Predictive control of supply temperature in district heating systems

Torben Skov Nielsen
Henrik Madsen
Jan Holst
Henning T. Søgaard

15th November 2002

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

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.

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

Contents

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

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 network[1] 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 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. (7.3)) 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 face 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, Mohtadi & Tuffs 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

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

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 distributer).* 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 an 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 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 2.1.

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.

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

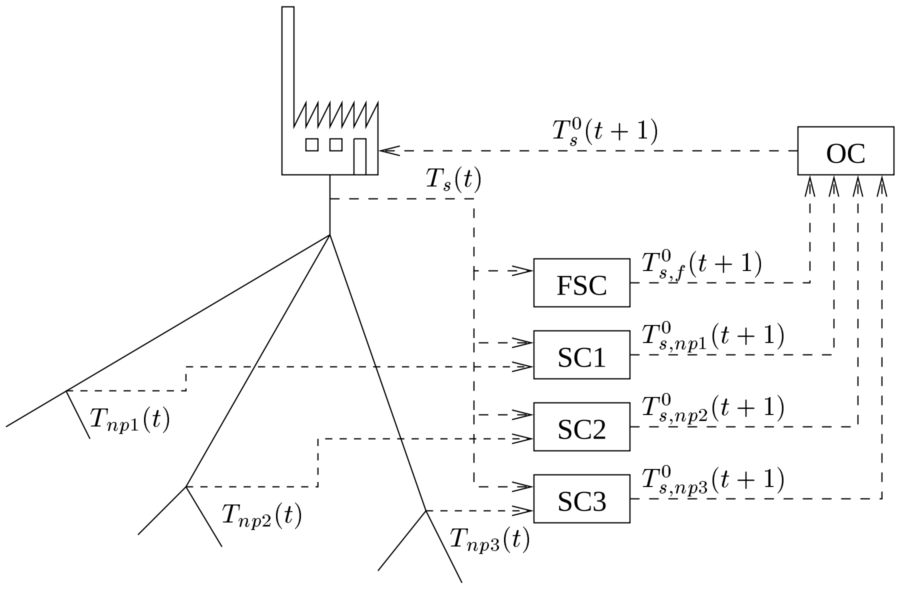

Figure 2.1: *Schematic overview of a district heating network, with three critical net-points and the corresponding controllers. OC is the overall controller, FSC is the flow sub-controller, SC# are the supply temperature sub-controllers,* $T_{np\#}$ *are the supply temperatures in the network,* T_s *is the supply temperature from the plant, and* $T^0_{s,\#}$ *are the supply temperatures required by the sub-controllers.*

specified ARIMAX model

$$A(q^{-1})\,y(t) \quad = \quad B(q^{-1})\,u(t) + \frac{C(q^{-1})\,e(t)}{\nabla} \quad , \tag{3.1}$$

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 (3.1) 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, Palsson, Sejling & Søgaard 1992)).

In (Palsson, Madsen & Søgaard 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 parameters

$$A_t(q^{-1})y(t) = B_t(q^{-1})u(t) + C_t(q^{-1})e(t) \ , \tag{3.2}$$

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 known[2] 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 (3.2) 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

$$\begin{aligned} \hat{y}(t+j|t) &= -\sum_{i=1}^{n} a_{i,t+j}\hat{y}(t+j-i|t) + \sum_{i=1}^{m} b_{i,t+j}u(t+j-i) \\ &\quad + \sum_{i=1}^{r} c_{i,t+j}\hat{e}(t+j-i|t) \ , \quad j \geq 1 \end{aligned} \tag{3.3}$$

where

$$\hat{y}(t+j|t) = y(t+j) \ , \quad j < 1 \ , \tag{3.4}$$

$$\hat{e}(s|t) = \begin{cases} e(s) = y(s) - \hat{y}(s|s-1) & \text{if} \quad s \leq t \\ 0 & \text{if} \quad s > t \end{cases} \ . \tag{3.5}$$

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 (3.3) 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 and the $\hat{e}$'s in (3.3) recursively using (3.4) and (3.5). If $v_j(t)$ denotes the sum of all terms on the right-hand side of (3.3) involving known data ($y(t), y(t-1), \ldots$ and $u(t-1), u(t-2), \ldots$), the j-step-ahead predictor can be expressed as

$$\hat{y}(t+j|t) = \sum_{i=1}^{j} \bar{h}_{j,i,t}u(t+j-i) + v_j(t) \ , \quad j = 1, 2, 3, \ldots \ , \tag{3.6}$$

[2] In 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.

where $\bar{h}_{j,i,t}$ $(i = 1, 2, 3, \ldots, j)$ 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 (3.3) recursively, but there is an easier way.

Write the model (3.2) as

$$y(t) = H_t(q^{-1})u(t) + G_t(q^{-1})e(t) \ , \tag{3.7}$$

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 $h_{0,t}, h_{1,t}, h_{2,t}, \ldots$ 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 (3.7) the j-step ahead output prediction is found by conditional expectation:

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

Comparison with (3.6) shows that

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

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

$$\hat{y}(t+j|t) = \sum_{i=1}^{j} h_{i,t+j} u(t+j-i) + v_j(t) \ . \tag{3.8}$$

Equation (3.8) shows that the $h_{i,t+j}$-values and $v_j(t)$ can be calculated as follows:

1. To find $v_j(t)$ $(j = 1, 2, 3, \ldots)$: Let $u(t+j-i) = 0$ for $i = 1, 2, \ldots, j$ and compute $\hat{y}(t+j|t)$ using (3.3). This "prediction" becomes $v_j(t)$.

2. To find $h_{i,t+j}$ $(j = 1, 2, \ldots$ and $i = 1, \ldots, j)$: Let

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

and $y(t-l) = 0$, $e(t-l) = 0$ for $l = 0, 1, 2, \ldots$ (corresponding to $v_j(t) = 0$) and compute $\hat{y}(t+j|t)$ using (3.3). 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 (3.8) it is found that the j-step-ahead predictions, j running from 1 up to N, can be written as a linear matrix expression:

$$\hat{\boldsymbol{y}}(t) = \boldsymbol{H}(t)\boldsymbol{u}(t) + \boldsymbol{v}(t) \; , \tag{3.9}$$

where

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

In (3.9) 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 jth row of $\boldsymbol{H}(t)$ and the jth element of $\boldsymbol{v}(t)$ 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 predetermined 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

$$\begin{aligned}
&\min_{\nabla u(t+j-1)_{j=1,\ldots,N_2}} J[N_1, N_2, N_u, \lambda_1, \ldots, \lambda_{N_u}; t, \nabla u(t+j-1)_{j=1,\ldots,N_2}] \\
&= E_t\left[\textstyle\sum_{j=N_1}^{N_2} [y(t+j) - y^0(t+j)]^2 + \sum_{j=1}^{N_2} \lambda_j [\nabla u(t+j-1)]^2\right]
\end{aligned} \tag{3.10}$$

subject to the model in (3.1) and

$$\nabla u(t+j-1) = 0 \; , \;\; j > N_u \; , \tag{3.11}$$

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 (3.10 may be rewritten as

$$\begin{aligned}\min_{\boldsymbol{\nabla u}(t)} J(\boldsymbol{\Lambda}; t, \boldsymbol{\nabla u}(t)) &= \\ &E_t[(\boldsymbol{y}(t) - \boldsymbol{y}^0(t))^T(\boldsymbol{y}(t) - \boldsymbol{y}^0(t)) + \boldsymbol{\nabla u}(t)^T \boldsymbol{\Lambda} \boldsymbol{\nabla u}(t)] \ , \end{aligned} \tag{3.12}$$

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

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

For the XGPC the control strategy is based on minimization of a more general cost function than the ordinary GPC cost function in (3.10):

$$\begin{aligned}&\min_{\boldsymbol{u}(t)} J(\boldsymbol{\Gamma}(t), \boldsymbol{\Lambda}(t), \boldsymbol{\omega}(t); t, \boldsymbol{u}(t)) \\ &\quad = E_t[(\boldsymbol{y}(t) - \boldsymbol{y}^0(t))^T \boldsymbol{\Gamma}(t)(\boldsymbol{y}(t) - \boldsymbol{y}^0(t)) + \\ &\quad\quad \boldsymbol{u}(t)^T \boldsymbol{\Lambda}(t) \boldsymbol{u}(t) + 2\boldsymbol{\omega}(t)^T \boldsymbol{u}(t)] \ , \end{aligned} \tag{3.13}$$

where

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

and

N is the maximum prediction horizon, which corresponds to N_2 in (3.10),

$\boldsymbol{\Gamma}(t)$ is a positive semidefinite[3] and symmetric matrix which weights the control errors,

$\boldsymbol{\Lambda}(t)$ is a positive semidefinite[3] 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 (3.13) the cost function of the GPC controller may be obtained, cf. page 13.

3.3 Choice of the design parameters for the XGPC controller

Apart from the fact that the $\boldsymbol{\Gamma}(t)$-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. $\boldsymbol{\Gamma}(t) = \mathrm{Var}[\boldsymbol{e}(t)]^{-1}$, where $\boldsymbol{e}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t)$. This choice means that less

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

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, $\mathbf{\Lambda}(t) = \lambda \boldsymbol{I}$ ($\lambda > 0$) 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

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

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 $\boldsymbol{u}(t)$ but dependent on $u^0(t)$. This leads to

$$\begin{aligned}&\mathbf{\Lambda}(t) = \\ &\begin{pmatrix} \lambda_{1,1}+\lambda_{2,1}+\lambda_{2,2} & -\lambda_{2,2} & \cdots & 0 & 0 \\ -\lambda_{2,2} & \lambda_{1,2}+\lambda_{2,2}+\lambda_{2,3} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & \lambda_{1,N-1}+\lambda_{2,N-1}+\lambda_{2,N} & -\lambda_{2,N} \\ 0 & 0 & \cdots & -\lambda_{2,N} & \lambda_{1,N}+\lambda_{2,N} \end{pmatrix}\end{aligned} \tag{3.14}$$

and

$$\boldsymbol{\omega}(t) = -\begin{pmatrix} \lambda_{1,1}u^0(t) + \lambda_{2,1}u(t-1) \\ \lambda_{1,2}u^0(t+1) \\ \lambda_{1,3}u^0(t+2) \\ \vdots \\ \lambda_{1,N}u^0(t+N-1) \end{pmatrix} . \tag{3.15}$$

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 (3.10).

