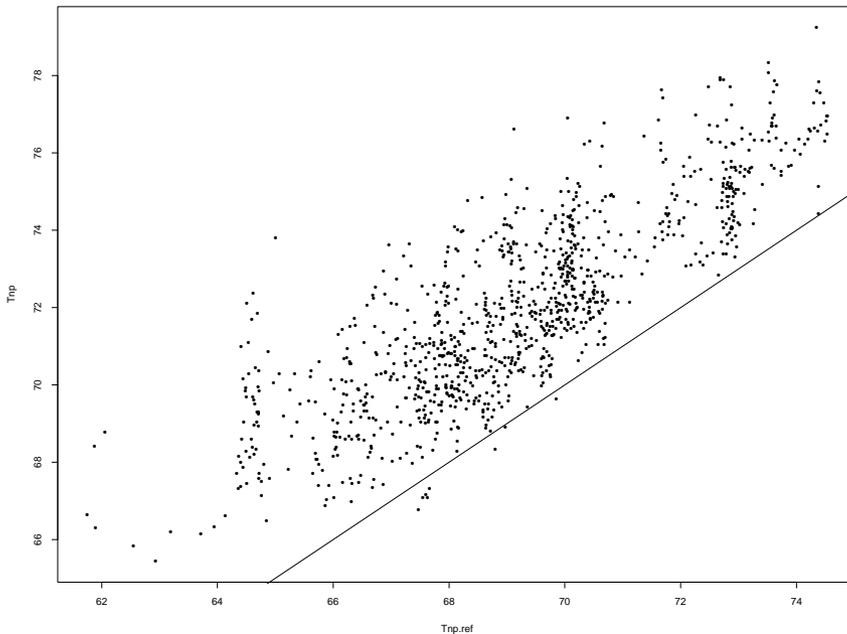


Figure 17: Observed net-point temperature (T_{np}) versus net-point reference temperature (T_{np}^{ref}) for the net-point at Egevangshusene (4) during periods, where the supply temperature has been governed by the corresponding net-point temperature controller. For observations above the full line the controller has observed the net-point temperature restrictions.

has been active. In Figure 16 and 17 the observed net-point temperature has been plotted versus the reference net-point temperature during periods, where the supply temperature has been governed by the corresponding net-point temperature controller for the critical points 2 and 4, respectively. From the plots it is readily seen that the proportion of the observations above the temperature requirement for the critical point (indicated by the full line) with some margin exceeds the 90% specified in Section 8.2. It is therefore concluded that the PRESS controller has been capable of observing the imposed net-point temperature restrictions.

Two predictive controllers – a Predictive Controller based on a physical relation and the Extended Generalized Predictive Controller – have been presented and aspects of their practical implementation in a software system, PRESS, for control of district heating systems have been discussed. PRESS has been tested at the district heating utility in Høje Tåstrup, a suburb of Copenhagen. It has there been demonstrated, that the proposed algorithm is

capable of lowering the supply temperature while keeping a high security of supply for the consumers.

References

- Abraham, B. & Ledolter, J. (1983), *Statistical methods for forecasting*, Wiley, New York.
- Bjerre, T. (1992), Generalized predictive control of energy systems, Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. In Danish.
- Brockwell, P. J. & Davis, R. A. (1987), *Time Series: Theory and Methods*, Springer-Verlag, Berlin/New York.
- Clarke, D. W., Mohtadi, C. & Tuffs, P. S. (1987), 'Generalized predictive control – part I. The basic algorithm', *Automatica* **23**, 137–148.
- Hansen, K. E. & Højlund, C. (1997), Dynamisk beregning af temperaturbilleder, Technical report, Rambøll, Virum, Denmark. In Danish.
- Hansen, K. K. & Bøhm, B. (1996), Nødvendig fremløbstemperatur for eksisterende fjernvarmetilslutningsanlæg - med tilhørende afkøling, Technical Report ET-ES 96-04, Department of Energy Engineering, Technical University of Denmark, Lyngby, Denmark. In Danish.
- Ljung, L. & Söderström, T. (1983), *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA.
- Madsen, H. (1989), *Tidsrækkeanalyse*, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. In Danish.
- Madsen, H., Palsson, O. P., Sejling, K. & Søgaaard, H. (1990), *Models and Methods for Optimization of District Heating Systems*, Vol. I: Models and Identification Methods, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Madsen, H., Palsson, O. P., Sejling, K. & Søgaaard, H. (1992), *Models and Methods for Optimization of District Heating Systems*, Vol. II: Models and Control Methods, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

-
- Nielsen, H. A. & Madsen, H. (2000), Forecasting the heat consumption in district heating systems using meteorological forecasts, Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Oliker, I. (1980), ‘Steam turbines for cogeneration power plants’, *Journal of Engineering for Power* **102**, 482–485.
- Palsson, O. P., Madsen, H. & Søgaaard, H. (1994), ‘Generalized predictive control for non-stationary systems’, *Automatica* **30**, 1991–1997.
- Sejling, K. (1993), Modelling and Prediction of Load in District Heating Systems, PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Søgaaard, H. (1988), Identification and adaptive control of district heating systems, Master’s thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. In Danish.

PAPER E

Control of supply temperature in district heating systems

E

Control of supply temperature in district heating systems

Torben Skov Nielsen and Henrik Madsen

Abstract

The paper describes a concept for controlling the supply temperature in district heating systems using stochastic modelling, prediction and control. The controller minimizes the supply temperature under the restriction that the consumer requirements to inlet temperature are fulfilled and that the flow rate in the system is kept within acceptable limits. The controller is implemented as a set of sub-controllers which operates the system as close to the minimum supply temperature as possible and such that the probability of violating the restrictions is small. The proposed controller has been implemented in a software package - PRESS - which are used operationally at Roskilde Varmeforsyning. The results obtained for the Roskilde district heating utility are evaluated with respect to obtained savings as well as to security of supply.

1 Introduction

District heating plays an important part in covering the heating demands in the Nordic countries, hence the subject of optimal operation of district heating systems has a huge economical potential. This is by no means a trivial subject though, as district heating systems are inherently non-linear and non-stationary, and the issue is further complicated by the fact, that district heating systems are very diverse with respect to production facilities, operational requirements and so forth.

A district heating system can be seen as consisting of three primary parts: one or more central heat producing units, a distribution network and finally the consumer installations for space heating and hot tap water production. The heat production units and the distribution network are often owned by different utilities, thus it makes economic sense from a company point of view to optimize the operation of heat production units and distribution network separately.

Traditionally supply temperature in a district heating system is determined without any feedback from the distribution network. Thus the supply temperature control is in fact an open loop control, and the supply temperature

has to be determined conservatively to ensure a sufficiently high temperature in the district heating network at all times.

This paper considers optimal operation of the distribution network and it is argued that optimal operation is achieved by minimizing network supply temperature under certain restrictions. The proposed control scheme has two objectives. First of all it optimizes the operation of the distribution network with respect to operational costs. Secondly it brings the supply temperature control into a closed loop context thereby making the control a more objective matter compared to the traditional ad hoc approach.

The paper is organized as follows. Section 2 describes the control objectives and restrictions and the control criterion is derived. Hereafter the controller implementation is described in Section 3. In Section 4 the results obtained for Roskilde Varmeforsyning are evaluated, and finally some conclusions are made in Section 5.

2 Control problem

The objectives of the present section is to identify the main conditions under which an optimization of the operational costs for a distribution system is carried out. The operational costs are separated into heating and distribution costs.

For a time interval, $]t - 1, t]$ indexed t , the integrated energy balance for the distribution network can be formulated as

$$E_t^{supply} = E_t^{loss} + E_t^{cons} + \nabla E_t^{netw} \quad (1)$$

where E_t^{supply} is the energy feed into the network at the supply points, E_t^{loss} is the energy loss in the network, E_t^{cons} is the energy delivered to the consumers, and ∇E_t^{netw} is the energy accumulated in the network. The use of an external energy storage – a heat accumulator – is not included in (1) as heat accumulators typically are part of the production system. Any redistributions of heat load within the distribution system must rely on energy storage in the distribution network, which is useful for smoothing of peak loads, but normally does not allow larger rescheduling of heat load.

Following (1) and disregarding energy storage in the distribution network the

accumulated heating costs C^{heat} over a period T can be written as

$$C^{heat} = \sum_T (E_t^{loss} + E_t^{cons}) * P_t \quad (2)$$

where P_t is the cost per unit energy during time interval t . In (2) only E_t^{loss} and for some systems P_t is controllable whereas E_t^{cons} is given.

The heat loss in the network is a (complex) function of the supply temperature, but experience shows that down to a certain limit a decrease in supply temperature implies a lower temperature in the network in general and consequently a decrease in the heat loss from the network. Below the limit the return temperature will increase with decreasing supply temperature (see e.g. Figure 5).

P_t will for some systems be fixed but for other systems increasing levels for supply and return temperature and peak load will imply a higher price per unit energy.

The distribution costs are dominated by the cost of the electricity consumption for the pumps in the distribution network. The supply temperature has direct impact on the pumping costs as flow rate and thus pumping costs will increase with decreasing supply temperature. For most district heating utilities in Denmark the pumping costs are an order of magnitude less than the energy costs associated with the heat loss in the distribution network – hence pumping costs are left out of the optimization. It should be noted though, that an optimization of the pumping strategy may be carried out independent of the control strategy proposed in the following.

The optimization of production cost is carried out under restrictions imposed by the distribution network and consumer installations. The restrictions are mainly due to a maximum limit on the flow rate as well as requirements to a minimum inlet temperature at the consumers installations. Both of these restrictions can be fulfilled by maintaining a sufficiently high supply temperature.

The operation of a district heating utility has a direct impact on the maintenance costs for the network. Large and frequent variations in supply temperature (and pressure) will increase the maintenance costs compared to a more steady operation, hence large and frequent fluctuations in the supply temperature should be avoided.

Based on the above considerations the operational costs of the distribution

network can be optimized by minimizing the supply temperature under the restriction that flow rate, consumer inlet temperatures and variations of supply temperature are kept within acceptable bounds.

It is here assumed that diurnal peak load and return temperature are not adversely affected by the optimization.

3 Controller implementation

In the following a control scheme for optimal operation of a certain class of distribution networks is proposed; namely distribution networks which primarily are supplied from a single supply point. The optimization is implemented as a set of controllers, which operates the system as close to the minimum supply temperature as possible without actually violating the restrictions. The flow rate is monitored by a single controller whereas the consumer inlet temperature is monitored by introducing a set of *critical points* in the distribution network. The critical points are selected so that if the temperature requirements for the critical points are satisfied then the temperature requirements for all consumers are satisfied. Thus, as illustrated in Figure 1, the control system consists of a flow controller and a net-point temperature controller for each critical net-point. At a given time the supply temperature implemented is then selected as the maximum of the recommended supply temperatures from the individual controllers. The sub-controller determining the supply temperature at a given time is called the *active* controller.

The restrictions on the variability of supply temperature are fulfilled by a tuning of the controller design parameters in the flow rate and net-point temperature controllers.

The controller implements a number of additional features:

- Rate of change for supply temperature is restricted. In Roskilde $\nabla T_s^{max} = 1.5^\circ\text{C}$.
- Minimum and maximum values for supply temperature. In Roskilde $T_s^{min} = 70.0^\circ\text{C}$ and $T_s^{max} = 95.0^\circ\text{C}$.
- Diurnal increase for supply temperature in order to reduce peak loads. In Roskilde T_s is increased with 2.0°C from 05:00 to 08:00 in the morning.

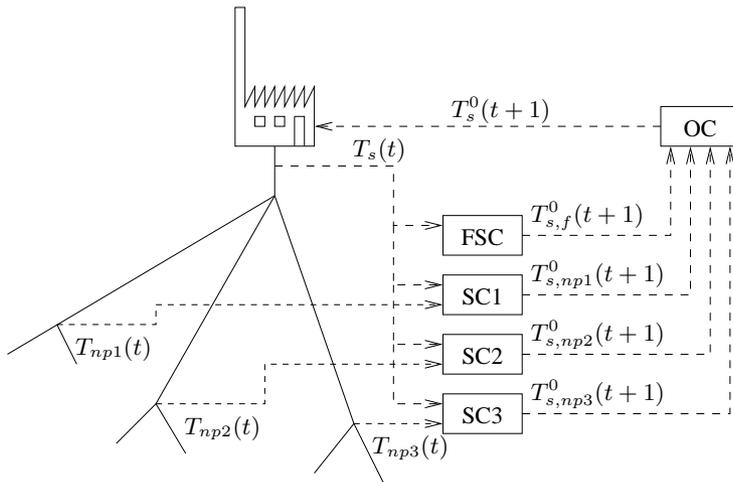


Figure 1: Overview of a district heating network, with 3 critical points and the controllers. OC is the overall controller, FSC is the flow sub-controller, SC# are the supply temperature sub-controllers, $T_{np\#}$ are the supply temperatures in the network, T_s is the supply temperature from the plant, and $T_{s,\#}^0$ are the supply temperatures required by the sub-controllers.

The dynamic relationships between supply temperature and flow rate and net-point temperatures are time-varying and difficult to establish due to the time-varying heat load in the system. Hence the control problem calls for methods, which will operate reliable under these circumstances. The flow and net-point temperature controllers are described in Section 3.1 and 3.2, respectively, whereas the subject of determining reference values for controllers monitoring restrictions are referred to Section 3.3.

3.1 Net-point temperature sub-controller

The net-point temperature controller depends on a model describing the dynamic relationship between supply temperature and net-point temperature(s). Due to the changing heat load this relationship exhibits a diurnal as well as an annual variation. In (Søgaard 1988) a stochastic model describing the relationship between (hourly) observations of supply and net-point temperatures

is identified. The model is given as

$$\begin{aligned}
 A_t(q^{-1})T_{np,t} &= B_t^1(q^{-1})T_{s,t} + \\
 &B_t^2(q^{-1})\cos\frac{2\pi t}{24}T_{s,t} + \\
 &B_t^3(q^{-1})\sin\frac{2\pi t}{24}T_{s,t} + e_t \quad (3) \\
 A_t(q^{-1}) &= 1 + a_t^1q^{-1} \\
 B_t^i(q^{-1}) &= q^{-\tau}(b_t^{0,i} + b_t^{1,i}q^{-1} + b_t^{2,i}q^{-2})
 \end{aligned}$$

where T_t^s and T_t^{np} are supply temperature and net-point temperature given at time t , respectively, q^{-1} is the back-shift operator, a_t^1 and $b_t^{0..2,1..3}$ are time-varying model parameters and τ is a time delay. The diurnal variation in the system dynamics is incorporated directly in the model, whereas the slow drift in the model parameters caused by the annual changes are accommodated by estimating the model parameters adaptively using a Recursive Least-Squares algorithm. The model formulated in (3) is a 1-step prediction model, where j -step predictions are obtained by recursive use of the 1-step prediction model.

The time delay specified in the model is time-varying and has to be estimated. Here the time delay is determined using a scheme based on maximizing the cross correlation between the supply and net-point temperature time series. More details are found in (Madsen et al. 1996).

The model (3) gives raise to a number of requirements on the net-point temperature controller. It must be robust toward non-minimum phase system (due to the possibility of wrongly specified time delays in the model) as well as being capable of handling time-varying systems. The controller should also be reasonably easy and robust to derive since the controller parameters are likely to change hourly as the model parameters are updated. The net-point temperature control is based on the Extended Generalized Predictive Controller (XGPC) proposed in (Palsson et al. 1994), which is a further development of the Generalized Predictive Controller (GPC) presented by Clarke et al. (1987). The main difference between the XGPC and GPC algorithms is found in the derivation of the control law. The XGPC uses conditional expectation to separate model output into a term with a linear dependency on future input values (control values) and a term depending on past input and output values, where a similar separation for the GPC is achieved by recursively solving a Diophantine equation. Furthermore the formulation of the GPC depends on a specific model structure (ARIMAX) whereas the only requirement on the model structure posed by the XGPC is that the future model output is separable as described above.

