Predicting the Heat Consumption in District Heating Systems using Meteorological Forecasts

Henrik Aalborg Nielsen and Henrik Madsen

Department of Mathematical Modelling
Technical University of Denmark
Lyngby
December 2000

IMM

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

Acknowledgements

First of all the authors wish to thank the Danish Energy Agency for financial support of this project under contract 1323/98-0025, Danish Energy Research Program. We also wish to thank Henrik R. Hansen, Vestegnens Kraftvarmeselskab I/S, Albertslund, Svend E. Jensen, Department of Agricultural Sciences, The Royal Veterinary and Agricultural University, Copenhagen, and Jørgen Olsen, Elkraft System, Ballerup for providing the data.

Special thanks to Henrik R. Hansen, Vestegnens Kraftvarmeselskab I/S for also providing us with information about the district heating system and for fruitful discussions and to Torben S. Nielsen at our department for permission to use his S-PLUS library for adaptive estimation. We would also like to thank Jørgen Jensen, Vestegnens Kraftvarmeselskab I/S, Dorthe R. Jensen, Centralkommunernes Transmissionsselskab I/S, and Keld Oksbjerg and Ole J. Olesen, Elkraft A.m.b.A. for fruitful discussions during the startup phase of the project.

Summary

Methods for on-line prediction of heat consumption in district heating systems hour by hour for horizons up to 72 hours are considered in this report. Data from the district heating system Vestegnens Kraftvarmeselskab I/S is used in the investigation. During the development it has been assumed that meteorological forecasts are available on-line. Such a service has recently been introduced by the Danish Meteorological Institute. However, actual meteorological forecasts has not been available for the work described here. Assuming the climate to be known the mean absolute relative prediction error for 72 hour predictions is 3.8% for data in November, 1995 (17% when no climate information is used). However, at some occasions large deviations occur and in January 1996 a value of 5.5% is obtained. The relative prediction error tends to increase with decreasing heat consumption. Approaches to implementation are suggested in a separate chapter of the report.

The methods of prediction applied are based on adaptive estimation, whereby the methods adapt to slow changes in the system. This approach is also used to track the transition from e.g. warm to cold periods. Due to different preferences of the households to which the heat is supplied this transition is smooth. By simulation, combined with theory known from the literature, it is shown that it is crucial to use the actual meteorological forecasts and not the observations of climate when estimating the parameters of the model. To our knowledge, this is somewhat contrary to practice.

The work presented is a demonstration of the value of the so called gray box approach where theoretical knowledge about the system under consideration is combined with information from measurements performed on the system in order to obtain a mathematical description of the system. Furthermore it is also demonstrated that it is important to select the estimation method depending on the particular application. Maximum likelihood estimates are often considered optimal, but here they prove to be inferior to output error estimates for long-term prediction. This is because the optimality of the maximum likelihood estimates are related to the properties of the estimates,

whereas for prediction purposes the properties of the prediction errors should be in focus.

Resumé

Nærværende rapport beskriver metoder til on-line prædiktion af det timevise varmeforbrug i fjernvarmesystemer for tidshorisonter på op til 72 timer. Data fra Vestegnens Kraftvarmeselskab I/S er blevet anvendt i undersøgelsen. Det er en forudsætning for anvendelse af den udviklede metode, at meteorologiske forudsigelser er tilgængelige on-line, og en sådan service er for relativt nylig blevet oprettet af Danmarks Meteorologiske Institut. Dog har faktiske meteorologiske forudsigelser ikke været tilgængelige for arbejdet beskrevet i denne rapport. Når klimaet antages kendt, bliver gennemsnittet af den relative absolutte prædiktionsfejl 3.8% for november 1995 (17% når information om klimaet ikke bruges). Store afvigelser herfra forekommer dog, og f.eks. i januar 1996 fås værdien 5.5%. Den relative prædiktionsfejl tenderer til at stige med aftagende varmeforbrug. Forslag til implementeringer af metoden beskrives i et selvstændigt kapitel af rapporten.

De anvendte prædiktionsmetoder er baseret på adaptiv estimation, hvorved metoden automatisk tilpasses langsomme ændringer i systemet. På denne måde identificeres overgange f.eks. fra varme til kolde perioder også. Ved en kombination af simulation og teori kendt fra litteraturen vises det i rapporten, at det er nødvendigt at bruge de faktiske meteorologiske forudsigelser og ikke de faktiske observationer af klima, når parametrene i modellen estimeres. Såvidt vi ved, er dette i nogen grad ikke i overensstemmelse med praksis.

Det præsenterede arbejde demonstrerer værdien af den såkaldte *gray box* metode, hvor teoretisk viden om det betragtede system kombineres med målinger fra systemet for at opnå en matematisk beskrivelse af dette. Det demonstreres også, at det er vigtigt at vælge estimationsmetode afhængig af den faktiske anvendelse. Maksimum likelihood estimater betragtes ofte som værende optimale. Når langtidsforudsigelser betragtes, viser det sig dog her at være bedre at bruge *output error* estimater. Dette skyldes, at optimaliteten af maksimum likelihood estimater er relateret til disses egenskaber, hvor man af hensyn til prædiktion bør fokusere på prædiktionfejlenes egenskaber.

Contents

Chapter 1

Introduction

This report deals with methods for on-line prediction of the heat load hour by hour in district heating systems where it is assumed that meteorological forecasts are available on-line. Such a service has recently been introduced by the Danish Meteorological Institute under the term "SAFE-Energy" (Petersen & Hilden 1999), where the forecasts are provided via the Internet. The prediction horizons considered in this report range from 1 to 72 hours, but the maximal prediction horizon are determined by a combination of the quality/existence of meteorological forecasts and the requirements on the precision of the predictions of heat load.

For prediction horizons up to 24 hours on-line prediction of heat consumption is implemented in the PRESS software (Madsen & Nielsen 1997). The prediction methods used in PRESS were developed at a time when on-line meteorological forecast were not readily available and therefore the software implicitly generates predictions of the climate variables. However, these predictions do only take into account local climate conditions and the methods used are extremely simple compared to meteorological models. Hence, it is beyond doubt that meteorological forecasts will improve the quality of the predictions of heat consumption.

The purpose of the work described here is to derive methods for on-line prediction of the heat load in district heating systems, which can be implemented on e.g. a normal PC. The predictions are based on information from a SCADA system and meteorological forecasts via the Internet. Furthermore, local observations of climate variables from a climate station might be valuable, especially for short term prediction. The methods must be easy to apply in practice and therefore they should be able to adapt to slow changes in the district heating system. Such methods also imply that the system need

not be calibrated each time it is installed at a new location, but it might be beneficial to tune some parameters related to the adaptive ability.

The approach taken for constructing such methods is first to use relations known from the theory of heat transfer to arrive at an initial model structure. Based on this structure, data in combination with statistical methods is used to from an actual mathematical model of the heat consumption. Finally, based on this model adaptive methods of prediction are investigated.

In Chapter 2 the data used and the preparation of the data are described. The statistical methods are described in Chapter 3. Chapter 4 deals with the initial model structure, which are then used together with the data in Chapter 5 to derive the final model. Adaptive methods of prediction are considered in Chapter 6, while Chapter 7 presents two alternative suggestions for implementation. Finally, in Chapter 8 the results obtained are outlined and discussed.

The majority of readers the will probably be most interested in the conclusion and discussion in Chapter 8 and in the results presented in Chapters 6 and 7 in which the prediction methods are investigated and suggestions for implementation are presented. To obtain a detailed understanding of the work and the selected methods the reader should also read Chapters 4 and 5 in which the basic model structure is investigated. Furthermore, the description of the data in Chapter 2 might be relevant. Readers not familiar with the statistical methods may also want to consult Chapter 3, but it is noted that references to specific sections in this chapter is used extensively in the report. In Section 3.5 simple summary statistics are described. The conclusion and discussion in Chapter 8 lists some possible extensions of the prediction method.

Chapter 2

Data

The data used in this report consists of hourly measurements of heat consumption and climate for the period July, 1995 to June, 1996. Furthermore, data on the types of the individual days are used. The measurements of heat consumption are supplied by Henrik R. Hansen, Vestegnens Kraftvarmeselskab I/S (VEKS) and consists of the heat supplied from the VEKS transmission system to the local distributors over the past hour. The unit of measurement is GJ and since the measurement is related to the past hour the unit GJ/h will be used in this report. The climate measurements consists of recordings of the air temperature in ^oC, wind speed in m/s, and global radiation in W/m^2 as averages over the past ten minutes up to the full hour. The measurements are performed at Højbakkegård in Taastrup by the Department of Agricultural Sciences, The Royal Vet. and Agric. Univ., Copenhagen (Jensen 1996). The wind speed measurements are performed two meters above ground level, which explains the relative low recordings. With the purpose of taking into account the different pattern of consumption for different types of days these are grouped into "working", "half-holy", and "holy" days. Half-holy days includes Saturdays. The data has been supplied by Jørgen Olsen, Elkraft System, Ballerup.

Plots of the raw data are included in Appendix A starting on page 113. The plot of the raw measurements of heat consumption shown in Figure A.1 reveals some problems with the quality of the data. For the climate data shown in Figure A.2 the quality of data seems to be higher. However, on plots such as those just mentioned only gross outliers can be detected in that e.g. a diurnal variation in the measurements will tend to hide some outliers of moderate size. For this reason the data are checked for outliers more thoroughly in Section 2.1, and in general values found to be outliers will be treated as missing values. For some of the methods applied in this report missing

values can not be handled satisfactory therefore series where the missing values (and outliers) are replaced by appropriate values are also generated. This is described in Section 2.2. Finally, in Section 2.3 it is explained how the global radiation is transformed to a variable presumably more adequate for describing the amount of solar radiation entering into the buildings. If not mentioned otherwise the date and time values used in this report refers to the standard time used in Denmark, i.e. one hour ahead of GMT and one hour of daylight savings from the last Sunday in March until the last Sunday in September (October for 1996 and onwards). However, in this chapter the climate variables are investigated using time values in which daylight savings are disregarded.

The method of decomposition mentioned in this chapter is described later in Section 3.3, for the climate variables the diurnal variation is not allowed to vary between the different, previously mentioned, types of days. In general three iterations in the so-called outer loop are allowed for and in none of the cases it is enough to obtain convergence. However, it corresponds to the number of robustness iterations performed by default by the function `lowess` of S-PLUS (Statistical Sciences 1995*a*), which applies exactly the same method for assigning weight to the observations. For each iteration in the outer loop the backfitting iterations are performed until convergence, i.e. until for both the trend and the seasonal component the largest change from one iteration to the next is less than 0.1% of the range of the most recent estimate of the component, cf. (3.18) on page 25. The trend is approximated locally using a window spanning 10 days and a second order polynomial. For the diurnal variation a window spanning 100 days and a first order polynomial is used. In this way relatively fast and irregular variations can be absorbed by the trend estimate. The estimate of the diurnal variation at each full hour and the estimate of trend is calculated at 400 equidistant points, i.e. approximately one point per day, and linear interpolation is used between these points.

2.1 Detection of outliers

2.1.1 Climate

The general approach is to investigate each of the series individually to identify possible outliers. Hereafter the three series are inspected visually and in parallel for the union of these time points and based on this it is decided which values to treat as outliers.

The measurements of air temperature is decomposed into a trend and a diurnal variation which is allowed to change slowly over time. The remainder of this decomposition

seems to be well described by an AR(2) model. This is confirmed by fitting an AR(2) model to the series and estimating the autocorrelation function of the residuals of the fit. However, a time-plot of the residuals shows that the variance is non-constant. To be able to identify outliers based on these residuals they are standardized using a smooth estimate of the standard deviation. This is obtained by smoothing the squared residuals against time using a local quadratic approximation and a nearest neighbour bandwidth corresponding to 60 days of non-missing observations being within the bandwidth. The 1% of the residuals having largest absolute values receive weight zero in the smoothing. The residuals are then divided with the so obtained time-varying standard deviation and finally scaled to have an overall standard deviation of one. For the residuals from the AR(2) model, standardized in the way described, it is decided to treat time points for which the absolute value is larger than four as possible outliers. This corresponds to 40 time points.

The measurements of wind speed are handled as the measurements of air temperature. However, the variance of the residuals from the AR(2) model seems to be approximately constant. For this reason the residuals are just scaled to have unit standard deviation. For 22 time points the absolute value is larger than four and therefore treated as possible outliers.

Decomposition of the measurements of the global radiation seems to be infeasible due to the special characteristics of this climate variable. Instead visual inspection is applied. Disregarding daylight savings, the measurements are plotted against the number of the day in the year for each hour of the day separately. On these plots 23 possible outliers are identified. A similar sequence of plots of the increase in global radiation over the past hour reveals 33 possible outliers. The possible outliers so identified correspond to 50 time points.

Based on the above, 52 separate periods exists for which the neighbour of the series of climate variables are inspected visually. Measurements of air temperature performed at VEKS are used as a guidance during the visual inspection. These measurements are shown in Figure A.3 on page 116. The decisions based on the visual inspection are summarized in Table 2.1. It is seen that only very few outliers are found. The remaining measurements of climate are depicted in Figure 2.1.

2.1.2 Heat consumption

A time-series plot of the raw recordings of heat consumption is shown in Figure A.1 on page 114. Some of the outliers are easily identified and therefore a number of steps

Year	Date	Time	
1995	December 9	13:00 – 20:00	Measurements of wind speed treated as outliers.
1995	December 11	22:00	Measurement of wind speed treated as outliers.
1996	March 28	14:00 – 23:00	Measurements of wind speed treated as outliers.
1996	April 22	02:00 – 03:00	Measurements of air temperature treated as outliers.

Table 2.1: Outliers in measurements of the climate. Daylight savings are disregarded.

are applied. In each step some outliers are identified and the corresponding values are treated as missing values in following steps. The steps can be summarized as

1. Extreme values which can be identified either directly or after removing a trend are regarded as outliers.
2. The log-transformed series is decomposed into trend, seasonal, and remainder components and extreme values are identified manually based on the remainder component.
3. Approximate time points of possible outliers are identified by plotting the heat consumption against the air temperature. Hereafter the actual outliers are identified by inspecting both series in regions around these time points.
4. Finally, the log-transformed series is decomposed again, but with the outliers identified above treated as missing values. The remainder component is well described by an autoregressive model of order 2, but the residual variance seems to be non-constant. For this reason the residuals are standardized using the same method as for the air temperature in Section 2.1.1. Hereafter, the heat consumption data are visually inspected in regions around time points having standardized residuals with an absolute value above 4.

Here details about which or how many observations are identified in each step will not be given. The log-transformation used in steps 2 and 4 is applied since it seems to be a

good way to identify fast peaks, especially occurring during summer. Exactly the same decomposition method and settings as used for the air temperature in Section 2.1.1 is applied in steps 2 and 4, with the exception that the seasonal component is allowed to vary between "working", "half-holy", and "holy" days. For the standardization of the residuals from the AR(2) model the nearest neighbour bandwidth corresponds to 30 days of non-missing observations. This value is smaller than the value used for the air temperature, but had to be used in order to stabilize the variance of the standardized residuals.

In total 354 values are identified as outliers and hence treated as missing values. Originally 56 recordings were missing and hence in total 410 values are missing.

2.2 Replacement of missing values

Figure 2.2 and Table 2.2 summarize the missing values including those arising from the detection of outliers. Only very few climate data are missing, but for the heat consumption near 5% of the data are missing. For many of the methods applied in this report handling of missing values among the climate variables is considerably more complicated than handling of missing values of the heat consumption. This is because climate variables typically are low-pass filtered before they are used as explanatory variables in a model for the heat consumption. Hence, it is crucial to replace the missing values with appropriate ones for the climate data and since only very few are missing the actual values are not very crucial.

Actually, it is only when `proc arima` of SAS/ETS (SAS Institute Inc. 1993) is used in Chapter 5 and when performing on-line selection of prediction method in Chapter 6, that missing values of the heat consumption pose a problem. However, missing values in the heat consumption series is replaced with appropriate ones also, but these are only used in the situations mentioned above.

Air temperature

There are two periods of missing values of air temperature both in 1996; 01:00 – 18:00 on January 27 and 03:00 – 04:00 on April 22. The missing values are replaced with the fitted values when regressing the air temperature on the air temperature measured by VEKS, cf. Figure A.3. For the missing values in January data from the period January

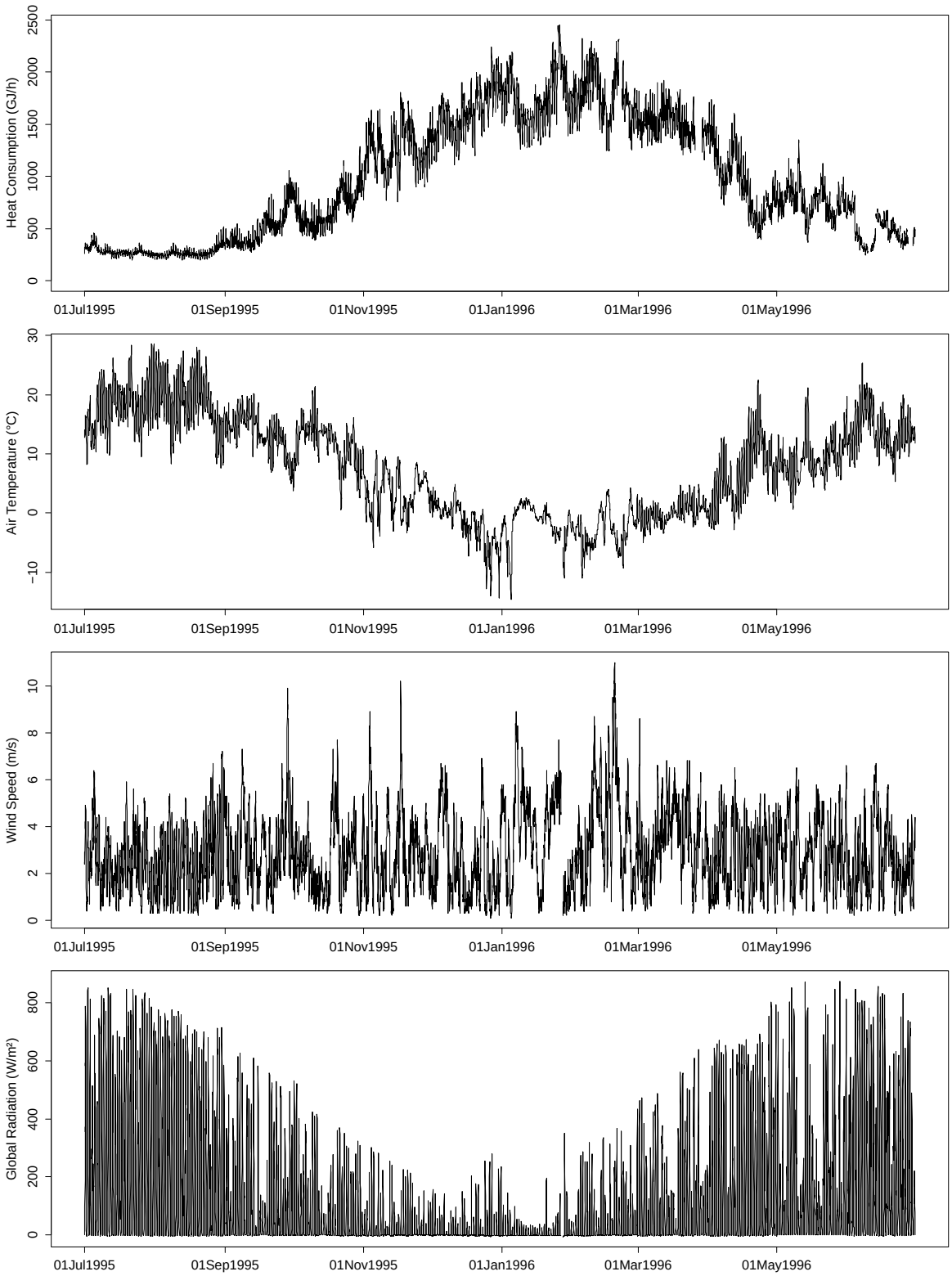

Figure 2.1: Actual data used in the report.