It is also possible to choose $\mathbf{\Lambda}(t)$ and $\boldsymbol{\omega}(t)$ so that filtered controls, $u^f(t)$, are weighted in (3.13). 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

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

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

where the x's represent non-zero elements.

From the above discussion one can conclude that the cost function in (3.13) 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 (3.13) 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 ($j = 1, \ldots, N$),

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

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

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

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 $\hat{\boldsymbol{y}}(t)$ from (3.9) the cost function becomes:

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

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

$$
\begin{aligned}
\tilde{J}(t, \boldsymbol{u}(t)) \quad &= \\
&(\boldsymbol{H}(t)\boldsymbol{u}(t) + \boldsymbol{\beta}(t))^T \mathbf{\Gamma}(t)(\boldsymbol{H}(t)\boldsymbol{u}(t) + \boldsymbol{\beta}(t)) + \boldsymbol{u}(t)^T \mathbf{\Lambda}(t) \boldsymbol{u}(t) + 2\boldsymbol{\omega}(t)^T \boldsymbol{u}(t) \, , \qquad (3.16)
\end{aligned}
$$

where

$$\boldsymbol{\beta}(t) = \boldsymbol{v}(t) - \boldsymbol{y}^0(t) \; . \tag{3.17}$$

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

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

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

$$\boldsymbol{u}(t) = -\left[\boldsymbol{H}(t)^T\boldsymbol{\Gamma}(t)\boldsymbol{H}(t) + \boldsymbol{\Lambda}(t)\right]^{-1}\left[\boldsymbol{H}(t)^T\boldsymbol{\Gamma}(t)^T\boldsymbol{\beta}(t) + \boldsymbol{\omega}(t)\right] \; . \tag{3.18}$$

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

$$u(t) = -\boldsymbol{l}(t)\left[\boldsymbol{H}(t)^T\boldsymbol{\Gamma}(t)^T\boldsymbol{\beta}(t) + \boldsymbol{\omega}(t)\right] \; , \tag{3.19}$$

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

The elements in the vector $\boldsymbol{\beta}(t)$ defined in (3.17) are predicted values of the control errors given future controls equal zero, i.e. the predictions if $\boldsymbol{u}(t) = \boldsymbol{0}$, cf. (3.9). Hence the optimal control in (3.19) 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 (3.13) (or the equivalent in (3.16)). Hence consider the following set of constraints:

$$\boldsymbol{S}(t)\boldsymbol{u}(t) = \boldsymbol{d}(t) \; , \tag{3.20}$$

where the matrix $\boldsymbol{S}(t)$ is $M \times N$ ($M \leq N$) and has full rank ($= M$). The vector $\boldsymbol{d}(t)$ is a column with M elements. To solve a minimization problem being subject to (3.20), the M equations can simply be solved with respect to a subset of the elements in $\boldsymbol{u}(t)$. 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 $\boldsymbol{u}$-vector[4] into two parts, $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$:

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

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

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

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

$$\boldsymbol{S} = [\boldsymbol{S}_1 \ \boldsymbol{S}_2] \ , \tag{3.21}$$

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

Now (3.20) can be written as

$$\boldsymbol{S}_1\boldsymbol{u}_1 + \boldsymbol{S}_2\boldsymbol{u}_2 = \boldsymbol{d} \ . \tag{3.22}$$

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

$$\boldsymbol{u}_2 = \boldsymbol{S}_2^{-1}(\boldsymbol{d} - \boldsymbol{S}_1\boldsymbol{u}_1) \ . \tag{3.23}$$

Matrices and vectors in (3.16) are partitioned so that they match the partitioning of $\boldsymbol{u}$:

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

Using this (3.16) can be rewritten as

$$\begin{aligned} \tilde{J}([\boldsymbol{u}_1^T \ \boldsymbol{u}_2^T]^T) &= (\boldsymbol{H}_1\boldsymbol{u}_1 + \boldsymbol{H}_2\boldsymbol{u}_2 + \boldsymbol{\beta})^T\boldsymbol{\Gamma}(\boldsymbol{H}_1\boldsymbol{u}_1 + \boldsymbol{H}_2\boldsymbol{u}_2 + \boldsymbol{\beta}) + \\ &\quad \boldsymbol{u}_1^T\boldsymbol{\Lambda}_{11}\boldsymbol{u}_1 + 2\boldsymbol{u}_1^T\boldsymbol{\Lambda}_{12}\boldsymbol{u}_2 + \boldsymbol{u}_2^T\boldsymbol{\Lambda}_{22}\boldsymbol{u}_2 + 2\boldsymbol{\omega}_1^T\boldsymbol{u}_1 + 2\boldsymbol{\omega}_2^T\boldsymbol{u}_2 \ , \end{aligned} \tag{3.24}$$

where the fact that $\boldsymbol{\Lambda}$ is symmetric has been utilized (i.e. $\boldsymbol{\Lambda}_{12} = \boldsymbol{\Lambda}_{21}^T$). Substituting $\boldsymbol{u}_2$ in (3.24) by the right-hand side of (3.23), it is found that

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

where

$$\begin{aligned} \boldsymbol{H}^* &= \boldsymbol{H}_1 - \boldsymbol{H}_2\boldsymbol{S}_2^{-1}\boldsymbol{S}_1 \\ \boldsymbol{\beta}^* &= \boldsymbol{H}_2\boldsymbol{S}_2^{-1}\boldsymbol{d} + \boldsymbol{\beta} \\ \boldsymbol{\Lambda}^* &= \boldsymbol{\Lambda}_{11} + (\boldsymbol{S}_1^T(\boldsymbol{S}_2^{-1})^T\boldsymbol{\Lambda}_{22} - 2\boldsymbol{\Lambda}_{12})\boldsymbol{S}_2^{-1}\boldsymbol{S}_1 \\ \boldsymbol{\omega}^* &= (\boldsymbol{\Lambda}_{12} - \boldsymbol{S}_1^T(\boldsymbol{S}_2^{-1})^T\boldsymbol{\Lambda}_{22})\boldsymbol{S}_2^{-1}\boldsymbol{d} + \\ &\quad \boldsymbol{\omega}_1 - \boldsymbol{S}_1^T(\boldsymbol{S}_2^{-1})^T\boldsymbol{\omega}_2 \ . \end{aligned} \tag{3.25}$$

Thus the optimal control is found by minimizing

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

with respect to $\boldsymbol{u}_1$. The structure of this cost function is similar to the structure of the cost function in (3.16). Therefore, the solution to the constrained optimization problem is obtained by replacing $\boldsymbol{H}$, $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\omega}$ in (3.18) by $\boldsymbol{H}^*$, $\boldsymbol{\beta}^*$, $\boldsymbol{\Lambda}^*$ and $\boldsymbol{\omega}^*$, respectively.

3.5.1 The constraints used by Clarke *et al.*

Below it is shown how to choose $\boldsymbol{H}^*$, $\boldsymbol{\beta}^*$, $\boldsymbol{\Lambda}^*$ and $\boldsymbol{\omega}^*$ in order to obtain the GPC method presented by (Clarke et al. 1987). According to (3.11), 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 (3.20). If, for instance, $N = 5$ and $N_u = 2$, then $M = N - N_u = 3$ and

$$\boldsymbol{S}(t) = \left(\begin{array}{cc|ccc} 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{array}\right), \quad \boldsymbol{d}(t) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

with the first two columns forming $\boldsymbol{S}_1(t)$ and the last three forming $\boldsymbol{S}_2(t)$.

If $N_1 = 1$ and $N_2 = N = 5$ in the cost function in (3.10) it is found that (cf. (3.14) and (3.15)):

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

For the time-invariant ARIMAX model used by (Clarke et al. 1987) the "impulse response matrix", $\boldsymbol{H}$, introduced in (3.9) becomes time-invariant

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

Observing that

$$\boldsymbol{S}_2^{-1}\boldsymbol{S}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 0 & -1 \\ 0 & -1 \end{pmatrix} , \tag{3.26}$$

it is readily found from (3.25) that

$$\begin{aligned}
\boldsymbol{H}^* &= \boldsymbol{H}_1 + \boldsymbol{H}_2 \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} h_1 & 0 \\ h_2 & h_1 \\ h_3 & h_1+h_2 \\ h_4 & h_1+h_2+h_3 \\ h_5 & h_1+h_2+h_3+h_4 \end{pmatrix} \\
\boldsymbol{\beta}^*(t) &= \boldsymbol{\beta}(t) \\
\boldsymbol{\Lambda}^* &= \begin{pmatrix} \lambda_1+\lambda_2 & -\lambda_2 \\ -\lambda_2 & \lambda_2 \end{pmatrix} \\
\boldsymbol{\omega}^*(t) &= \boldsymbol{\omega}_1(t) = (-\lambda_1 u(t-1)\ \ 0)^T ,
\end{aligned} \tag{3.27}$$

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

where the vector $\boldsymbol{\beta}(t)$ contain the predicted values of the input-free control errors (defined in (3.17)).