Formulation of the XGPC control law is illustrated in the following for an ARMAX model. Consider a ARMAX model with time varying parameters

$$A_t(q^{-1}) y_t = B_t(q^{-1}) u_t + C_t(q^{-1}) e_t \quad (4)$$

Using conditional expectation the j -step output prediction is easily calculated as

$$\begin{aligned} \hat{y}_{t+j|t} = & - \sum_{i=1}^n a_{i,t+j} \hat{y}_{t+j-i|t} + \sum_{i=1}^m b_{i,t+j} u_{t+j-i} + \\ & \sum_{i=1}^r c_{i,t+j} \hat{e}_{t+j-i|t} \quad , \quad j-i \geq 1 \end{aligned} \quad (5)$$

where

$$\begin{aligned} \hat{y}_{t+j|t} = & - \sum_{i=1}^n a_{i,t+j} \hat{y}_{t+j-i|t} + \sum_{i=1}^m b_{i,t+j} u_{t+j-i} + \\ & \sum_{i=1}^r c_{i,t+j} \hat{e}_{t+j-i|t} \quad , \quad j-i \geq 1 \end{aligned} \quad (6)$$

$$\hat{y}_{t+j|t} = y_{t+j} \quad , \quad j < 1$$

$$\hat{e}_{s|t} = \begin{cases} e_s = y_s - \hat{y}_{s|s-1} & \text{if } s \leq t \\ 0 & \text{if } s > t \end{cases}$$

and n, m, r in (5) are the order of the A_t, B_t, C_t polynomials in (4). The separation of model output is achieved by using conditional expectation:

- The system impulse response denoted $h_t(q^{-1})$ is calculated as the model output conditioned on a unit impulse control at time t and otherwise zero.
- The input free system response denoted v_t is calculated as the predicted output conditioned on future controls equal zero.

The conditional j -step prediction is now written as

$$\hat{y}_{t+j|t} = \sum_{i=1}^j h_{i,t+j} u_{t+j-i} + v_{j,t} \quad (7)$$

Using matrix notation the output predictions (7) for horizons between 1 and N is written as

$$\hat{\mathbf{y}}_t = \mathbf{H}_t \mathbf{u}_t + \mathbf{v}_t \quad (8)$$

where

$$\begin{aligned}\hat{\mathbf{y}}_t &= (\hat{y}_{t+1|t}, \dots, \hat{y}_{t+N|t})^T \\ \mathbf{u}_t &= (u_t, \dots, u_{t+N-1})^T \\ \mathbf{v}_t &= (v_{1,t}, \dots, v_{N,t})^T \\ \mathbf{H}_t &= \begin{pmatrix} h_{1,t+1} & 0 & \cdots & 0 \\ h_{2,t+2} & h_{1,t+2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ h_{N,t+N} & h_{N-1,t+N} & \cdots & h_{1,t+N} \end{pmatrix}\end{aligned}$$

The XGPC utilizes a cost function of the form

$$\begin{aligned}\min_{\mathbf{u}_t} J(\mathbf{\Gamma}_t, \mathbf{\Lambda}_t, \boldsymbol{\omega}_t; t, \mathbf{u}_t) &= \\ E[(\mathbf{y}_t - \mathbf{y}_t^0)^T \mathbf{\Gamma}_t (\mathbf{y}_t - \mathbf{y}_t^0) + \mathbf{u}_t^T \mathbf{\Lambda}_t \mathbf{u}_t + 2\boldsymbol{\omega}_t^T \mathbf{u}_t] & \quad (9)\end{aligned}$$

where \mathbf{y}_t is a vector of future outputs, \mathbf{y}_t^0 is a vector of future reference values, $\mathbf{\Gamma}_t$ is a positive semidefinite and symmetric matrix weighting the control errors, $\mathbf{\Lambda}_t$ is a positive semidefinite and symmetric matrix weighting the squared control values, and $\boldsymbol{\omega}_t$ is a vector weighting the control values linearly.

Inserting $\mathbf{y}_t = \hat{\mathbf{y}}_t + \mathbf{e}_t$ and (8) into (9) and minimizing with respect to \mathbf{u}_t results in the XGPC control law

$$\begin{aligned}\mathbf{u}_t &= \\ -[\mathbf{H}_t^T \mathbf{\Gamma}_t \mathbf{H}_t + \mathbf{\Lambda}]^{-1} [\mathbf{H}_t^T \mathbf{\Gamma}_t (\mathbf{v}_t - \mathbf{y}_t^0) + \boldsymbol{\omega}_t] & \quad (10)\end{aligned}$$

The choice of the control parameters $\mathbf{\Gamma}_t$, $\mathbf{\Lambda}_t$ and $\boldsymbol{\omega}_t$ determines the behavior of the XGPC. Selection of control parameters is treated in (Madsen et al. 1996).

3.2 Flow sub-controller

A prediction model relating the future mass flow to past and future supply temperatures has to take the future heat load into account, i.e. it will have to depend on heat load predictions. Instead of identifying a mass flow model which depends on output from a heat load model, the control scheme proposed in the following employs the heat load predictions directly to calculate the minimum necessary supply temperature imposed by the mass flow restrictions. Using the energy balance equation

$$p_t = c_w q_t (T_{s,t} - T_{r,t}) \quad , \quad (11)$$

where c_w is the specific heat of water, p_t is the heat load, q_t is the flow rate and $T_{s,t}$, $T_{r,t}$ are supply and return temperature, respectively, and considering only the system response to the next time point this leads to the following control

$$T_{s,t+1}^1 = \hat{T}_{r,t+1|t} + \frac{\hat{p}_{t+1|t}}{c_w q^0} \quad (12)$$

where the observed return temperature $T_{r,t}$ and mass flow q_t in (11) have been replaced by the predicted return temperature $\hat{T}_{r,t+1|t}$ and the maximum value for the mass flow q^0 , respectively.

A change of supply temperature at time t will affect the mass flow in the system until the introduced temperature gradient has reached the most remote (distant) consumers which in a large district heating utility will take several hours. The flow control should therefore be based on the heat load predictions for the horizons mostly affected by a change in supply temperature as opposed to (12), where only the heat load prediction to time $t + 1$ is considered. The controller is based on (12) for the considered horizons, and then calculating the supply temperature as a weighted sum of the individual $T_{s,t+1}^{(j)}$

$$T_{s,t+1}^{(j)} = \hat{T}_{r,t+j|t} + \frac{\hat{p}_{t+j|t}}{c_w q_{t+j|t}^0} \quad (13)$$

$$T_{s,t+1} = \sum_{j=N_1}^{N_2} \gamma_j T_{s,t+1}^{(j)}, \quad \sum_{j=N_1}^{N_2} \gamma_j = 1$$

The weights γ_j is found as the fraction of the heat produced at time t , which is consumed at time $t + j$. In (13) $\hat{T}_{r,t+j|t} = T_{r,t}$ have been used as a predictor for the return temperature – a simplification which seems reasonable due to the small variations in $T_{r,t}$.

The heat load predictions in (13) are calculated using the model

$$A_t^k(q^{-1})p_t = B_{1,t}^k(q^{-1})\nabla T_{s,t} + B_{2,t}^k(q^{-1})T_{a,t} + \mu_{1,t}^k + I_{a,t}\mu_{2,t}^k + l^k + e_t^k \quad (14)$$

where $A_t^k(q^{-1})$, $B_{1,t}^k(q^{-1})$ and $B_{2,t}^k(q^{-1})$ are time-varying polynomials in the back-shift operator (q^{-1}), ∇ is the difference operator, $I_{a,t}$ is an indicator function which is 0 on work days and 1 otherwise, $\mu_{1,t}^k$ and $\mu_{2,t}^k$ are diurnal profiles for working and non-working days, respectively, l^k is a mean value and e_t^k is a noise term.

3.3 Controller reference value

The XGPC control law (10) requires, that an output reference is specified. The reference may be constant over time or given as a function of one or more explanatory variables. In the latter case predictions of the explanatory variable(s) is needed in order to calculate the future reference values.

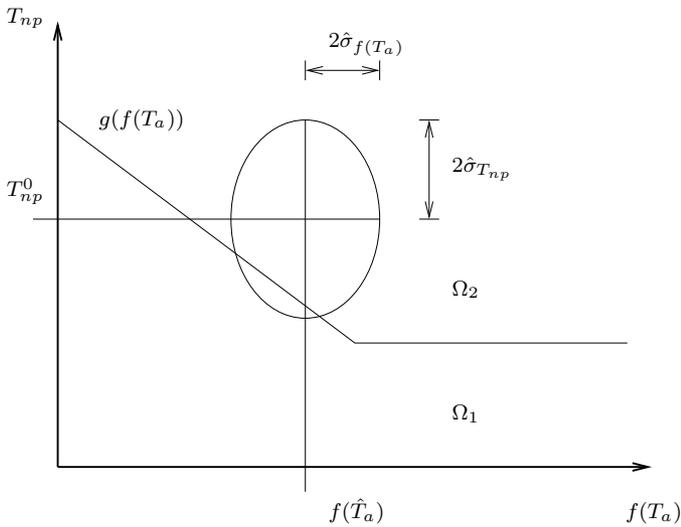


Figure 2: Reference net-point temperature curve. The reference curve determines the required net-point temperature as a function of the low pass filtered air temperature – $f(T_a)$.

In a traditionally controlled district heating system the supply temperature is often determined as a function of the current air temperature, and it seems reasonable to let the minimum acceptable net temperature in the critical net-points be governed by a similar function as illustrated in Figure 2. The increasing net temperature with decreasing air temperature reflects the limited capacity in the consumers room heating installations, whereas the minimum is determined by the hot tap water installations. This is also in accordance with (Hansen & Bøhm 1996), where the requirements on supply temperature is investigated for a number of building.

The heat capacity of the total thermal mass of the buildings acts as a low pass filter on the influence of air temperature on heat demand, hence the required net-point temperature is determined as a function of the filter air temperature. The filter function $f()$ is implemented as a rectangular weight

function of the air temperature for the past 24 hours

$$f(T_a, t) = \frac{1}{24} \sum_{i=0}^{23} T_{a,t-i} \quad . \quad (15)$$

The use of (15) reduces the diurnal variation of the net-point temperature significantly and reflects the expected filtering of the buildings reasonably well. A more physical motivated filter is proposed in (Nielsen & Madsen 2000), where model studies have indicated, that the ambient air temperature should be filtered through a simple first-order filter

$$f(T_a, t) = \frac{0.06}{1 - 0.94q^{-1}} T_a(t) \quad , \quad (16)$$

before the relationship between heat load and ambient air temperature is estimated, and it seems reasonable to expect the required net-point temperature to exhibit a similar dependency on ambient air temperature. (16 has recently been implemented in PRESS.

The XGPC cost function (9) penalizes the quadratic control errors, hence positive and negative control errors are weighted equally in the control criterion. This implies, that the controller will aim at minimizing the control errors, but not discriminate between realizations above or below the reference signal. In many situations this is as intended, but in some applications the reference signal acts as an output restriction and the uncertainty in the predictions of system output and explanatory variable(s) must be taken into account, when the output reference values are determined. In PRESS the reference values are determined so that the probability of future net-point temperature observations below a value given as a simple function of the future air temperature is fixed (and small).

Figure 2 establishes the “legal” area Ω_2 as the area above $g(f(T_a))$ as well as the “illegal” area Ω_1 below $g(f(T_a))$, where in the latter case the consumer inlet temperature restriction is violated. The reference net-point temperature $T_{np,t+j|t}^0$ is determined so that

$$\begin{aligned} P\{ (f(T_{a,t+j}), T_{np,t+j}) \in \Omega_2 \mid I_t \} &= \pi \quad , \quad j > 0 \\ \{ (f(T_{a,t}), T_{np,t}) \mid T_{np,t} \geq g(f(T_a)) \} &= \Omega_2 \end{aligned} \quad (17)$$

where $f(T_{a,t+j})$ and $T_{np,t+j}$ are future values of filtered air temperature and net-point temperature, respectively, $1 - \pi$ is the probability of violating the restriction, and I_t is the information set at time t . Given the distribution of

the prediction errors for $f(T_{a,t+j})$ and $T_{np,t+j}$ (17) and inserting

$$\begin{aligned} f(T_{a,t+j}) &= f(\hat{T}_{a,t+j|t}) + e_{f(T_a),t+j|t} \\ T_{np,t+j} &= T_{np,t+j|t}^0 + e_{T_{np,t+j|t}} \end{aligned}$$

into (17) the resulting equation is readily solved with respect to $T_{np,t+j|t}^0$ by numerical methods.

4 Results obtained in Roskilde

The models and controllers are implemented in a software system called PRESS described in (Madsen et al. 1996) and (Nielsen, Madsen & Nielsen 2001). PRESS is installed at Roskilde Varmeforsyning – a district heating utility supplying Roskilde City and suburbs. Heat is supplied from the VEKS transmission system, which distributes the heat production from CHP and waste incineration plants in the eastern part of Zealand. A peak load boiler is installed in the distribution network, but is rarely used.

The annual heat purchase is 1.700.000 GJ with a maximum heat load of 110 MW. The supply area of Roskilde Varmeforsyning consists of two separate distribution network, where PRESS controls the larger of these. The controlled area corresponds to 55% of the supply area.

PRESS has been used operationally at Roskilde Varmeforsyning since the beginning of January 2001. Prior to the installation of PRESS the supply temperature was controlled manually based on the experience of the operators. In Section 4.1 the savings obtained by PRESS compared to the previously used control strategy is assessed whereas the controllers ability to observe the imposed restrictions is evaluated in Section 4.2.

4.1 Savings

In order to evaluate how PRESS has influenced the operational costs two data periods are examined: The first period prior to the installation of PRESS covers nine months from January 1st 2000 to September 30th 2000 whereas the second period after installation of PRESS consists of the data from the similar months in 2001. The district heating system has not been affected by notable changes during the compared periods except for the installation of

PRESS, hence it seems reasonable to attribute any differences found in heat and electricity purchases to the introduction of PRESS.

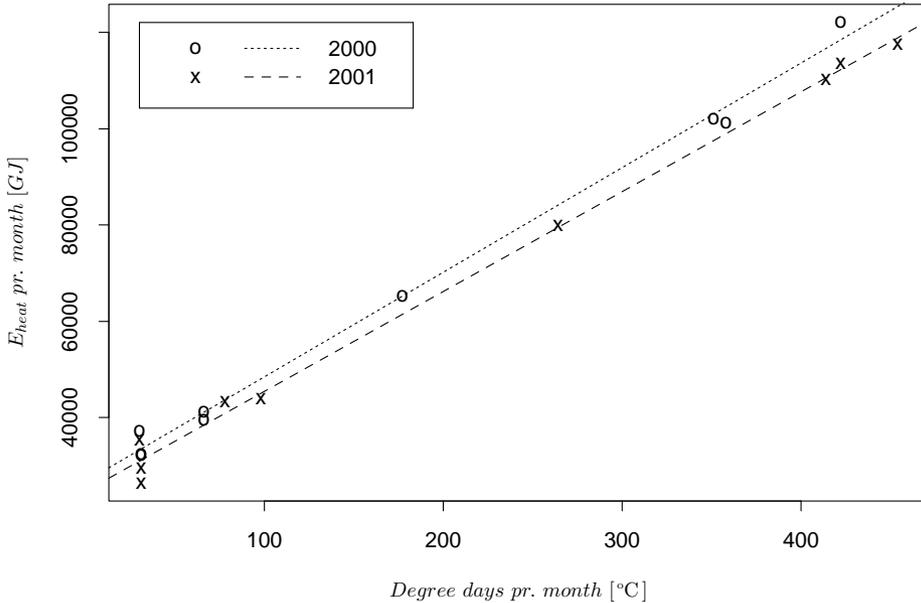


Figure 3: Heat purchase per month versus monthly degree days for the nine first months of year 2000 and 2001. The lines are the Ordinary Least Squares estimate of the relationships for the two periods.

Figure 3 shows the heat purchase per month versus the monthly degree days for the two periods. The monthly degree days T_{mon}^{dd} are calculated as

$$T_{mon}^{dd} = \sum_{\text{Days in month}} \max(0, 17 - \bar{T}_a^{diur})$$

where \bar{T}_a^{diur} is the diurnal average temperature for the days in the month.

From the figure it seems reasonable to model the relationship between heat purchase and degree days by a straight line for both periods. The Ordinary Least Squares fit of the relationship is given as

$$\begin{aligned} 2000 : \quad E_{mon}^{heat} &= 217 \frac{GJ}{^\circ C} T_{mon}^{dd} + 26700 GJ & (18) \\ 2001 : \quad E_{mon}^{heat} &= 208 \frac{GJ}{^\circ C} T_{mon}^{dd} + 24700 GJ . \end{aligned}$$

Using (18) for each of the first nine months of a normal year¹ the total difference in heat purchase before and after the installation of PRESS is calculated to -37,400 GJ corresponding to a reduction in heating costs of 1,760,000 Dkr.