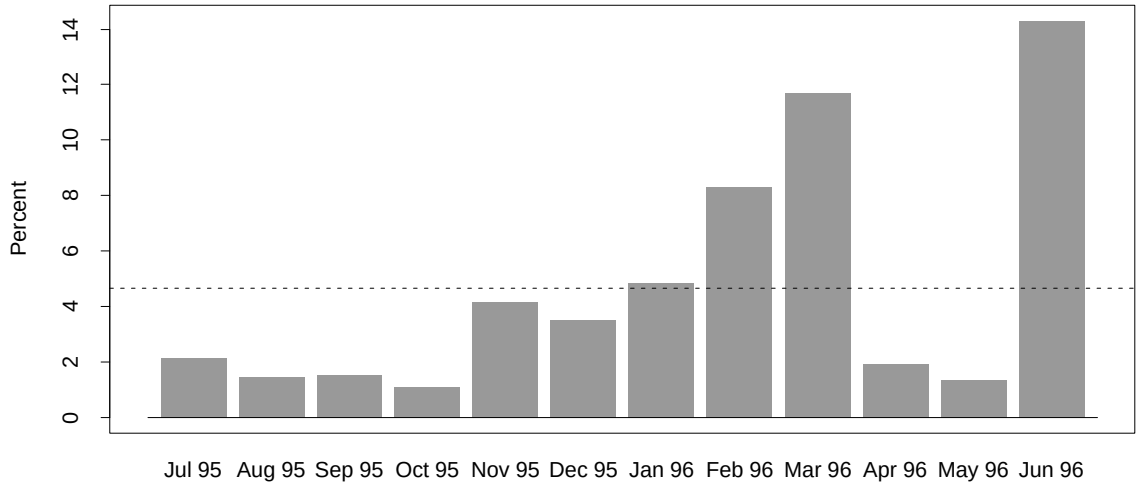

Figure 2.2: Percent missing values of data on heat consumption by month. Outliers are treated as missing values. The dotted line indicate the overall fraction of missing values.

Date	Time
09 Dec 1995	13:00 – 20:00
11 Dec 1995	22:00
27 Jan 1996	01:00 – 18:00
28 Mar 1996	14:00 – 23:00
22 Apr 1996	03:00 – 04:00

Table 2.2: Dates and times for which one or more climate variables are missing.

1 until February 20 is used in the regression since the data are then restricted to a winter situation and since the VEKS recordings seems to be erroneous just after February 20. For the missing values in April data from April and May are used since plots of the air temperature indicate that the spring has started on April 22. For regression the non-parametric method described by Cleveland (1981) and implemented in the function `lowess` in S-PLUS (Statistical Sciences 1995*a*) is used with three robustness iterations and a nearest neighbour bandwidth of 2/3.

Wind Speed

There are four periods of missing values of wind speed; two in 1995; 13:00 – 22:00 on December 9 and 22:00 on December 11, and two in 1996; 01:00 – 18:00 on January 27 and 14:00 – 23:00 on March 28. Since this variable is not clearly related to any of the other variables the missing values are reconstructed by first decomposing the series as described in Section 2.1.1, but with only one iteration in the outer loop since the outliers is treated as missing values. An AR(2) model is then fitted to the remainder series. We would then prefer to use maximum likelihood estimation and the fixed-point smoother by Harvey & Pierse (1984) to reconstruct the missing values in the remainder and then add these to the trend and seasonal components. However, the fixed-point smoother is not implemented in software available to us and we therefore chose to use generalized M-estimates for the autoregressive parameters and approximate conditional mean type robust forward and backward filters as it is implemented in the functions `ar.gm` and `acm.ave` of S-PLUS (Statistical Sciences 1995*a*) to reconstruct the missing values in the remainder component. Since the functions can not handle missing values these are replaced with values equal to the sum of the mean remainder and 1000 times the range of the remainder. Tukey's bisquare weights (Statistical Sciences 1995*a*) are then used in order to remove the influence of these large values.

Global Radiation

Only for 01:00 – 18:00 on January 27, 1996 the global radiation is missing. Plot of the air temperature indicates that on January 27, 1996 it was cloudy. This is also true for January 26, 1996 and therefore the values on January 26 is used to fill the values on January 27.

Heat Consumption

Overall, when outliers are treated as missing values, approximately 5% of the data on heat consumption is missing. In Figure 2.2 the fraction of missing values is shown by month. Most of the missing values corresponds to periods with only a few hours of missing values. To replace the missing values the log-transformed series is decomposed as described in items 2 and 4 on page 6, but since outliers are treated as missing values, with only one iteration in the outer loop. For the short periods of missing values the AR(2) auto correlation structure of the remainder is used to reconstruct the missing values as for the wind speed. There are three periods of length 4 and one period of each of the lengths 5, 8, 12, 31, 32, 48, and 72. In lag 8 the sample auto correlation function of the remainder of the decomposition is 0.19 and therefore we use the method outlined above for periods of length five or lower.

Fortunately, for the remaining periods of missing values, the air temperature is not missing and therefore the relation between heat consumption and air temperature is exploited. Specificly, a conditional parametric model with a separate response for each hour of the day and for working and non-working days is fitted to the data using 2nd order approximations and a nearest neighbour bandwidth of 70%. The residuals of this fit contain a slow irregular variation and therefore they are smoothed against time using a local quadratic regression with a fixed bandwidth of 100 hours (the longest period of missing values is 72 hours). The residuals of this last fit are then reconstructed as for periods of length five or lower, after it is verified that an AR(2) model is appropriate. Hereafter, the missing values for periods of length eight or more are reconstructed by adding the two series of fitted values to the reconstructed residuals.

2.3 Solar radiation on building walls

The measurement of global radiation is, loosely speaking, the solar radiation on a horizontal plane. For predicting the heat consumption the solar radiation hitting the walls of the houses is presumably more adequate. Here a square pillar facing the four quarters of the globe will be considered. For each of the walls the solar radiation per square meter is calculated as the sum of the direct, diffuse, and reflected contributions. The average over the four walls is used as a measure of the contribution of the solar radiation to the heating of the buildings. The coefficient of reflection is taken to be 0.2 since this is correct for most surfaces (Hansen, Kjerulf-Jensen & Stampe 1987, p. 48).

The calculations are complicated by the fact that no measurements of both direct and diffuse radiation are available. Instead the cloud cover is calculated using the measurements of global radiation and an equation (Madsen 1985) describing the dependence of global radiation on cloud cover. The cloud cover is taken to be the integer value between zero and eight resulting in the best match. In some sense this solves the problem of multiple solutions, but it also reveals that the determination of cloud cover is somewhat uncertain. Hereafter the direct and the diffuse radiation is calculated using equations by Madsen (1985). Quantities such as the height of the sun and the angle of incidence is calculated based on the equations in (Øivin Holter, Ingebretsen & Parr 1979, pp. 45-46). It is noted that the declination is calculated as an approximation. The calculations are performed for the geographical location of Taastrup in Denmark (latitude: 55.67, longitude: 12.32).

Chapter 3

Statistical Methods

This chapter describes the most of the methods used in the analysis of the corrected data. The methods used for data correction are mainly referred to in Chapter 2. Parametric models, i.e. models in which the properties of the response (heat consumption in this case) is assumed to be controlled by a finite set of parameters, are described in Section 3.1. Often it is difficult to specify an appropriate parametric structure. In this case non- or semi-parametric models can be applied, such models are described in Section 3.2 and in Section 3.3 a related method for decomposition of time series is described. In general statistical properties of the estimates such as confidence intervals or t-tests are not applied in this report. Therefore only estimation in the models are described, the exception being that a likelihood ratio test is described in the end of Section 3.1.

For generation of predictions on-line it is desirable to let the estimates adapt to slow changes in the system. For this purpose the adaptive recursive least squares method with exponential forgetting is applied. This method is described in Section 3.4. Some simple summary statistics are listed in Section 3.5. With the purpose of evaluating different models or selecting a smoothing parameter K-fold cross validation can be used. This method is described in Section 3.6. Finally, in Section 3.7 it is described how sample correlation functions and sample cross correlation functions are calculated in case of missing observations.

3.1 Parametric Models

Below we consider least squares estimation in linear and non-linear regression models and maximum likelihood estimation in linear regression models where the model error is an ARMA process. Furthermore, the concept of transfer functions are described and, finally, a likelihood ratio test is introduced.

Linear regression models

The linear regression model is one of the models used in statistics for which most is known about the properties of the estimates etc., see e.g. (Myers 1986, Jørgensen 1993). The term "linear" refers to the fact that, under the model, the expected value of the response y_s, where s denotes the observation number, is a linear combination of some, possibly transformed, deterministic explanatory variables $\mathbf{x}$

$$y_s = \mathbf{x}_s^T \boldsymbol{\theta} + e_s, \tag{3.1}$$

where $\boldsymbol{\theta}$ is the parameter vector and e_s is the model error. When the model errors is independently identical distributed and Gaussian with zero mean, the estimates $\hat{\boldsymbol{\theta}}$ presented below can be considered maximum likelihood estimates (Jørgensen 1993). If the assumptions of independent and/or Gaussian observations are dropped the estimates are termed least squares estimates and the estimates are still unbiased.

Given N observations numbered $s = 1, 2, \ldots, N$, we introduce the design matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix}$$

and the vector of responses

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

Assuming that $\mathbf{X}$ has full rank the estimates of $\boldsymbol{\theta}$ can be written

$$\hat{\boldsymbol{\theta}} = [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}\mathbf{y}. \tag{3.2}$$

These estimates minimize the sum of the squared residuals $\sum_{s=1}^{N}(y_s - \mathbf{x}_s^T\boldsymbol{\theta})^2$ and correspondingly they are denoted least squares estimates. Note that for the actual numerical calculations (3.2) should not be applied directly. One solution is to use methods such as those presented by Miller (1992). However, most software packages for statistics use appropriate numerical methods.

As noted above the linear regression model does not imply that the response on the explanatory variables is linear. A simple example of this is the following model where the response y is measured with noise and evolves over time t as a first order harmonic with *known* period length, set to 2π for simplicity

$$y_s = a_0 + a_1\cos(t_s) + a_2\sin(t_s) + e_s.$$

This is a linear regression model and, based on observations, the parameters a_0, a_1, and a_3 can be estimated as described above with $\mathbf{x} = [1\ \cos(t_s)\ \sin(t_s)]^T$ and $\boldsymbol{\theta} = [a_0\ a_1\ a_3]^T$. Simple as it is, this example points towards much richer ways of generating the design matrix $\mathbf{X}$. If the data is generated by the system

$$y_s = f(u_s) + e_s$$

the function $f(\cdot)$ can be estimated under the assumption that it has continuous derivatives up to order two by approximating it with a piecewise cubic polynomial with continuous derivatives up to order two. This can be accomplished by using (3.2) with

$$\mathbf{x}_s = \begin{bmatrix} 1 \\ u_s \\ u_s^2 \\ u_s^3 \\ (u_s - u_{(1)})_+^3 \\ (u_s - u_{(2)})_+^3 \\ \vdots \\ (u_s - u_{(K)})_+^3 \end{bmatrix},$$

where the subscript "+" denotes truncation of negative values and $u_{(j)};\ j = 1, 2, \ldots, K$ is the points at which the third order derivative is allowed to change. These points are called the knots and the basis defined by $\mathbf{x}_s$ is called a cubic spline. Quadratic splines are defined similarly, except that the discontinuity occurs in the second order derivatives. A B-spline basis will provide the same estimate of $f(\cdot)$ and numerically it is more stable (de Boor 1978). Also, using a B-spline basis, it is possible to estimate the function $f(\cdot)$ under the additional assumption that it is periodic with a specified period length (de Boor 1978).

It is noted that by use of spline bases one or more of the elements in $\boldsymbol{\theta}$ in (3.1) may be replaced by functions which can be estimated together with the remaining parameters by use of the methods outlined above.

Non-linear regression models

Opposed to linear regression models, in non-linear regression models the *parameters* enter the model in a non-linear way so that the expected value of the response y can not be written as a linear combination of some, possibly transformed, explanatory variables $\mathbf{x}$. Instead, the expected value of the response is a function of the explanatory variables. Except for the value of the parameters $\boldsymbol{\theta}$ the function is known. This may be expressed as

$$y_s = f(\mathbf{x}_s, \boldsymbol{\theta}) + e_s. \tag{3.3}$$

Often e_s is assumed to be independently identical distributed with zero mean. Below estimation in (3.3) is outlined, for further reading see e.g. (Gallant 1987).

Given N observations numbered $s = 1, 2, \ldots, N$, the parameters $\boldsymbol{\theta}$ are estimated by least squares, if the estimates $\hat{\boldsymbol{\theta}}$ are chosen to minimize

$$V(\hat{\boldsymbol{\theta}}) = \sum_{s=1}^{N} (y_s - f(\mathbf{x}_s, \hat{\boldsymbol{\theta}}))^2, \tag{3.4}$$

i.e. in principle the same criterion as is used for the linear regression model. However, in this case a closed-form solution does not exist. Therefore starting values of the estimates must be supplied, where after a general optimization routine is applied with the purpose of finding the values $\hat{\boldsymbol{\theta}}$ which minimize the criteria above. However, in general, only a local minima can be guaranteed. In this report the function `nlminb` in S-PLUS 3.4 is applied (Statistical Sciences 1995*a*). This function applies the algorithm described by Gay (1983).

ML Estimation in Regression Models with correlated errors

If model (3.1) are used when the observations s are obtained sequentially in time t the noise will often be auto-correlated. In this case the least squares estimates are unbiased but have increased variance. Furthermore, the least squares estimates provide no estimates corresponding to the correlation structure of the noise. One solution is

to use maximum likelihood (ML) estimation under an assumption of a particular noise model. In this report models of the following type is used:

$$y_t = \mathbf{x}_t^T \boldsymbol{\theta} + z_t, \tag{3.5}$$

where observations are indexed by t to indicate that they are obtained sequentially in time. $\{z_t\}$ is a zero mean, stationary, and invertible $ARMA(p, q)$ process, i.e.

$$A(q^{-1})z_t = C(q^{-1})e_t, \tag{3.6}$$

in which $A(q^{-1})$ and $C(q^{-1})$ are polynomials of order p and q, respectively, in the backward shift operator q^{-1}. Both polynomials may be replaced by products of polynomials and some of the parameters in the polynomials may be fixed at zero. The process $\{e_t\}$ is assumed to be zero mean Gaussian white noise. The (complex) solutions to $A(q^{-1}) = 0$ and $C(q^{-1}) = 0$ should all have modulus less than one in order to ensure stationarity and invertibility (Brockwell & Davis 1986).

In this report the ML estimates under the model are calculated using the S-PLUS function `arima.mle` (Statistical Sciences 1995*a*, Section 17.3). For the data considered in this report y_t is missing for some t and therefore a state space formulation is used together with the Kalman filter in order to evaluate the likelihood function (Kohn & Ansley 1986, Madsen 1995). Since, for the data considered, there is no missing observations in the beginning of the series the Kalman filter recursions are conditioned on the first p observations. Furthermore, the regression parameters are concentrated out of the likelihood (Kohn & Ansley 1985) and the estimates of the parameters in the polynomials are transformed to ensure stationarity and invertibility (Jones 1980).

Transfer functions

Transfer functions are related to the mathematical description of a dynamical system and essentially describe how the input process $\{x_t\}$ is filtered to produce some output $\{y_t\}$. One example of a transfer function is the dependence of indoor air temperature on the outdoor temperature. This dependence is dynamic in that, during the heating season, only slow variations in the outdoor temperature are reflected in the indoor temperature or in the heat needed to maintain a particular indoor temperature. In the following rational transfer functions for discrete time models are described.

A rational transfer function can be written

$$H(q) = q^{-d}\frac{B(q^{-1})}{A(q^{-1})} \tag{3.7}$$

where $q^{-d}x_t = x_{t-d}$, and d is an integer denoting a delay. $A(q^{-1})$ and $B(q^{-1})$ are polynomials in the backward shift operator q^{-1}, i.e.

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} \\ B(q^{-1}) &= b_0 + b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \end{aligned}$$

Thus the notation

$$y_t = H(q)x_t$$

refers to the recursions

$$\begin{aligned} y_t &= -a_1 y_{t-1} - \ldots - a_{n_a} y_{t-n_a} \\ &\quad + b_0 x_{t-d} + b_1 x_{t-d-1} + \ldots + b_{n_a} x_{t-d-n_b}. \end{aligned}$$

The transfer function is stable, in the sense that a limited input results in a limited output, if the solutions to

$$z^{n_a} + a_1 z^{n_a - 1} + \ldots + a_{n_a} = 0$$

lies strictly within the unit circle in the complex plane, i.e. if the solutions all fulfill $|z| < 1$. The stationary gain, also called the DC-gain, of the transfer function is

$$H(1) = \frac{\sum_{i=0}^{n_b} b_i}{1 + \sum_{i=1}^{n_a} a_i}.$$

See (Ljung 1987) or (Madsen 1995) for a more thorough description of the subject. Transfer functions can be used in connection with linear regression models or models as defined by (3.5) and (3.6), whereby the parameters of the transfer functions can be estimated from data, see e.g. (Abraham & Ledolter 1983, Ljung 1987, Madsen 1995).

Testing hypothesis

When applying statistical tests the support in the data for different hypotheses are investigated. This is done by formulating a null hypothesis H_0 and an alternative hypothesis H_1 about the data. Normally, H_0 should be a statement about parameters in H_1. For instance, two "treatments" A and B may be assigned randomly to a number of experimental units, where after some response is measured on each unit. Hereafter it is desirable to investigate if a difference in the effects of "treatments" on the expected responses μ_A and μ_B of an experimental unit can be detected. This can be formulated as a null hypothesis in which $\mu_A = \mu_B$ and an alternative hypothesis in which $\mu_A \neq \mu_B$. The hypothesis testing then proceeds by calculating a quantity called

the test statistic, which, assuming H_0 to be correct, follows a particular distribution. Then, loosely speaking, if the test statistic is unlikely to have originated from the particular distribution H_0 is rejected and the expected values are said to be significantly different. Strictly, if H_0 is not rejected it can not be inferred that $\mu_A = \mu_B$ since the procedure outlined above does not control the error rate of accepting H_0 when H_1 is true. In this case it is recommended to calculate a confidence interval for the difference in means $\mu_A - \mu_B$. For further reading see e.g. (Moses 1986, Chapters 5 and 6).

In this report likelihood ratio tests (Jørgensen 1993, p. 32) are used to test hypotheses about $\boldsymbol{\theta}$ in the model (3.5) and (3.6). Let H_1 be the hypothesis corresponding to the full model and let H_0 be the hypothesis corresponding to a subset of the full model. Furthermore, let L_1 and L_0 be the corresponding likelihoods as calculated by `arima.mle` (Statistical Sciences 1995*a*) and let n_1 and n_0 be the number of parameters in the two models. The likelihood ratio statistic is then defined as

$$G = 2\log(L_1/L_0), \tag{3.8}$$

where, under H_0, the asymptotic distribution of the statistic is a χ^2 distribution with $n_1 - n_0$ degrees of freedom. Thus, if $G > \chi^2_{1-\alpha}(n_1 - n_0)$, where α is the level of significance and $\chi^2_{1-\alpha}(n_1 - n_0)$ is the $1 - \alpha$ quantile in the χ^2 distribution, then H_0 is rejected. Before the test is applied the fulfillment of the assumptions for maximum likelihood estimation in the model corresponding to H_1 should be investigated.

3.2 Non- and Semi-parametric models

In pure non-parametric regression the observations y_s are assumed to be generated by a system having the following mean structure:

$$E[y_s] = f(\mathbf{x}_s), \tag{3.9}$$

where $\mathbf{x}_s$ is a vector containing the explanatory variables corresponding to observation s. Several methods are available for estimation in (3.9), see e.g. (Hastie & Tibshirani 1990). For the application considered in this report the number of explanatory variables is high, making estimation in (3.9) difficult, see e.g. (Hastie & Tibshirani 1990, pp. 83-84). For this reason and since we have some approximate knowledge of the structure of the appropriate models, cf. Chapter 4, the models described below are applied.

Conditional parametric models are linear regression models in which the parameters are replaced by functions of one or more explanatory variables. This kind of models are

also called varying-coefficient models, but here we prefer to reserve this term to models in which the coefficient-functions do not share the same argument(s) as in (Hastie & Tibshirani 1993). Estimation in conditional parametric models may be accomplished by use of a method similar to local regression (Cleveland & Devlin 1988). The method is described below.

If *some* of the coefficient-functions in a conditional parametric model are required to be constant across all values of the argument(s) the resulting model is called a semi-parametric model as by Hastie & Tibshirani (1990, p. 118). Estimation in semi-parametric models are also described in this section. Note that other authors term other types of models as "semi-parametric" also, but here the terminology of (Hastie & Tibshirani 1990) will be followed.

Conditional parametric models

Below estimation in conditional parametric models is described. The model is of the form

$$y_s = \mathbf{x}_s^T \boldsymbol{\theta}(\mathbf{u}_s) + e_s; \quad s = 1, \ldots, N, \tag{3.10}$$

where the response y_s is a stochastic variable, $\mathbf{u}_s$ and $\mathbf{x}_s$ are explanatory variables, e_s is i.i.d. $N(0, \sigma^2)$, $\boldsymbol{\theta}(\cdot)$ is a vector of unknown but smooth functions with values in $\mathbb{R}^p$, and $s = 1, \ldots, N$ are the observation numbers. When $\mathbf{u}_s$ is constant across the observations the model reduces to an ordinary parametric linear model.

Estimation in (3.10) aims at estimating the functions $\boldsymbol{\theta}(\cdot)$ within the space spanned by the observations of $\mathbf{u}_s; s = 1, \ldots, N$. The functions are only estimated for distinct values of the argument $\mathbf{u}$. Below $\mathbf{u}$ denotes one of these fitting points and $\hat{\boldsymbol{\theta}}(\mathbf{u})$ denotes the estimates of the coefficient-functions, when the functions are evaluated at $\mathbf{u}$.