Note that the first column of $\boldsymbol{H}^*$ is equal to the first column of $\boldsymbol{H}$, and the second column is the sum of columns 2 to 5 of $\boldsymbol{H}$. As a general rule it can be established that the first $N_u - 1$ columns of $\boldsymbol{H}^*$ are copied from $\boldsymbol{H}$ while the N_uth column is the sum of the remaining columns of $\boldsymbol{H}$. This result was also heuristically proposed by (Bjerre 1992). Note also that $\boldsymbol{\Lambda}^*$ 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 (3.13) minimized in Section 3 in order to obtain the optimal controller (3.19) 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.

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 4.1. In other words $y^0(t+j)$ should be determined so that

$$P\{\,(x(t+j), y(t+j)\,) \in \Omega_2 \,|\, I_t\} \quad = \quad \pi \;\; , 0 < j < N \; , \tag{4.1}$$

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

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

The future values of x and y can be written as a sum of predictions and prediction errors, i.e.

$$\begin{aligned} x(t+j) &= \hat{x}(t+j|t) + e_x(t+j|t) \\ y(t+j) &= \hat{y}(t+j|t) + e_y(t+j|t) \; , \end{aligned} \tag{4.2}$$

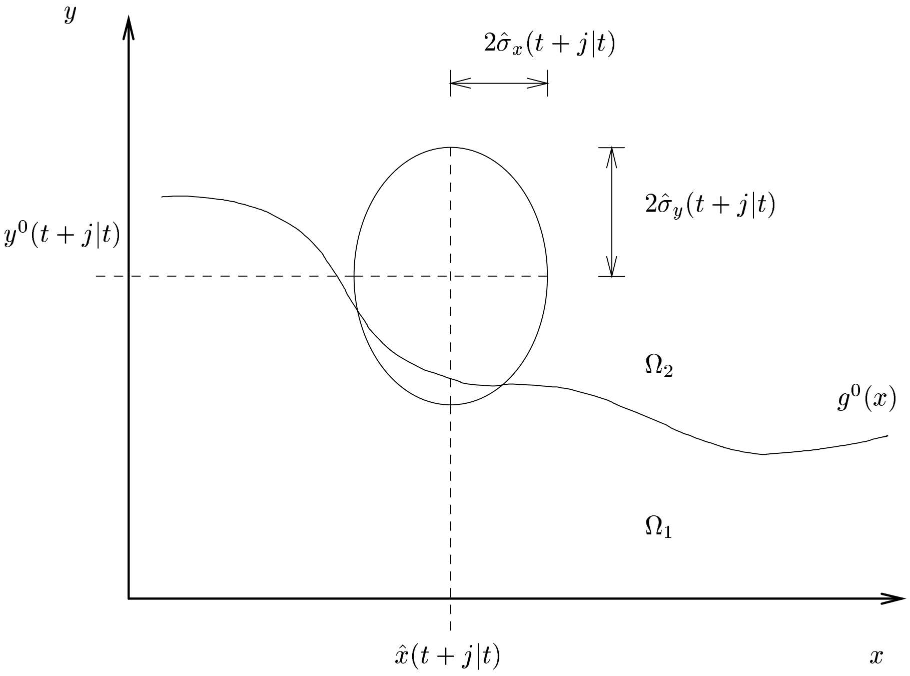

Figure 4.1: *Inference of an output reference based on model uncertainty for the case where the reference or restriction curve, $g^0(x)$, is a function of one explanatory variable, x. The model uncertainty is specified by estimated standard deviation $\hat{\sigma}_y()$ and $\hat{\sigma}_x()$ for the 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 $g^0(x)$.*

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 (4.2) by $y^0(t+j)$ before inserting (4.2) into (4.1) thereby obtaining

$$\begin{aligned} P\{ \, (\, \hat{x}(t+j|t) + e_x(t+j|t), y^0(t+j) + e_y(t+j|t) \,) \\ \in \Omega_2 \, \} = \pi \quad , \; j > 0 \; . \end{aligned} \tag{4.3}$$

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, (4.3) is rewritten as

$$1 - \pi = \int_{-\infty}^{\infty} \int_{-\infty}^{g^0(x)} f_{t+j}(x(t+j), y(t+j)) \, dx \, dy \tag{4.4}$$

Using (4.2) 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,

IMM
Informatics and Mathematical Modelling
Technical University of Denmark

(4.4) is reformulated as

$$\begin{aligned}1-\pi &= \int_{-\infty}^{\infty}\int_{-\infty}^{g^0(x)} \phi\left(\frac{x-\hat{x}(t+j|t)}{\sigma_x^j(t+j)}\right)\phi\left(\frac{y-y^0(t+j)}{\sigma_y^j(t+j)}\right) dy\, dx \\ &= \int_{-\infty}^{\infty} \phi\left(\frac{x-\hat{x}(t+j|t)}{\sigma_x^j(t+j)}\right)[\int_{-\infty}^{g^0(x)} \phi\left(\frac{y-y^0(t+j)}{\sigma_y^j(t+j)}\right) dy\,]\, dx \\ &= \int_{-\infty}^{\infty} \phi\left(\frac{x-\hat{x}(t+j|t)}{\sigma_x^j(t+j)}\right)\Phi\left(\frac{g^0(x)-y^0(t+j)}{\sigma_y^j(t+j)}\right) dx\ , \qquad (4.5)\end{aligned}$$

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

Since (4.5) 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 determinating $y^0(t+j)$ 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, $g^0(x_1,\cdots,x_p)$, then the output reference, $y^0(t+j)$, is found as the solution to

$$\begin{aligned}&1-\pi = \\ &\int_{-\infty}^{\infty} \phi\left(\frac{x_1-\hat{x_1}(t+j|t)}{\sigma_{x_1}^j(t+j)}\right)\cdots\int_{-\infty}^{\infty} \phi\left(\frac{x_p-\hat{x_p}(t+j|t)}{\sigma_{x_p}^j(t+j)}\right) \qquad (4.6)\\ &\Phi\left(\frac{g^0(x_1,\cdots,x_p)-y^0(t+j)}{\sigma_y^j(t+j)}\right) dx\ .\end{aligned}$$

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 $\sigma_y^j(t+j), \sigma_{x_1}^j(t+j), \cdots \sigma_{x_p}^j(t+j)$. The solution to (4.6) is found as for (4.5). 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

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

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

$$\begin{aligned}
\hat{\sigma}_e(k) &= \sqrt{\frac{1}{N_{obs}-1}\sum_{t=1}^{N_{obs}}(e(t|t-k)-\hat{\mu}_e(k))^2} \qquad (5.1)\\
\hat{\mu}_e(k) &= \frac{1}{N_{obs}}\sum_{t=1}^{N_{obs}} e(t|t-k)\\
e(t|t-k) &= Y(t)-\hat{Y}(t|t-k)
\end{aligned}$$

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 5.1.

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 *mean value predictor*. 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 *persistent or naive predictor*.
$\nabla_{24}y(t)=e_{\nabla_{24}}(t)$	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 *persistent diurnal predictor*.

Table 5.1: *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 suburb[5] with heat and hot tap water. The system is divided into several sub-networks, which are operated individually. The results described

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

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

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 6.1 shows an overview over the Gasværksvej sub-network including the branches supplying three net-points, which have been selected as critical points (cf. page 3).

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.

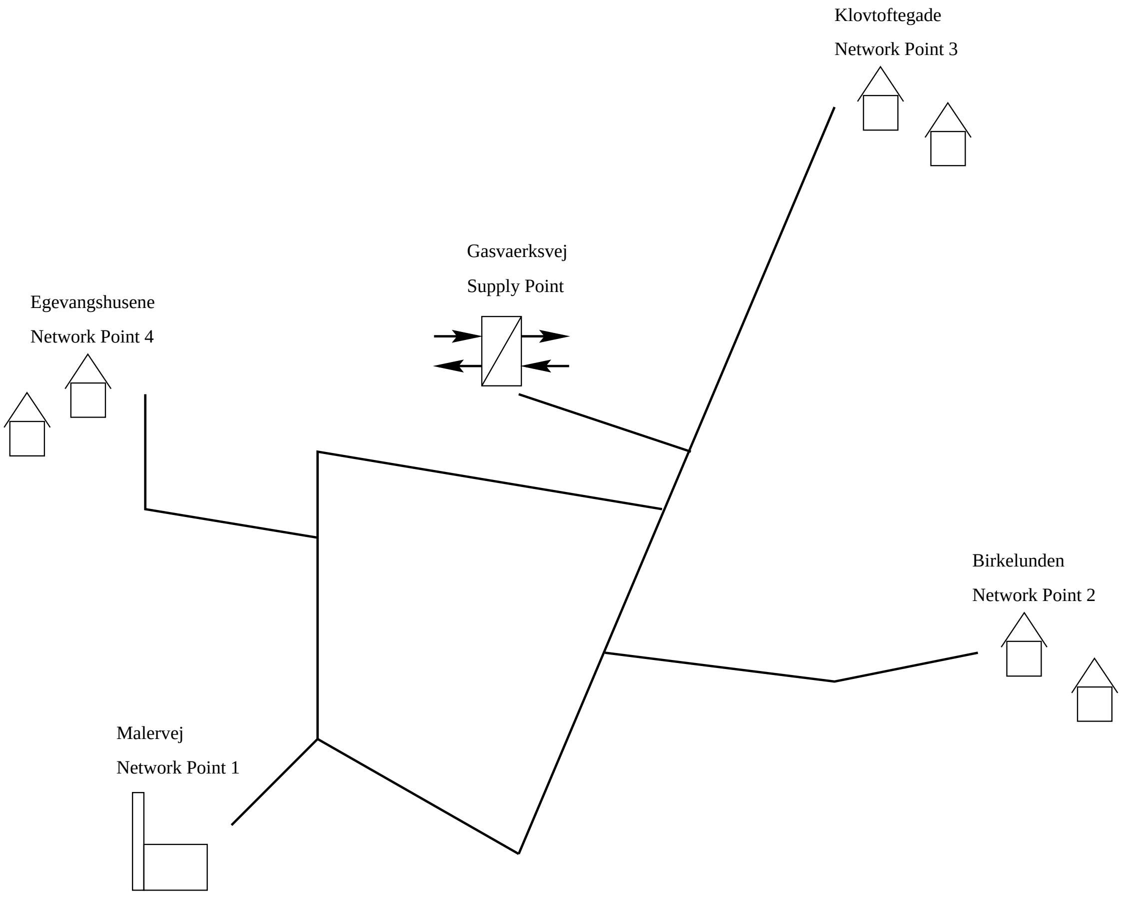

Figure 6.1: *An overview of the Gasværksvej network in the Høje Tåstrup district heating utility.*