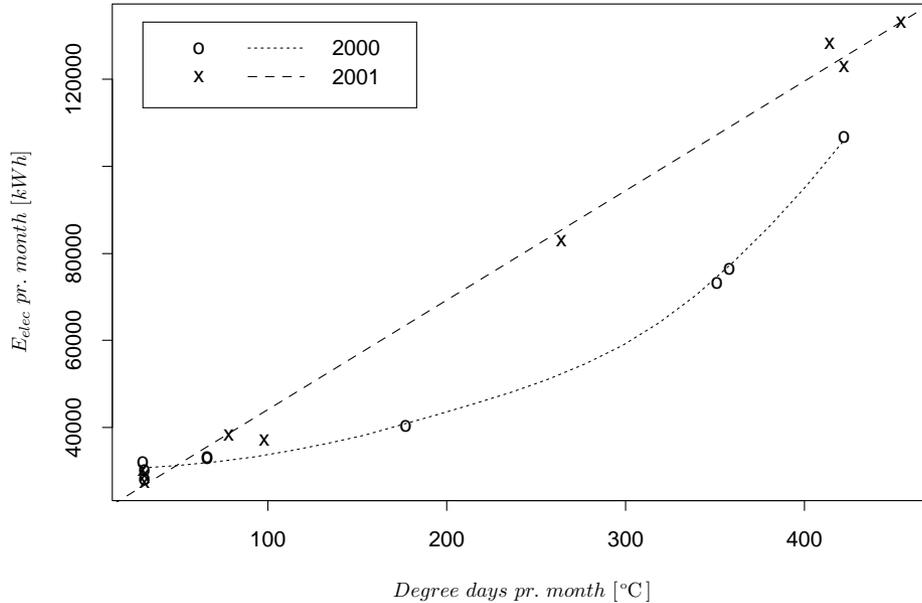


Figure 4: Electricity purchase per month versus monthly degree days for the nine first months of year 2000 and 2001. For year 2000 the relationship between degree days and electricity purchase is estimated using local polynomial regression whereas Ordinary Least Squares regression is used for year 2001.

Figure 4 shows the electricity purchase per month versus the monthly degree days for the two periods. For year 2001 the relationship between electricity purchase and degree days is modelled by a straight line, but for year 2000 this is not a reasonably model. Instead a non-parametric line is estimated using local polynomial regression with a second order approximation and nearest neighbor bandwidth of 100%. The relationships are given as

$$\begin{aligned}
 2000 : \quad E_{mon}^{elec} &= \hat{f}(T_{mon}^{dd}) & (19) \\
 2001 : \quad E_{mon}^{elec} &= 251 \frac{kWh}{^{\circ}C} T_{mon}^{dd} + 19000 kWh .
 \end{aligned}$$

where $\hat{f}()$ is the estimated local regression line.

Using (19) for each of the first nine months of a normal year the total difference in electricity purchase before and after the installation of PRESS is calculated

¹In the Roskilde area a normal year corresponds to 2805 degree days.

to 149,000 kWh corresponding to an increase in electricity costs of 194,000 Dkr.

For district heating utilities supplied from the VEKS transmission system excessive return temperatures and peak loads are penalized by an increase in the cost per unit energy.

According to Roskilde Varmeforsyning use of the peak load boiler has been reduced after the installation of PRESS, hence PRESS seems to have reduced the peak loads.

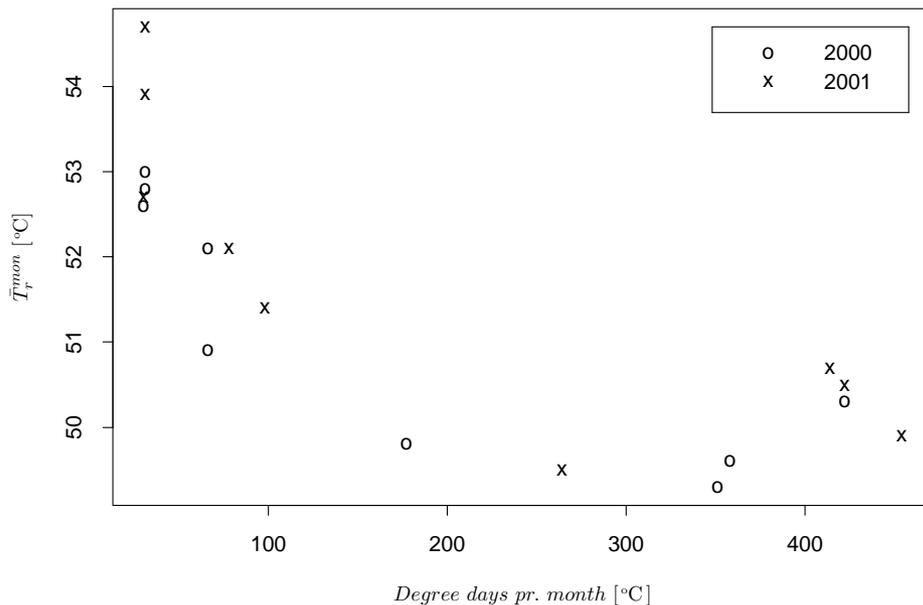


Figure 5: Monthly averages of return temperature versus monthly degree days for the nine first months of year 2000 and 2001.

From Figure 5 it is seen that the return temperature mostly seems to be unaffected by PRESS and only for the lowest observations of degree days has the use of PRESS resulted in an increase in the return temperature. The increase is too small to imply any noticeable penalty in the energy costs, but could otherwise be countered by a minor increase of the minimum value of the net-point temperature reference curve in Figure 2.

Hereafter it may be concluded that PRESS has not adversely affected the cost per unit energy. In total the installation of PRESS has resulted in a reduction of the operational costs corresponding to 1,566,000 Dkr.

4.2 Quality of control

The quality of control is evaluated with respect to the ability of the controller to observe the restrictions on maximum flow rate and minimum net-point temperatures. In Roskilde PRESS consists of a flow controller and three net-point controllers. Experience has shown that the flow sub-controller and one of the net-point temperature sub-controllers – Haraldsborg – alternate in being the active controller. During cold periods at winter time ($T_a < 0^\circ\text{C}$) the flow controller is active most of the time; during early spring and late autumn the flow controller is active only during the morning peak load and the remaining part of the year the Haraldsborg net-point controller is active.

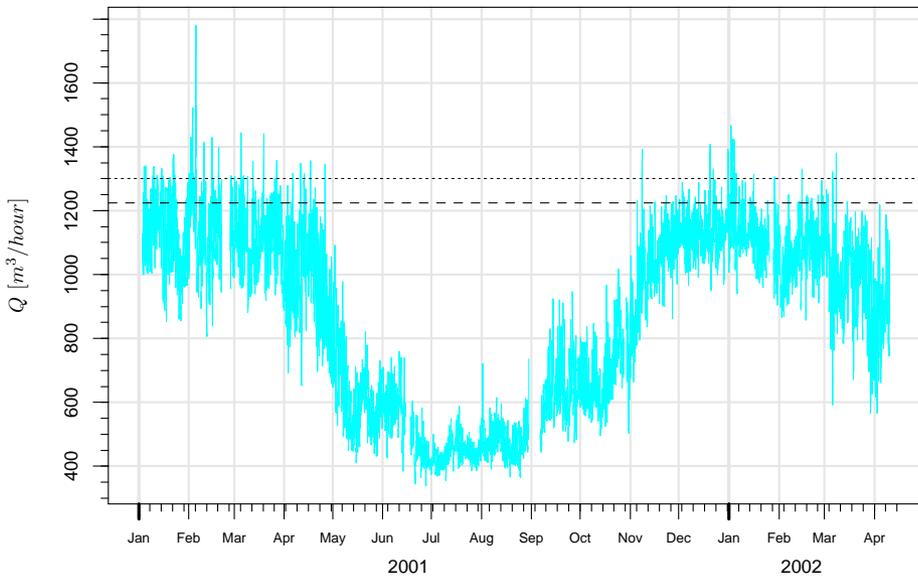


Figure 6: Hourly averages of flow rate versus time for a period from 3rd January 2001 to 10th April 2002. Unto the beginning of January 2002 the maximum flow rate of 1300 m³/hour (dotted line) was used, hereafter the maximum has been lowered to 1225 m³/hour.

Hourly average values of flow rate are plotted in Figure 6 together with the maximum flow rate(s). From the figure it is seen that PRESS in general has kept the flow rate below the maximum limit, but also that some violations of the limit occur. In Roskilde the main supply point is equipped with two pumps and the maximum flow rate observed by PRESS has been selected to avoid starting the second pump but during prolonged periods with high heat load this is not possible. This and technical problems at the supply

point explains most of the violations seen in Figure 6 and in general it is the assessment of Roskilde Varmeforsyning that PRESS is capable of observing the flow rate limit.

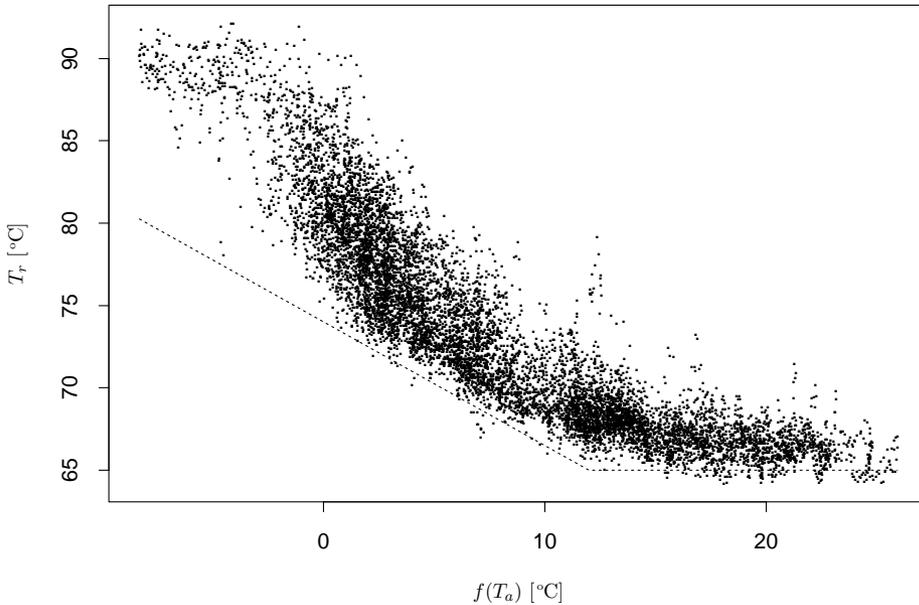


Figure 7: Hourly averages of net-point temperature at Haraldsborg versus filtered air temperature for a period from 3rd January 2001 to 10th April 2002. The dotted line is the minimum reference curve for minimum net-point temperature used at Haraldsborg.

In Figure 7 hourly average values of net-point temperature at Haraldsborg have been plotted versus filtered air temperature, where the filter is given by (15). From the plots it is readily seen that the part of the observations above the temperature requirement for the critical point (indicated by the dotted line) exceeds the specified $\pi = 95\%$ with some margin. It is therefore concluded that the PRESS controller has been capable of observing the imposed net temperature restrictions.

5 Conclusion

A new concept for controlling the supply temperature in a district heating system has been presented. The controller optimizes the operational costs

for the distribution network by minimizing the supply temperature without compromising supply or consumers temperature requirements.

The controller has been installed at the district heating utility supplying Roskilde City and suburbs and compared to the previously used control strategy savings corresponding to approximately 5% of the operational costs has been estimated for a nine months period.

It is shown that the above savings have been obtained without adversely affecting the operation of the distribution network or sacrificing security of supply.

References

- Clarke, D. W., Mohtadi, C. & Tuffs, P. S. (1987), 'Generalized predictive control – part I. The basic algorithm', *Automatica* **23**, 137–148.
- Hansen, K. K. & Bøhm, B. (1996), Nødvendig fremløbstemperatur for eksisterende fjernvarmetilslutningsanlæg - med tilhørende afkøling, Technical Report ET-ES 96-04, Department of Energy Engineering, Technical University of Denmark, Lyngby, Denmark. In Danish.
- Madsen, H., Nielsen, T. S. & Søgaaard, H. (1996), *Control of Supply Temperature*, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Nielsen, H. A. & Madsen, H. (2000), Forecasting the heat consumption in district heating systems using meteorological forecasts, Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Nielsen, T. S., Madsen, H. & Nielsen, H. A. (2001), 'Intelligent control', *Danish Board of District Heating – News from DBDH* (3), 14–16.
- Palsson, O. P., Madsen, H. & Søgaaard, H. (1994), 'Generalized predictive control for non-stationary systems', *Automatica* **30**, 1991–1997.
- Søgaaard, H. (1988), Identification and adaptive control of district heating systems, Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. In Danish.

PAPER F

A New Reference for Wind Power Forecasting

F

A New Reference for Wind Power Forecasting

Torben S. Nielsen¹, Alfred K. Joensen¹, Henrik Madsen¹, Lars Landberg²
and Gregor Giebel²

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 et al. 1994, Landberg & Watson 1994) and statistical models as in (Joensen 1997, Nielsen & Madsen 1997, Madsen et al. 1996). Up to now the reference for these models, and many other meteorological forecasting models, has been the persistence model given by

$$p_{t+k} = p_t + \varepsilon_{t+k} \quad (1)$$

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

$$\hat{p}_{t+k} = p_t \quad (2)$$

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

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

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 naïve 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 et al. 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

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

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

$$RMS_p = \sqrt{MSE_p} \quad (4)$$

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

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

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

$$\bar{p} = \frac{1}{N} \sum_{t=1}^N p_t \quad (6)$$

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}

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

where

$$\tilde{p}_t = p_t - \bar{p} \quad (8)$$

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.

Summer period

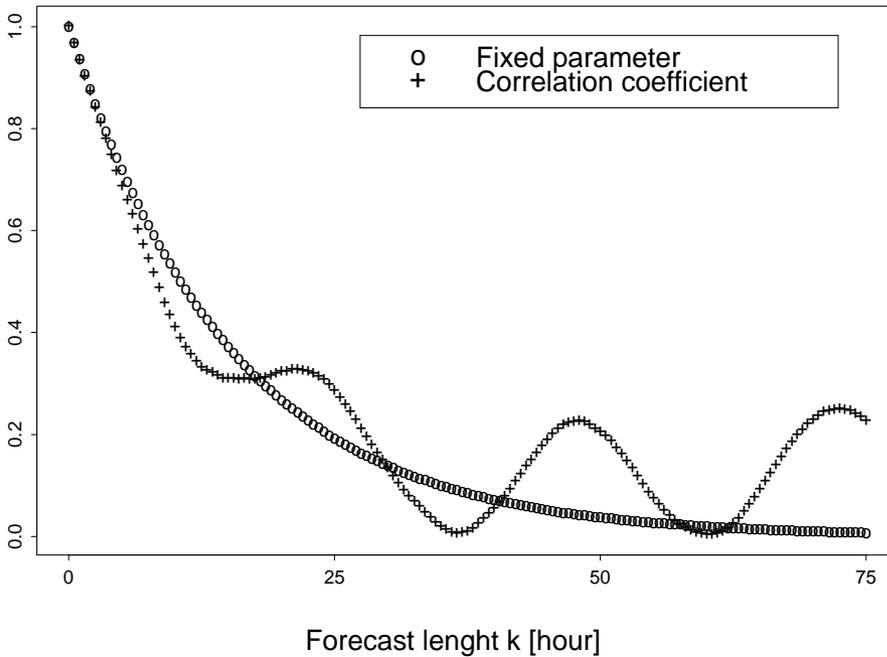


Figure 1: Estimated correlation as a function of the forecast length for 4380 half hourly mean values of observed wind power in a summer period, and the values of the fixed parameter function.