One solution to the estimation problem is to replace $\boldsymbol{\theta}(\mathbf{u}_s)$ in (3.10) with a constant vector $\boldsymbol{\theta}_u$ and fit the resulting model locally to $\mathbf{u}$, using weighted least squares. Below two similar methods of allocating weights to the observations are described, for both methods the weight function $W : \mathbb{R}_0 \to \mathbb{R}_0$ is a nowhere increasing function, $\mathbb{R}_0$ denotes the non-negative real numbers. In this report the tricube weight function

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

is used. Hence, $W : \mathbb{R}_0 \to [0, 1]$.

In the case of a so-called spherical kernel the weight on observation s is determined by the Euclidean distance $||\mathbf{u}_s - \mathbf{u}||$ between $\mathbf{u}_s$ and $\mathbf{u}$, i.e.

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

A so-called product kernel is characterized by distances being calculated for one dimension at a time, i.e.

$$w_u(\mathbf{u}_s) = \prod_j W\left(\frac{|u_{j,s} - u_j|}{\hbar(\mathbf{u})}\right),$$

where the multiplication is over the dimension of $\mathbf{u}$. The scalar $\hbar(\mathbf{u}) > 0$ is called the bandwidth. If $\hbar(\mathbf{u})$ is constant for all values of $\mathbf{u}$ it is denoted a fixed bandwidth. If $\hbar(\mathbf{u})$ is chosen such that a certain fraction (α) of the observations fulfill $||\mathbf{u}_s - \mathbf{u}|| \leq \hbar(\mathbf{u})$ then α is denoted a nearest neighbour bandwidth. If $\mathbf{u}$ has dimension of two or larger, scaling of the individual elements of $\mathbf{u}_s$ before applying the method should be considered, see e.g. (Cleveland & Devlin 1988). Rotating the coordinate system in which $\mathbf{u}_s$ is measured may also be relevant.

If the bandwidth $\hbar(\mathbf{u})$ is sufficiently small the approximation of $\boldsymbol{\theta}(\cdot)$ as a constant vector near $\mathbf{u}$ is good. However, this implies that a relatively low number of observations is used to estimate $\boldsymbol{\theta}(\mathbf{u})$, resulting in a noisy estimate or large bias if the bandwidth is increased. See also the comments on kernel estimates in (Anderson, Fang & Olkin 1994).

It is, however, well known that locally to $\mathbf{u}$ the elements of $\boldsymbol{\theta}(\cdot)$ may be approximated by polynomials, and in many cases these will be good approximations for larger bandwidths than those corresponding to local constants. Local polynomial approximations are easily included in the method described. Let $\theta_j(\cdot)$ be the j'th element of $\boldsymbol{\theta}(\cdot)$ and let $\mathbf{p}_{d(j)}(\mathbf{u})$ be a column vector of terms in the corresponding d-order polynomial evaluated at $\mathbf{u}$, if for instance $\mathbf{u} = [u_1\ u_2]^T$ then $\mathbf{p}_2(\mathbf{u}) = [1\ u_1\ u_2\ u_1^2\ u_1u_2\ u_2^2]^T$. Furthermore, let $\mathbf{x}_s = [x_{1s} \dots x_{ps}]^T$. With

$$\mathbf{z}_s^T = \left[x_{1s}\mathbf{p}_{d(1)}^T(\mathbf{u}_s) \dots x_{js}\mathbf{p}_{d(j)}^T(\mathbf{u}_s) \dots x_{ps}\mathbf{p}_{d(p)}^T(\mathbf{u}_s)\right] \tag{3.11}$$

and

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

where $\hat{\boldsymbol{\phi}}_j(\mathbf{u})$ is a column vector of local constant estimates at $\mathbf{u}$ corresponding to $x_{js}\mathbf{p}_{d(j)}(\mathbf{u}_s)$, estimation is handled as described above, but fitting the linear model

$$y_s = \mathbf{z}_s^T\boldsymbol{\phi}_u + e_s; \quad i = 1, \dots, N, \tag{3.12}$$

locally to $\mathbf{u}$, indicated by the subscript on the parameter-vector. Hereafter the elements of $\boldsymbol{\theta}(\mathbf{u})$ are estimated by

$$\hat{\theta}_j(\mathbf{u}) = \mathbf{p}_{d(j)}^T(\mathbf{u})\,\hat{\boldsymbol{\phi}}_j(\mathbf{u}); \quad j = 1, \ldots, p. \tag{3.13}$$

When $\mathbf{x}_s = 1$ for all s, i.e. $p = 1$, this method is identical to the method by Cleveland & Devlin (1988), with the exception that they center the elements of $\mathbf{u}_s$ used in $\mathbf{p}_{d(j)}(\mathbf{u}_s)$ around $\mathbf{u}$ and hence $\mathbf{p}_{d(j)}(\mathbf{u}_s)$ must be recalculated for each value of $\mathbf{u}$ considered.

The local constant (intermediate) estimates for arguments $\mathbf{u}_s;\ s = 1, 2, \ldots, N$ can be expressed as

$$\hat{\boldsymbol{\phi}}(\mathbf{u}_s) = [\mathbf{Z}^T\mathbf{W}_s\mathbf{Z}]^{-1}\mathbf{Z}^T\mathbf{W}_s\mathbf{y}, \tag{3.14}$$

where $\mathbf{Z}$ is the design matrix corresponding to the local constant estimates, i.e. the rows of $\mathbf{Z}$ consists of (3.11). $\mathbf{W}_s$ is a diagonal weight matrix corresponding to the fitting point $\mathbf{u}_s$

$$\mathbf{W}_s = \text{diag}\{w_{u_s}(\mathbf{u}_1), w_{u_s}(\mathbf{u}_2), \ldots, w_{u_s}(\mathbf{u}_N)\}$$

and $\mathbf{y}$ is a vector containing the observations of the response. From (3.14) and (3.12) it is clear that the fitted values can be expressed as

$$\hat{y}_s = \mathbf{z}_s^T\hat{\boldsymbol{\phi}}(\mathbf{u}_s) = \mathbf{z}_s^T[\mathbf{Z}^T\mathbf{W}_s\mathbf{Z}]^{-1}\mathbf{Z}^T\mathbf{W}_s\mathbf{y}. \tag{3.15}$$

Consequently the fitted values are linear combinations of the observations $\mathbf{y}$, unless $\mathbf{y}$ is used to select the weights in $\mathbf{W}_s$. The fitted (smoothed) values $\hat{\mathbf{y}}$ can be written

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}, \tag{3.16}$$

where $\mathbf{S}$ is called the smoother matrix and when $\mathbf{S}$ does not depend on $\mathbf{y}$ the smoother is called a linear smoother (Hastie & Tibshirani 1990). When nearest neighbour or fixed bandwidths are used for estimation in conditional parametric models as described above the resulting smoother is linear. Note that $\mathbf{S}$ is equivalent to the hat-matrix in linear regression (Jørgensen 1993).

Semi-parametric models

A semi-parametric model in the sense used in this report is a mix between a linear regression model and a conditional parametric model, i.e.

$$y_s = \mathbf{z}_s^T\boldsymbol{\phi} + \mathbf{x}_s^T\boldsymbol{\theta}(\mathbf{u}_s) + e_s; \quad s = 1, \ldots, N, \tag{3.17}$$

where $\mathbf{z}$ is a vector of explanatory variables for which the corresponding coefficients are global constants and where $\mathbf{x}$ is a vector of explanatory variables for which the corresponding coefficients depend on one or more explanatory variables $\mathbf{u}$.

As shown above, if estimation in a conditional parametric model is performed as described above, the resulting smoother is linear, in the sense that the fitted values are linear combinations of the observations of the response. For this reason the results described by Hastie & Tibshirani (1990, p. 118) apply to estimation in (3.17), when the method described in the previous subsection is used for estimation of the conditional parametric part of the model.

Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

and

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \vdots \\ \mathbf{z}_N^T \end{bmatrix}$$

and let $\mathbf{S}$ denote the smoother matrix corresponding to the conditional parametric model and the estimation method used. An estimate of the global parameters $\boldsymbol{\phi}$ is then

$$\hat{\boldsymbol{\phi}} = [\mathbf{Z}^T(\mathbf{I}-\mathbf{S})\mathbf{Z}]^{-1}\mathbf{Z}^T(\mathbf{I}-\mathbf{S})\mathbf{y}.$$

Note that $(\mathbf{I}-\mathbf{S})\mathbf{Z}$ can be obtained as the residuals after smoothing each column of $\mathbf{Z}$. Hereafter, estimates of the coefficient-functions can be obtained applying the smoother to $\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\phi}}$.

Note also that if both $\mathbf{z}$ and $\mathbf{x}$ has one element which is constant across the observations the estimates are not uniquely defined. In this case we will require that $\mathbf{x}_s^T\hat{\boldsymbol{\theta}}(\mathbf{u}_s)$ sums to zero over the observations. This can be accomplished by applying the conditional parametric model in the usual way, where after the average of the fitted values are subtracted from the fitted values. It is noted that the average of the fitted values is, like the fitted values them self, a linear combination of the response to which the conditional parametric model is applied. For this reason the properties of the smoother is preserved.

3.3 Decomposing Time Series

This section describes a method for decomposition of time series in a trend, seasonal, and remainder components. A method was chosen that results in smooth estimates of the trend component and allows for non-stationary, but smooth, changes in the seasonal component. Furthermore, the method allows for different seasonal components for different types of days.

The method is very similar to the STL procedure (Seasonal-Trend decomposition procedure based on Loess) described in (Cleveland, Cleveland, McRae & Terpenning 1990). The STL procedure uses locally weighted polynomial regression smoothers iteratively to decompose a time series. For readers familiar with the STL procedure we mention that the difference is that we apply a convergence criterion in both the inner and outer loop. Furthermore, we do not perform the low-pass filtering of the smoothed cycle-subseries and the post smoothing of the seasonal component. Finally, due to the type of data, we must allow for different seasonal components for different type of days (weekends, working days).

The algorithm consists of an outer and an inner loop. At startup an estimate of the trend component is calculated by smoothing the time series plot of the data, using a bandwidth well above the period length of the seasonal component and ignoring this component. The outer loop down weights observations with large residuals, exactly as described in Section 2.4 of (Cleveland et al. 1990). The inner loop is essentially the back-fitting algorithm (Hastie & Tibshirani 1990). It is implemented as follows

Step 1 Subtract the trend component and do seasonal smoothing (see below) on the result.

Step 2 Detrend the seasonal component by smoothing the time series plot of the seasonal with the same smoother as used to estimate the trend component and subtracting the smooth from the seasonal obtained in step 1.

Step 3 Subtract the seasonal component obtained in step 2 from the data and smooth the resulting time series as for the initial estimate of the trend component.

Step 4 Check for convergence and go to step 1 if convergence has not been obtained.

Let x_t denote the original time series with an estimate of the trend component subtracted. Let $h(t)$ denote the time of day at the running time t, and let $d(t)$ denote the

type of day corresponding to t. The seasonal smoothing is accomplished by smoothing (t, x_t) for each individual element of the pair $(h(t), d(t))$, i.e. no smoothing is performed over neighbouring values of $h(t)$. Consequently, the seasonal component is allowed to change slowly.

For smoothing locally weighted polynomial regression (Cleveland & Devlin 1988) (with fixed bandwidth, tricube window, and linear interpolation between equally spaced points) is used. Thus, the estimates are mainly determined by (i) the span in days of the windows used for the trend and the seasonal smoother, (ii) degrees of the polynomials fitted locally, and (iii) a user defined grouping of days. Furthermore, the user might decide not to distinguish between days for some of the 24 hours of the diurnal cycle. As long as the number of equally spaced points in which to calculate the smooth is not too low it is of minor importance. We have found that 50 are appropriate in most situations. However, small window spans require more points.

Let $\widehat{C}_{new}(t_j)$ be the most resent estimate of the component (seasonal or trend) evaluated in the fitting point t_j and let $\widehat{C}_{old}(t_j)$ be the corresponding value at the second most resent iteration. For both seasonal and trend components the following value is calculated

$$\frac{\max_j\{|\widehat{C}_{new}(t_j) - \widehat{C}_{old}(t_j)|\}}{\max_j\{\widehat{C}_{new}(t_j)\} - \min_j\{\widehat{C}_{new}(t_j)\}} \tag{3.18}$$

and when both of these are below a certain value the iterations in the inner loop are stopped. Unless otherwise stated a value of 0.001, corresponding to a maximum change of 0.1%, is used. For the outer loop max in the numerator of (3.18) is replaced by a quantile calculation, where by default the 99% quantile is used.

3.4 Adaptive Estimation

Normally, when estimating the parameters of a model based on a set of observations one set of estimates are obtained. However, in on-line situations new observations arrive at (approximately) known time points and it is thus desirable to update the estimates using this new data. Also, it may be desirable to down weight, and eventually exclude, old data. Reasons for this includes (i) the basic system changes slowly over time, e.g. due to more houses being connected to the district heating network and (ii) some parts of the system which are not modelled induces *slow* changes in the estimates corresponding to the best approximation. An example of (ii) is the transition from the summer- to the winter-situation. During the summer period people turn

off their heating equipment and during the early autumn people gradually start to turn the equipment on again. Due to different "thresholds" the transition for a large group of houses is smooth. Results presented in this report show that the transition occurs in a way that, for overlapping temperature intervals, the slope on plots of heat consumption versus temperature is markedly different for August and September in 1995, cf. Figure 5.2 on page 45.

In this report adaptive recursive estimation, with exponential forgetting, in linear regression models and autoregressive models is used. The method is described by Ljung & Söderström (1983). Using the method the estimates at time t for model (3.1) is obtained as

$$\hat{\boldsymbol{\theta}}_t = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{s=1}^{t} \lambda^{t-s}(y_s - \mathbf{x}_s^T\boldsymbol{\theta})^2, \tag{3.19}$$

where $0 < \lambda < 1$ is called the forgetting factor. Thus, the method is a special case of weighted linear regression. Auto regressive models can also be handled using this concept. For instance the parameters μ and ϕ_{24} in the model $y_t = \mu + \phi_{24}y_{t-24} + e_t$, where e_t is independently identical distributed random variables with zero mean, can be estimated adaptively using $\mathbf{x}_s = [1\ y_{t-24}]^T$ and $\boldsymbol{\theta} = [\mu\ \phi_{24}]^T$.

The estimates (3.19) can be calculated using the updating formulas

$$\mathbf{R}_t = \lambda\mathbf{R}_{t-1} + \mathbf{x}_t\mathbf{x}_t^T \tag{3.20}$$

and

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \mathbf{R}_{u,t}^{-1}\mathbf{x}_t\left[y_t - \mathbf{x}_t^T\hat{\boldsymbol{\theta}}_{t-1}\right]. \tag{3.21}$$

Initial estimates must be supplied. In the work described in this report a vector of zeros is used, although in on-line situations a qualified guess should be applied to avoid large prediction errors at startup. Similarly, the estimates can not be updated until the matrix $\mathbf{R}_t$ is non-singular. Here we do not try to invert $\mathbf{R}_t$ until information corresponding to 14 days of hourly observations is included in the matrix.

In general, after the initial period, it can not be guaranteed that $\mathbf{R}_t$ will stay non-singular at every point in time. However, in the work described here no such problems are encountered, but in on-line situations methods to deal with the problem should be implemented. Solutions to the problem include (i) not to update the estimates if $\mathbf{R}_t$ is close to being singular and (ii) use a formulation of (3.20) and (3.21) in which no matrix inversion is required (Ljung 1987). However, using method (ii) may produce highly inadequate estimates in situations corresponding to $\mathbf{R}_t$ being singular. A general solution to the problem is to use a method called selective forgetting (Parkum 1992) in

which, loosely speaking, only the parts of $\mathbf{R}_t$ in which new information becomes available is updated with new information and correspondingly old information forgotten. However, in many situations the simple solution (i) may be sufficient.

3.5 Summary Statistics

In this section the summary statistics used to evaluate the model errors are presented. The squared degree of determination or coefficient of determination R^2, in linear regression also called the squared multiple correlation coefficient, is a single number which describes the model error. The value express how large a part of the variation in the dependent variable is explained by the model. For a particular model R^2 is defined as

$$R^2 = 1 - \frac{\sum_{s=1}^{N} \hat{e}_s^2}{\sum_{i=1}^{N} (y_s - \bar{y})^2}, \tag{3.22}$$

where y_s is the observation of the dependent variable, $\bar{y}$ is the average of $y_s, s = 1, \ldots, N$, and $\hat{e}_s, s = 1, \ldots, N$ is the residuals or model errors.

Note that R^2 is calculated based on the residuals from the data set for which the parameters of the model are estimated. Hence, a model for which a large R^2 is obtained is not necessarily desirable, since it might be due to an over fitting of the data.

The root mean square error RMS is defined as

$$RMS = \sqrt{\frac{1}{N} \sum_{s=1}^{N} \hat{e}_s^2} \tag{3.23}$$

and the mean absolute error is defined as

$$MAE = \frac{1}{N} \sum_{s=1}^{N} |\hat{e}_s|. \tag{3.24}$$

For a linear regression model with zero mean Gaussian independent identical distributed model errors RMS is the maximum likelihood estimate of the standard deviation of the model error e_s. Compared to the MAE the RMS assigns more weight to large model errors.

For both RMS and MAE a small value is desirable but as for R^2 close to one this may be achieved at the cost of over fitting when the residuals are calculated based on the same data set as used for estimation. Over fitting can be corrected for by use of cross-validation, cf. Section 3.6.

3.6 Cross Validation

Cross validation aims at giving an unbiased estimate of the predictive ability of a model. This is archived by excluding some observations from estimation and using the obtained estimates to predict the observations excluded. Often the term *cross validation* is used in the sense of leaving one observation out at a time. Here we will use the term *leave-one-out cross validation* to distinguish it from the more general *K-fold cross validation*. Assuming the observations are numbered $1, 2, \ldots, N$, using leave-one-out cross validation the cross validated sums of squares CV_N is defined by

$$CV_N = \frac{1}{N} \sum_{s=1}^{N} \left(y_s - \hat{y}_s^{(-s)} \right)^2, \tag{3.25}$$

where $\hat{y}_s^{(-s)}$ is the prediction of y_s calculated without the observation s.

In K-fold cross validation the data are split in K parts which are excluded one at a time. The K-fold cross validated sums of squares CV_K is defined by

$$CV_K = \frac{1}{K} \sum_{j=1}^{K} \frac{1}{N_j} \sum_{s \in I_j} \left(y_s - \hat{y}_s^{(-j)} \right)^2, \tag{3.26}$$

where I_j is the set of observation indexes corresponding to the j'th part of the data, N_j is the number of observations in the j'th part of the data, and $\hat{y}_s^{(-j)}$ is the prediction of y_s calculated without the j'th part of the data. For $K = N$ (3.26) reduces to (3.25).

If more than one observation is left out at a time there is numerous ways to split the data set. In the work presented in this report all data are related to a particular date and time. Furthermore, the main problem seems to be that the underlying model changes over time. Therefore, we shall split the data along the time-axis, making the K-fold cross validation considerably more simple to perform than leave-one-out cross validation.

As noted by Breiman & Spector (1992) and by Shao (1993) leave-one-out cross validation will not lead to the selection of the model with the best predictive ability in that it tends to select an unnecessarily large model. Also, since we expect correlation of model errors close in time leave-one-out cross validation can lead to even more over fitting in this case. Breiman & Spector (1992) showed that in a pure regression setting 5-fold cross validation is superior to leave-one-out cross validation.

If K is small it can be informative to consider each $1/N_j \sum_{s \in I_j} (y_s - \hat{y}_s^{(-j)})^2$ in (3.26) separately, or more generally the RMS or MAE of the validated errors for each part

j of the data. In this way heterogeneity among the different parts of the data can be highlighted, cf. Figure 5.5 on page 49.

3.7 Sample Correlation Functions

This section describes how missing observations are treated when estimating the autocovariance/correlation function (ACF), the partial autocorrelation function, and the inverse autocorrelation function. Furthermore, it is described how the sample cross correlation function (SCCF) between two time series are calculated in case of some observations being missing. For the sample autocorrelation function (SACF) the approach described in (Dunsmuir & Robinson 1981) is used, but the actual formulas are changed slightly to make the estimators produce the usual estimates in the case of no missing observations. The sample inverse and partial autocorrelation functions (SIACF and SPACF) are based on the values of SACF. The sample cross correlation function (SCCF) between two time series are calculated using methods similar to the methods used for SACF.