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, Palsson, Sejling & Søgaard 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

$$\begin{aligned}
&\hat{p}(t+1|t) = \\
&\quad a_1\,p(t) + a_2\,p(t-1) + a_{24}\,p(t-23) + a_{25}\,p(t-24) + a_{26}\,p(t-25) + \\
&\quad b_{1,0}\,\nabla T_s(t+1) + b_{1,1}\,\nabla T_s(t) + b_{1,24}\,\nabla T_s(t-23) + b_{1,25}\,\nabla T_s(t-24) + \\
&\quad b_{2,1}\,T_a(t) + b_{2,2}\,T_a(t-1) + b_{2,24}\,T_a(t-23) + b_{2,25}\,T_a(t-24) + \\
&\quad \sum_{i=1}^{3}\left[c_{1,i}\,\sin\frac{2\,i\,\pi\,(t+1)}{24} + d_{1,i}\,\cos\frac{2\,i\,\pi\,(t+1)}{24}\right] + \\
&\quad I_{\{A\}}(t)\sum_{i=1}^{3}\left[c_{2,i}\,\sin\frac{2\,i\,\pi\,(t+1)}{24} + d_{2,i}\,\cos\frac{2\,i\,\pi\,(t+1)}{24}\right] + l\ ,
\end{aligned} \tag{7.1}$$

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 (7.1) 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

$$
\begin{aligned}
&\hat{p}(t+j|t) = \\
&\quad a_j\, p(t) + a_{j+1}\, p(t-1) + \\
&\quad a_{24}\, p(t+j-24) + a_{25}\, p(t+j-25) + a_{26}\, p(t+j-26) + \\
&\quad b_{1,0}\, \nabla T_s(t+j) + b_{1,1}\, \nabla T_s(t+j-1) + b_{1,2}\, \nabla T_s(t+j-2) + \\
&\quad b_{1,3}\, \nabla T_s(t+j-3) + b_{1,24}\, \nabla T_s(t+j-24) + b_{1,25}\, \nabla T_s(t+j-25) + \\
&\quad b_{2,j}\, T_a(t) + b_{2,j+1}\, T_a(t-1) + b_{2,j+2}\, T_a(t-2) + b_{2,j+3}\, T_a(t-3) + \\
&\quad b_{2,24}\, T_a(t+j-24) + b_{2,25}\, T_a(t+j-25) + \\
&\quad \sum_{i=1}^{3} \left[c_{1,i} \sin \frac{2\, i\, \pi\, (t+j)}{24} + d_{1,i} \cos \frac{2\, i\, \pi\, (t+j)}{24} \right] + \\
&\quad I_{\{A\}}(t) \sum_{i=1}^{3} \left[c_{2,i} \sin \frac{2\, i\, \pi\, (t+j)}{24} + d_{2,i} \cos \frac{2\, i\, \pi\, (t+j)}{24} \right] + l
\end{aligned}
\tag{7.2}
$$

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.

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 7.1: *The estimated standard deviation of the predictions errors the heat load model (7.1, 7.2) ($\hat{\sigma}_e$) 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.*

The results obtained at the Høje Tåstrup district heating utility by the heat load prediction model (7.1, 7.2) are shown in Table 7.1 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 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 7.1 and 7.2 show plots of observed and predicted heat load as well as the observed prediction errors for k equal 12 and 24 hours.

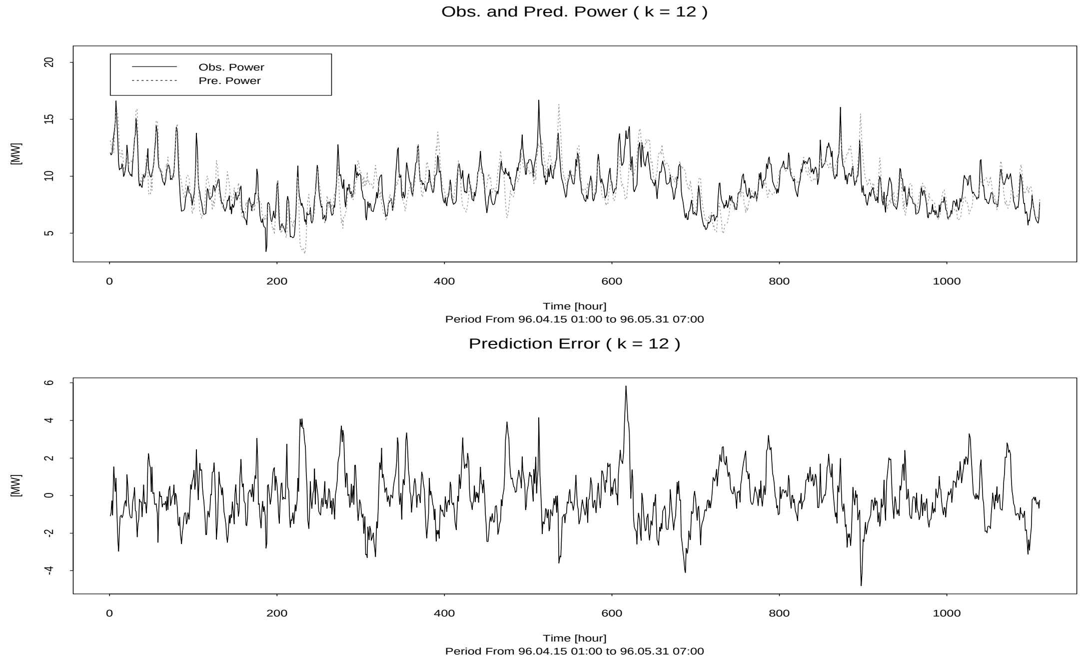

Figure 7.1: *Top) Observed heat load (full line) and the 12 hour predictions of heat load (dotted line) at the Høje Tåstrup district heating utility in the period from April 15 1996 to May 31 1996. Bottom) The 12 hour prediction error in the same period.*

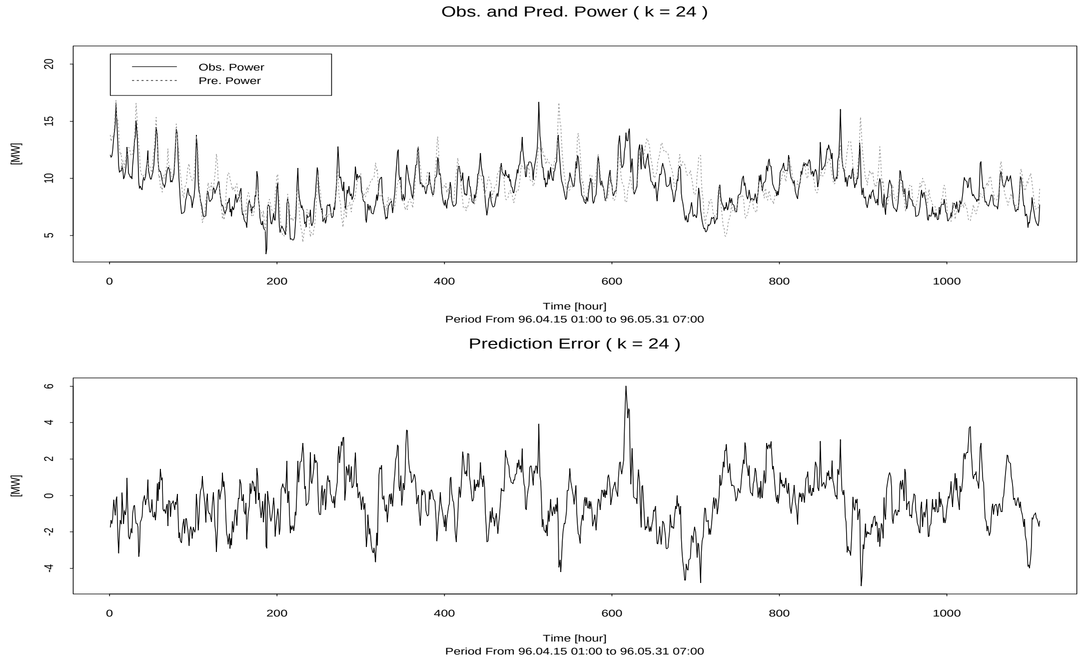

Figure 7.2: *Top) Observed heat load (full line) and the 24 hour predictions of heat load (dotted line) at the Høje Tåstrup district heating utility in the period from April 15 1996 to May 31 1996. Bottom) The 24 hour prediction error in the same period.*

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:

$$p(t) = c_w q(t)(T_s(t) - T_r(t)) \ , \tag{7.3}$$

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

$$P\{q(t+j) \leq q_{\max}\} = \pi \ , \quad 0 < N_1 \leq j \leq N_2 \ , \tag{7.4}$$

where N_1 and N_2 are chosen so that $q(t+N_1), \ldots, q(t+N_2)$ 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 (7.3) is used to substitute for $q(t+j)$ in (7.4) we obtain

$$P\{p(t+j) \leq c_w q_{\max}(T_s(t+j) - T_r(t+j))\} = \pi \ . \tag{7.5}$$

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

$$\hat{T}_r(t+j|t) = T_r(t) \ , \quad j = 1, 2, 3, \ldots \ . \tag{7.6}$$

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

$$\begin{aligned} p(t+j) &= \hat{p}(t+j|t) + e_p(t+j|t) \\ T_r(t+j) &= \hat{T}_r(t+j|t) + e_{T_r}(t+j|t) \ , \end{aligned} \tag{7.7}$$

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 (7.7) into (7.5) gives (after a few manipulations)

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

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

$$\frac{c_w q_{\max}(T_s(t+j) - \hat{T}_r(t+j|t)) - \hat{p}(t+j|t)}{\sqrt{\sigma_p^2(j) + c_w^2 q_{\max}^2 \sigma_{T_r}^2(j)}} = u_\pi \ , \tag{7.8}$$

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 (7.1, 7.2) 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 (7.1, 7.2).