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

$$f(k) = \phi^k \quad (9)$$

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

Winter period

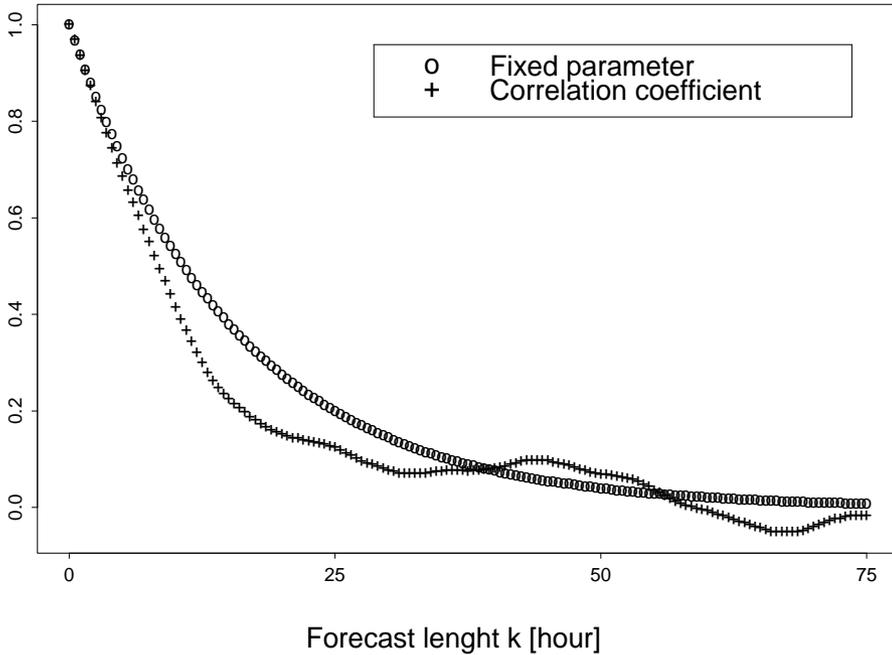


Figure 2: Estimated correlation as a function of the forecast length for 4380 half hourly mean values of observed wind power in a winter period, and the values of the fixed parameter function.

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.

Prediction performance

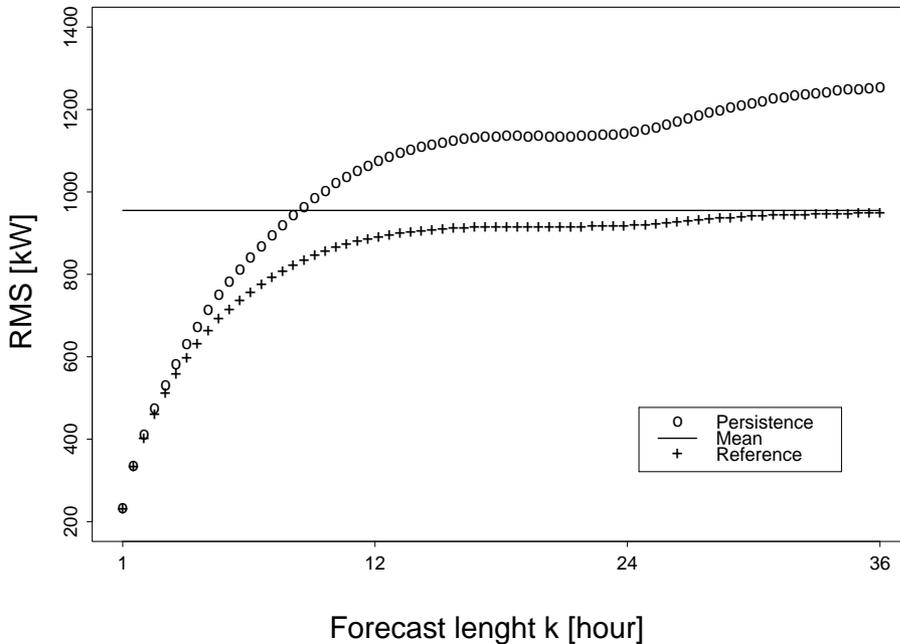


Figure 3: The *RMS* for the three simple forecast models: the mean, the persistence and the new reference. Calculated using one year of half hourly mean values of measured wind power.

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 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

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

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

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

we see that the MSE for this model is

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

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, Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. In Danish.

- Landberg, L. (1999), ‘Short-term prediction of power production from wind farms’, *Journal of Wind Engineering and Industrial Aerodynamics* **80** (1-2), 207–220.
- Landberg, L. & Watson, S. J. (1994), ‘Short-term prediction of local wind conditions’, *Boundary-Layer Meteorology* **70** (1-2), 171–195.
- Landberg, L., Watson, S. J., Halliday, J., Jørgensen, J. U. & Hilden, A. (1994), Short-term prediction of local wind conditions, Technical Report Risø –R–702(EN), Risø, Risø National Laboratory, Roskilde, Denmark.
- Madsen, H., Sejling, K., Nielsen, H. A. & Nielsen, T. S. (1996), *Models and Methods for Predicting Wind Power*, ELSAM, Fredericia, Denmark.
- Nielsen, T. S. & Madsen, H. (1997), Using meteorological forecasts in on-line predictions of wind power, Technical Report IMM-REP-1997-31, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

PAPER G

Statistical Methods for Predicting Wind Power

G

Statistical Methods for Predicting Wind Power

Torben Skov Nielsen, and Henrik Madsen

Abstract

This paper describe some statistical methods for predicting the power production for individual wind farms based on measurements and meteorological forecasts for the wind farms. The time varying dynamics of the wind means that traditional linear least square techniques are not well suited thus calling for different approaches. In this paper the problem with time varying dynamics has been addressed by using linear ARX models where the model parameters are estimated adaptively using recursive least squares techniques.

1 Introduction

During these years the world experiences a renewed interest in power production from renewable sources, and it must be expected, that power production from wind turbines will be of increasing importance in the future. In Denmark, for instance, it is foreseen that within a time span of ten to fifteen years as much as 10 % of the annual electricity consumption will be provided by wind turbines. Such level of wind energy penetration clearly calls for reliable methods for predicting the future wind power production not only for planning purposes (minimization of spinning reserves, maintenance schedules for fossil fueled power units etc.), but also the marked value of wind energy on Europe's future free energy markets will depend on the availability of such methods.

It has previously been demonstrated in ((Landberg et al. 1997), (Landberg 1999)) that physical models describing the wind farm layout and the influence of the surroundings can be used in combination with meteorological forecasts of wind speed and direction to make reliable predictions of power production with a horizon of 12 to 36 hours. It has also become clear though, that for prediction horizons less than 12 hour the suggested method performs less convincing. In ((Madsen et al. 1995) ,(Madsen et al. 1996)) it is shown how statistical models can be used in combination with online measurements of wind power and wind speed to predict power production from wind farms with a horizon of upto 12 hours ahead. From the above it seems almost self evident that an improvement of both methods can be achived by combining online measurements and meteorological forecasts in one model using statistical methods.

The approach taken in this paper is to use statistical methods to determine the optimal weight between the on-line measurements and the meteorological forecasted variables in ARX (Auto-Regressive with eXogenous input) type models. Under this approach the power production from a wind farm is described by a stochastic process, which by nature is non-linear as well as time varying (and consequently non-stationary). Models describing this kind of stochastic processes can not be handled successfully using ordinary least squares estimation techniques. In our case the non-linearities and non-stationarities are taken into account by using adaptive estimation techniques based on the recursive least squares algorithm.

The paper is organized as follows. Section 2 describes the data material used in this work. In Section 3 the estimation techniques are outlined. The developed prediction models for the reference wind farms are described in Section 4 and finally some conclusions are drawn in Section 5.

2 Data

The reported results are based on data from the Vedersø Kær wind farm at the west coast of Jutland in the ELSAM supply area (ELSAM is the power utility covering the Jutland and Fuen area in Denmark) for which the modeling have been done as a part of an EU project, “Implementing short-term wind power predictions at utilities”. The wind farm consists of 27 Vestas 225 *kW* wind turbines placed in a quadratic grid.

Data covers a full one year period starting from May 1st 1994 consisting of 5 minute averages of the power production and the wind speed as well as meteorological forecasts of the 10 *m* wind speed and wind direction.

The meteorological forecasts have been provided by the Danish Meteorological Institute (DMI) using their HIgh Resolution Limited Area Models (HIRLAM) system. The forecasts are updated every 12 hour and cover a horizon of 36 hours ahead with a 3 hour resolution (See (Landberg et al. 1994) for details).

The observations and meteorological forecasts have been subsampled, respectively interpolated, to form the 30 minute values used in the model estimation.

3 Model estimation

When modeling systems containing non-stationary or non-linear dynamics two extremes in the modeling approach can be outlined as follows: A model containing all the physical relations governing the system is drawn up. Such a model will be capable of describing the complex dynamics and should ideally have constant (and known) parameters. Alternatively a model approximating the complex dynamics by linearization around the current working point is applied. Such a model will only give a reasonable description of the system dynamics in a region around the current working point and the model parameters will be time varying as the models working point changes over time. This paper describes how the latter approach can be applied using a recursive least squares method to estimate the model parameters instead of direct linearization of a physical model.

3.1 Recursive Least Square Estimation

The estimation method used is often referred to as recursive least squares estimation with exponential forgetting ((Ljung 1987)). This method requires, that the model is linear in the parameters, i.e. it can be formulated as

$$y_t = \phi_t^T \theta + e_t \quad (1)$$

where θ is the parameter vector, ϕ_t is a vector of regressors and e_t is a independent identically distributed (*i.i.d.*) noise sequence.

The least squares method is based on minimizing a criterion of the form

$$\begin{aligned} V(\theta) &= \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_{t|t-1}(\theta))^2 \\ &= \frac{1}{N} \sum_{t=1}^N (\tilde{y}_{t|t-1}(\theta))^2 \end{aligned} \quad (2)$$

where N is the number of observations, y_t is the observation at time t and $\hat{y}_{t|t-1}(\theta)$ is the prediction of the observation at time t given observations upto time $t - 1$.

For the recursive least squares method with exponential forgetting the crite-

riterion (2) is changed to a criterion of the form

$$\begin{aligned} V(\theta_t) &= \frac{1}{N} \sum_{s=1}^t \lambda^{t-s} (y_s - \hat{y}_s(\theta_t))^2 \\ &= \frac{1}{N} \sum_{s=1}^t \lambda^{t-s} (\tilde{y}_s(\theta_t))^2 \end{aligned} \quad (3)$$

where λ is a forgetting factor ($0 < \lambda \leq 1$). It is seen that the adaptivity is obtained by multiplying the older observations with an exponentially decreasing weight function.

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

$$N_{eff} = \frac{1}{1 - \lambda} \quad (4)$$

For $\lambda = 1$ and $t = N$ it is noticed that the least squares estimate in (2) is obtained.

3.2 1-Step Predictions

Conventionally the 1-step prediction errors are used in the recursive least squares method, see e.g. (Ljung 1987). Later on in this section it is shown how the method can be extended to provide k-step predictions.

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

1. Calculation of 1-step prediction error using the estimate of θ at time $t - 1$:

$$\tilde{y}_{t|t-1} = y_t - \varphi_{1,t}^T \hat{\theta}_{t-1} \quad (5)$$

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

$$P_t = \frac{1}{\lambda} \left(P_{t-1} - \frac{P_{t-1} \varphi_{1,t} \varphi_{1,t}^T P_{t-1}}{\lambda + \varphi_{1,t}^T P_{t-1} \varphi_{1,t}} \right) \quad (6)$$

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

3. Update of the parameter estimates:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \varphi_{1,t} \tilde{y}_{t|t-1} \quad (7)$$

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

The 1-step prediction of y_{t+1} at time t is calculated as

$$\hat{y}_{t+1|t} = \varphi_{1,t+1}^T \hat{\theta}_t. \quad (8)$$

3.3 k-Step Predictions

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

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

$$\tilde{y}_{t|t-k}^{pseudo} = y_t - \varphi_{k,t}^T \hat{\theta}_{t-1}, \quad (9)$$

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

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

$$\hat{y}_{t+k|t} = \varphi_{k,t+k}^T \hat{\theta}_t. \quad (10)$$

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

4 Models for Predicting Wind Power

Previously it has been shown ((Madsen et al. 1995),(Madsen et al. 1996)) that a reasonable ARX model based solely on measured data for predicting the power production from a wind farm is given by

$$\begin{aligned}\sqrt{p_{t+k}} &= a_1\sqrt{p_t} + b_1\sqrt{w_t} + b_2w_t + m_{t+k} + e_{t+k} \\ m_t &= m + c_1 \sin\left[\frac{2\pi t}{48}\right] + c_2 \cos\left[\frac{2\pi t}{48}\right]\end{aligned}\quad (11)$$

where p_t denotes the measured power production at time t , w_t is the measured wind speed at time t , e_{t+k} is an *i.i.d.* noise sequence and m_t is a function describing both a level and the diurnal variation in the power production. Model (11) will also be referred to as the *WPPT model*.

The square root transformation of power and wind speed is motivated by the skew density of power and wind speed. In (Madsen et al. 1995) it is shown that the square root transformation leads to distributions of the prediction errors, which can be approximated by the Gaussian distribution. Note that the transfer function from $\sqrt{w_t}$ to $\sqrt{p_{t+k}}$ is formulated using a second order polynomial expression.

The model given by (11) has a good performance for a prediction horizon up to 12 hours ($k = 24$), but in (Nielsen & Madsen 1997) it is made clear that for longer prediction horizons meteorological forecasts of wind speed and direction must be taken into account. Section 4.1 and 4.2 describes two different approaches for including meteorological forecasts into model (11).

The results obtained is assessed by comparison to the well known persistence predictor

$$\hat{p}_{t+k|t} = p_t \quad (12)$$

which simply states, that what you see now is what you will get in the future.

Furthermore the results are compared with a new statistical reference proposed in (Nielsen & Madsen 1997). The reference proposed in (Nielsen & Madsen 1997) addresses the fact that the persistence predictor performs very badly for prediction horizons larger than 12 to 18 hours.

In both cases the estimated standard deviation for the prediction error (denoted S.E. from here on) is used as a reference.

4.1 Polynomial Extension to WPPT Model

As a first approach the meteorological forecasts have been included in model (11) using the same polynomial expression as used for the observed wind speeds. This leads to a model of the form

$$\begin{aligned} \sqrt{p_{t+k}} &= a_1\sqrt{p_t} + b_1\sqrt{w_t} + b_2w_t + \\ &\quad b_3\sqrt{w_{t+k|t}^{HIR}} + b_4w_{t+k|t}^{HIR} + \\ &\quad m_{t+k} + e_{t+k} \\ m_t &= m + c_1 \sin\left[\frac{2\pi t}{48}\right] + c_2 \cos\left[\frac{2\pi t}{48}\right] \end{aligned} \quad (13)$$

where $w_{t+k|t}^{HIR}$ is the forecasted wind speed at time $t+k$ given at time t .

This approach results in a simple extension, but it is a bit problematic if any dependency of wind direction exists, as both b_3 and b_4 must depend on the wind direction.

Some of the parameter estimates in the full model are not significant and especially b_1 and b_2 are found to be of little significance for most values of k . To assess the potential improvements by tailoring the models to the different prediction horizons, a model without the $b_1\sqrt{w(t)}$ term was estimated. A small reduction in S.E. is found for all values of k indicating that it is advantageous to tailor the prediction models to the individual prediction horizons. The reason probably is that the reduction in the parameter set makes the adaptive model less sensitive to noise thereby improving the prediction performance. The observed prediction performance is illustrated in Figure 1 as “ P_{pol} ”.

4.2 Power Curve Extension to WPPT Model

The model suggested in the previous section relied on a simple polynomial relationship between forecasted wind speed and power production. An alternative approach is to include the meteorological forecasts in model (11) through a power curve model.

$$\begin{aligned} G(w_t, \phi_t) &= \exp[-b \exp[-k(\phi_t) w_t]] \\ k(\phi_t) &= k_0 + \sum_{i=1}^{N_{tric}} \left[k_{i1} \sin[i\phi_t] + k_{i2} \cos[i\phi_t] \right] \end{aligned} \quad (14)$$

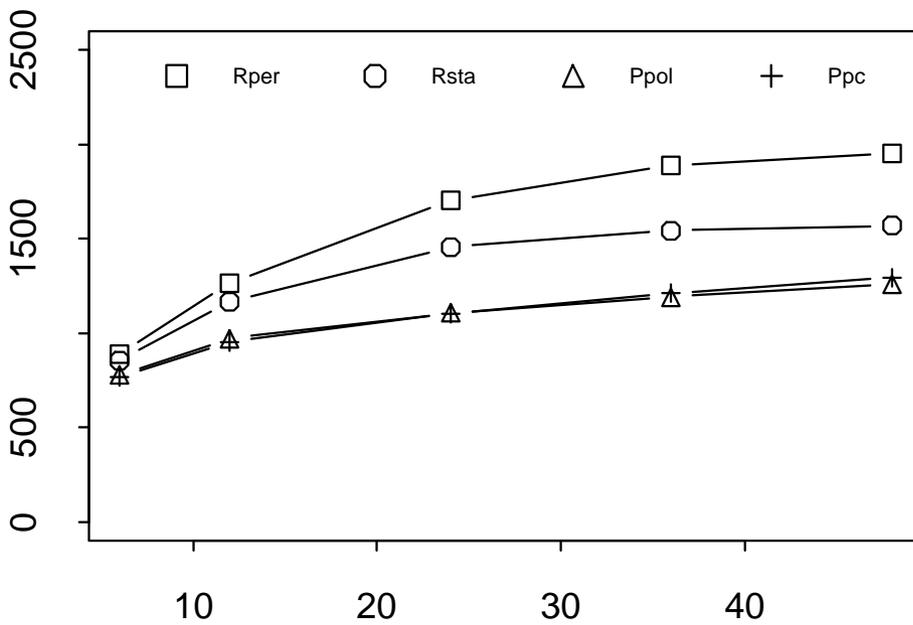


Figure 1: *S.E.* [kW] versus prediction horizon [$\frac{1}{2}$ hour] for the persistence reference (R_{per}), the statistical reference (R_{sta}), the WPPT model with polynomial extension (P_{pol}) and the WPPT model with power curve extension (P_{pc}).

where a Gompertz parameterization with a wind direction dependent power curve model is used.