The autocorrelation function

Let $x_1, x_2, \ldots, x_N$ be the observations from the process $\{x_t\}$, where some observations may be missing. Define

$$a_t = \begin{cases} 0, & \text{if } x_t \text{ is missing} \\ 1, & \text{otherwise} \end{cases} . \tag{3.27}$$

Using the temporary quantities

$$C_a(k) = \frac{1}{N - |k|} \sum_{t=1}^{N-|k|} a_t a_{t+|k|},$$

and

$$C^{\square}(k) = \frac{1}{N} \sum_{t=1}^{N-|k|} a_t a_{t+|k|} (x_t - \bar{x})(x_{t+|k|} - \bar{x}),$$

where

$$\bar{x} = \sum_{t=1}^{N} a_t x_t / \sum_{t=1}^{N} a_t.$$

The estimate of the autocovariance in lag k is

$$C(k) = \frac{C^{\square}(k)}{C_a(k)}. \tag{3.28}$$

It is seen that the method just skips the summands which cannot be calculated due to missing observations, and adjusts the number of observations accordingly. When no observations are missing $C_a(k) = 1$ and

$$C(k) = \frac{1}{N} \sum_{t=1}^{N-|k|} (x_t - \bar{x})(x_{t+|k|} - \bar{x}),$$

which is the usual estimator, cf. (Chatfield 1984, p. 60) or (Madsen 1995, p. 147). If the definition of $C_a(k)$ in (Dunsmuir & Robinson 1981) is used, then $C_a(k) = (N - |k|)/N$ when no observations are missing. Consequently, $1/N$ in the usual estimator of autocovariance is replaced by $1/(N-|k|)$, and the estimator leads no longer to a positive semi-definite autocovariance function, cf. (Priestley 1981, p. 323) or (Madsen 1995, p. 131).

Based on the estimates of autocovariance (3.28) the sample autocorrelation function is calculated as

$$\hat{\rho}(k) = \frac{C(k)}{C(0)}. \tag{3.29}$$

The inverse autocorrelation function

The inverse autocorrelation function is calculated as in `proc arima` of SAS/ETS (SAS Institute Inc. 1993), except that missing values are allowed. A high order autoregressive process is fitted to the series using the Yule-Walker equations and the estimates of the autocorrelation function (3.29). Hereafter the autocorrelation function is calculated for the dual process, i.e. a moving average process of the same order as the autoregressive process.

Let k_{max} be the maximum lag for which the sample inverse autocorrelation function is to be calculated. The order of the process is then

$$p = \min\left(\left[\frac{N}{2}\right], k_{max}\right),$$

where $[\cdot]$ denotes the integer value. If $k_{max} > N/2$, the estimates for the lags after $N/2$ is zero.

The partial autocorrelation function

The partial autocorrelation function is estimated based on (3.29) and calculated as described in (Abraham & Ledolter 1983, p. 212).

The cross correlation function

Let $x_1, x_2, \ldots, x_N$ and $y_1, y_2, \ldots, y_N$ be observations from the processes $\{X_t\}$ and $\{Y_t\}$, respectively. Some of the observations may be missing. Along with (3.27) for x_t we define

$$b_t = \begin{cases} 0, & \text{if } y_t \text{ is missing} \\ 1, & \text{otherwise} \end{cases} . \tag{3.30}$$

For $k \geq 0$ the temporary quantities

$$\begin{aligned} C_{ab}(k) &= \frac{1}{N-k} \sum_{t=1}^{N-k} a_t b_{t+k} \\ C_{ab}(-k) &= \frac{1}{N-k} \sum_{t=1}^{N-k} a_{t+k} b_t \end{aligned}$$

and

$$\begin{aligned} C_{xy}^{\square}(k) &= \frac{1}{N} \sum_{t=1}^{N-k} a_t b_{t+k} (x_t - \bar{x})(y_{t+k} - \bar{y}) \\ C_{xy}^{\square}(-k) &= \frac{1}{N} \sum_{t=1}^{N-k} a_{t+k} b_t (x_{t+k} - \bar{x})(y_t - \bar{y}), \end{aligned}$$

where $\bar{y}$ is defined similarly to $\bar{x}$ above, are defined. For both negative and positive k

$$C_{xy}(k) = \frac{C_{xy}^{\square}(k)}{C_{ab}(k)} \tag{3.31}$$

is used as an estimate of the cross covariance function at lag k and the sample cross correlation function at lag k is calculated as

$$\hat{\rho}_{xy}(k) = \frac{C_{xy}(k)}{\sqrt{C_{xx}(0) C_{yy}(0)}}, \tag{3.32}$$

where $C_{xx}(0)$ and $C_{yy}(0)$ are the estimates of the variances as in (3.28). When no observations are missing the formulae reduce to the well known ones (Chatfield 1984).

Variance of estimates

It is well known (Chatfield 1984, Madsen 1995) that if the observations originates from a white noise process the variance of $SACF$ and $SPACF$ are both approximately $1/N$. This is also true for $SIACF$ although it is usually not mentioned in textbooks (one exception being (Madsen 1995)). If the distribution of the white noise has finite moments the fact can be derived from Theorem 3.6 in (Bhansali 1980), which is valid for the method of estimation mentioned above. It is thus natural to adjust N to account for the number of observations which cannot be used due to missing values, i.e.

$$V[\hat{\rho}(k)] \approx [NC_a(k)]^{-1}\,, \tag{3.33}$$

which equals $1/N$ if no observations are missing.

For the SIACF and SPACF this approach is not directly applicable since it is not possible directly to relate the estimates to a specific lag. The maximum or minimum of $C_a(k)$ over the lags for which the estimates are calculated could be used, depending on the particular application. However, in this report the usual $1/N$ is used, since the method will newer be used anyway if a substantial fraction of the observations is missing.

For the cross correlation function it is well known (Chatfield 1984) that the estimates have variance $1/N$ if the series are uncorrelated. We therefore proceed as for the auto correlation function and use

$$V[\hat{\rho}_{xy}(k)] \approx [NC_{ab}(k)]^{-1}\,, \tag{3.34}$$

which also equals $1/N$ if no observations are missing.

Chapter 4

Considerations on model structure

In this chapter some physical characteristics of the system are used select an appropriate structure for the statistical model. First the stationary heat transfer trough a wall, the stationary heat and radiation transfer trough a window, and heat loss due to ventilation are described. A simple model for the dynamic response on climate is postulated and special considerations related to the heat consumption of a larger group of buildings are described. In all of the above aspects actual values of physical constants, e.g. thermal conductivity, are not used. Instead the structure of the model obtained is used to obtain a structure of an appropriate statistical model. It turns out that for a winter situation and given a particular value of the wind speed a linear regression model with dynamic response on the ambient air temperature and the solar radiation is appropriate. This kind of model is called a transfer function model in the statistical literature (Box & Jenkins 1976), but in this case strong relations among the parameters exist. However, since the wind speed is assumed to affect the convection heat coefficient on the outside of the buildings, most of the parameters of the model must be replaced by functions of either the wind speed or a low-pass filtered wind speed. Furthermore, in the autumn and spring there will be a period of time in which the consumers start/stop reacting on climate changes. For these reasons the appropriate statistical model turns out to be a varying-coefficient model (Hastie & Tibshirani 1993) with a dynamic response on the climate.

4.1 Stationary relations

In this section the stationary heat transfer trough a wall and a window are described. For further reading see e.g. (Incropera & DeWitt 1985, Hansen et al. 1987).

4.1.1 Heat transfer trough a wall

Figure 4.1 shows a sketch of a wall with indoor air temperature T_i, ambient air temperature T_a, indoor surface temperature $T_{s,i}$, outdoor surface temperature $T_{s,o}$, solar radiation orthogonal to the wall R_0, and the direction of a positive heat flux $\dot{Q}$.

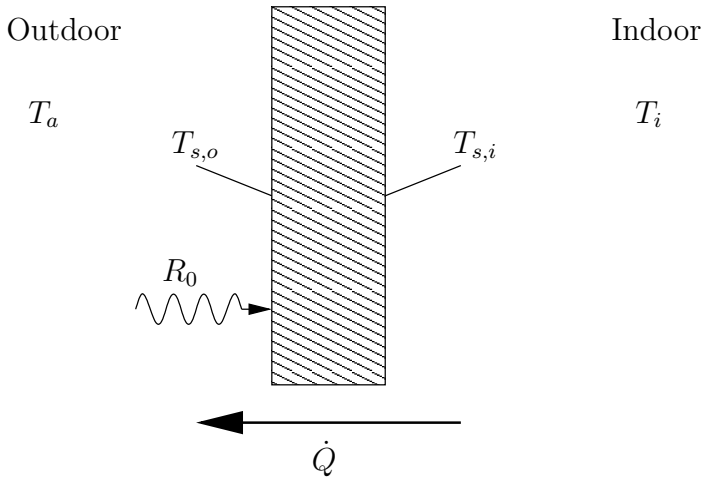

Figure 4.1: Heat transfer trough an outer wall with notation indicated.

For the convection and solar heating on the outside of the wall the stationary relation yields

$$\dot{Q} = h_o(T_{s,o} - T_a) - \epsilon R_0, \tag{4.1}$$

where h_o is the convection heat coefficient on the outside of the wall and ϵ is the fraction of solar radiation not reflected. The convection coefficient will be influenced by the wind speed and direction. For the inside of the wall only convection is considered, and the stationary relation thus becomes

$$\dot{Q} = h_i(T_i - T_{s,i}), \tag{4.2}$$

where the convection heat coefficient on the inside of the wall h_i will be nearly constant due to the to the fairly constant indoor environment. Furthermore, the stationary heat

conduction trough the wall is described by

$$\dot{Q} = U_w(T_{s,i} - T_{s,o}), \tag{4.3}$$

where U_w is the thermal conductivity divided by the wall thickness. Note that (4.3) is also valid when the wall consists of layers of different materials. In this case U_w can be found using the thermal conductivity and thickness of each layer (Incropera & DeWitt 1985).

Simple arithmetics lead to the following stationary relation between the ambient air temperature, the indoor air temperature, the solar radiation orthogonal to the wall, and the stationary heat flux trough the wall.

$$\dot{Q} = U(T_i - T_a) - U\frac{\epsilon}{h_o}R_0, \tag{4.4}$$

where $U = (1/U_w + 1/h_i + 1/h_o)^{-1}$.

4.1.2 Energy transfer trough a window

The energy transfer trough a window consists of conduction/convection as described above and solar radiation directly transmitted trough the window. Figure 4.2 shows a window with indoor air temperature T_i, ambient air temperature T_a, solar radiation orthogonal to the window R_0, and the direction of a positive energy flux $\dot{Q}$. Energy is also transfered trough the window by conduction/convection as described in the previous section. However, it is assumed that all solar radiation not reflected is transfered trough the window as radiation, i.e. for application in this context (4.1) and (4.4) must be modified by omitting terms containing ϵ. This is because these terms would relate to solar radiation heating the window itself. Consequently, the following equation is obtained for the energy flux trough the window

$$\dot{Q} = -\epsilon R_0 + U(T_i - T_a), \tag{4.5}$$

where the ϵ is the fraction of solar radiation not reflected and $U = (1/U_{win} + 1/h_i + 1/h_o)^{-1}$, here U_{win} is the thermal conductivity divided by the window thickness. The equation is valid also when the window consists of multiple layers of glass separated by a gas or atmospheric air.

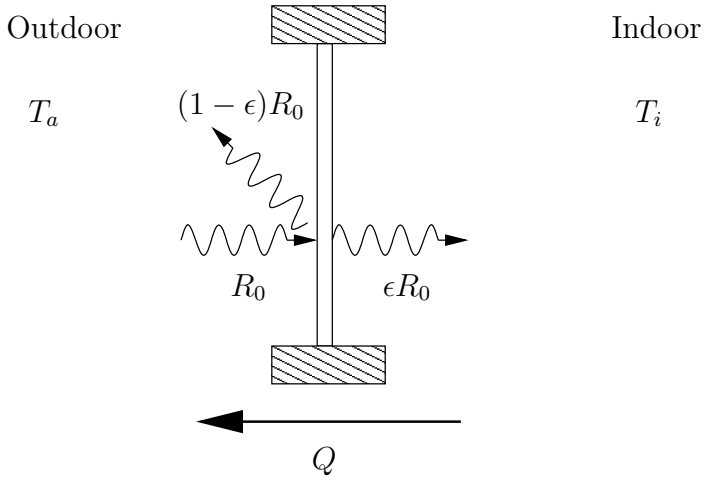

Figure 4.2: Solar radiation trough a window.

4.1.3 Ventilation

The warm air from the buildings is gradually replaced by cold air from the surroundings. Here we call this process ventilation, although it is often called infiltration in the technical literature. The heat flux needed to heat the air from the ambient air temperature T_a to the indoor temperature T_i can be expressed as

$$\dot{Q} = \mathcal{C}\dot{V}(T_i - T_a), \tag{4.6}$$

where $\dot{V}$ is the flow of air trough the building and $\mathcal{C}$ is the product of the specific heat capacity of the air and the mass density of the air. It is evident that for some buildings $\dot{V}$ depends on the wind speed. Also, the humidity of the air might influence $\mathcal{C}$. Since the amount of water vapour corresponding to a specific relative humidity is strongly dependent on the temperature it is plausible that the variation in $\mathcal{C}$ to some extend can be explained by the ambient air temperature.

4.2 Approximate dynamics

In this section the stationary relations described in Section 4.1 will be modified to take into account the dynamic response on changing climate conditions. However, the actual building dynamics will not be entirely modelled. This seems reasonable since for instance a constant indoor temperature will effectually eliminate the heat storage

capacity of the floor and internal walls. Filtering of climate variables are considered in this section. As described in e.g. (Åström 1970), when a linear dynamic system is sampled and if the input can be considered constant within the sampling interval the input-output relation of the discrete system can be described by rational transfer functions $H(q) = B(q^{-1})/A(q^{-1})$, where A and B are polynomials and q^{-1} is the backward shift operator ($q^{-1}x_t = x_{t-1}$). For this reason it seems reasonable to use rational transfer functions when filtering the climate variables. Furthermore, it will be required that the stationary gain of the filters $H(1)$ is one. Finally, the dot above Q will be dropped since it is only a matter of a constant if we consider the average heat flux from $t-1$ to t or the heat consumption over the time interval $t-1$ to t.

The heat transfer related to a wall consists of convection in the boundary layers and conduction trough the wall. It is assumed that when the sampling period is one hour as in this study the dynamics of the boundary layers can be neglected. Furthermore it is clear that the ambient air temperature T_a, wind speed W, and solar radiation orthogonal to the wall R_0 must be low-pass filtered. Also, since this filter is related to the materials of the wall the same filter can be used for all three variables. Furthermore, it is reasonable to assume that the dynamics of the walls behaves such that the transfer function $H_1(q)$ will have real and positive poles only. Assuming a fairly constant indoor temperature results in the following equation describing the approximate dynamics of a wall

$$Q_{1,t} = U_1(H_1(q)W_t)[T_{i,t} - H_1(q)T_{a,t}] - \epsilon_1 \frac{U_1(H_1(q)W_t)}{h_{01}(H_1(q)W_t)} H_1(q) R_{0,t}^{wall}, \tag{4.7}$$

where $Q_{1,t}$ is the energy transfered trough the wall during the hour starting at $t-1$ and ending at t, $T_{i,t}$ is the (unknown) indoor temperature, $T_{a,t}$ is the ambient air temperature, W_t is the wind speed, and $R_{0,t}^{wall}$ is the solar radiation orthogonal to the wall. The climate variables are measured as averages over the preceding 10 minutes.

For the energy transfer trough a window it evident that the dynamics consists of the dynamics of the glass and air in the window and on the dynamics of indoor building elements (floor, indoor walls, etc.) which heats the air after being heated by the solar radiation. Consequently, two transfer functions are needed to describe the approximate dynamics of the energy transfer trough a window. However, since the sampling interval is one hour the dynamics related to the window glass is neglected. Again, assuming a fairly constant indoor temperature results in the following equation describing the approximate dynamics related to a window

$$Q_{2,t} = -\epsilon_2 H_2(q) R_{0,t}^{win} + U_2(W_t)[T_{i,t} - T_{a,t}], \tag{4.8}$$

where $Q_{2,t}$ is the energy which, due to the windows, is needed during the hour starting at $t-1$ and ending at t to maintain the indoor temperature at $T_{i,t}$ and $R_{0,t}^{win}$ is the solar

radiation orthogonal to the window. The remaining variables are described above. Note that the solar radiation must enter trough the window and heat e.g. the floor before it affects the energy needed to maintain the indoor temperature. For this reason $Q_{2,t}$ is *not* the energy transfered trough the window during the time period ranging from $t-1$ to t, but the energy low-pass filtered by the floor, indoor walls etc.

It seems plausible that the slowest dynamics are related to $H_1(q)$ followed by $H_2(q)$. Using values in (Hansen et al. 1987) it can be deduced that the heat conduction trough a window can easily be 4-8 times as large at the heat conduction trough a wall of the same size. For this reason it must be expected that the functions $U_1(\cdot)$ and $U_2(\cdot)$ are comparable in magnitude.

With respect to the heat loss due to ventilation it will be assumed that the dynamics are negligible when a sampling interval of one hour is used

$$Q_{3,t} = \mathcal{C}_3(T_{a,t})\dot{V}_3(W_t)[T_{i,t} - T_{a,t}], \tag{4.9}$$

where $Q_{3,t}$ is the energy which, due to the windows, is needed during the hour starting at $t-1$ and ending at t to maintain the indoor temperature at $T_{i,t}$. The remaining variables are described above.

Note that because the climate measurements are averages over the last 10 minutes, since these are only measured at one location (Taastrup), and since the heat consumption is related to a large geographical area (København, Roskilde, Solrød) it is possible that, for both (4.8) and (4.9), low-pass filtering will be advantageous. An obvious first step would be to replace each climate variable X_t with $(X_t + X_{t-1})/2$.

4.3 The heat consumption of an area

The heat consumption of an area consists of both heat loss to the surroundings $Q_{L,t} = Q_{1,t} + Q_{2,t} + Q_{3,t}$, i.e. the sum of (4.4), (4.5), and (4.6), "free" heat $Q_{F,t}$ coming from e.g. electrical equipment but also from humans, and energy needed for hot tap water $Q_{W,t}$. The energy needed for room heating $Q_{H,t}$ can be expressed as

$$Q_{H,t} = [Q_{L,t} - Q_{F,t}]_+. \tag{4.10}$$

The truncation of negative values is used since when the quantity inside the squared brackets gets negative the indoor temperature will increase or ventilation will be used to prevent this.

Since an area is considered it seems natural to define $R_{0,t}^{wall}$ and $R_{0,t}^{win}$, using the fraction of walls $\delta_r^{wall}(\phi)$ with azimuth angle ϕ, the corresponding fraction of windows $\delta_r^{win}(\phi)$, and the solar radiation $R_{0,t}(\phi)$ on a vertical plane with azimuth angle ϕ:

$$R_{0,t}^{wall} = \frac{1}{360^o}\int_0^{360^o} \delta_r^{wall}(\phi)R_{0,t}(\phi)d\phi \tag{4.11}$$

$$R_{0,t}^{win} = \frac{1}{360^o}\int_0^{360^o} \delta_r^{win}(\phi)R_{0,t}(\phi)d\phi \tag{4.12}$$

It must be expected that $\delta^{wall}(\cdot)$ is more constant than $\delta^{win}(\cdot)$.