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

$$T_s(t+j) = T_s(t+1) \ , \quad j = 2,3 \tag{7.9}$$

($\Rightarrow \nabla T_s(t+j) = 0$ for $j = 2,3$). Under this restriction, (7.1, 7.2) can be expressed as

$$\begin{aligned} \hat{p}(t+j|t) &= \beta_j \nabla T_s(t+1) + \gamma_j(t) \\ &= \beta_j (T_s(t+1) - T_s(t)) + \gamma_j(t) \ , \end{aligned} \tag{7.10}$$

where

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

and, respectively,

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

and

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

Here the diurnal profiles in (7.1, 7.2) of the heat load during working and non-working days have been abbreviated μ_1 and μ_2, respectively.

Inserting (7.6), (7.9) and (7.10) into (7.8) 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 $T_s(t+1)$,

$$T_s(t+1) = \frac{S_j + K T_r(t) - \beta_j T_s(t) + \gamma_j(t)}{K - \beta_j} \ , \tag{7.11}$$

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), (7.11) results in a value, $T_{s,j}(t+1)$, of $T_s(t+1)$, which is a solution to (7.4), 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 $T_{s,3}(t+1)$:

$$T_s(t+1) = \boldsymbol{\chi}^T \boldsymbol{T}_s(t+1) \ , \tag{7.12}$$

where

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

and

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

Equation (7.12) 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 (7.12), 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 3).

The net-point temperatures are monitored by a set of controllers – one controller for each net-point – based on the XGPC optimal controller (3.18). 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.

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 8.1. As it is seen from the figure the controller requires input from three different sources:

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

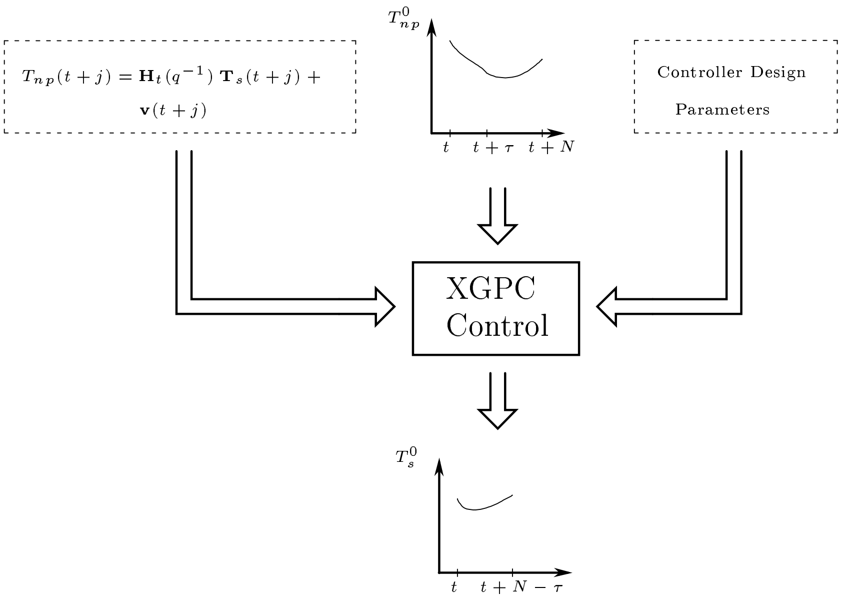

Figure 8.1: *Overview of XGPC net-point temperature controller. From left to right the XGPC requires, that the following informations are supplied: A system model describing how past and future supply temperatures will affect the future net-point temperature, a reference for the future net-point temperature and a specification of the full set of controller design parameters.*

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 contributers 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.

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

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

$$\hat{\gamma}_{T_s,T_{np}}(\tau,t) \quad = \quad \frac{1}{t-\tau} \sum_{i=\tau+1}^{t} \left(T_s(i) - \frac{1}{t}\sum_{j=1}^{t} T_s(j) \right) \left(T_{np}(i-\tau) - \frac{1}{t}\sum_{j=1}^{t} T_{np}(j) \right) \ , \tag{8.1}$$

In (8.1) 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 (8.1) is replaced by the recursive estimation algorithm

$$\begin{aligned}
\hat{\gamma}_{Ts,Tnp}(\tau, t) &= (1-\lambda_\gamma)\big(T_s(t)-\bar{T}_s(t)\big)\big(T_{np}(t-\tau)-\bar{T}_{np}(t-\tau)\big) + \\
&\quad \lambda_\gamma \hat{\gamma}_{Ts,Tnp}(\tau, t-1) \\
\bar{T}_s(t) &= (1-\lambda_m)T_s(t) + \lambda_m \bar{T}_s(t-1) \\
\bar{T}_{np}(t) &= (1-\lambda_m)T_{np}(t) + \lambda_m \bar{T}_{np}(t-1)
\end{aligned} \tag{8.2}$$

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 (8.2) it has been used that

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

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

$$\begin{aligned}
T_{np}(t) &= a_1\, T_{np}(t-1) + b_0(t-\tau)\, T_s(t-\tau) + b_1(t-\tau-1)\, T_s(t-\tau-1) + \\
&\quad b_2(t-\tau-2)\, T_s(t-\tau-2) + e(t) \\
b_0(t-\tau) &= b_{0,1} + b_{0,2}\sin\tfrac{2\pi(t-\tau)}{24} + b_{0,3}\cos\tfrac{2\pi(t-\tau)}{24} \\
b_1(t-\tau-1) &= b_{1,1} + b_{1,2}\sin\tfrac{2\pi(t-\tau-1)}{24} + b_{1,3}\cos\tfrac{2\pi(t-\tau-1)}{24} \\
b_2(t-\tau-2) &= b_{2,1} + b_{2,2}\sin\tfrac{2\pi(t-\tau-2)}{24} + b_{2,3}\cos\tfrac{2\pi(t-\tau-2)}{24}
\end{aligned} \tag{8.3}$$

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 (8.3) 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 (8.3) *and* the time delay estimation scheme previously described in Section 8.1.1.

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 8.1: *The estimated standard deviation of the predictions errors for the net-point temperature model (8.3) ($\hat{\sigma}_e$) 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.*

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 8.1. From the table it is seen, that net-point model (8.3) 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 8.2, 8.3 and 8.4 show plots of observed supply temperature, observed net-point temperature, predicted net-point temperature and prediction error for Birkelunden, Klovtoftegade and Egevangshusene, respectively. As expected from Table 8.1 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 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 8.4, which shows, that the pattern in supply temperature is hardly recognizable in net-point temperature for Egevangshusene.

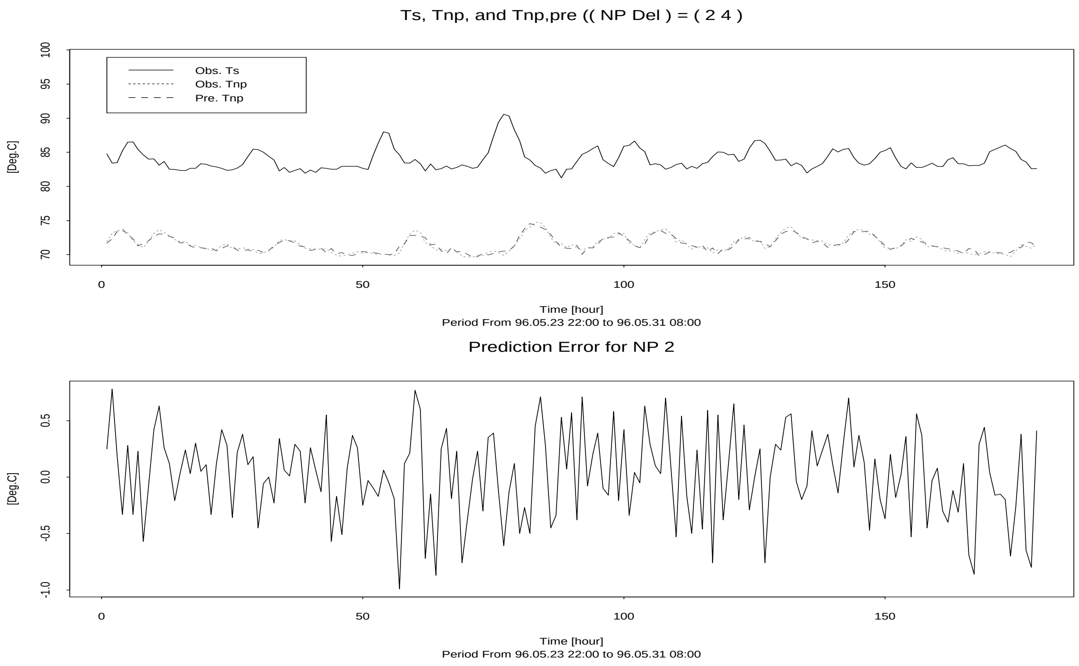

Figure 8.2: *Top) Observed supply temperature (full line), observed net-point temperature (dashed line) and the 1 hour predictions of net-point temperature (dotted line) for the net-point at Birkelunden during the period from May 23 1996 to May 31 1996. Bottom) The 1 hour prediction error in the same period.*