Replacing the observed wind speed and direction in (14) by the forecasted wind speed and direction leads to the following power prediction model

$$\begin{aligned}\sqrt{p_{t+k}} &= a_1\sqrt{p_t} + b_1\sqrt{w_t} + b_2w_t + \\ & b_3\sqrt{G(w_{t+k|t}^{HIR}, \phi_{t+k|t}^{HIR})} + \\ & m_{t+k} + e_{t+k} \\ m_t &= m + c_1 \sin\left[\frac{2\pi t}{48}\right] + c_2 \cos\left[\frac{2\pi t}{48}\right]\end{aligned}\tag{15}$$

where $\phi_{t+k|t}^{HIR}$ is the forecasted wind direction at time $t+k$ given at time t .

Contrary to model (13) such a model is already prepared for wind direction dependencies but, as the power curve model has to be estimated separately, at the cost of a higher model complexity.

The observed performance for model (15) can be found in Figure 1 as “ P_{pc} ”.

4.3 Summary

Comparing “ P_{pol} ” to “ P_{pc} ” in Figure 1 it is seen that for $k \leq 24$ model “ P_{pc} ” is slightly superior to model “ P_{pol} ” whereas the opposite is the case for $k > 24$. It is evident though that the individual prediction horizons should be handled by different models. Selecting the best model for the individual prediction horizons the reductions achieved in S.E. by introducing meteorological forecasts into model (11) are ranging from 13.7% for $k = 6$ to 35.6% for $k = 48$ when using the persistence predictor as a reference. When comparing with the statistical reference the similar reductions in S.E. range from 10.3% for $k = 6$ to 19.8% for $k = 48$.

An analysis of the weighting of the various input variables to the model shows, that for short prediction horizons (less than 4 hours) meteorological forecasts are of little value compared to on-line measurements of power production and climatic variables. For prediction horizons larger than 4 hours the meteorological forecasts become of increasing value as the prediction horizon increases.

5 Conclusion

This paper has shown how to combine meteorological forecasts with actual measurements in models for predicting the wind power production. It is suggested to use the meteorological forecasts as input to the statistical models and let the estimation procedure put a weight on the forecasts that accounts for the actual precision of the meteorological forecasts.

Our work has made it clear, that statistical methods based on on-line measurements with advantage can be combined with meteorological forecasts when predicting power production from wind turbines. For shorter prediction horizons, i.e. less than 4 hours, the suggested methods are superior to methods based purely on meteorological forecasts, but also for longer prediction horizons very promising results have been shown.

6 Acknowledgements

This work was partially funded by the European Commission under JOULE III which is hereby gratefully acknowledged.

References

- Landberg, L. (1999), 'Short-term prediction of power production from wind farms', *Journal of Wind Engineering and Industrial Aerodynamics* **80** (1-2), 207–220.
- Landberg, L., Hansen, M. A., Vesterager, K. & Bergstrøm, W. (1997), Implementing wind forecasting at a utility, Technical Report Risø –R-929(EN), Risø, Risø National Laboratory, Roskilde, Denmark.
- Landberg, L., Watson, S. J., Halliday, J., Jørgensen, J. U. & Hilden, A. (1994), Short-term prediction of local wind conditions, Technical Report Risø –R-702(EN), Risø, Risø National Laboratory, Roskilde, Denmark.
- Ljung, L. (1987), *System Identification, Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ.
- Madsen, H., Sejling, K., Nielsen, H. A. & Nielsen, T. S. (1996), *Models and Methods for Predicting Wind Power*, ELSAM, Fredericia, Denmark.

- Madsen, H., Sejling, K., Nielsen, T. S. & Nielsen, H. A. (1995), Wind power prediction tool in control dispatch centres, Technical report, ELSAM, Fredericia, Denmark.
- Nielsen, T. S. & Madsen, H. (1997), Using meteorological forecasts in on-line predictions of wind power, Technical Report IMM-REP-1997-31, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

PAPER H

Non-parametric Statistical Methods for Wind Power Prediction

Non-parametric Statistical Methods for Wind Power Prediction

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 dependant 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 (NWP) of wind speed ω_t and direction ϕ_t from HIRLAM ((Ed.) 1988) 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.

The data from Hollandsbjerg covers one year, and it is therefore not possible

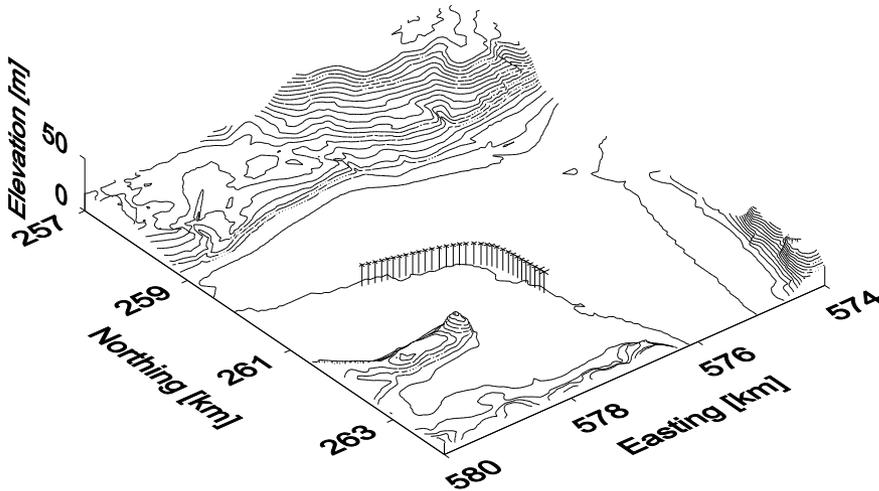


Figure 1: Layout of wind farm and the terrain surface

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, Nielsen & Madsen 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

$$y_i = f(\mathbf{x}_i) + e_i, \quad i = 1, \dots, n \quad (1)$$

where y_i are observations of a response, \mathbf{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 $P(\mathbf{x}, \mathbf{x})$ model within a sliding window, and parameterized such that

$$f(\mathbf{x}) = P(\mathbf{x}, \mathbf{x}) \quad (2)$$

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

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

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

$$\hat{f}(\mathbf{x}) = \hat{P}(\mathbf{x}, \mathbf{x}) \quad (4)$$

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.

$$w_i(\mathbf{x}) = W\left(\frac{x_{i1} - x}{h_1}\right) \dots W\left(\frac{x_{id} - x}{h_d}\right) \quad (5)$$

where the tri-cube weight function given by

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

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 parametric 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

$$p_t = g(w_t, \phi_t) + e_t \quad (7)$$

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.

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 leaves on the

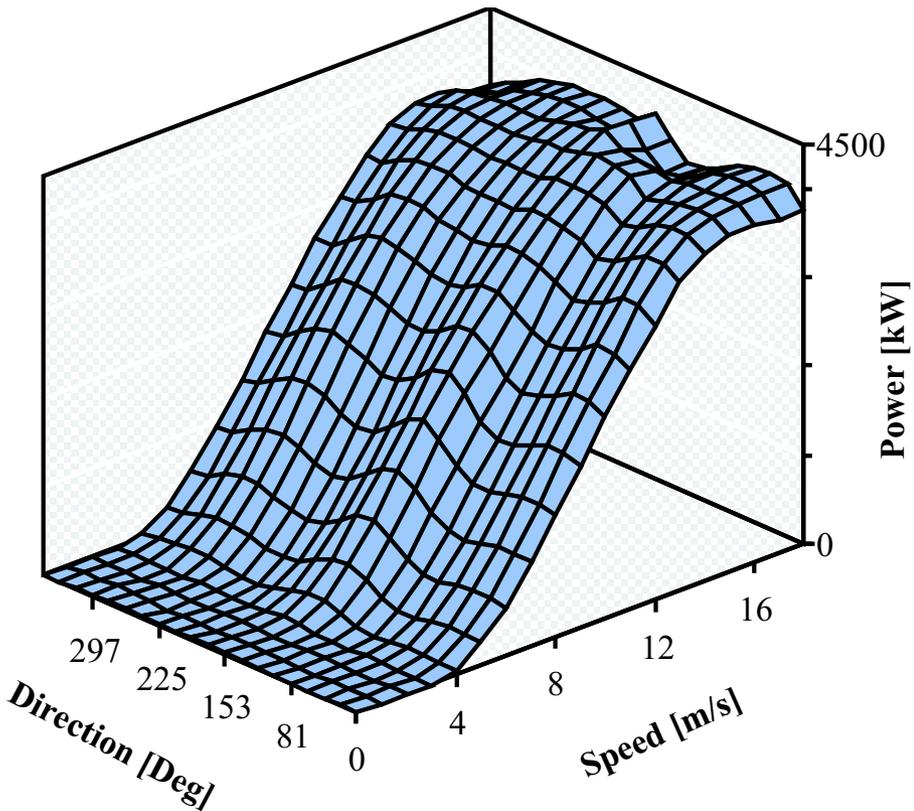


Figure 2: The power curve

trees and to wintertime where there are no leaves. 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 continuous 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.

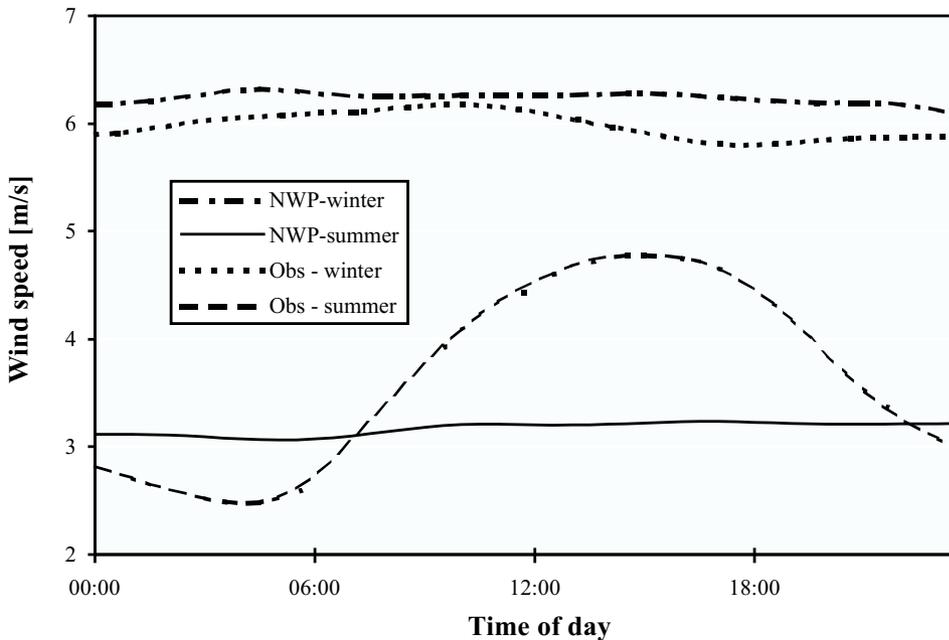


Figure 3: Diurnal variation of observed and predicted wind speed

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 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.

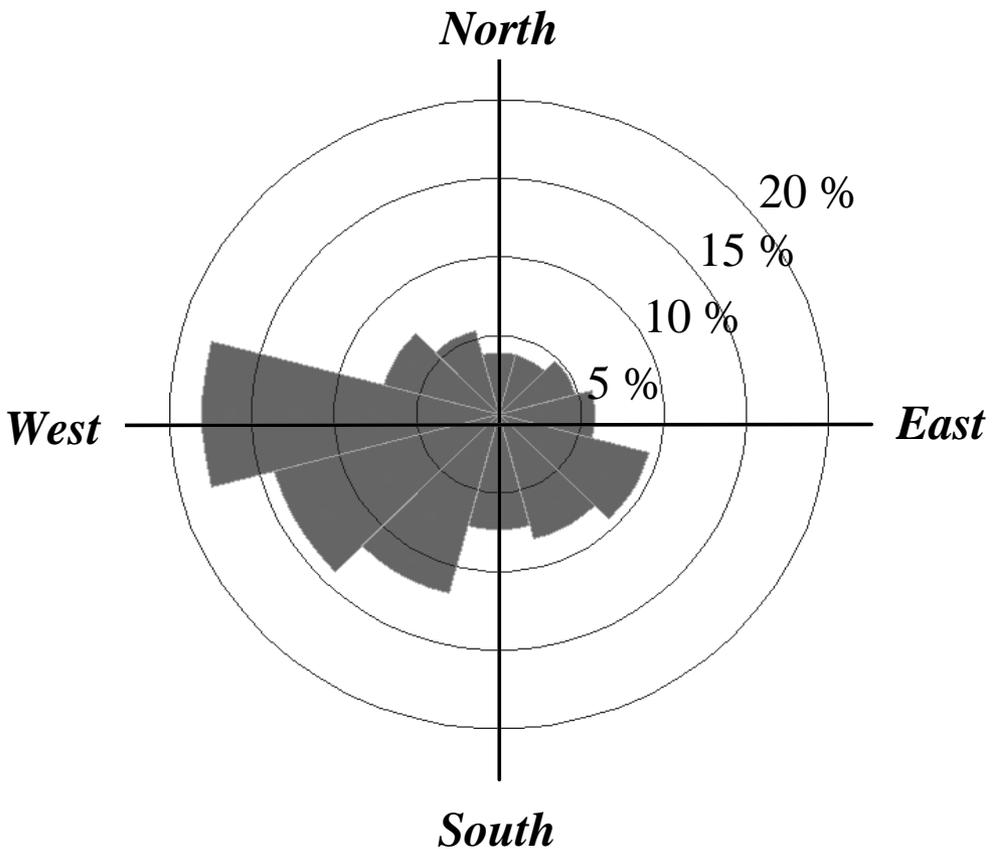


Figure 4: The distribution of the NWP wind direction

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 is therefore redundant.

3 The Model

There are several ways to formulate multi-step prediction models (Joensen 1997, Nielsen & Madsen 1997). 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

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

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 NWP's.

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 1997). 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.