Let $\delta_{c,t}$ be the fraction of consumers reacting on the climate, and let $\delta_{p,t}$ be the fraction of the potential consumption active at time t, i.e. $\delta_{p,t}$ accounts for holidays and $\delta_{c,t}$ accounts for the fact that during the summer almost no consumers react on the climate and that they do not all start / stop reacting on the climate at the same time of year. The dependence on holidays might be negligible since it will to a large extend only affect the demand for hot tap water. If assuming that $\delta_{p,t}$ only affects the consumption of hot tab water and free heat the total heat consumption Q_t over the hour starting at $t-1$ and ending at t can be expresed as

$$Q_t = \delta_{c,t}[Q_{L,t} - \delta_{p,t}Q_{F,t}]_+ + \delta_{p,t}Q_{W,t}. \tag{4.13}$$

4.4 Identifiable model

The quantities of the model defined by (4.7), (4.8), (4.9) (4.11), (4.12), and (4.13) can not be estimated with the measurements available, which are the total heat consumption Q_t, the ambient air temperature $T_{a,t}$, the wind speed W_t, and the global radiation $R_{g,t}$. The aim of this section is to reach a model structure containing quantities which can be estimated from the data. It will be assumed that the truncation in (4.13) can be neglected since the truncation will only become active a few times during the spring ($\delta_{c,t}$ will be close to zero during the summer period). Therefore the model for the total heat consumption can be expressed as

$$Q_t = \delta_{c,t}Q_{L,t} - \delta_{c,t}\delta_{p,t}Q_{F,t} + \delta_{p,t}Q_{W,t}. \tag{4.14}$$

In the model $R_{0,t}^{wall}$ and $R_{0,t}^{win}$ are unknown. If $R_{0,t}(\phi)$ are known then in principle $\delta_r^{wall}(\phi)$ and $\delta_r^{win}(\phi)$ can be estimated by approximating the integral by a finite sum and using local regression together with backfitting. However, it is assumed that this is far too ambitious for practical use. The approach in this project has therefore been

to replace both $R_{0,t}^{wall}$ and $R_{0,t}^{win}$ with the global radiation or with a variable which can be calculated from the global radiation, solar evaluation, and time of the year. In the following this quantity will be denoted R_t. Note that this will distort ϵ_1 and ϵ_2 by a factor. Using (4.7), (4.8), (4.9), and (4.14) the total heat consumption can be expressed as

$$\begin{aligned} Q_t &= \delta_{p,t}Q_{W,t} - \delta_{c,t}\delta_{p,t}Q_{F,t} - \delta_{c,t}\epsilon_2 H_2(q)R_t \\ &+ \delta_{c,t}U_1(H_1(q)W_t)T_{i,t} - \delta_{c,t}U_1(H_1(q)W_t)H_1(q)T_{a,t} \\ &- \delta_{c,t}\epsilon_1 \frac{U_1(H_1(q)W_t)}{h_{01}(H_1(q)W_t)}H_1(q)R_t \\ &+ \delta_{c,t}U_2(W_t)T_{i,t} - \delta_{c,t}U_2(W_t)T_{a,t} \\ &+ \delta_{c,t}\mathcal{C}_3(T_{a,t})\dot{V}_3(W_t)T_{i,t} - \delta_{c,t}\mathcal{C}_3(T_{a,t})\dot{V}_3(W_t)T_{a,t}. \end{aligned} \tag{4.15}$$

If the dependence of $\mathcal{C}_3$ on $T_{a,t}$ is neglected, and if it is assumed that the heat needed for hot tap water $Q_{W,t}$ and the "free" heat $Q_{F,t}$ depend on the time of day h_t^{24} and on the type of day Υ_t the following model follows from (4.15)

$$\begin{aligned} Q_t &= \delta_{h,t}\mu(h_t^{24}, \Upsilon_t) + \delta_{c,t}a_{20}H_2(q)R_t \\ &+ \delta_{c,t}a_{11}(H_1(q)W_t) + \delta_{c,t}a_{12}(H_1(q)W_t)H_1(q)T_{a,t} \\ &+ \delta_{c,t}a_{10}(H_1(q)W_t)H_1(q)R_t \\ &+ \delta_{c,t}a_{21}(W_t) + \delta_{c,t}a_{22}(W_t)T_{a,t}, \end{aligned} \tag{4.16}$$

Alternatively if the dependence of $\mathcal{C}_3$ on $T_{a,t}$ is included then $a_{21}(W_t)+a_{22}(W_t)T_{a,t}$ must be replaced by $g(W_t, T_{a,t})$ and possibly the solar radiation should also be included as an argument. In (4.16) $\delta_{h,t}$ is used instead of $\delta_{c,t}$ and $\delta_{p,t}$ in $\delta_{p,t}Q_{W,t} - \delta_{c,t}\delta_{p,t}Q_{F,t}$. If $\mu(\cdot,\cdot)$ is modelled by a linear model and if the parameters of the transfer functions $H_1(q)$ and $H_2(q)$ are known, then (4.16) is a varying-coefficient model as described by Hastie & Tibshirani (1993). It is noted that the coefficient-functions are time-varying. Table 4.1 summarizes the interpretation of the coefficient-functions of (4.16).

The initial investigation of the model (4.16) will be complicated if the variation over time is to be taken into account. Instead a (winter) period in which it is reasonable to neglect the time variation will be identified and the initial investigation will be performed on data from this period only. Without the time-variation the model will still be somewhat complicated to investigate since, as mentioned above, there do not seem to be an obvious parametrization for the dependence of the coefficients on the wind speed. If the transfer-functions are known the model is a varying-coefficient model as described by Hastie & Tibshirani (1993) and it can be separated into a linear model and two conditional parametric models, see Sections 3.1 and 3.2. For this reason it seems obvious to start the analysis with the part of the model which is expected to be most important. It is expected that the conditional parametric model describing

Coefficient	Interpretation	Except for a
$\delta_{h,t}\mu(h_t^{24}, \Upsilon_t)$	$\delta_{p,t}Q_{W,t} - \delta_{c,t}\delta_{p,t}Q_{F,t}$	level
$\delta_{c,t}a_{20}$	$-\delta_{h,t}\epsilon_2$	factor (pos.)
$\delta_{c,t}a_{11}(H_1(q)W_t)$	$\delta_{c,t}U_1(H_1(q)W_t)T_i$	level
$\delta_{c,t}a_{12}(H_1(q)W_t)$	$-\delta_{c,t}U_1(H_1(q)W_t)$	–
$\delta_{c,t}a_{10}(H_1(q)W_t)$	$-\delta_{c,t}\epsilon_1\frac{U_1(H_1(q)W_t)}{h_{01}(H_1(q)W_t)}$	factor (pos.)
$\delta_{c,t}a_{21}(W_t)$	$\delta_{c,t}(U_2(W_t) + \mathcal{C}_3\dot{V}_3(W_t))T_i$	level
$\delta_{c,t}a_{22}(W_t)$	$-\delta_{c,t}(U_2(W_t) + \mathcal{C}_3\dot{V}_3(W_t))$	–

Table 4.1: Interpretation of the coefficient-functions in (4.16).

the heat transfer trough the wall will be predominant. Hereafter, the model can be extended with the linear model, followed by the conditional parametric model intended for the description of the ventilation, i.e. the last line of (4.16). Finally, the model may be extended further based on an analysis of the model error.

Chapter 5

Modelling the heat consumption

In Chapter 4 the basic structure of the relation between climate and heat consumption was derived based on physical considerations. However, many aspects of the model was left unknown. This chapter deals with identification of these parts of the model, based on data from a period of time in which it is reasonable to assume that all consumers are reacting on the climate, cf. Section 4.3. Section 5.1 is concerned with the identification of such a period. Based on this period an initial model only considering the heat transfer trough walls is considered in Section 5.2. Hereafter, the model is extended in Sections 5.3 and 5.4 to include all terms considered in Chapter 4. Based on an investigation of the model errors the correlation structure of the error term is considered in Section 5.5. In Section 5.6 a non-causal model is considered. Estimates obtained using different methods of estimation are compared in Section 5.7 and to guide the selection of the final model some statistical hypothesis tests are considered in Section 5.8. Finally, in Section 5.9, the diurnal variation is re-modelled with the aim of reducing the number of parameters used in this part of the model.

It is noted that the time-variation of the parameters are not considered in this chapter. Instead it is handled by the adaptive estimation used in Chapter 6. If not stated otherwise in this chapter outliers for the heat consumption data are treated as missing values and for the climate data the missing values and outliers are replaced by appropriate values, cf. Chapter 2.

5.1 Period without time-variation

To a large extend the heat consumption follows the ambient air temperature. From previous studies (Nielsen & Madsen 1997, Section 8.3.2) it is known that the dynamic response on the temperature can be well described by a transfer-function with one pole at approximately 0.93. Here it is further assumed that the transfer-function has no zeros. Figure 5.1 shows a cave-plot (Becker, Clark & Lambert 1994) of $\frac{0.07}{1-0.93q^{-1}}T_{a,t}$ and a smoothed version of the heat consumption. The smoothed heat consumption is created using a local quadratic approximation in time and a fixed bandwidth of three days. Using this smoother noise and diurnal variation are removed without introducing a phase-shift. The cave-plot is constructed so that if the dependence of the smoothed heat consumption on the filtered air temperature is linear and constant over time then the vertical distance between the curves is constant. This is accomplished by regressing the smoothed heat consumption on the filtered air temperature and obtaining the transformed temperature as the fitted values[1].

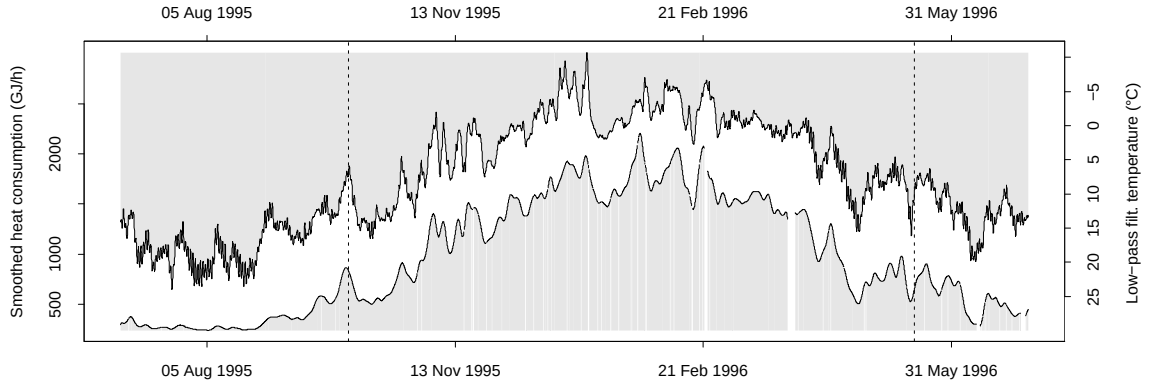

Figure 5.1: Cave plot of low-pass filtered air temperature and smoothed heat consumption. The vertical lines indicate 10Oct95:00:00 and 15May96:24:00, respectively.

From the plot it is seen that during the summer period there is very limited response on the temperature but at some stage during the autumn the heat consumption starts to follow the ambient air temperature. Except for a few occasions the difference between the curves are approximately constant during the period ranging from 1 October 1995 until 15 May 1996, indicating that in this period the time variation can be ne-

[1]More details can be found under the heading "Cave Plots" at the URL http://www.imm.dtu.dk/∼han/pub/caveplot

glected. This is further confirmed by Figure 5.2 which shows monthly scatter-plots of temperature and consumption and by Figure 5.3 in which estimates of the time-varying intercept $c_0(t)$ and slope $c_1(t)$ in the model

$$Q_t = c_0(t) + c_1(t)\frac{0.07}{1 - 0.93q^{-1}}T_{a,t} + e_t, \tag{5.1}$$

are depicted. Q_t is the heat consumption at time t and e_t is the corresponding model error. The estimates are based on local linear approximations and a fixed bandwidth of 30 days. The estimates are calculated using LFLM (Nielsen 1997). In conclusion the period ranging from 1 October 1995 until 15 May 1996 will be used to build a model in which the time-variation is neglected.

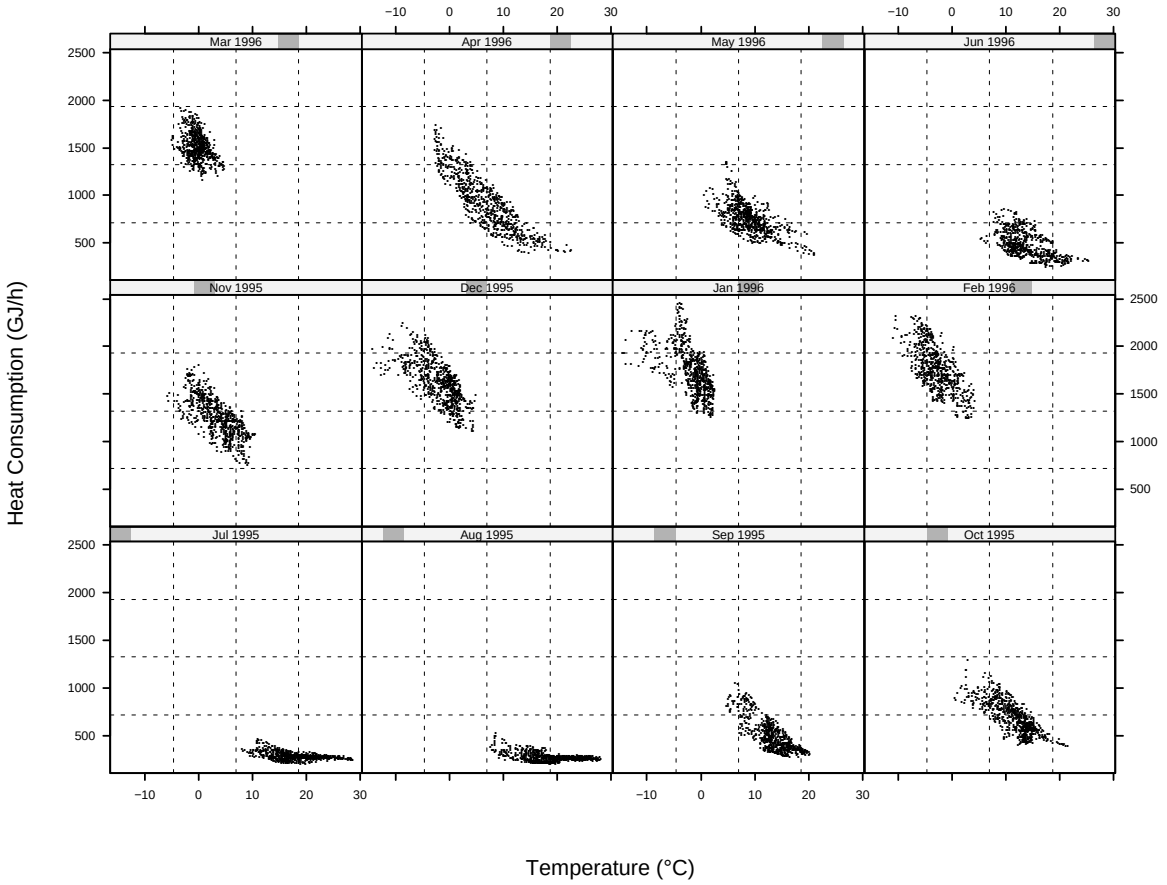

Figure 5.2: Heat consumption versus ambient air temperature, separately for each month.

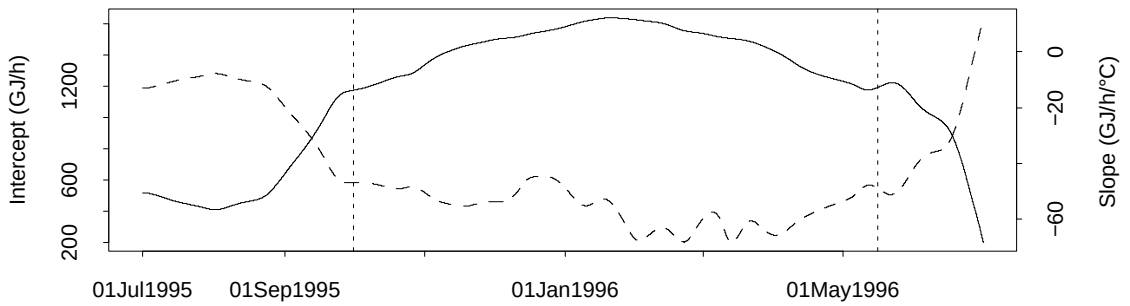

Figure 5.3: Estimates of intercept (solid) and slope (dashed) in model (5.1). The vertical lines indicate 1Oct95:00:00 and 15May96:24:00, respectively.

5.2 Initial model

As described in the last part of Section 4.4 the most important part of the model is expected to be the one primarily related to heat transfer trough a wall. In this section the rest of model (4.16) will be neglected and thus the section will be based on the model

$$\begin{aligned} Q_t &= a_{11}(H_1(q)W_t) + a_{12}(H_1(q)W_t)H_1(q)T_{a,t} \\ &+ a_{10}(H_1(q)W_t)H_1(q)R_t + e_t, \end{aligned} \tag{5.2}$$

where Q_t is the heat consumption at time t, W_t is the wind speed, $T_{a,t}$ is the ambient air temperature, R_t is the solar radiation on a square pillar as described in Chapter 2, e_t is the model error, $H_1(q)$ is a rational transfer function, and $a_{..}(\cdot)$ are unknown coefficient-functions which will be assumed to be smooth. Note that the interpretation of the coefficient-functions as indicated by Table 4.1 on page 41 is not strictly valid since some parts of the original model not included in (5.2) will affect the estimates of $a_{..}(\cdot)$. However, the data is approximately balanced with respect to the diurnal variation and therefore the diurnal variation will not affect the estimates to any large extend. The data is not balanced with respect to the type of day and therefore the overall level of the heat consumption included in the estimate of the function $a_{11}(\cdot)$ will be a weighted average over the types of days. Only data from the period ranging from 1 October 1995 until 15 May 1996 is used in this section.

When $H_1(q^{-1})$ is known (5.2) is a conditional parametric model and the functions $a_{..}(\cdot)$ can be estimated using local regression techniques. As mentioned in Section 5.1 previous work has indicated that

$$H_1(q) = \frac{1-\phi}{1-\phi q^{-1}}, \tag{5.3}$$

with $\phi = 0.93$, is an appropriate choice. This particular transfer-function will be used when investigating the estimated functions for different choices of bandwidth and local approximating polynomial. From model (4.16) and Table 4.1 it is expected that $a_{10}(\cdot)$ and $a_{12}(\cdot)$ are negative and that $a_{11}(\cdot)$ is positive. Furthermore, the exclusion of parts of (4.16) in (5.2) is unlikely to affect the sign of the estimated functions.

Figure 5.4 shows the estimates of the intercept and slopes in model (5.2) versus $\frac{0.07}{1-0.93q^{-1}}W_t$, using local linear approximations and nearest neighbour bandwidths 10%, 20%, ..., 90%. For the lower bandwidths the estimates show a very erratic behaviour, probably due to (local) collinearity in the underlying linear models which is fitted locally to the filtered wind speed.

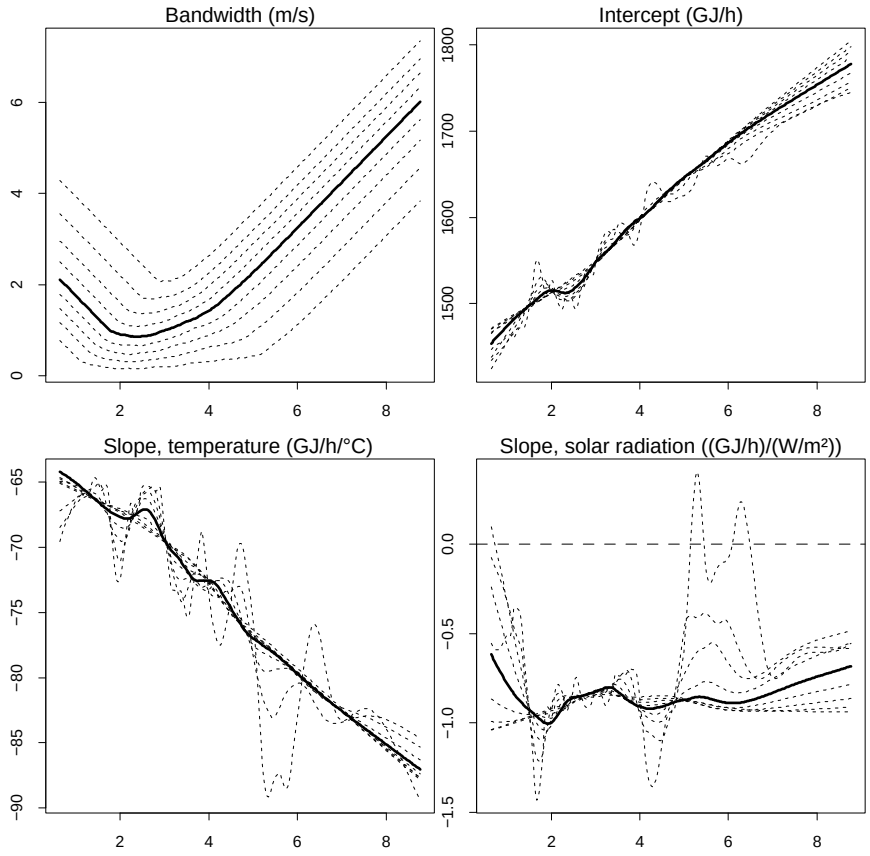

Figure 5.4: Bandwidth and estimates of intercept and slopes in model (5.2) versus low-pass filtered wind speed, using local linear approximations and nearest neighbour bandwidths 10%, 20%, ..., 90%. The curves corresponding to 50% are indicated by a bold line.

To guide the selection of the bandwidth both 5-fold and 100-fold cross-validation is performed, cf. Section 3.6, and the results are shown in Figure 5.5. For the 5-fold cross-validation the set number is indicated on the plot and the individual sets are connected across bandwidths. For the 100-fold cross-validation the individual sets are also connected. The mean and median scores of the 100-fold cross-validation are shown in Figure 5.6.

The results from the 5-fold cross-validation indicate that the validation sets are inhomogeneous, probably because (5.2) with the particular choice of transfer-function only describes part of the variation in the data. Both the 5-fold and 100-fold cross-validation indicate that low bandwidths are inappropriate and that for almost any of the validation sets bandwidths of approximately 40% or larger has equivalent performance. The mean scores of the 100-fold cross-validation depicted in Figure 5.6 indicate that on average a bandwidth of 90% (or larger) will be appropriate. However, due to the inhomogeneity of the validation sets it is desirable to neglect large deviations at this stage. This is accomplished by the median scores which indicate that a bandwidth of 60% is near-optimal with respect to both RMS and MAE. However, to explore the influence of the pole ϕ in (5.3) a bandwidth of 50% will be used since for other poles than $\phi = 0.93$ a slightly more flexible smoother might be appropriate.