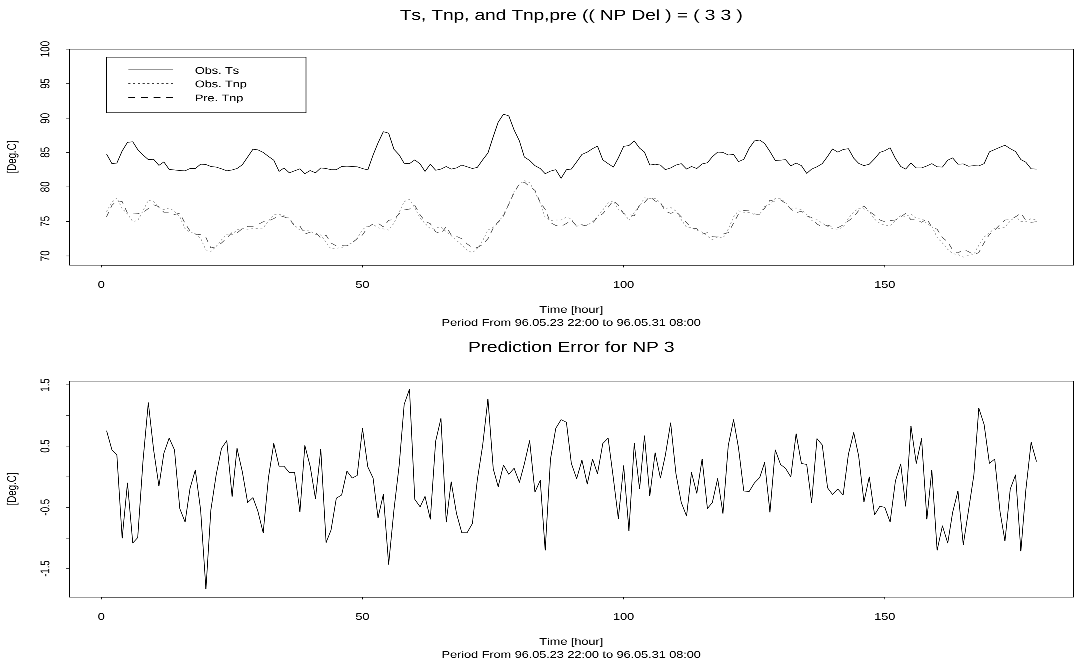

Figure 8.3: *Top) Observed supply temperature (full line), observed net-point temperature (dashed line) and the 1 hour predictions of net-point temperature (dotted line) for the net-point at Klovtoftegade during the period from May 23 1996 to May 31 1996. Bottom) The 1 hour prediction error in the same period.*

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

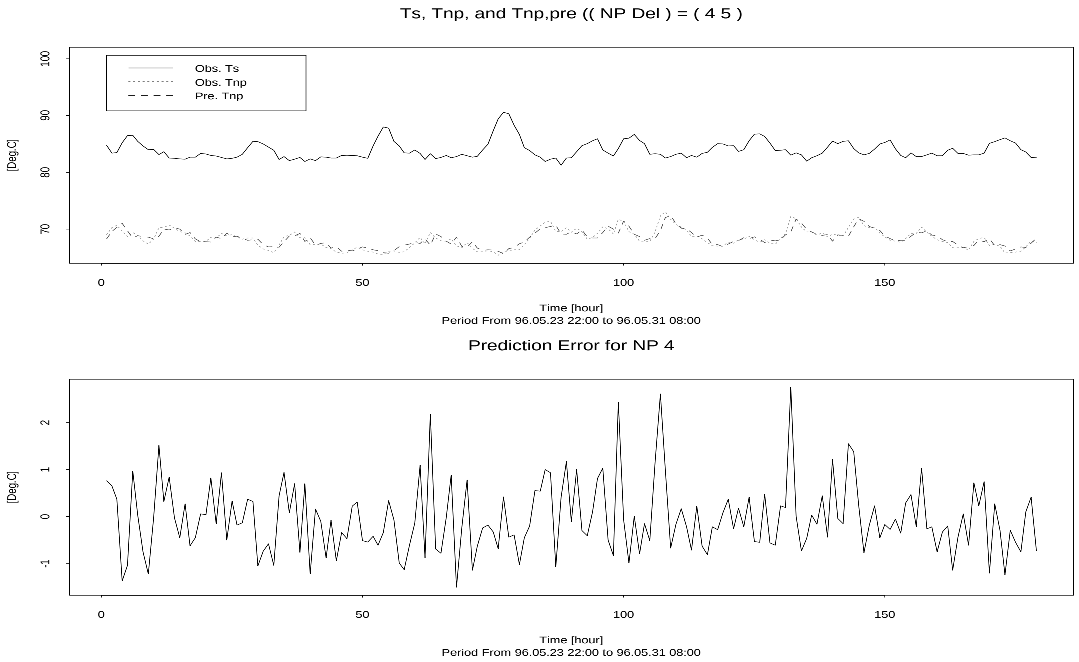

Figure 8.4: *Top) Observed supply temperature (full line), observed net-point temperature (dashed line) and the 1 hour predictions of net-point temperature (dotted line) for the net-point at Egevangs-husene during the period from May 23 1996 to May 31 1996. Bottom) The 1 hour prediction error in the same period.*

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

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.

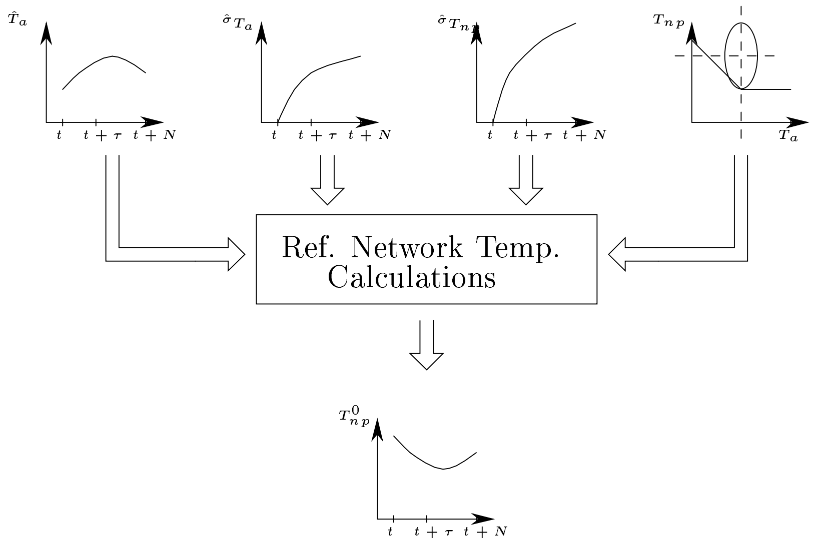

Figure 8.5: *Reference net-point temperature overview. From left to right the reference net-point temperature calculations require, that the following informations are supplied: A prediction of ambient temperature covering the output horizon ($\hat{T}_a$), estimates of the standard deviation for the 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 8.5, 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 ambient 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

$$S_k^2(t) = (1-\lambda)\left(y(t) - \hat{y}(t|t-k)\right)^2 + \lambda S_k^2(t-1)\ , \tag{8.4}$$

  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.

IMM

Informatics and Mathematical Modelling
Technical University of Denmark

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 8.6. 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.

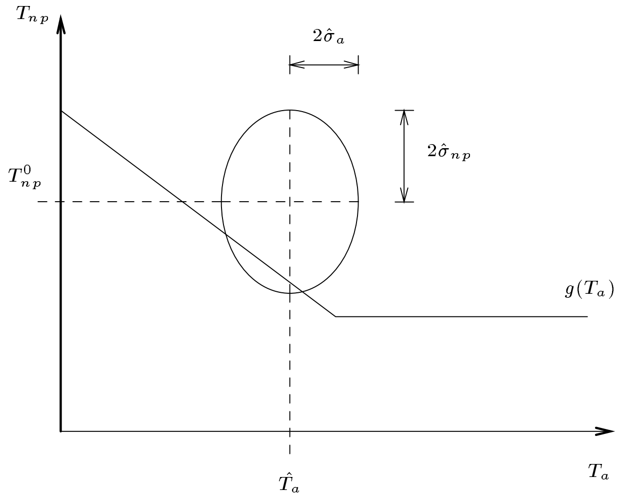

Figure 8.6: *Reference net-point temperature curve. The reference curve g^0 is a function of the ambient air temperature T_a. For a given (observed) ambient air temperature the corresponding net-point temperature should be in the area above g^0.*

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 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 8.6 – 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 8.7 illustrates the method using the net-point at Birkelunden as an example.