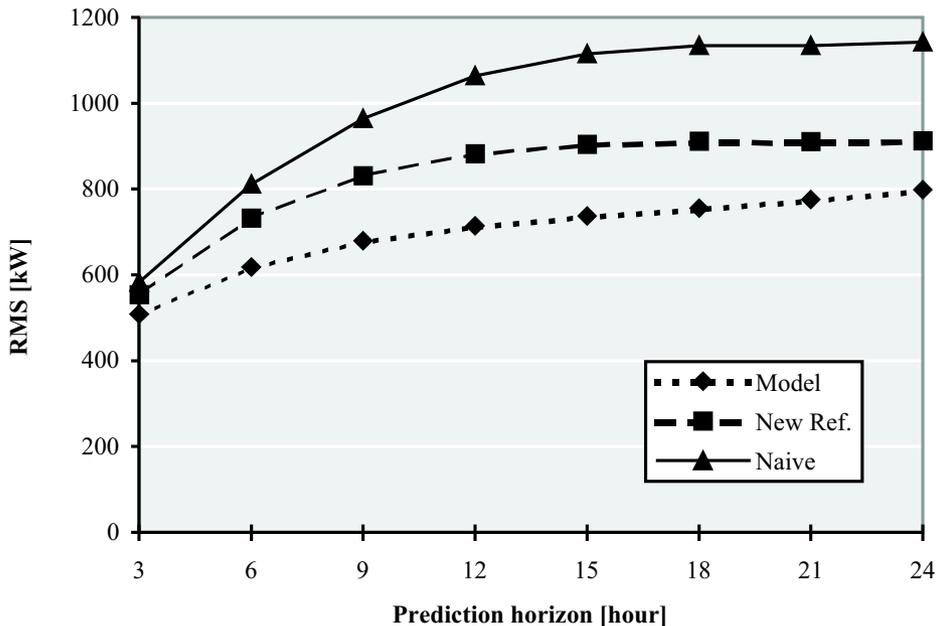


Figure 5: The RMS for different prediction horizons

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 a 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

- (Ed.), B. M. (1988), Hirlam final report, Technical report, Danish Meteorological Institute, Copenhagen, Denmark.
- 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, Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. In Danish.
- Nielsen, H. A., Nielsen, T. S. & Madsen, H. (1997), Arx-models with parameter variations estimated by local fitting, *in* 'Proceedings of 11th IFAC Symposium on System Identification', Vol. 2, Kitakyushu, Japan, pp. 475–480.
- Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L. & Giebel, G. (1999), 'A new statistical reference for predicting wind power', *Wind Energy* **1**, 29–34.
- Nielsen, T. S. & Madsen, H. (1997), Using meteorological forecasts in on-line predictions of wind power, Technical Report IMM-REP-1997-31, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

PAPER I

Experiences With Statistical Methods for Predicting Wind Power

Experiences With Statistical Methods for Predicting Wind Power

Torben Skov Nielsen, John Tøfting and Henrik Madsen

Abstract

This paper describes a tool for predicting the power production from wind turbines in an area - the Wind Power Prediction Tool (WPPT). The predictions are based on on-line measurements of power production for a selected set of reference wind farms in the area as well as numerical weather predictions covering the locations of the reference wind farms. WPPT is in operational use in the Western part of Denmark and the utilities experiences with the tool is presented.

Keywords: Wind Power Forecasts, Operating Experience, Statistics.

1 Introduction

During these years the employment of wind energy undergoes a rapid development and today the power production from wind turbines have a substantial impact on the operation for some power utilities. In the Western part of Denmark, for instance, the total capacity of the wind turbines amount to approximately 1000 MW which according to plans will be increased up to 1500 MW by year 2005.

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. Some key figures for the power production and consumption in the Western part of Denmark may be listed as:

- The total power production capacity is 6650 MW.
- By the end of 1998 the total rated wind power was 1000 MW.
- In 1997 the annual minimum and maximum load was approximately 1200 MW and 3700 MW, respectively.

It is clear, however, that in order to incorporate such a substantial wind power production efficiently into the existing production system reliable short-term (i.e. up to 48 hour) predictions of the available wind power is a necessity. This paper describes a tool, Wind Power Prediction Tool (WPPT), for providing such predictions as well as the results obtained by using WPPT in the daily planning in the Elsam/Eltra control centers.

The paper is organized as follows. The implementation of WPPT is outlined in Section 2. Section 3 describes the prediction model used in WPPT. The utility experiences using WPPT is described in Section 4 and finally some conclusions are drawn in Section 5.

2 Implementation

In WPPT 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 represented by a wind farm. Predictions of wind power with a lead time from half an hour up to 39 hours are then formed for the individual wind farms using local measurements of power production 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.

WPPT requires the following input:

- Measured power production for a number of wind farms (reference wind farms) in the area.
- Measured wind speed and direction at the location of the reference wind farms (Optional).
- Forecasts from a numerical weather model of wind speed and direction at the location of the reference wind farms. The forecasts should as a minimum cover the next 48 hours.
- Rated power and utilization time for both reference wind farms and wind turbines outside the reference wind farms.

A schematic presentation of WPPT is found in Figure 1.

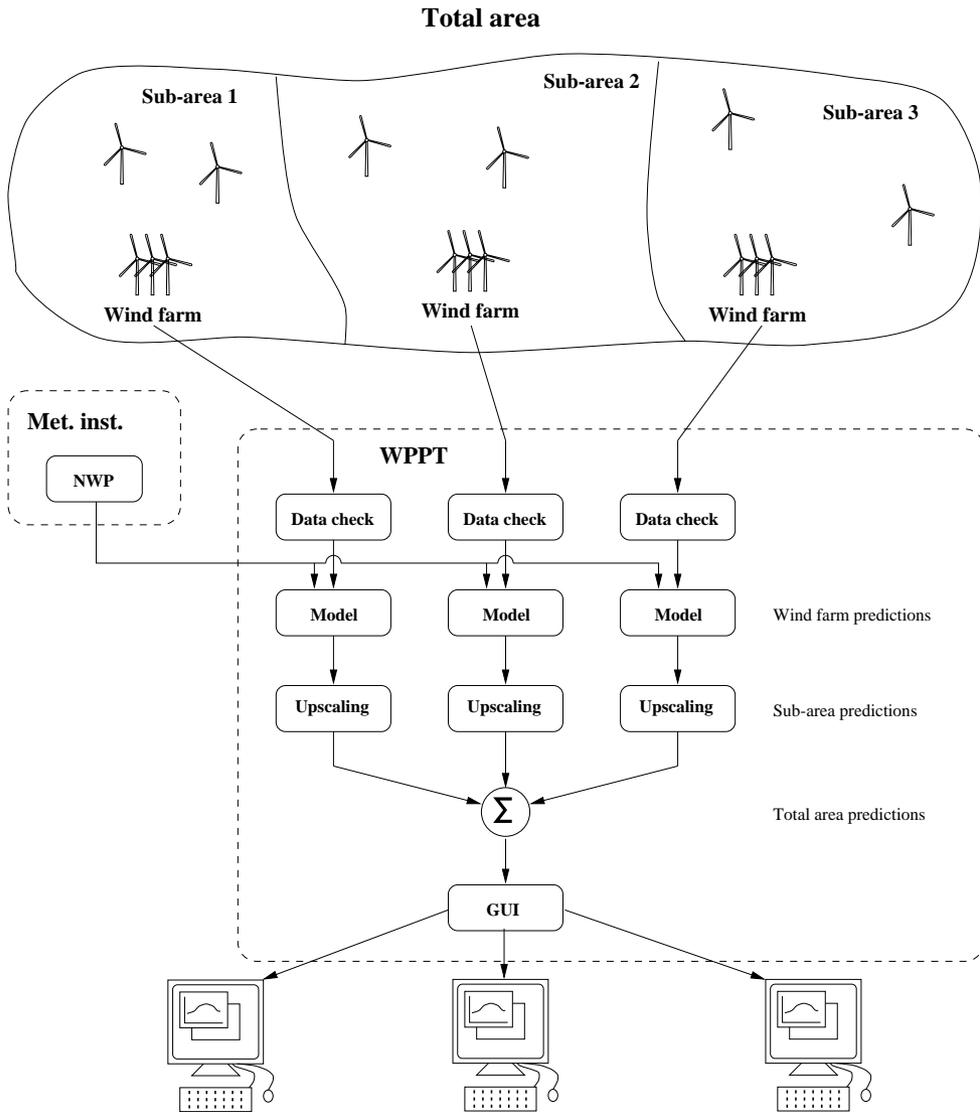


Figure 1: In WPPT the prediction of the wind power in the total area is calculated on the basis of power predictions for a number of wind farms each representing a sub-area within the total area. The predictions for the individual wind farms are calculated using local measurements of power production as well as meteorological forecasts of wind speed and direction. The wind farm power predictions for each sub-area are up-scaled to cover all wind turbines in the sub-area and then summarized to form a prediction for the total area.

WPPT is implemented as two fairly independent parts, a numeric part (WPPT-N) and a presentation part or graphical user interface (WPPT-P/GUI). A file based interface between the two sub-systems as well as between the SCADA system and WPPT has been chosen for portability reasons as it is foreseen

that WPPT will have to run on wide range of platforms. Currently WPPT have been successfully tested on HP Unix systems, Digital VMS systems as well as PC systems running Linux. WPPT is expected to run on any platform which meets the following requirements:

- X-Windows version X11R5 or better and Motif version 1.2 is available.
- ANSI compliant C and C++ compilers are available.

The numeric part of WPPT is meant to be running continuously whereas any number of GUI's can be running at a given time, i.e. WPPT is not restricted to be used by only one user at a time. In the following two sections WPPT-N and WPPT-P is presented in more detail.

2.1 WPPT-N

WPPT-N can by and large be considered to consist of four major modules: data input (measurements and meteorological forecasts), data check, model estimation and prediction and finally the up-scaling module. The measurements are given as 5 minute average values and the data validation is carried out on the 5 minute values before these are subsampled to the 30 minute values used by the models. The meteorological forecasts are given as hourly values which are interpolated to form 30 minute values before being used in the models. A brief description of the functionality within each module is given in the following:

- *Data input.* The data interface for exchanging measurements and meteorological forecasts between the local SCADA system and WPPT is established via a set of plain ASCII files. The measurements are updated every 5 minute and the meteorological forecasts are updated 4 times a day.
- *Data check.* Experience have shown that despite large efforts on-line measurements are prone to failures (errors). It is therefore essential to have some sort of automatic error classification of the measurements in order to protect the models against the influence of erroneous measurements.
 - Range check. The measurements are checked versus predetermined minimum and maximum values.

- Stationarity check. The measurements are checked for stationarity, i.e. are hung on a constant value. Measurements of wind speed and power are allowed to become stationary around 0 for longer periods of time but otherwise stationary measurements are discarded as erroneous.
- Confidence check. Here the output models describing the relationship between related measurements, e.g. wind speed and power production, are compared with the actual measurements. If a measurement falls outside some predefined confidence bands provided by the model, it is classified as erroneous.

Only the measurements are subject to the data validation methods described above and the validation of the meteorological forecasts is left with quality control of the national weather service.

- *Model estimation and prediction.* Every 30 minutes a new prediction with a lead time from half an hour up to 39 hours is calculated for the power production of each wind farm. During periods where model input is marked as erroneous the model estimation is inhibited in order to protect the model from the influence of bad data and the predictions for the actual park are marked as being unavailable.
- *Upscaling.* Both power production measurements and forecasts for the selected wind farms are up-scaled and summarized so as to calculate an estimate for the power production in the total area. For each reference wind farm a number of substitution wind farms have been defined and in case the values for a wind farm becomes marked as unavailable, the wind farm in question is replaced by one of its predefined substitutes in the up-scaling.

2.2 WPPT-P

The GUI in WPPT has to serve several purposes:

1. Display the 39 hour forecast of the total wind production in the total area.
2. Provide an overview over the climatical conditions and the power production throughout the total area.
3. Provide an overview of the current status for the measurement equipment installed in the reference wind farms.

4. Display detailed information for each reference wind farm for diagnostic purposes, e.g. if an forecast seems to be unrealistic the detailed plot for the wind farms can be used to determine the reason.
5. Act as an interface to the upscaling algorithm.

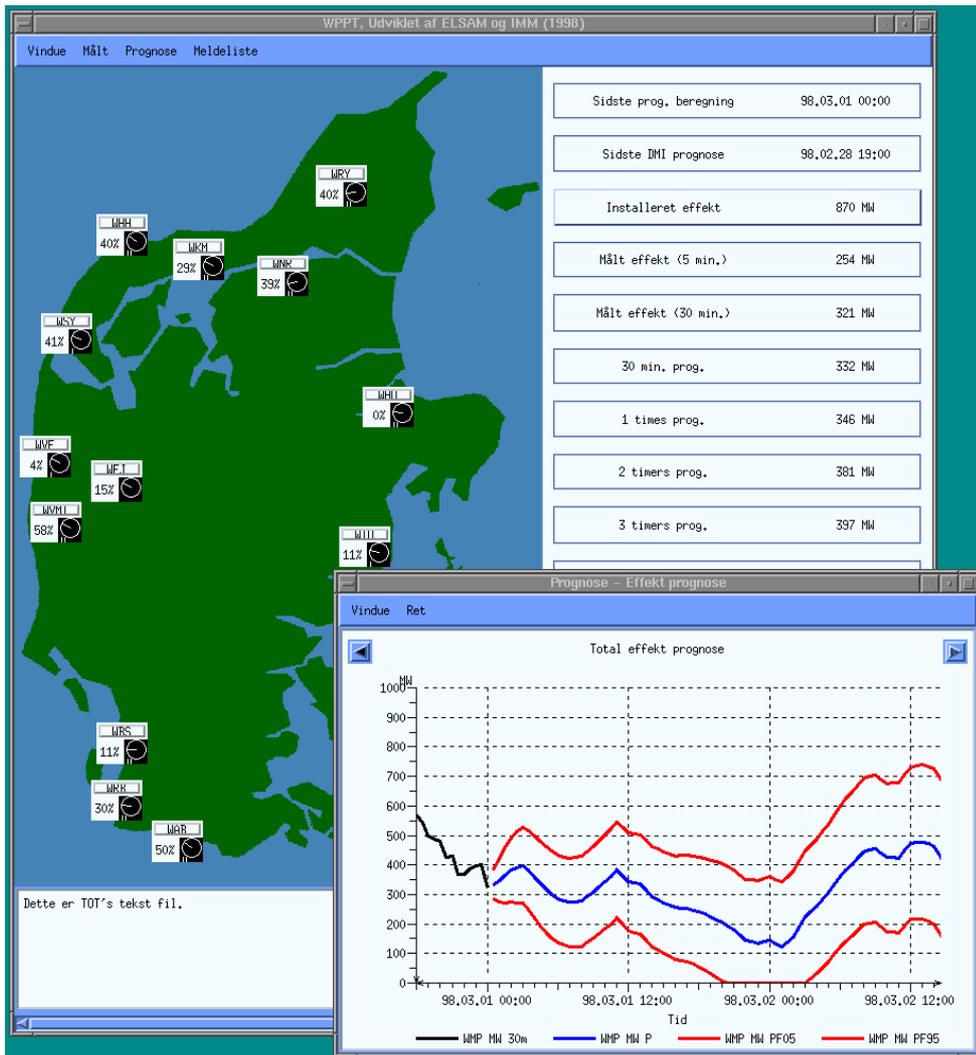


Figure 2: The need for providing both an overview as well as detailed information is reflected in the design of GUI in WPPT. The main window together with a number of plots directly accessible from the menu bar on the main window provides the operators with an overview of the system state whereas the system engineer has access to more detailed information through a set of sub-windows dedicated to the individual wind farms.

2.3 Elsam/Eltra installation

In the Elsam/Eltra installation of WPPT the wind power predictions are calculated on basis of input from the following sources:

- 14 wind farms in the Western part of Denmark, which have been selected as reference wind farms. The total rated power for the reference wind farms is 117 MW.
- Meteorological forecasts provided by the Danish Meteorological Institute using their fine resolution HIRLAM model. The forecasts are updated 4 times a day and each update covers the next 48 hours in steps of one hour.

3 The prediction model

In WPPT statistical methods are used to determine the optimal weight between the on-line measurements and the meteorological forecasted variables in ARX (Auto-Regressive with eXogenous input) type models. Under this approach the power production from a wind farm is described by a stochastic process, which by nature is non-linear as well as time varying (and consequently non-stationary). Models describing this kind of stochastic processes can not be handled successfully using ordinary least squares estimation techniques. In our case the non-linearities and non-stationarities are taken into account by using adaptive estimation techniques based on the Recursive Least Squares (RLS) algorithm. This furthermore makes the models self calibrating and allows WPPT to compensate for slow changes in the system.

Each wind farm has a set of models covering the prediction horizon (30 minutes up to 39 hours) in steps of 30 minutes. Each model is a linear k-step prediction model.

The prediction model used can be expressed as

$$\begin{aligned}
 \sqrt{p_{t+k}} &= a_t^1 \sqrt{p_t} + b_t^1 w_t + \\
 &\quad b_t^2 \sqrt{w_{t+k|t}^{HIR}} + b_t^3 w_{t+k|t}^{HIR} + \\
 &\quad m_{t+k} + e_{t+k} \\
 m_t &= c_t^0 + c_t^1 \sin\left[\frac{2\pi t}{48}\right] + c_t^2 \cos\left[\frac{2\pi t}{48}\right]
 \end{aligned} \tag{1}$$

where p_t denotes the measured power production at time t , w_t is the measured wind speed at time t , $w_{t+k|t}^{HIR}$ is the forecasted wind speed at time $t+k$ given at time t , e_{t+k} is noise sequence, m_t is a function describing both a level and the diurnal variation in the power production and finally $a_t^1, b_t^1, b_t^2, b_t^3, c_t^0, c_t^1, c_t^2$ are the time-varying model parameters to be estimated.