Using a local linear smoother and a nearest neighbour bandwidth of 50% the mean absolute error (MAE) and root mean square error (RMS) is calculated for different values of the pole ϕ in (5.3) and the result is shown in Figure 5.7. On the figure "o" marks the pole with the minimum RMS and "$\times$" marks $\phi = 0.94$. Based on the figure 0.94 is taken as an estimate of ϕ in (5.3). With respect to MAE this pole is near-optimal, although MAE indicates that 0.95 is more appropriate. However, the two poles have virtually the same MAE. The value of R^2 for the resulting pole is 90.8%. The estimates of the coefficient-functions are not shown since they are very similar to the ones shown in bold on Figure 5.4.

5.3 Adding diurnal and direct solar radiation

In this section the model (5.2) with $H_1(q) = \frac{0.06}{1-0.94q^{-1}}$ is extended to include the part of (4.16) which, except for $H_2(q)$, is linear in the parameters:

$$\begin{aligned} Q_t &= \mu(h_t^{24}, \Upsilon_t) + a_{20}H_2(q)R_t \\ &+ a_{11}(H_1(q)W_t) + a_{12}(H_1(q)W_t)H_1(q)T_{a,t} \\ &+ a_{10}(H_1(q)W_t)H_1(q)R_t + e_t, \end{aligned} \tag{5.4}$$

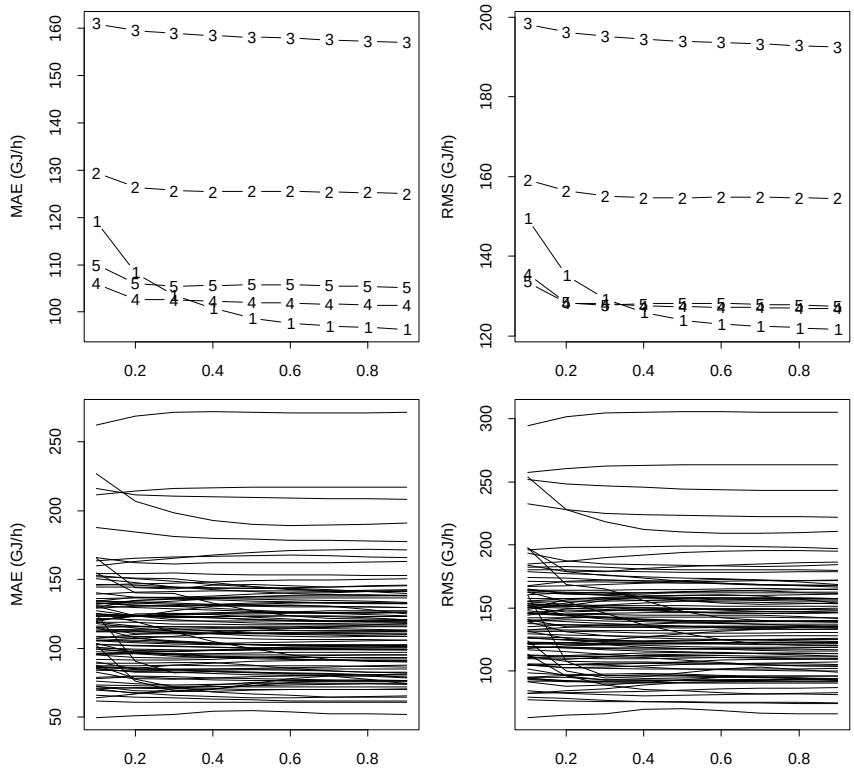

Figure 5.5: Mean Absolute Error (MAE) and Root Mean Square error (RMS) versus bandwidth for 5-fold (top) and 100-fold (bottom) cross-validation of predictions from (5.2) using nearest neighbour bandwidths 10%, 20%, ..., 90%.

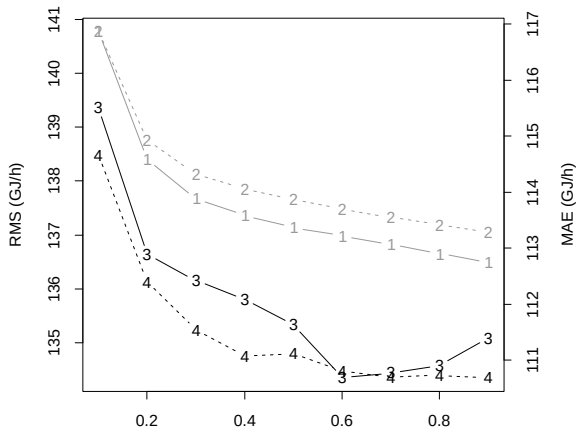

Figure 5.6: Mean RMS ("1"), mean MAE ("2"), median RMS ("3"), and median MAE ("4") versus bandwidth for the 100-fold cross-validation of predictions from (5.2).

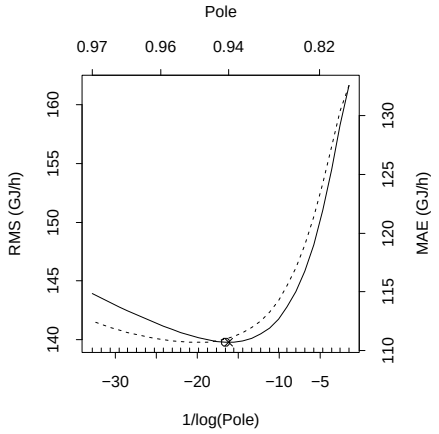

Figure 5.7: RMS (solid) and MAE (dotted) for different values of the pole ϕ in (5.3) and using local linear approximations and a 50% nearest neighbour bandwidth. The rugs on the horizontal axis indicate the values of the poles evaluated.

where the diurnal variation and the overall level is combined into $\mu(\cdot,\cdot)$ and modelled with a free parameter for each hour of the day and with separate parameters for (1) working days, (2) half-holy days, and (3) holy days. The actual parametrization is obtained as helmet contrasts (Statistical Sciences 1995*a*) corresponding to main effects and interaction of factors describing the time of day (0,1,...,23) and the type of day (1,2,3). The remaining quantities are defined as for (5.2).

Initially $H_2(q)$ is modelled as a Finite Impulse Response (FIR) filter and the parameter a_{20} is included in the filter

$$a_{20}H_2(q) = \sum_{i=0}^{24} h_{2,i}q^{-i} \tag{5.5}$$

In this way the model becomes semi-parametric, cf. the definition in Section 3.2, and the linear part being $\mu(h_t^{24}, \Upsilon_t) + a_{20}H_2(q)R_t$. A maximum lag of 24 is chosen based on (Nielsen & Madsen 1997, Section 8.3.3). As for (5.2) the coefficient-functions will be estimated using local linear approximations and a 50% nearest neighbour bandwidth. The estimates of $h_{2,0}, h_{2,1}, \ldots, h_{2,24}$ are depicted in Figure 5.8. Figure 5.9 shows the estimated coefficient-functions together with the estimates obtained using (5.2) and the estimates obtained under the assumption that the coefficient-functions are linear (in which case the model becomes a linear model). The root mean square (RMS) of the residuals of the semi-parametric fit is 91.53 GJ/h, and when the coefficient-functions are restricted to be linear this increase by less that 0.5 GJ/h to 92.00 GJ/h. This can hardly justify the complexity of the estimated coefficient-functions. The only software known to us which can estimate in models like (5.2) and which is used as an element

for estimation in models like (5.4) is LFLM 1.0 (Nielsen 1997). When the number of observations is large it is not feasible to use the software for calculation of the degrees of freedom of the smooth. However, by observing that the degrees of freedom is independent of the response (Hastie & Tibshirani 1990, Nielsen 1997) it is possible to get an approximation, using the function `loess()` in S-PLUS 3.4 (Statistical Sciences 1995*b*). A local line smoother with a 50% nearest neighbour bandwidth is applied to estimate a non-parametric regression function with $H_1(q)W_t$ as the explanatory variable and an arbitrary response. The approximate degrees of freedom of this smooth is 3.8. As an approximation the degrees of freedom of the smoother used in (5.4) is then taken to be $3 \times 3.8 - 1 = 10.4$, note that one is subtracted since the smooth is required to sum to zero over the observations. Requiring, the coefficient-functions to be linear reduces the degrees of freedom to $1 + 2 + 2 = 5$, i.e. $(92.00 - 91.53)/(10.4 - 5) \approx 0.09$ units of root mean square per degree of freedom.

The estimated functions, together with the values of RMS mentioned above indicate that the estimates from the semi-parametric model has high variance for values of $H_1(q)W_t$ above 5 m/s. However, the model is not complete and therefore it does not make sense to bootstrap the residuals (Efron & Tibshirani 1993) and re-fit the semi-parametric model 100 or 200 times to obtain pointwise confidence intervals of the coefficient-functions. Furthermore, this seems unpractical in that with our current implementation and equipment[2] it typically takes about 20 minutes to fit the model.

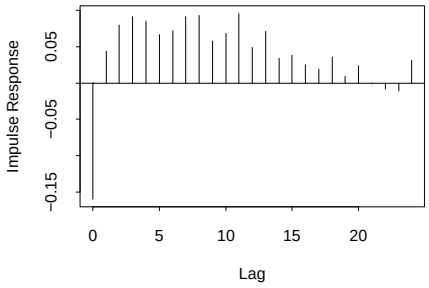

Figure 5.8: Estimated impulse response of $a_{20}H_2(q)$ based on the model consisting of (5.4) and (5.5).

Turning to the estimated impulse response depicted in Figure 5.8 it is noted that this may be well approximated by the following rational transfer-function

$$a_{20}H_2(q) = \frac{b_{200} + b_{201}q^{-1} + b_{202}q^{-2}}{1 + a_{201}q^{-1} + a_{202}q^{-2}}. \tag{5.6}$$

[2]HP 9000/KF460 with HPPA 8100 CPUs (Specfp $\approx$ 18) and 1 GB of memory.

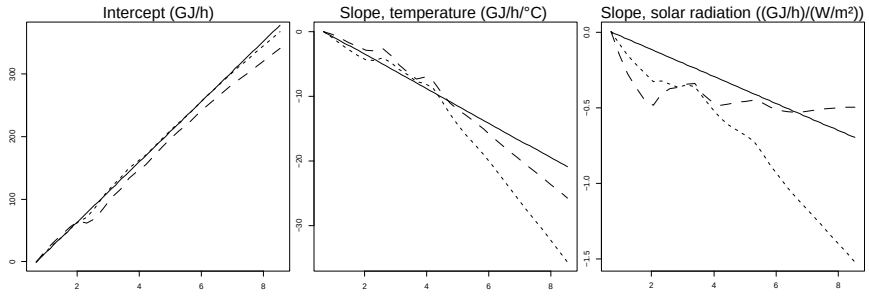

Figure 5.9: Non-parametric estimates (50% local lines) of the functions $a_{11}(\cdot)$, $a_{12}(\cdot)$, and $a_{10}(\cdot)$ in the model consisting of (5.4) and (5.5) (dotted), globally linear functions (solid), and non-parametric estimates corresponding to model (5.2) (dashed). The estimates are shifted vertically to start at the same point.

To fit this transfer-function first the partial residuals (Hastie & Tibshirani 1993) $Q_{p,t}$ are calculated as

$$\begin{aligned} Q_{p,t} &= Q_t - \hat{\mu}(h_t^{24}, \Upsilon_t) \\ &- \hat{a}_{11}(H_1(q)W_t) - \hat{a}_{12}(H_1(q)W_t)H_1(q)T_{a,t} \\ &- \hat{a}_{10}(H_1(q)W_t)H_1(q)R_t, \end{aligned} \tag{5.7}$$

for both the semi-parametric model and for the linear model, separately. Hereafter the transfer-functions are estimated using the model

$$Q_{p,t} = \frac{\theta_{20} + \theta_{21}q^{-1} + \theta_{22}q^{-2}}{1 + \phi_{21}q^{-1} + \phi_{22}q^{-2}} R_t + e_t. \tag{5.8}$$

Since the software available, `proc arima` of SAS/ETS (SAS Institute Inc. 1993), can not handle missing values the partial residuals are calculated using data Q_t in which missing values have been replaced as described in Chapter 2. The impulse response of the two estimates of the transfer-function are shown in Figure 5.10 together with the estimates obtained using the FIR-filter (5.5). It is seen that the approximations are good and that restricting the coefficient-functions to be linear results in more weight on lag zero and less on higher lags, as compared to the semi-parametric model.

After fixing the filters $H_1(q)$ and $H_2(q)$ obtained so far, the model (5.4) is re-fitted to the data (note that this adjusts the stationary gain of $a_{20}H_2(q)$). Using a semi-parametric model the root mean square (RMS) of the residuals is 91.78 GJ/h and for the linear model the corresponding value is 92.30 GJ/h, i.e. a difference of 0.52 GJ/h, which is a small increase compared to the estimated FIR-filter. Table 5.1 summarizes the values of RMS obtained for the models. It is noted that the total increase in RMS

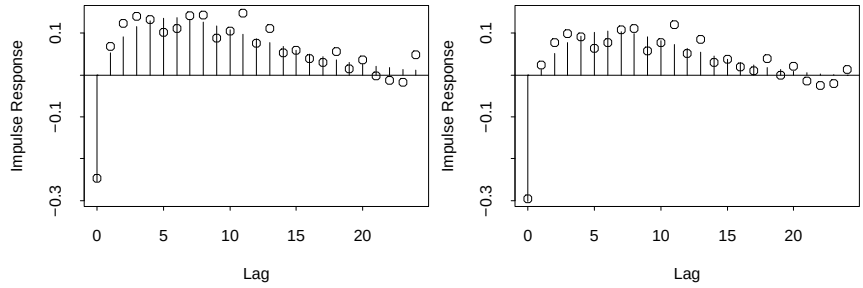

Figure 5.10: Estimated impulse response using (5.8) for the semi-parametric fit (left) and the linear model (right). Circles indicate estimates obtained using a FIR-filter (5.5).

when comparing the a semi-parametric model using the FIR-filter to a linear model using the estimated filter (5.8) is 0.77 GJ/h, whereas there is a huge difference in the degrees of freedom of the two models. Neglecting the difference in degrees of freedom originating from the coefficient-functions results in approximately 0.04 GJ/h for each degree of freedom. It therefore seems reasonable to use the linear model with the filters $H_1(q)$ and $H_2(q)$ fixed as described above. However, since fitting the linear model is quite fast and since $H_1(q)$ is fully characterized by its pole it is feasible to calculate RMS of the residuals for other values of the pole than 0.94 as used above. It is chosen to investigate values of the pole in the range 0.92 to 0.96 in steps of 0.001. For 0.934 the minimal RMS of 92.12 GJ/h is obtained. This pole will be used in the following, although the improvement is small. It is noted that the pole is also near-optimal with respect to the mean absolute error.

	SP	LM	Increase
FIR	91.53	92.00	0.47
RTF	91.78	92.30	0.52
Increase	0.25	0.30	0.77

Table 5.1: Cross-tabulation of values of RMS obtained using finite impulse response (FIR) and rational transfer function (RTF) filters and semi-parametric (SP) and linear (LM) models. In both cases the pole 0.94 is used in $H_1(q)$.

The parameters of the linear model and the parameters of the filters can be estimated simultaneously using non-linear optimization. It is noted that when the coefficient-functions of (5.4) are restricted to be linear, then given the transfer-functions $H_1(q)$ and $H_2(q)$ the model can be parametrized so that it is linear in the parameters. For this reason, given the transfer-functions, the least squares estimate of the parameter

vector has a closed-form solution and can be calculated quickly and reliably. It is therefore obvious to find the least squares estimate of the model including transfer-functions by letting the non-linear minimization search trough the possible values of the parameters of the transfer-functions and find the least squares estimates of the remaining parameters using the least squares solution of the linear model given the transfer-functions.

It is important that the transfer-functions are parametrized so that e.g. the stationary gain can be adjusted by the linear model. For this technical reason the following parametrizations are used

$$H_1(q) = \frac{1}{1 - \phi_{11} q^{-1}}, \tag{5.9}$$

and

$$H_2(q) = \frac{1 + \theta_{21} q^{-1} + \theta_{22} q^{-2}}{1 + \phi_{21} q^{-1} + \phi_{22} q^{-2}}. \tag{5.10}$$

Initial estimates of $\theta_{..}$ and $\phi_{..}$ are obtained rescaling coefficients in the filters identified above. It is noted that since the coefficient-functions are linear and since the solar radiation are filtered both trough $H_1(q)$ and $H_2(q)$ a singularity may exist for some values of the parameters of the transfer-functions. However, near the initial estimates the dynamics of $H_1(q)$ and $H_2(q)$ are very different and a singularity is unlikely to be found using these initial estimates.

The non-linear minimization is performed using the function `nlminb()` of S-PLUS 3.4 (Statistical Sciences 1995*b*) and since the possible correlation of the noise is not modelled missing values of Q_t is just excluded from the calculations. The minimization routine is not able to find any direction away from the initial estimates in which the gradient of the least squares criterion is different from zero and hence the initial estimates can be considered the estimates corresponding to the full non-linear model.

In conclusion the model selected so far is (5.4) with the coefficient-functions restricted to be linear and the filters

$$H_1(q) = \frac{0.066}{1 - 0.934 q^{-1}}, \tag{5.11}$$

and

$$H_2(q) = \frac{-0.350 + 0.612 q^{-1} - 0.226 q^{-2}}{1 - 1.703 q^{-1} + 0.739 q^{-2}}, \tag{5.12}$$

where both are normalized to have unit stationary gain. This model has an value of RMS of 92.12 GJ/h.

The estimated diurnal variation is depicted in Figure 5.11 for the model just mentioned

and for the semi-parametric model with

$$H_1(q) = \frac{0.06}{1 - 0.94q^{-1}}$$

and

$$H_2(q) = \frac{-0.163 + 0.311q^{-1} - 0.117q^{-2}}{1 - 1.691q^{-1} + 0.721q^{-2}},$$

where $H_2(q)$ corresponds to the semi-parametric fit on Figure 5.10. It is noted that with respect to the diurnal variation the models result in practically the same estimates. Furthermore, all curves have practically common values at 00:00 (24:00) although this restriction is not implied on the estimates. This is reassuring since in the real system no discontinuity is expected when going from, say, Friday 24:00 to Saturday 00:00.

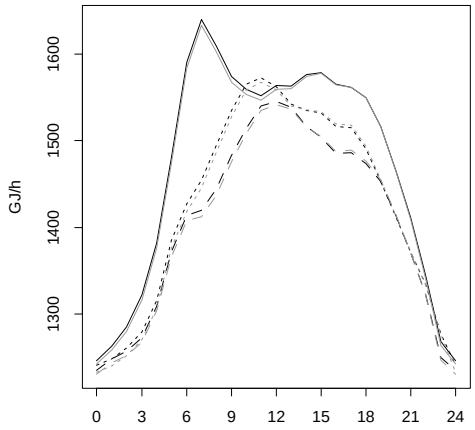

Figure 5.11: Diurnal variation for working days (solid), half-holy days (dotted), and holy days (dashed) for the linear version of (5.4) with (5.11) and (5.12) and for the semi-parametric model using $H_1(q) = 0.06/(1 - 0.94q^{-1})$ and $H_2(q) = (-0.163 + 0.311q^{-1} - 0.117q^{-2})/(1 - 1.691q^{-1} + 0.721q^{-2})$.

5.4 Adding heat transfer trough windows and ventilation

In this section the linear model described in the end of Section 5.3 is modified to take into account the heat loss due to ventilation and heat transfer by convection/conduction trough windows. Of cause a part of this effect is already described by the low-pass

filtered climate variables in the model. As argued in Section 4.2 the dynamic response is expected to be very fast. For this reason the first part of this section will consider a non-dynamic response and examine if the corresponding coefficient-functions can be linearized. Since the climate is measured at only one location (Taastrup) and since the heat consumption is related to a large geographic area it may however be advantageous to filter the climate variables. This is also considered in this section.

5.4.1 Non-dynamic response

The linearized version of (5.4) with an added possible non-linear static response on wind speed and temperature related to heat transfer trough windows and ventilation is described by the model

$$\begin{aligned} Q_t &= \mu(h_t^{24}, \Upsilon_t) + a_{20}H_2(q)R_t \\ &+ a_{111}H_1(q)W_t + a_{120}H_1(q)T_{a,t} + a_{121}H_1(q)W_tH_1(q)T_{a,t} \\ &+ a_{100}H_1(q)R_t + a_{101}H_1(q)W_tH_1(q)R_t \\ &+ a_{21}(W_t) + a_{22}(W_t)T_{a,t} + e_t, \end{aligned} \tag{5.13}$$

where $H_1(q)$ and $H_2(q)$ are defined in Eqs. (5.11) and (5.12). Note that the constants from the linearization of the coefficient functions are now merged with the overall level contained in the diurnal variation $\mu(\cdot, \cdot)$.