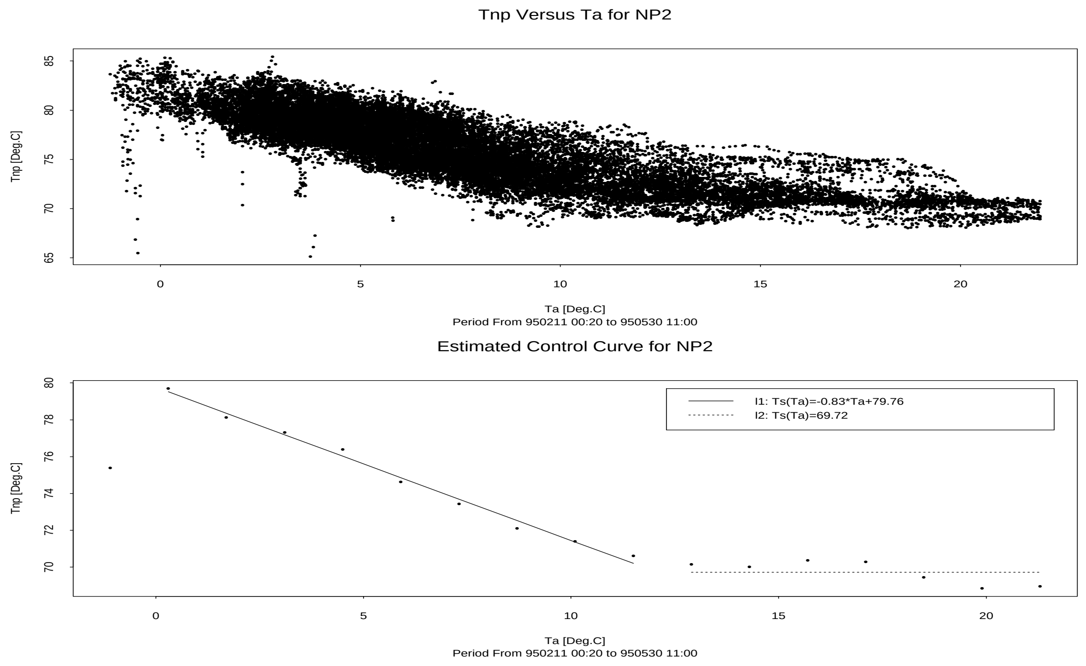

Figure 8.7: *Estimated Control Curve for the net-point at Birkelunden (2). Top) Observed net-point temperature versus observed ambient air temperature from a period – February 11 1995 to May 30 1995 – prior to the installation of PRESS. Bottom) The estimated control curve based on the 10 % quantile for the observed net-point temperature. The dots indicate the 10 % quantiles in each of the sub-intervals (defined by the ambient air temperature),* l1 *is the linear regression line for quantiles in intervals below 13 °C and* l2 *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 costumer 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 8.2.

$Net - point$	$T_{np,min}$	$T_{np,0}$	$T_{a,disc}$
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 8.2: *The final control curve parameters used at Høje Tåstrup from May 1996 and onwards. Here $T_{np,min}$ corresponds to* l2 *in Figure 8.7, $T_{np,0}$ 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* l1 *and* l2.

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

$$T_a(t) \quad = \quad a_k\, T_a(t-k) + a_{24}\, T_a(t-24) + \sum_{i=1}^{2} [c_{i,1} \sin \frac{2i\pi t}{24} + c_{i,2} \cos \frac{2i\pi t}{24}] + \mu + e(t) \qquad (8.5)$$

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 (8.5), implemented as a weighting between the three simple prediction models from Table 5.1 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 8.3: *The estimated standard deviation of the predictions errors for the ambient air temperature model (8.5) ($\hat{\sigma}_e$) 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 8.3 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 8.8 and 8.9 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.

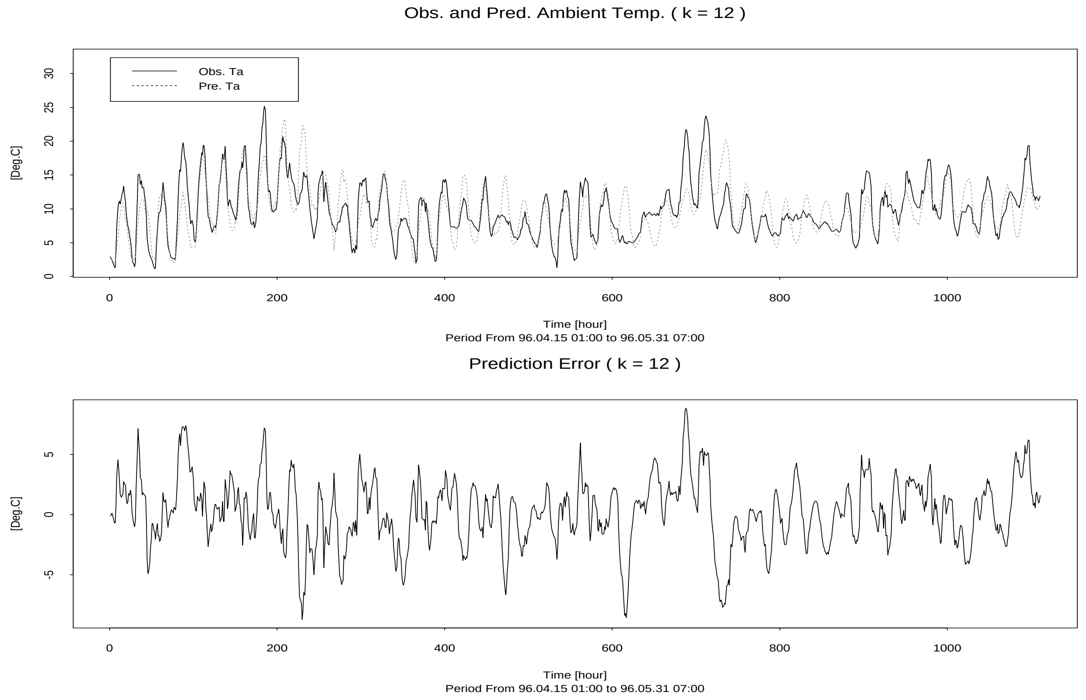

Figure 8.8: *Top) Observed ambient air temperature (full line) and the 12 hour predictions of ambient air temperature (dotted line) at the Høje Tåstrup district heating utility in the period from April 15 1996 to May 31 1996. Bottom) The 12 hour prediction error in the same period.*

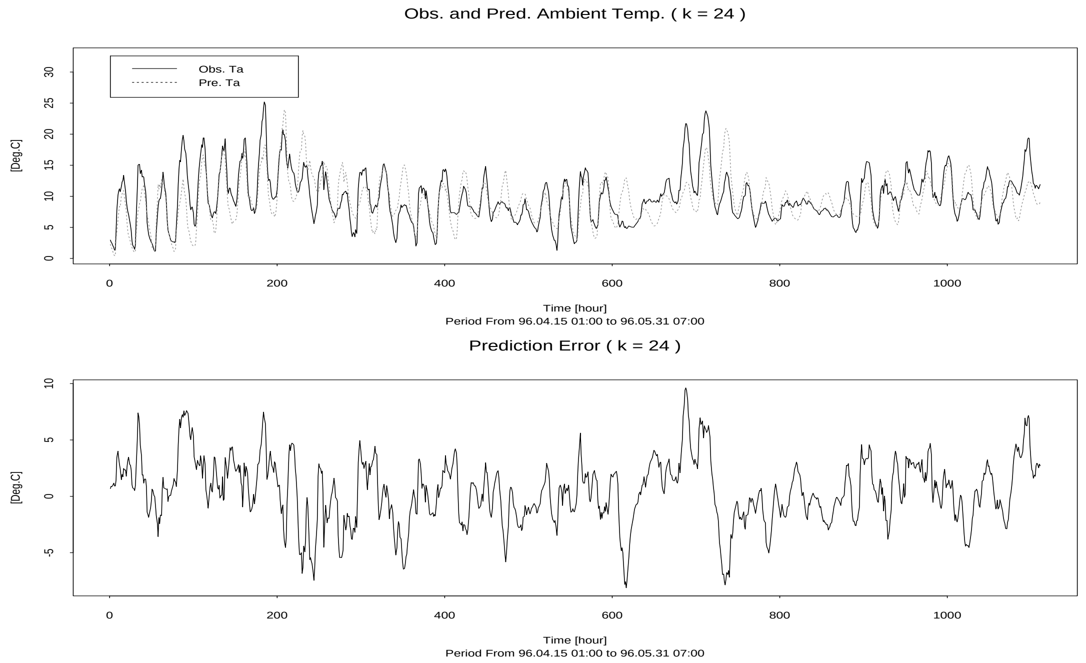

Figure 8.9: *Top) Observed ambient air temperature (full line) and the 24 hour predictions of ambient air temperature (dotted line) at the Høje Tåstrup district heating utility in the period from April 15 1996 to May 31 1996. Bottom) The 24 hour prediction error in the same period.*

Informatics and Mathematical Modelling
Technical University of Denmark

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:

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

Here

$\boldsymbol{u}(t)$ is a vector of future controls (supply temperatures)

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

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

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

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

$\boldsymbol{H}(t)$ is a matrix containing the time-varying impulse response of the system

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

$\boldsymbol{\beta}(t)$ is a vector containing the expected control errors from the input free system

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

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

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

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

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

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

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

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

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

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

and the partitioning of $\boldsymbol{H}(t)$, $\boldsymbol{\beta}(t)$, $\boldsymbol{\Gamma}(t)$, $\boldsymbol{\omega}(t)$, $\boldsymbol{S}(t)$ and $\boldsymbol{d}(t)$ 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 x 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 $\boldsymbol{S}_1(t)$, $\boldsymbol{S}_2(t)$ and $\boldsymbol{d}(t)$ linear equality constraint matrices and vectors.* These matrices define the linear equality constraints imposed on $\boldsymbol{u}(t)$. 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 $\boldsymbol{u}(t)$. In that case $\boldsymbol{S}_1(t)$, $\boldsymbol{S}_2(t)$ and $\boldsymbol{d}(t)$ become time-invariant and – as stated in Section 3.5.1 – are then given by

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

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

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

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

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

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

  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-negative[6] signal upon the required supply temperature as determined by

Hour of day	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 8.4: *Additional diurnal supply temperature signal.*

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 8.4.

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 8.1, 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 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

$$f(T_a, t) = \frac{1}{24}\sum_{i=0}^{23} T_{a,t-i} \quad . \tag{8.6}$$

The use of (8.6) 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 (8.6) are replaced by predictions, cf. page 34. 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

$$f(T_a, t) = \frac{0.06}{1 - 0.94q^{-1}} T_a(t) \ , \tag{8.7}$$

before the relationship between heat load and ambient air temperature is estimated, and it seems reasonable to expected the required net-point temperature to exhibit a similar dependency on ambient air temperature. (8.7 has recently been implemented in PRESS.

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

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.