Further details regarding model and estimation methods are found in (Nielsen & Madsen 1997) and (Nielsen et al. 1999).

4 Utility experiences

WPPT was installed in the control centers 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 three cases have been analysed in order to illustrate how the predictions are used and with which consequences.

4.1 Case 1 (October 17th, 1998)

On October 17th 1998 the wind power production was characterized by large fluctuations (See Figure 3). The deviation between the actual production and the prediction given at 10.30 AM the day before was too large to be covered by the running reserve in the period from 5 PM to 7:30 PM and the missing power would have had to be purchased from NordPool. The prediction given at 4.30 PM the day before 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.

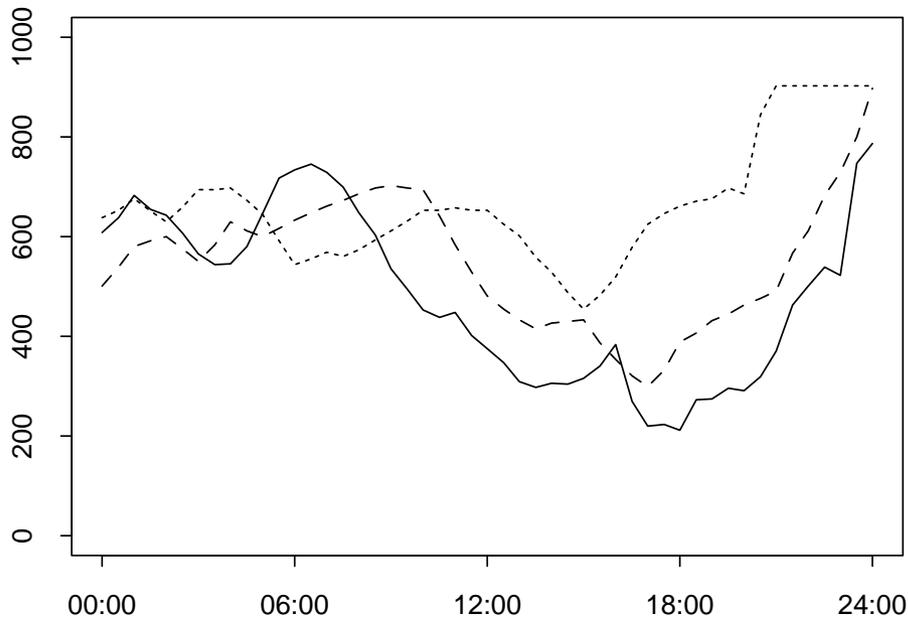


Figure 3: Measured (up-scaled) wind-based electricity production October 17, 1998 (full line) compared with the WPPT predictions from 10:30 AM (dotted line) and 4:30 PM (dashed line) the day before.

4.2 Case 2 (October 24, 1998)

On October 24th the wind power production stayed around 100 MW during the first part of the day. From noon the wind power began to increase and at midnight maximum production was reached (See Figure 4). The prediction from WPPT given at 10.30 AM the day before had indicated a less changeable course of the wind production. At 8 PM the deviation was approximately 250 MW which was covered by the running reserve and by changing the heat to power ratio on some of the CHP units. Consequently the deviation was handled without considerable costs. The situation was worsened by the fact that the operator had written up the WPPT prediction by 100 MW at the time.

At midnight the production was approximately 250 MW higher than predicted, which could be handled the normal means of regulation without any additional costs compared to a perfect forecast.

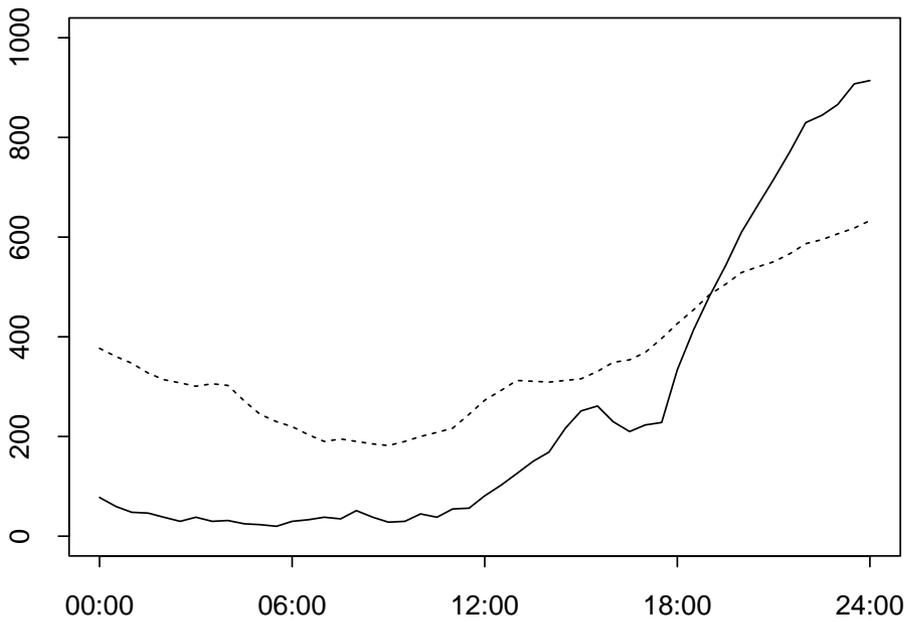


Figure 4: Measured (up-scaled) wind-based electricity production October 24th, 1998 (full line) compared with the WPPT prediction from 10:30 AM the day before (dotted line).

4.3 Case 3 (November 9, 1998)

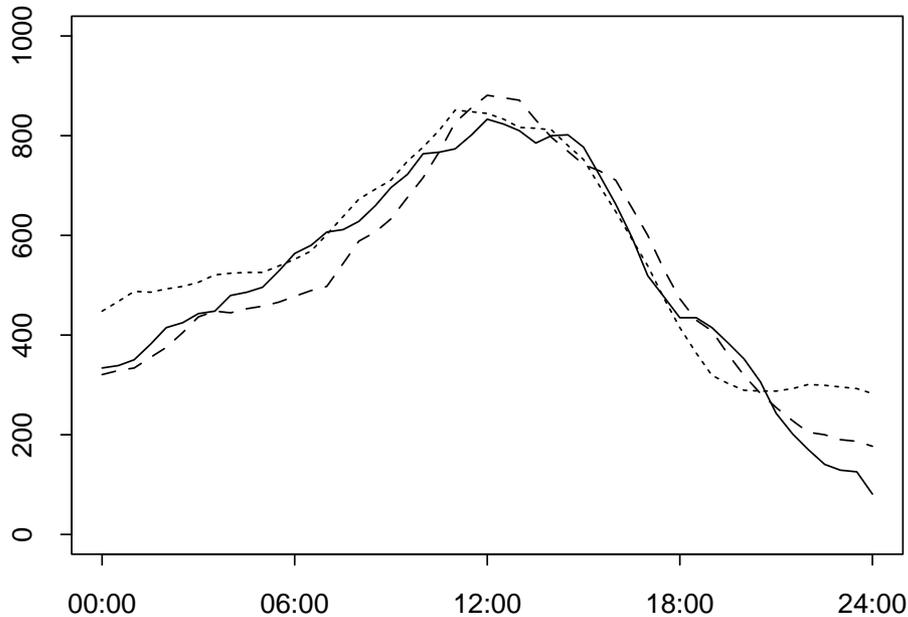


Figure 5: Measured (up-scaled) wind-based electricity production November 9, 1998 (full line) compared with the WPPT predictions from 9:30 AM (dotted line) and 9:30 PM (dashed line) the day before.

On November 9th the wind power production was marked by a large peak around noon (See Figure 5). This course of the wind power production was accurately predicted the day before and consequently the full economical value could be gained.

5 Conclusion

As indicated by the examples above the operators rely on the wind power production from WPPT in the daily planning. 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.

6 Acknowledgements

WPPT has been developed as a part of two EU projects - "Wind Power Prediction Tool in Central Dispatch Centres" founded under JOULE II, contract JOU2-CT92-0083, and "Implementing Short-term Wind Power Predictions at Utilities", founded under JOULE III, contract JOR3-CT95-0008. The results obtained and experiences gained during the two projects are described in (Madsen et al. 1995) and (Landberg, Joensen, Giebel, Watson, Madsen, Nielsen, Laursen, Jørgensen, Dalas, Tøfting, Ravn, MacCarty, Davis & Chapman 1999), respectively.

References

- Landberg, L., Joensen, A., Giebel, G., Watson, S. J., Madsen, H., Nielsen, T. S., Laursen, L., Jørgensen, J. U., Dalas, D. P., Tøfting, J., Ravn, H., MacCarty, E., Davis, E. & Chapman, J. (1999), Implementing short-term prediction at utilities, Technical report, Risø, Risø National Laboratory, Roskilde, Denmark. Confidential.
- Madsen, H., Sejling, K., Nielsen, T. S. & Nielsen, H. A. (1995), Wind power prediction tool in control dispatch centres, Technical report, ELSAM, Fredericia, Denmark.
- Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L. & Giebel, G. (1999), 'A new statistical reference for predicting wind power', *Wind Energy* **1**, 29–34.
- Nielsen, T. S. & Madsen, H. (1997), Using meteorological forecasts in on-line predictions of wind power, Technical Report IMM-REP-1997-31, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.

PAPER J

Prediction of wind power using time-varying coefficient-functions

Prediction of wind power using time-varying coefficient-functions

Torben Skov Nielsen, Henrik Madsen, Henrik Aalborg Nielsen

Abstract

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

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

1 Introduction

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

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

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

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

The proposed models are implemented in an on-line application for wind power prediction - the Wind Power Prediction Tool (WPPT) - which is used operationally by several of the Danish electrical utilities. WPPT has previously used more traditional (linear) parametric models for power prediction but it will be shown that conditional parametric models implies a significant improvement of the prediction performance compared to more traditional parametric models. The two models for short-term prediction are outlined and the performances are compared for six different wind farms - five in Denmark and one from the Zaragoza region in Spain (La Muela). The wind farm at La Muela is investigated further in (Marti et al. 2001), where the performance of the new power prediction model is evaluated for various wind forecasts.

2 Model and estimation method

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

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

where the response y_s is a stochastic variable, \mathbf{u}_s and \mathbf{x}_s are explanatory variables, e_s is i.i.d. $N(0, \sigma^2)$, $\boldsymbol{\theta}(\cdot)$ is a vector of unknown but smooth functions with values, and $s = 1, \dots, N$ are observation numbers.

Estimation in the model (1) aims at estimating the functions $\boldsymbol{\theta}(\cdot)$ within the space spanned by the observations of \mathbf{u}_s ; $s = 1, \dots, N$. The functions are only estimated for distinct values of the argument \mathbf{u} . Below \mathbf{u} denotes one single of these fitting points and $\hat{\boldsymbol{\theta}}(\mathbf{u})$ denotes the estimates of the coefficient-functions, when the functions are evaluated at \mathbf{u} .

One solution to the estimation problem is to replace $\boldsymbol{\theta}(\mathbf{u}_s)$ in (1) with a constant vector $\boldsymbol{\theta}_u$ and fit the resulting model locally to \mathbf{u} , using weighted least squares

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

Below two similar methods of allocating weights to the observations are described. For both methods the weight function $W : \mathbf{R}_0 \rightarrow \mathbf{R}_0$ is a nowhere increasing function. In this paper the tri-cube weight function

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

is used. Hence, $W : \mathbf{R}_0 \rightarrow [0, 1]$

In the case of a spherical kernel the weight on observation s is determined by the Euclidean distance $\|\mathbf{u}_s - \mathbf{u}\|$ between \mathbf{u}_s and \mathbf{u} , i.e.

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

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

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

where the multiplication is over the dimensions of \mathbf{u} . The scalar $h(\mathbf{u}) > 0$ is called the bandwidth. If $h(\mathbf{u})$ is constant for all values of \mathbf{u} it is denoted a fixed bandwidth. If $h(\mathbf{u})$ is chosen so that a certain fraction (α) of the observations fulfill $\|\mathbf{u}_s - \mathbf{u}\| \leq h(\mathbf{u})$ it is denoted a nearest neighbor bandwidth. If \mathbf{u} has the dimension two or larger, scaling of the individual elements of \mathbf{u}_s before applying the method should be considered, see e.g. (Cleveland & Devlin 1988). Rotating the coordinate system in which \mathbf{u}_s is measured may also be relevant. In this study the models have been estimated using a product kernel with a fixed bandwidth.

If the bandwidth $h(\mathbf{u})$ is sufficiently small the approximation of $\boldsymbol{\theta}(\cdot)$ as a constant vector near \mathbf{u} is good. This implies that a relatively low number of observations is used to estimate $\boldsymbol{\theta}(\mathbf{u})$, resulting in a noisy estimate or large bias if the bandwidth is increased. See also the comments on kernel estimates in (Cleveland & Devlin 1988).

It is, however, well known that locally to \mathbf{u} the elements of $\boldsymbol{\theta}(\cdot)$ may be approximated by polynomials, and in many cases these will be good approximations for larger bandwidths than those corresponding to local constants. Let us describe how local polynomial approximations are used in a local least squares setting. Let $\theta_j(\cdot)$ be the j 'th element of $\boldsymbol{\theta}(\cdot)$ and let $\mathbf{p}_d(\mathbf{u})$ be a column vector of terms in a d -order polynomial evaluated at \mathbf{u} , if for instance $\mathbf{u} = [u_1 \ u_2]^T$ then $\mathbf{p}_2(\mathbf{u}) = [1 \ u_1 \ u_2 \ u_1^2 \ u_1 u_2 \ u_2^2]^T$. Furthermore, let $\mathbf{x}_s = [x_{1s} \ \dots \ x_{ps}]^T$. With

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

and

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

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

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

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

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

This method is identical to the method described in (Cleveland & Devlin 1988) when $\mathbf{x}_j = 1$ for all j with the exception that in (Cleveland & Devlin 1988) the elements of \mathbf{u}_s used in $\mathbf{p}_d(\mathbf{u}_s)$ are centered around \mathbf{u} and hence $\mathbf{p}_d(\mathbf{u}_s)$ must be recalculated for each value of \mathbf{u} considered.

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

3 Adaptive estimation

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

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

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

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

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

Following (Nielsen, Nielsen, Joensen, Madsen & Holst 2000) the solution to

(11) can be found recursively as

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

where

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

It is observed that existing numerical procedures for recursive least squares estimation can be applied by replacing \mathbf{z}_t and y_t with $\mathbf{z}_t \sqrt{w_u(\mathbf{u}_t)}$ and $y_t \sqrt{w_u(\mathbf{u}_t)}$, respectively.

When \mathbf{u}_t is far from \mathbf{u} it is clear from (13) that $\mathbf{R}_{u,t} \approx \lambda \mathbf{R}_{u,t-1}$. This may result in abruptly changing estimates if \mathbf{u} is not visited regularly. This is considered a serious practical problem and consequently (13) has to be modified to ensure that the past is weighted down only when new information become available, i.e.