As usual the climate variables are filtered using observations in which missing values and outliers have been replaced by appropriate values, cf. Chapter 2, whereas observations (rows) corresponding to missing values of Q_t are just excluded. Since the response due to heat transfer trough windows and ventilation (the last line of (5.13)) is non-dynamic it is possible to estimated the model after exclusion of observations in which one or more of W_t, $T_{a,t}$, and Q_t are missing. However, to obtain values of RMS comparable with the results of the previous sections climate variables in which missing values are replaced with appropriate values will be used.

Linearization of $a_{21}(\cdot)$ and $a_{22}(\cdot)$ and subsequently estimation yields a value of RMS of 84.80 GJ/h. Compared to the model selected in Section 5.3, for which RMS = 92.12 GJ/h, this is a relatively large improvement obtained using only three extra parameters. Since the climate variables are only averages over the last 10 minutes of the hour the average of the two latest values might be more appropriate to use instead of W_t and $T_{a,t}$ in the last line of (5.13). However, the values of RMS obtained in this case is 86.41 GJ/h. This may be due to the averaging already obtained by the filtering of climate variables.

When excluding observations corresponding to missing values of W_t or $T_{a,t}$ a RMS of 82.79 GJ/h is obtained. This indicates the some of the observations in which missing values are replaced with appropriate values may be quite influential. To investigate this aspect the elements on the diagonal of the hat-matrix, cf. (Jørgensen 1993), are investigated. These values are the partial derivative of the fitted values with respect to the response, i.e. $\partial\hat{Q}_t/\partial Q_t$, and for a linear model these do not depend on the response, i.e. Q_t, see (Ye 1998, Section 2). In Figure 5.12 the values are displayed and values corresponding to missing values of W_t or $T_{a,t}$ are indicated by "$\times$" and by rugs on the horizontal axis. Venables & Ripley (1997, p. 204) state that large values are two or three times the average. Here no values are larger than three times the average, but quite a few are larger than two times the average. However, the observations corresponding to missing values as described above do not seem to be more influential than many other observations. Therefore the predictions of the model will not be seriously affected by whether these are included or not. To be able to compare results with the previous sections the observations considered are therefore not excluded.

A strong cyclic pattern is observed in Figure 5.12. Investigation show that practically all values above two times the average occur on half-holy days (mainly Saturdays, 778 in all) and on these days practically all values are above two times the average. Similarly the values just below two times the average is almost all related to holy days (mainly Sundays, 948 in all) and the low values are related to working days (3485 in all). There numbers are related to the total number of days in each group.

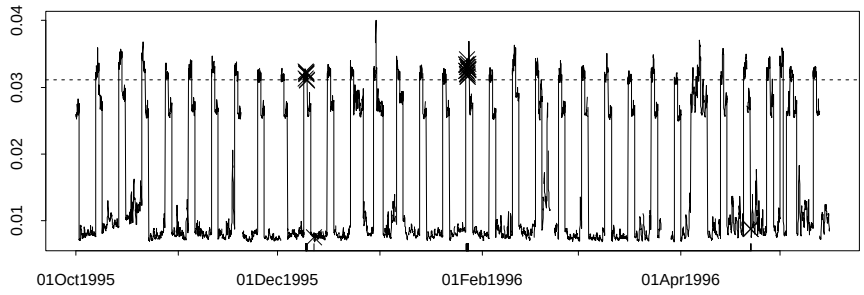

Figure 5.12: Elements on the diagonal of the hat-matrix. The dotted line marks two times the average.

To investigate the possible advantage of non-linear modelling of $a_{21}(\cdot)$ and $a_{22}(\cdot)$ in (5.13) these are estimated using a local linear smoother and a 30% nearest neighbour bandwidth. The equivalent number of parameters of $a_{21}(W_t) + a_{22}(W_t)T_{a,t}$ is approximately 9.5, calculated using the function `loess` of S-PLUS 3.4 (Statistical Sciences 1995*b*), compared to 3 degrees of freedom for the linerized model. However, the RMS of the residuals is only improved marginally to 84.18 GJ/h (compare with 84.80 GJ/h).

For this reason it is concluded that it is appropriate to use the linearized version of (5.13).

5.4.2 Dynamic response

As mentioned in the beginning of Section 5.4 a dynamic response corresponding to heat transfer trough windows and ventilation might be relevant. Again, it seems reasonable that both the wind speed and ambient air temperature should be filtered trough the same unit-gain transfer-function, called $H_3(q)$ below. As shown in the previous subsection (5.13) can be linearized with a very small loss in terms of RMS of the residuals. Consequently the model becomes

$$\begin{aligned} Q_t &= \mu(h_t^{24}, \Upsilon_t) + a_{20}H_2(q)R_t \\ &+ a_{111}H_1(q)W_t + a_{120}H_1(q)T_{a,t} + a_{121}H_1(q)W_tH_1(q)T_{a,t} \\ &+ a_{100}H_1(q)R_t + a_{101}H_1(q)W_tH_1(q)R_t \\ &+ a_{211}H_3(q)W_t + a_{220}H_3(q)T_{a,t} + a_{221}H_3(q)W_tH_3(q)T_{a,t} + e_t. \end{aligned} \quad (5.14)$$

Note that $H_3(q)W_tH_3(q)T_{a,t} \neq [H_3(q)]^2W_tT_{a,t}$. For this reason identification of $H_3(q)$ by use of a FIR-approximation is complicated. When excluding the term $W_tT_{a,t}$ from the linearized version of the previous model (5.13) the RMS of the residuals increase slightly by 0.04 GJ/h to 84.84 GJ/h. It therefore seems obvious to conclude that for identification of $H_3(q)$ the term $H_3(q)W_tH_3(q)T_{a,t}$ can be disregarded. Furthermore, if the wind speed and the ambient air temperature is filtered trough possibly different transfer-functions a FIR-approximation will result in a model linear in the parameters when $H_1(q)$ and $H_2(q)$ are assumed known. Here (5.11) and (5.12) will be used and the FIR-filters will both include lags 0–24. In Figure 5.13 the estimated coefficients of the two FIR-filters are depicted. It is evident that lags zero and one are far the most important ones, especially for the ambient air temperature.

Since only two lags are identified it is possible to reenter the interaction term and fit the model

$$\begin{aligned} Q_t &= \mu(h_t^{24}, \Upsilon_t) + a_{20}H_2(q)R_t \\ &+ a_{111}H_1(q)W_t + a_{120}H_1(q)T_{a,t} + a_{121}H_1(q)W_tH_1(q)T_{a,t} \\ &+ a_{100}H_1(q)R_t + a_{101}H_1(q)W_tH_1(q)R_t \\ &+ a_{211,0}W_t + a_{211,1}W_{t-1} + a_{220,0}T_{a,t} + a_{220,1}T_{a,t-1} \\ &+ a_{2,00}W_tT_{a,t} + a_{2,01}W_tT_{a,t-1} + a_{2,10}W_{t-1}T_{a,t} + a_{2,11}W_{t-1}T_{a,t-1} + e_t. \end{aligned} \quad (5.15)$$

The RMS of the residuals of this model is 81.46 GJ/h. The model has five degrees of freedom more than the linear model obtained in Section 5.4.1 for which RMS =

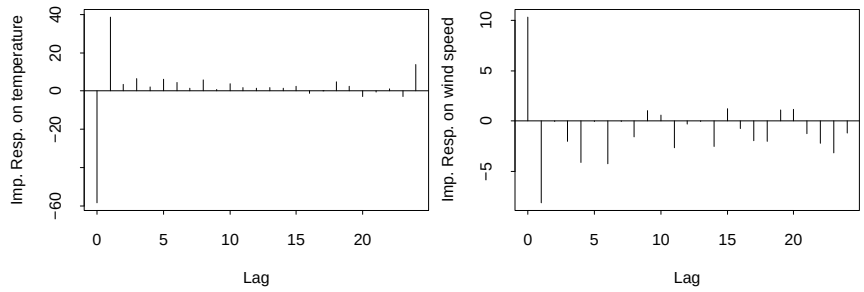

Figure 5.13: Estimated impulse response corresponding to heat transfer trough windows and ventilation.

84.80 GJ/h. Dropping the interaction terms in the last line of (5.15) results in RMS = 81.65 GJ/h, consequently these four terms contribute only marginally to the quality of the fit. However, to retain the physical interpretation of the model the terms will be kept in the model.

5.5 Noise model

Figure 5.14 shows the residuals and the sample autocorrelation function (SACF), the sample partial autocorrelation function (SPACF), and the sample inverse autocorrelation function (SIACF), cf. Section 3.7, of the residuals from (5.15). These estimates indicate that an autoregressive model can account for most of the correlation in the residuals. SPACF point towards an AR(3) model, whereas SIACF indicates that an AR(1) model should be sufficient. The estimated autocorrelations (SACF) of the residuals from AR-models of order 1–3 fitted to the residuals from model (5.15) via maximum likelihood (assuming the innovations to be iid. Gaussian) are depicted in Figure 5.15. It is seen that an AR(3) model is needed to remove the autocorrelation at the low lags, but in all cases correlation exists for higher lags and seems to be grouped around lags 24, 48, etc. In Figure 5.16 SACF, SPACF, and SIACF are shown for the residuals from the AR(3) model. These indicate that the residuals from model (5.15) may be more appropriately modelled by

$$(1 - a_1 q^{-1} - a_2 q^{-2} - a_3 q^{-3})(1 - a_{24} q^{-24}) r_t = (1 - c_1 q^{-1}) e_t, \tag{5.16}$$

where r_t is the residuals from model (5.15) and e_t is iid. $N(0, \sigma^2)$. The SACF of $\{\hat{e}_t\}$ from this model is shown in plot A in Figure 5.17. Many significant lags are found, e.g. lag 2, 23, and 25, this indicates that the mixing of parameters related to low and

high lags is inappropriate. This leads to the model

$$(1 - \sum_{i \in \{1:3,24:27\}} a_i q^{-i}) r_t = (1 - c_1 q^{-1}) e_t, \tag{5.17}$$

where e.g. $1:3$ denotes $1, 2, 3$. Plot B in the aforementioned figure shows the result. Still an significant value is observed at lag 23 and therefore the MA-part is replaced by an autoregressive term in lag 23:

$$(1 - \sum_{i \in \{1:3,23:27\}} a_i q^{-i}) r_t = e_t. \tag{5.18}$$

The result, depicted in plot C, is quite satisfactory. However, the model has many parameters at high lags. If deleting lags larger than 24 as in

$$(1 - \sum_{i \in \{1:3,23,24\}} a_i q^{-i}) r_t = e_t, \tag{5.19}$$

the result shown in plot D is obtained. Due to the fewer parameters this is of course slightly inferior C, but (5.19) will be used since it contains only five parameters. Note that also (5.16) contain five parameters.

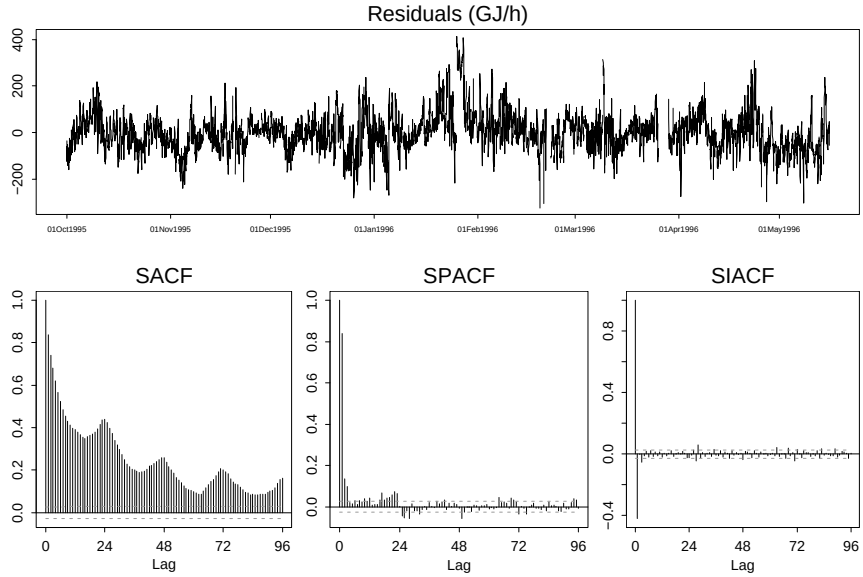

Figure 5.14: Residuals from model (5.15) together with estimated autocorrelation functions of the residuals. The dotted lines marks an approximate 95% confidence interval under the hypothesis of white noise.

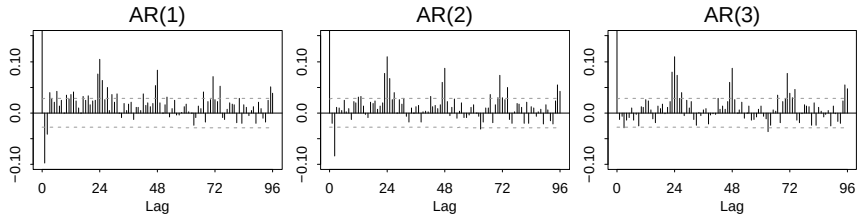

Figure 5.15: Estimated autocorrelations of the residuals from three AR-models fitted to the series shown in Figure 5.14. The dotted lines marks an approximate 95% confidence interval under the hypothesis of white noise.

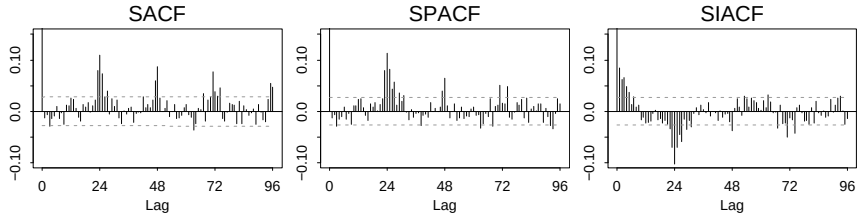

Figure 5.16: Estimated autocorrelation functions of the residuals from an AR(3) model fitted to the series shown in Figure 5.14. The dotted lines marks an approximate 95% confidence interval under the hypothesis of white noise.

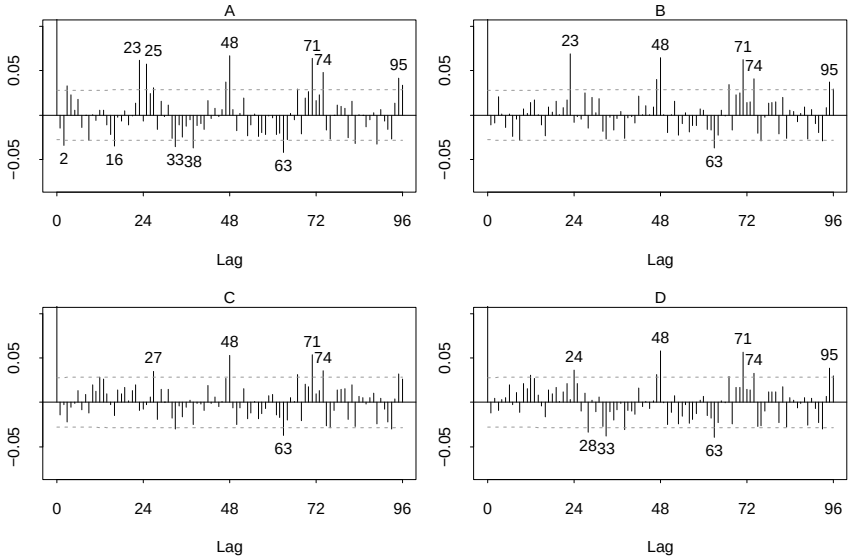

Figure 5.17: Estimated autocorrelation functions of the residuals from models A: (5.16), B: (5.17), C: (5.18), and D: (5.19) fitted to the series shown in Figure 5.14. The dotted lines marks approximate 95% confidence intervals under the hypothesis of white noise. Interesting significant lags are indicated.

From the above it is possible to formulate a complete model for the heat consumption Q_t in the VEKS area as:

$$\begin{aligned} Q_t &= \mu(h_t^{24}, \Upsilon_t) + a_{20}H_2(q)R_t \\ &+ a_{111}H_1(q)W_t + a_{120}H_1(q)T_{a,t} + a_{121}H_1(q)W_tH_1(q)T_{a,t} \\ &+ a_{100}H_1(q)R_t + a_{101}H_1(q)W_tH_1(q)R_t \\ &+ a_{211,0}W_t + a_{211,1}W_{t-1} + a_{220,0}T_{a,t} + a_{220,1}T_{a,t-1} \\ &+ a_{2,00}W_tT_{a,t} + a_{2,01}W_tT_{a,t-1} + a_{2,10}W_{t-1}T_{a,t} + a_{2,11}W_{t-1}T_{a,t-1} \\ &+ \frac{1}{1 - a_1q^{-1} - a_2q^{-2} - a_3q^{-3} - a_{23}q^{-23} - a_{24}q^{-24}}\, e_t, \end{aligned} \tag{5.20}$$

where $\{e_t\}$ is iid. $N(0, \sigma^2)$. The parameters of the model are estimated via maximum likelihood and by conditioning on the 24 first values of Q_t. Again, only data from the period identified in Section 5.1 are used. For the regressors on the right hand side of (5.20) data for which missing values and outliers are filled by appropriate values, cf. Chapter 2, are used. $H_1(q)$ and $H_2(q)$ are given by (5.11) and (5.12), respectively. For the heat consumption Q_t outliers are left missing and for maximum likelihood estimation a state-space formulation of (5.20) are used, see (Statistical Sciences 1995*b*) and Section 3.1. The root mean square (RMS) of the residuals from (5.20) is 38.46 GJ/h. Comparing with the RMS of the residuals from (5.19), which is 43.71 GJ/h, it is seen that there is a relatively large advantage from estimating the parameters simultaneously.

The sample autocorrelation function (SACF) of the residuals from (5.20) is shown in Figure 5.18, together with the SACF of the residuals from (5.19) fitted to the model error (residuals) from (5.15). It is seen that both SACFs are quite similar and satisfactory. However, the time series plots, shown in Figure 5.19, indicate that the variance is non-constant.

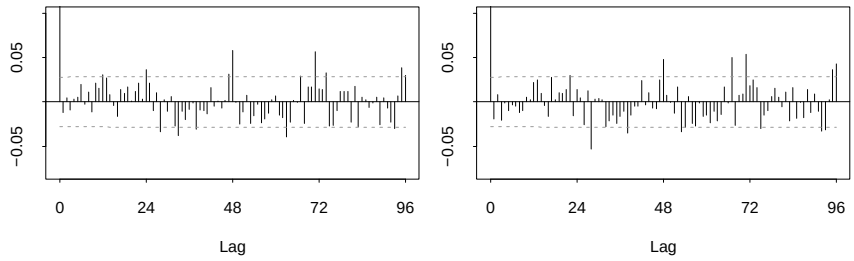

Figure 5.18: SACF of the residuals from (5.20) (right), and SACF of the residuals from (5.19) (left). The dotted lines marks an approximate 95% confidence interval under the hypothesis of white noise.

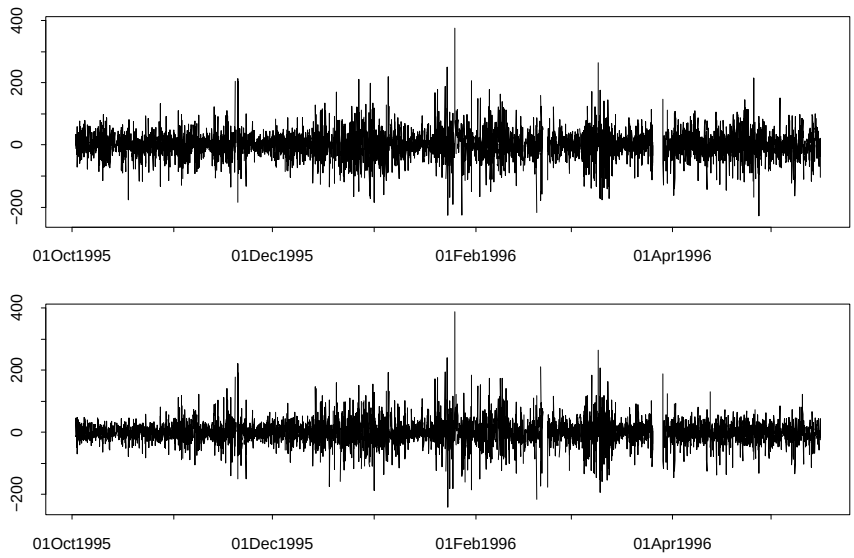

Figure 5.19: Time series plot of the residuals from (5.20) (bottom) and (5.19) (top).