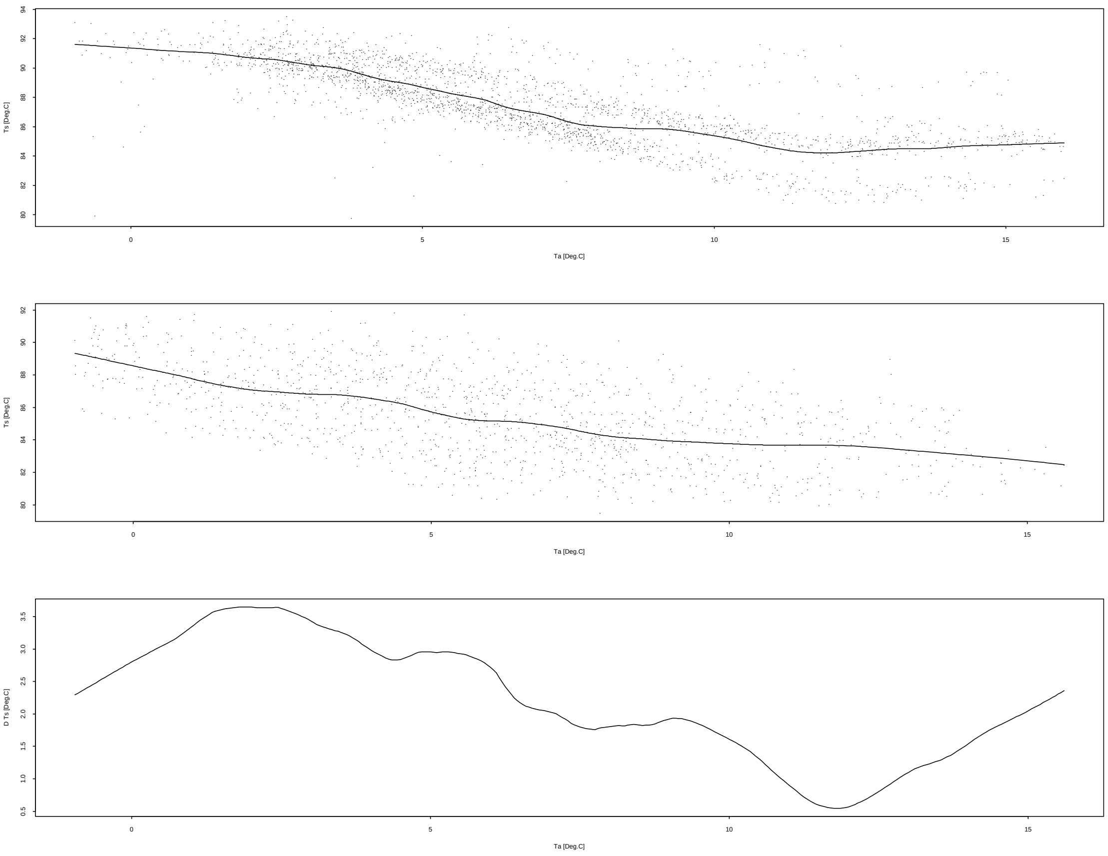

Figure 9.1: *Top) Supply temperature versus air temperature at Høje Tåstrup for a three month period during spring '95. Middle) Supply temperature versus air temperature at Høje Tåstrup for a three month period during spring '97. Bottom) The savings in supply temperature obtained by using PRESS versus air temperature.*

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 9.1 by estimating a non-parametric regression line for 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.

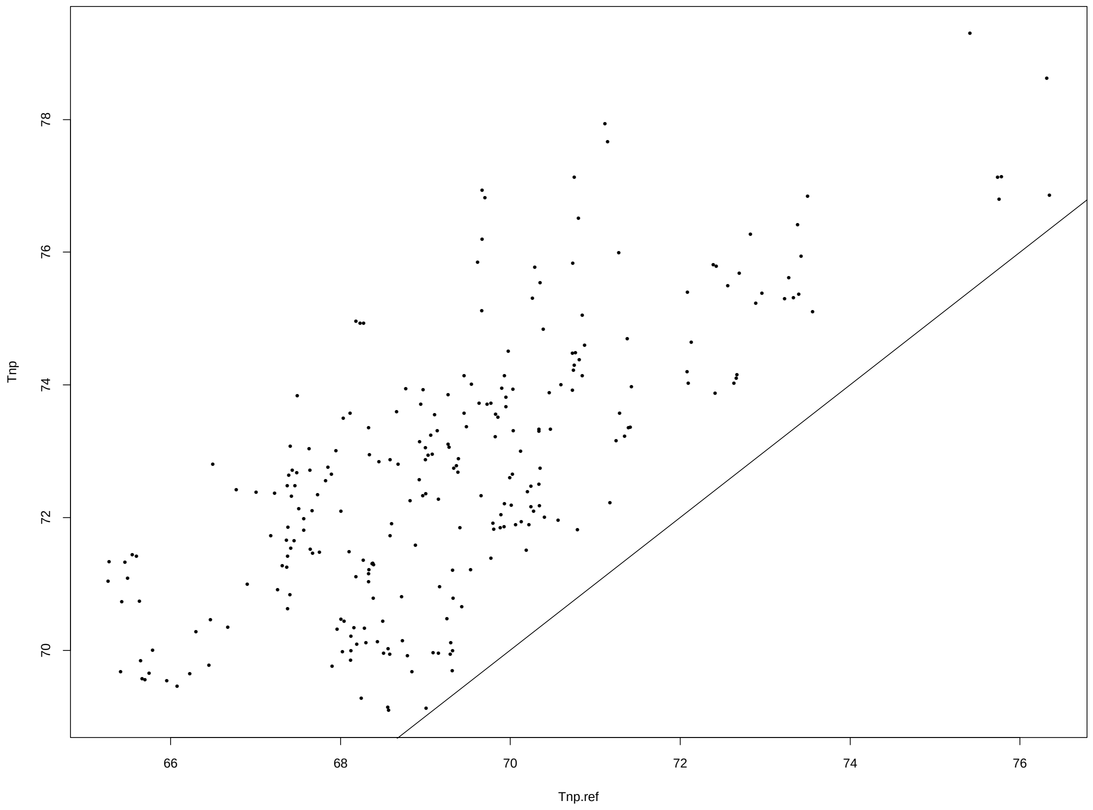

Figure 9.2: *Observed net-point temperature (T_{np}) versus net-point reference temperature (T_{np}^{ref}) for the net-point at Birkelunden (2) during periods, where the supply temperature has been governed by the corresponding net-point temperature controller. For observations above the full line the controller has observed the net-point temperature restrictions.*

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 temperature – 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 determinating 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

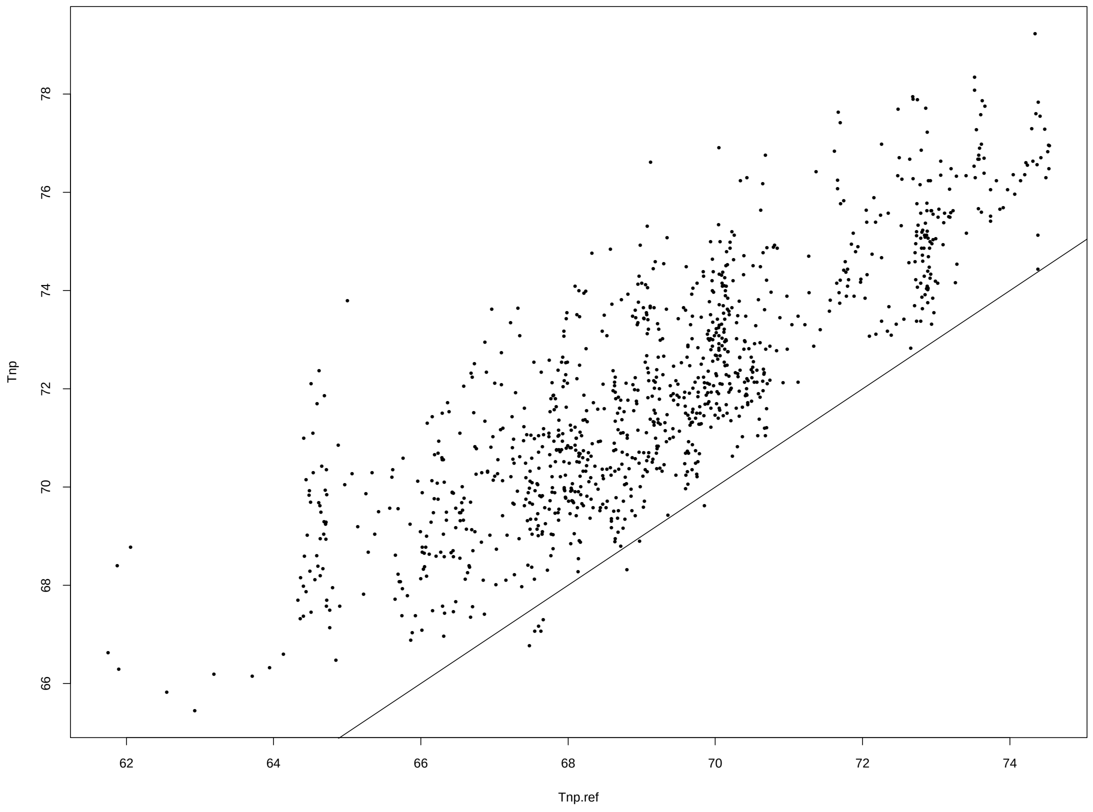

Figure 9.3: *Observed net-point temperature (T_{np}) versus net-point reference temperature (T_{np}^{ref}) for the net-point at Egevangshusene (4) during periods, where the supply temperature has been governed by the corresponding net-point temperature controller. For observations above the full line the controller has observed the net-point temperature restrictions.*

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 has been active. In Figure 9.2 and 9.3 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. & 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.

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. & Søgaard, 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.

Søgaard, 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.

IMM

Informatics and Mathematical Modelling
Technical University of Denmark