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

where $v(\cdot; \lambda)$ is a nowhere increasing function on $[0; 1]$ fulfilling $v(0; \lambda) = 1/\lambda$ and $v(1; \lambda) = 1$. Note that this requires that the weights span the interval ranging from zero to one. Here only the linear function $v(w; \lambda) = 1/\lambda - (1/\lambda - 1)w$ is considered. Thus (14) becomes

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

It is reasonable to denote

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

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

3.1 Summary of the method

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

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

- Construct the explanatory variables corresponding to local constant estimates using (6):

$$\mathbf{z}_t^T = [x_{1,t} \mathbf{P}_{d(1)}^T(\mathbf{u}_t) \dots x_{p,t} \mathbf{P}_{d(p)}^T(\mathbf{u}_t)].$$
- Calculate the weight using e.g. (4) and (3):

$$w_{u^{(i)}}(\mathbf{u}_t) = (1 - (\|\mathbf{u}_t - \mathbf{u}^{(i)}\|/h^{(i)})^3)^3, \text{ if } \|\mathbf{u}_t - \mathbf{u}^{(i)}\| < h^{(i)} \text{ and zero otherwise.}$$
- Find the effective forgetting factor using (16):

$$\lambda_{eff}^{(i)}(t) = 1 - (1 - \lambda)w_{u^{(i)}}(\mathbf{u}_t).$$
- Update $\mathbf{R}_{u^{(i)},t-1}$ using (15):

$$\mathbf{R}_{u^{(i)},t} = \lambda_{eff}^{(i)}(t) \mathbf{R}_{u^{(i)},t-1} + w_{u^{(i)}}(\mathbf{u}_t) \mathbf{z}_t \mathbf{z}_t^T.$$
- Update $\hat{\phi}_{t-1}(\mathbf{u}^{(i)})$ using (12):

$$\hat{\phi}_t(\mathbf{u}^{(i)}) = \hat{\phi}_{t-1}(\mathbf{u}^{(i)}) + w_{u^{(i)}}(\mathbf{u}_t) \mathbf{R}_{u^{(i)},t}^{-1} \mathbf{z}_t \left[y_t - \mathbf{z}_t^T \hat{\phi}_{t-1}(\mathbf{u}^{(i)}) \right].$$
- Calculate the updated local polynomial estimates of the coefficient-functions using (9):

$$\hat{\theta}_{j,t}(\mathbf{u}^{(i)}) = \mathbf{P}_{d(j)}^T(\mathbf{u}^{(i)}) \hat{\phi}_{j,t}(\mathbf{u}^{(i)}); \quad j = 1, \dots, p$$

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

4 Wind power prediction models

The development of the Wind Power Prediction Tool (WPPT) began in 1992 and the first test version was installed at a Danish power utility in 1995. WPPT went into operational use in 1998 and has since then been used operationally by most of the Danish power utilities. During WPPT's life time several studies have been carried out to improve the performance of the power prediction models. Much effort has been dedicated to make best possible use of the available meteorological forecasts e.g. by introducing wind direction dependency in the power curve and employing additional explanatory variables besides forecasted wind speed and wind direction. This section first gives an overview of the model used in the version of WPPT which is operational in Denmark today (WPPT version 2). Later on the new model (WPPT version 4) is outlined.

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

$$\begin{aligned}
 p_{t+k} = & a_1 p_t + a_2 p_{t-1} + b_1^m w_{t+k|t}^m + b_2^m (w_{t+k|t}^m)^2 + \\
 & \sum_{i=1}^3 [c_i^c \cos \frac{2i\pi h_{t+k}^{24}}{24} + c_i^s \sin \frac{2i\pi h_{t+k}^{24}}{24}] + \\
 & m + e_{t+k}
 \end{aligned} \tag{17}$$

where p_t is the observed power at time t , $w_{t+k|t}^m$ is the forecasted wind speed at $t+k$ given at time t , h_{t+k}^{24} is time of day at time $t+k$, e_{t+k} is a noise term, and a_1 , a_2 , b_1^m , b_2^m , $c_{1..3}^c$, $c_{1..3}^s$ and m are the time-varying model parameters which are estimated adaptively. For a specific prediction horizon k some of the terms in (17) may not be present in the model.

Predictions of the wind power with an prediction horizon from 1 hour up to 39 hours are updated every hour.

The new WPPT models (WPPT4) uses conditional parametric estimates of

wind direction dependent power curves in the transformation of forecasted wind speed and wind direction to power. The model is given as

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

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

where p_{t+k}^{pc} is the predicted power production from the power curve model, p_{t+k}^{pp} is the final power prediction where also autoregressive and diurnal effects are included, $\theta_{t+k|t}^m$ is the forecasted wind direction and f , a , b , c^c and c^s are smooth time-varying functions to be estimated as described previously.

Power curve predictions, p^{pc} , with an prediction horizon from 1 hour to 48 hours are updated every six hours whenever a new wind forecast becomes available. The final power prediction, p^{pp} , are updated every hour, but here the maximum prediction horizon depends on the calculation time of the last wind forecast received. At present the wind forecast from the Danish Meteorological Institute (DMI) is available two hours after the calculations are initiate, which means that the maximum prediction horizon for the final power prediction model varies between 46 hours and 40 hours.

5 The prediction performance

The performance of WPPT2 and WPPT4 has been compared for five wind farms in Denmark sited at Dræby, Fjaldene, Hollandsbjerg, Rejsby and Sydthy and for a wind farm in Spain sited at La Muela in the Zaragoza region.

For the five Danish wind farms the data set consists of hourly values of observed power production as well as forecasted wind speed and wind direction from the lowest model level (level 31) of the Danish HIRLAM DKV model (17km grid size) with a prediction horizon from 1 hour to 48 hours in steps of 1 hour. The data set covers almost an entire year from 1997-05-26 01:00 to 1998-05-18 00:00. In order to exclude effects of model initialization from the results only the data from 1998-01-19 00:00 and onward has been used in the model evaluation.

The Spanish data set consists of hourly values of observed power produc-

tion for five of the wind turbines from the wind farm at La Muela and forecasted values of the 10 meter wind speed and wind direction from the Spanish HIRLAM model (17km grid size) with a prediction horizon from 1 hour to 24 hours in steps of 1 hour. The data set covers the period from 2000-01-31 12:00 to 2000-08-16 18:00 and again only data from the last part of the period is used in the model evaluation – here from 2000-06-16 05:00 and onward.

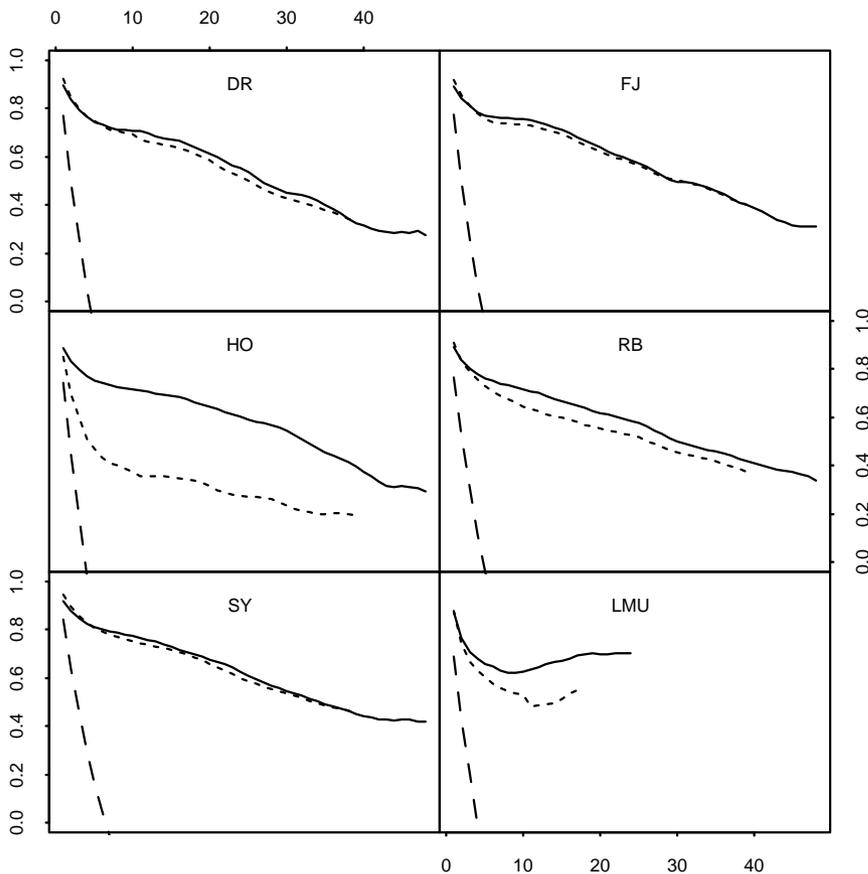


Figure 1: Degree of explanation for WPPT4 (full line), WPPT2 (dotted line) and the naive predictor (dashed line) as a function of prediction horizon [hours]. From top left to bottom right the results are for the wind farms at Dræby (DR), Fjaldene (FJ), Hollandsbjerg (HO), Rejsby (RB), Sydthy (SY) and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

Figure 1 summarizes the prediction performance obtained for the WPPT2 and WPPT4 models as well as the naive (what you see is what you get) predictor.

Degree of explanation (r^2), which describes how large a part of the variability of the observed value is explained by the prediction, is used as a performance measure. r^2 should be a number between 0 and 1 where 0 is the score obtained by the mean value predictor and 1 is the score of the perfect model, i.e. all variability of the observed value is explained by the model.

From the figure it is seen that for most of the wind farms the WPPT4 model gives a clear improvement compared to the WPPT2 model and for no wind farms does WPPT4 perform worse than WPPT2. r^2 range from approximately 0.9 for a prediction horizon of 1 hour down to 0.45 to 0.50 for a prediction horizon of 36 hours depending on the wind farm.

The Spanish wind farm at La Muela are situated in semi-complex terrain as opposed to the Danish wind farms which all are situated in rather flat terrain. Never the less the best performance of the WPPT4 model is found for La Muela. The reason for this, at first glance unexpected result, can be found in (Marti et al. 2001), which shows that it is clearly advantages to use the forecasts of the 10 meter winds as input to the WPPT4 models instead of the forecasts of the model level winds.

For the two wind farms at Hollandsbjerg and La Muela the score of the WPPT4 models is much better than the WPPT2 models. This can be explained by a very pronounced wind direction dependency in the estimated power curve for these two wind farms – see Figure 2 and 3, which only can be handled by the more advanced power curve model in WPPT4.

From Figure 1 it is seen that at La Muela the performance of the WPPT4 models gets better as the prediction horizon increases. Some of the improvement can be attributed to a slightly increasing performance of the wind forecasts as the prediction horizon increases, and some can be attributed to a strong diurnal variation in the wind speed (and power production) at La Muela. The model structure in the power prediction model is probably sub-optimal for a site with a strong diurnal variation and a model where p_t has been replaced with a weighted power prediction $p_{t+k}^w = w(k)p_t + (1 - w(k))p_{t+k-24}$ is likely to be better suited for such sites.

6 Summary

In this paper methods for adaptive and recursive estimation in a class of non-linear autoregressive models with external input are proposed. The model

HO - Estimated power curve

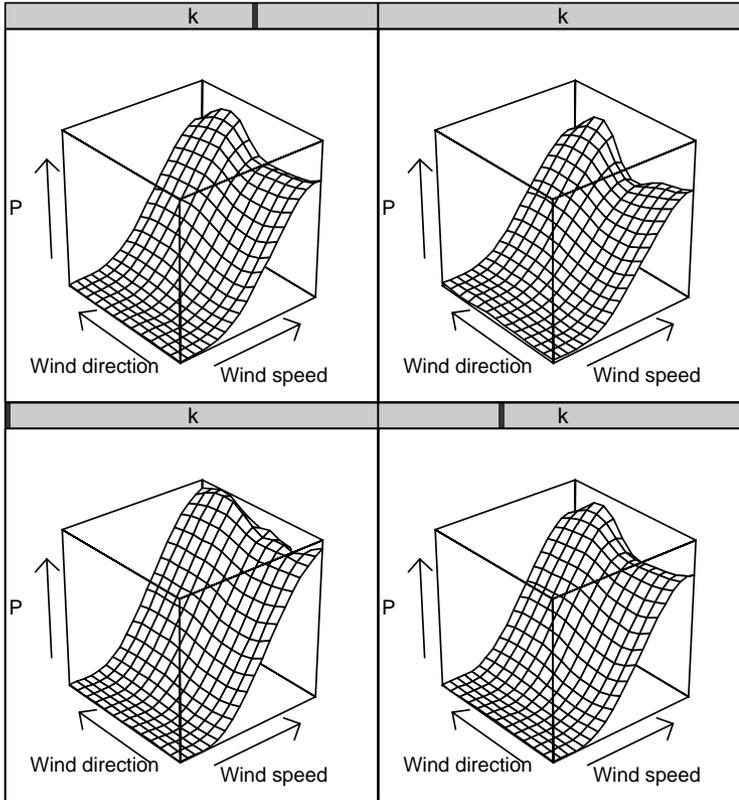


Figure 2: *The estimated power curve for Hollandsbjerg. From bottom left to top right the power curves correspond to prediction horizons of 0 hours (the analysis), 12 hours, 24 hours and 36 hours.*

class considered is conditionally parametric models, which is a conventional linear model in which the parameters are replaced by smooth, but otherwise unknown, functions of a low-dimensional input process. These functions are estimated adaptively and recursively without specifying a global parametric form.

The methods can be seen as generalizations or combinations of recursive least squares with exponential forgetting (Ljung & Söderström 1983), local polynomial regression (Cleveland & Devlin 1988), and conditional parametric fits

LMU - Estimated power curve

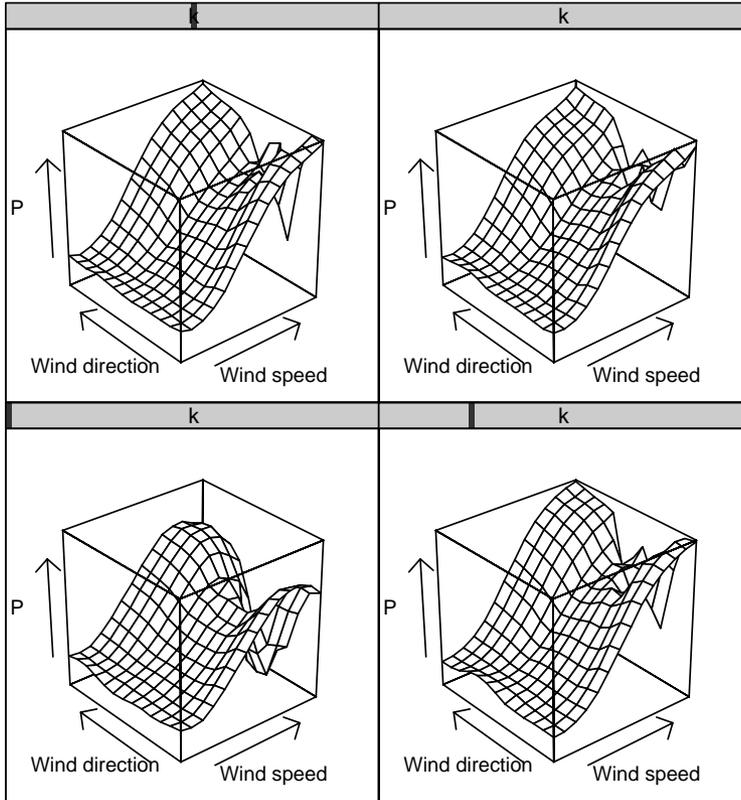


Figure 3: *The estimated power curve for La Muela. From bottom left to top right the power curves correspond to prediction horizons of 0 hours (the analysis), 6 hours, 12 hours and 24 hours.*

(Anderson, Fang & Olkin 1994). Hence, the methods constitutes an extension to the notion of local polynomial estimation.

The method is illustrated using power prediction for wind farms as an example. Both parametric and conditional parametric models are considered. The predictions based on conditional parametric models are shown to be superior to the predictions obtained by state-of-the-art parametric models. The degree of explanation varies from 0.90 for a one-hour prediction horizon to 0.45 to 0.50 for a 36 hour prediction horizon.

References

- Anderson, T. W., Fang, K. T. & Olkin, I., eds (1994), *Multivariate Analysis and Its Applications*, Institute of Mathematical Statistics, Hayward, chapter Coplots, Nonparametric Regression, and conditionally Parametric Fits, pp. 21–36.
- Chambers, J. M. & Hastie, T. J., eds (1991), *Statistical Models in S*, Wadsworth, Belmont, CA.
- Cleveland, W. S. & Devlin, S. J. (1988), ‘Locally weighted regression: An approach to regression analysis by local fitting’, *Journal of the American Statistical Association* **83**, 596–610.
- Joensen, A., Madsen, H., Nielsen, H. A. & Nielsen, T. S. (1999), ‘Tracking time-varying parameters using local regression’, *Automatica* **36**, 1199–1204.
- Ljung, L. & Söderström, T. (1983), *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA.
- Marti, I., Nielsen, T. S., Madsen, H., Navarro, J. & Barquero, C. G. (2001), Prediction models in complex terrain, in ‘Proceedings of the European Wind Energy Conference’, Copenhagen, Denmark.
- Nielsen, H. A., Nielsen, T. S., Joensen, A., Madsen, H. & Holst, J. (2000), ‘Tracking time-varying coefficient functions’, *International Journal of Adaptive Control and Signal Processing* **14**, 813–828.
- Nielsen, H. A., Nielsen, T. S. & Madsen, H. (1997), Arx-models with parameter variations estimated by local fitting, in ‘Proceedings of 11th IFAC Symposium on System Identification’, Vol. 2, Kitakyushu, Japan, pp. 475–480.
- Nielsen, T. S., Madsen, H., Nielsen, H. A. & Tøfting, J. (1999), Using meteorological forecasts in on-line predictions of wind power, Technical report, Eltra, Fredericia, Denmark.