Figure 5.20 shows the sample cross correlation function (SCCF) between the climate variables and the residuals from (5.20). The most pronounced feature is correlation of residuals with future values of the ambient air temperature a similar feature is observed for the solar radiation. Furthermore, there seem to be some correlation between the residuals and solar radiation at lags 12–24 (repeated 24 hours later). The SCCFs mentioned above can be used for model diagnostics but the transfer function which describes how the noise process $\{e_t\}$ enters (5.20) will complicate the interpretation of the SCCFs so that these are difficult to use for suggesting modifications of the model.

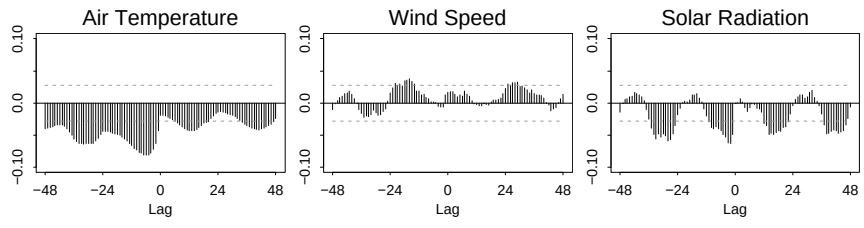

Figure 5.20: SCCF of the residuals from (5.20) and the three climate variables. The dotted lines marks an approximate 95% confidence interval under the hypothesis of uncorrelated series.

To see this consider the simple model

$$Y_t = H(q)X_t + \frac{1}{A(q^{-1})}e_t,$$

where Y_t is the output, X_t is the input, e_t is the noise process, $H(q)$ is a transfer-function with known structure, and $A(q^{-1})$ is a n_A-order order polynomial in the backshift operator q^{-1}. Now, if the SCCF between $\{\hat{e}_t\}$ and $\{X_t\}$ indicates that the structure of $H(q)$ should be changed, this can be thought of as trying to identify the structure of $H_0(q)$ in the model

$$Y_t = \widehat{H}(q)X_t + H_0(q)X_t + \frac{1}{\widehat{A}(q^{-1})}e_t,$$

where the hat indicates that the estimated coefficients are used. The model may also be written as

$$\widehat{A}(q^{-1})Y_t = \widehat{H}(q)\widehat{A}(q^{-1})X_t + H_0(q)\widehat{A}(q^{-1})X_t + e_t,$$

from which it is seen that the SCCF mentioned above is more helpful in identifying the structure of $H_0(q)\widehat{A}(q^{-1})$ than the structure of $H_0(q)$. However, the SCCF between $\{\hat{e}_t\}$ and $\{\widehat{A}(q^{-1})X_t\}$ can be used for structural identification of $H_0(q)$.

If the climate variables are filtered as outlined above the SCCFs shown in Figure 5.21 are obtained. It is noted that the correlation between the residuals and solar radiation at lags 12–24 is practically removed by the filtering of the solar radiation. On the figure the lags with the maximum absolute correlation is indicated. For the wind speed this maximum occurs at lag -2 and the value is considered small. Therefore the wind speed will not be considered further at this place. For the ambient air temperature and the solar radiation the maximum absolute correlation occurs at lag -1. Again, it is seen that future values of the ambient air temperature and solar radiation seem to affect the heat consumption. This is probably due to the geographically distributed consumption and the fact that the climate variables only are measured at one location. An other explanation could be that the consumers (local network operators) use some prediction device combined with a heat storage facility.

When using the results obtained in the study described in this report meteorological forecasts of the climate variables have to be used. It is natural to use, possibly weighted, averages of meteorological forecasts over the geographical area supplied by the VEKS transmission system. In this case non-causal models will probably be irrelevant and therefore in this study non-causal models will not be used. It is noted that, when using the data available in this study, this decision will lead to slightly inferior prediction results as would have been obtained had a non-causal model been used. For the sake of completeness, we will show the results obtained when using a simple non-causal model in the following section.

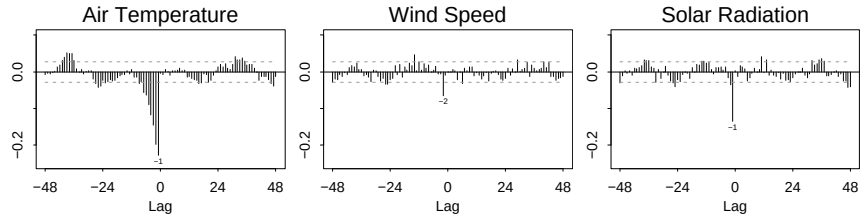

Figure 5.21: SCCF of the residuals from (5.20) and the three filtered climate variables. The dotted lines marks an approximate 95% confidence interval under the hypothesis of uncorrelated series.

5.6 A non-causal model

From Figure 5.21 for both the ambient air temperature and the solar radiation the decline of SCCFs towards zero for decreasing lags starting at lag -1, seems be of first order. If X_t denotes the climate variable and $\theta_i;\ i = 1, 2$ are some parameters then

$$X_t = q^{-1}\frac{\theta_1}{1-\theta_2 q^{-1}}\hat{e}_t,$$

will be able to approximate the impulse response. This can also be written as

$$\hat{e}_t = \frac{1}{\theta_1}X_{t+1} - \frac{\theta_2}{\theta_1}X_t.$$

Comparing with the terms already in the model (5.20) this means that terms containing $T_{a,t+1}$ and R_{t+1} should be added on the right hand side of (5.20). Therefore the following non-causal model is obtained

$$\begin{aligned}
Q_t &= \mu(h_t^{24}, \Upsilon_t) + a_{20}H_2(q)R_t \\
&+ a_{111}H_1(q)W_t + a_{120}H_1(q)T_{a,t} + a_{121}H_1(q)W_tH_1(q)T_{a,t} \\
&+ a_{100}H_1(q)R_t + a_{101}H_1(q)W_tH_1(q)R_t \\
&+ a_{211,0}W_t + a_{211,1}W_{t-1} + a_{220,0}T_{a,t} + a_{220,1}T_{a,t-1} \\
&+ a_{2,00}W_tT_{a,t} + a_{2,01}W_tT_{a,t-1} + a_{2,10}W_{t-1}T_{a,t} + a_{2,11}W_{t-1}T_{a,t-1} \\
&+ a_TT_{a,t+1} + a_RR_{t+1} \\
&+ \frac{1}{1-a_1q^{-1}-a_2q^{-2}-a_3q^{-3}-a_{23}q^{-23}-a_{24}q^{-24}}\ e_t.
\end{aligned} \tag{5.21}$$

The parameters of this model is estimated via maximum likelihood as for (5.20). Figure 5.22 shows the obtained SCCFs between the residuals and the climate variables filtered using the estimates of $a_i;\ i = 1, 2, 3, 23, 24$. Comparing with Figure 5.21 it is seen that most of the cross correlation is removed.

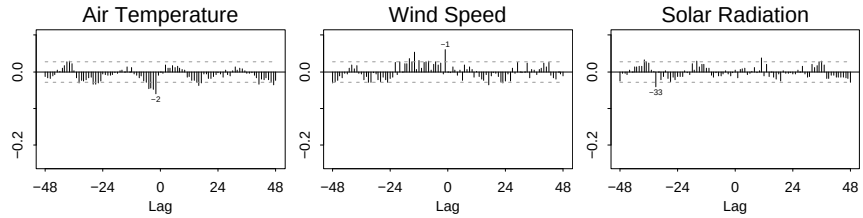

Figure 5.22: SCCF of the residuals from the non-causal (5.21) and the three filtered climate variables. The dotted lines marks an approximate 95% confidence interval under the hypothesis of uncorrelated series.

5.7 Comparison of models and estimates

In this section the maximum likelihood estimates obtained using the model (5.20) is compared with the least squares estimates obtained from model (5.15). It is noted that these models have the same deterministic structure, but in (5.20) the correlation of the noise is modelled. Results presented above indicates that the deterministic structure and the correlation of the noise is modelled quite adequately by (5.20).

Assuming the climate to be known the in-sample long-term prediction of Q_t from (5.20) is obtained using the deterministic part of the model only. The root mean square (RMS) of the corresponding prediction errors is 168.8 GJ/h. For (5.15) the in-sample long-term prediction errors is just $\hat{e}_t$ and the RMS of these is 81.5 GJ/h, i.e. less than half the value obtained for (5.20). For the in-sample one step prediction errors the opposite is true in that for (5.20) the RMS of these is 38.5 GJ/h and 81.5 GJ/h for (5.15). However, if the noise-model used in (5.20) is applied to the prediction errors from (5.15) the RMS of the one-step prediction errors can be reduced to 46.7 GJ/h.

The results above indicate that in on-line applications a lot will be lost in terms of RMS of long-term prediction errors using (5.20) instead of (5.15) and for short-term prediction the performance of (5.15) can be improved markedly by modelling the correlation of the prediction errors from (5.15). This is contrary to conventional statistical theory in which maximum likelihood estimates often are considered optimal in some sense. However, this theory is based on the assumption that the true system is of the same structure as the model. Obviously, in situations like the one considered in this report, some details of the true system can often not be handled by the models considered and therefore the central assumption leading to the optimality of the maximum likelihood estimates is violated.

The maximum likelihood estimates (MLE) in (5.20) are close to the estimates minimizing the sum of squared one step prediction errors; the only difference being due to some values of Q_t being missing. Following (Ljung 1987) the least squares estimates obtained based on (5.15) are called output error estimates (OEE) below. In Table 5.2 the OEEs and the MLEs of parameters related to climate variables are shown. Note that the regressor rather than the name of the parameter is used to identify the estimates. Except for two cases the absolute value of the MLE is closer to zero than the absolute value of the OEE. It is also notable that the regressors for both exceptions do *not* contain the ambient air temperature. In conclusion, the MLEs use the autocorrelation of the heat consumption instead of the climate variables. Since the OEEs do not allow for this they perform better for long-term predictions.

Regressor	OEE	MLE	
$H_2(q)R_t$	0.50	0.52	
$H_1(q)W_t$	39.15	38.00	×
$H_1(q)T_{a,t}$	-42.40	-29.20	×
$H_1(q)W_tH_1(q)T_{a,t}$	-2.67	-0.36	×
$H_1(q)R_t$	-1.14	-0.52	×
$H_1(q)W_tH_1(q)R_t$	-0.08	-0.11	
W_t	17.18	3.93	×
W_{t-1}	-10.43	0.52	×
$T_{a,t}$	-64.01	-20.17	×
$T_{a,t-1}$	45.03	3.71	×
$W_tT_{a,t}$	0.47	-0.21	×
$W_tT_{a,t-1}$	-2.01	0.25	×
$W_{t-1}T_{a,t}$	-3.14	-0.26	×
$W_{t-1}T_{a,t-1}$	4.26	-0.03	×

Table 5.2: Maximum likelihood estimates (MLE) and output error estimates (OEE) obtained using model (5.20) and (5.15), respectively. "×" in the rightmost column indicates that $|MLE| < |OEE|$.

The OEE of the overall level is 1403.7 GJ/h and the MLE is 1320.9 GJ/h. The diurnal variation around the overall level is depicted in Figure 5.23 also on the figure the average deviation due to the type of day is shown. No large difference between MLE and OEE is observed.

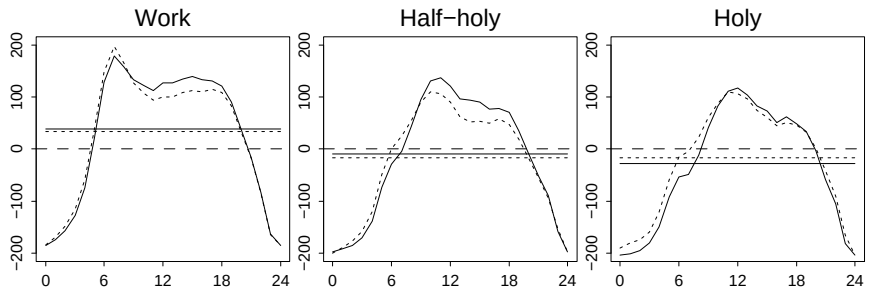

Figure 5.23: Diurnal variation around the overall level for OEE (solid) and MLE (dotted) and around the average for the particular type of day indicated by horizontal lines.

5.8 Likelihood ratio tests

Likelihood ratio tests in the model (5.20) are performed. The results of the tests are listed in Table 5.3. From the tests it is indicated that the short term interaction between wind speed and ambient air temperature, i.e. the parameters $a_{2,ij};\ i = 0, 1;\ j = 0, 1$, can be excluded from the model. This is consistent with the root mean square of the residuals from the regression model mentioned on page 59. The test for difference between half-holy and holy days is marginally insignificant at the 5% level. However, by use of appropriate approximations, cf. Section 5.9, the diurnal variation can be described by use of fewer degrees of freedom per day group. This could easily make the aforementioned test insignificant ($\chi^2(23) = 35.172$) and therefore the difference between half-holy and holy days are retained in the model. All the remaining tests are significant at the 5% level. Akaike's information criterion indicates that the model where $a_{2,ij} = 0;\ i = 0, 1;\ j = 0, 1$ should be used, i.e. (as expected) it is consistent with the likelihood ratio test.

5.9 Remodelling the diurnal

In all the models considered so far the diurnal variation is modelled with one free parameter for each combination of hour of the day $(0, 1, \ldots, 23)$ and type of day (working, half-holy, holy). This amounts to 72 parameters. In this section it will be investigated if the number of parameters used to model the diurnal variation can be reduced by use of an other parametrization of the diurnal variation. Since the models are to be used for prediction horizons up to 72 hours the investigation will be based on least squares

Hypothesis	LR statistic	df	p-value
$a_{2,00} = a_{2,01} = a_{2,10} = a_{2,11} = 0$	2.613	4	0.624
$a_{2,00} = a_{2,01} = a_{2,10} = a_{2,11} = a_{211,0} = a_{211,1} = 0$	13.794	6	0.032
$a_{2,00} = a_{2,01} = a_{2,10} = a_{2,11} = a_{220,0} = a_{220,1} = 0$	247.053	6	0.000
Wind speed excluded from model	127.451	9	0.000
Solar radiation excluded from model	271.913	3	0.000
Ambient air temperature excluded from model	435.288	8	0.000
No diff. between half-holy and holy days	35.520	24	0.061

Table 5.3: Likelihood ratio tests of selected null hypotheses; in all cases with (5.20) as the alternative hypothesis.

estimates in model (5.15) on page 58, but with an alternative parametrization of the diurnal variation and with $a_{2,ij} = 0;\ i = 0, 1;\ j = 0, 1$.

The parametrizations considered will be (i) harmonic expansions (Graybill 1976, Section 8.8) and (ii) periodic B-spline bases (de Boor 1978) and Section 3.1. This have the advantage that the resulting model is linear in the parameters. Hence, the parameters can be estimated reliably and relatively quickly. However, the main advantage in the context of on-line predictions is that an reliable adaptive and recursive method exists for this kind of models. A purely non-parametric estimation of the diurnal variation for each type of day is possible by use of a semi-parametric model, cf. Section 3.2 and 5.4, where the smoother of the kind described in (Nielsen 1997) is applied, but with an angular measure of distance. However, this approach will not be considered further since we are not aware of adaptive and recursive methods for estimation in such models which have been investigated thoroughly, although a method has been suggested in (Nielsen, Nielsen, Joensen, Madsen & Holst 1999).

Initially cubic and quadratic B-spline bases with equidistant knots is investigated. The first knot is placed at 00:00. If e.g. the diurnal variation for one particular type of day is to be modelled using a periodic cubic B-spline basis and six degrees of freedom the six internal knots will be placed at $00, 04, \ldots, 20$. For a cubic spline three knots will have to be placed beyond the range of the data resulting in the knots $-12, -8, -4, 0, 4, \ldots, 20, 24, 28, 32, 36$. A standard cubic B-spline basis is then generated using these knots and the last three columns of the basis is deleted, but first added to the first three columns. The resulting basis has six columns and is a flexible basis for estimation of a periodic function. In the following we shall refer to such a basis as a periodic cubic B-spline basis with six internal equidistant knots. Periodic quadratic B-spline bases are generated similarly, but only two knots are needed beyond the range of the data and only the last two columns of the basis is deleted, but first

added to the first two columns.

The top row of plots in Figure 5.24 shows the maximum absolute difference between each of the diurnal profiles estimated using periodic cubic B-spline bases, periodic quadratic B-spline bases, and harmonic expansions and the diurnal profile estimated using one free parameter for each combination of time and type of day. The middle row shows the mean of these absolute differences and the bottom row shows the RMS of the residuals. Note that the three plots in the bottom row all corresponds to the total RMS. The horizontal line in the bottom row indicates the RMS when one free parameter is used for each combination of time and type of day. The models with fewest degrees of freedom corresponds to a third order harmonic expansion and seven internal equidistant knots. Note that the full range of the RMS-values is not shown on the plot.

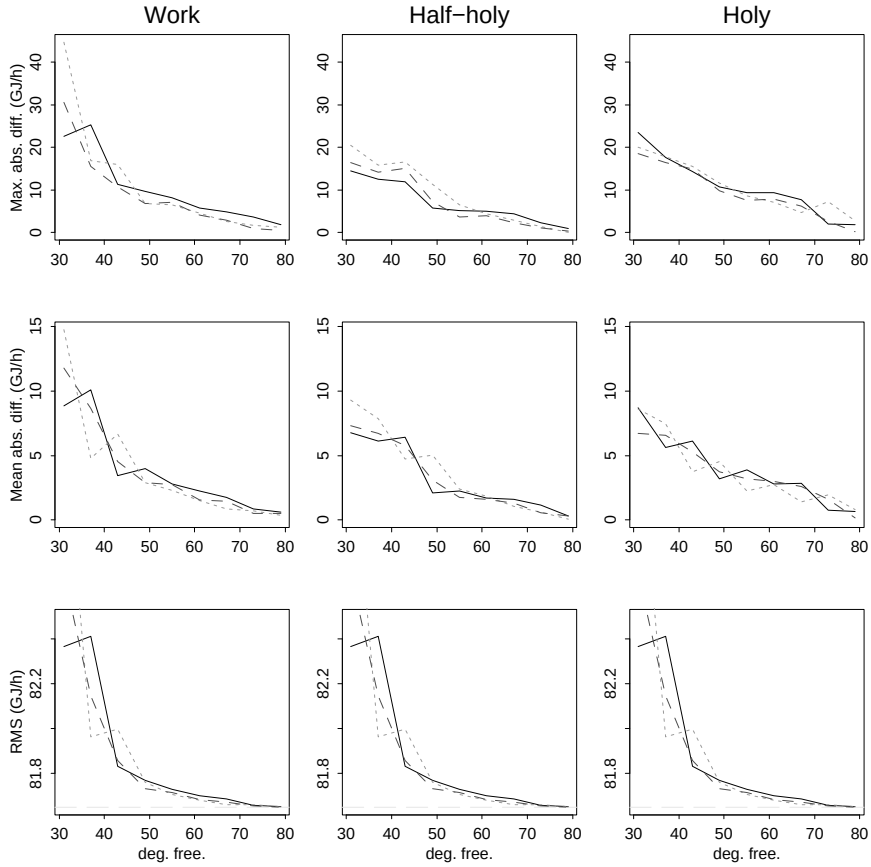

Figure 5.24: Summary statistics of fits using periodic cubic B-spline bases (solid), periodic quadratic B-spline bases (dotted), and harmonic expansions (dashed) plotted against the total degrees of freedom of the model.

Especially the plot of RMS shows a change in slope at 43 degrees of freedom, corresponding to 11 internal and equidistant knots and a fifth order harmonic expansion. For the model in which a periodic cubic B-spline basis is applied the RMS is 81.83 GJ/h, compared with 81.65 GJ/h when one free parameter is used for each combination of time and type of day. Generally the three types of parametrizations seems to be quite similar except for the case where the total degrees of freedom is low. In this case the periodic cubic B-spline basis seems preferable. When seven internal knots are used, corresponding to 31 degrees of freedom, this type of basis results in a RMS value of 82.37 GJ/h. The harmonic expansion with the same number of degrees of freedom results in an RMS value of 82.69 GJ/h.

It is noted that the curves in Figure 5.24 corresponding to spline bases are rather non-smooth and the RMS of the residuals is not decreasing when the total degrees of freedom is increasing. This is because the smaller models (in the sense of the total degrees of freedom) is not sub-models of the larger models.

In Figure 5.25 the estimated diurnal profiles using periodic cubic B-spline bases and harmonic expansions are shown. The location of the internal knots are indicated by rugs on the time-axis. In the top row the models using 43 degrees of freedom is addressed. For both the periodic cubic B-spline basis with 11 equidistant knots and the fifth order harmonic expansion the results are very close to the results obtained when using one free parameter for each combination of time and type of day. The bottom row of the figure shows the result when four of the internal knots are deleted so that the density of knots are largest near the morning peak load period. This model has 31 degrees of freedom, which is also the number of parameters in a model in which a third order harmonic expansion is used. For this reason the results obtained using a third order harmonic expansion is also displayed in the bottom row. In this case the cubic spline basis models the morning peak better than does the harmonic expansion; especially on working days. Both methods show some departure from the large model around midnight.

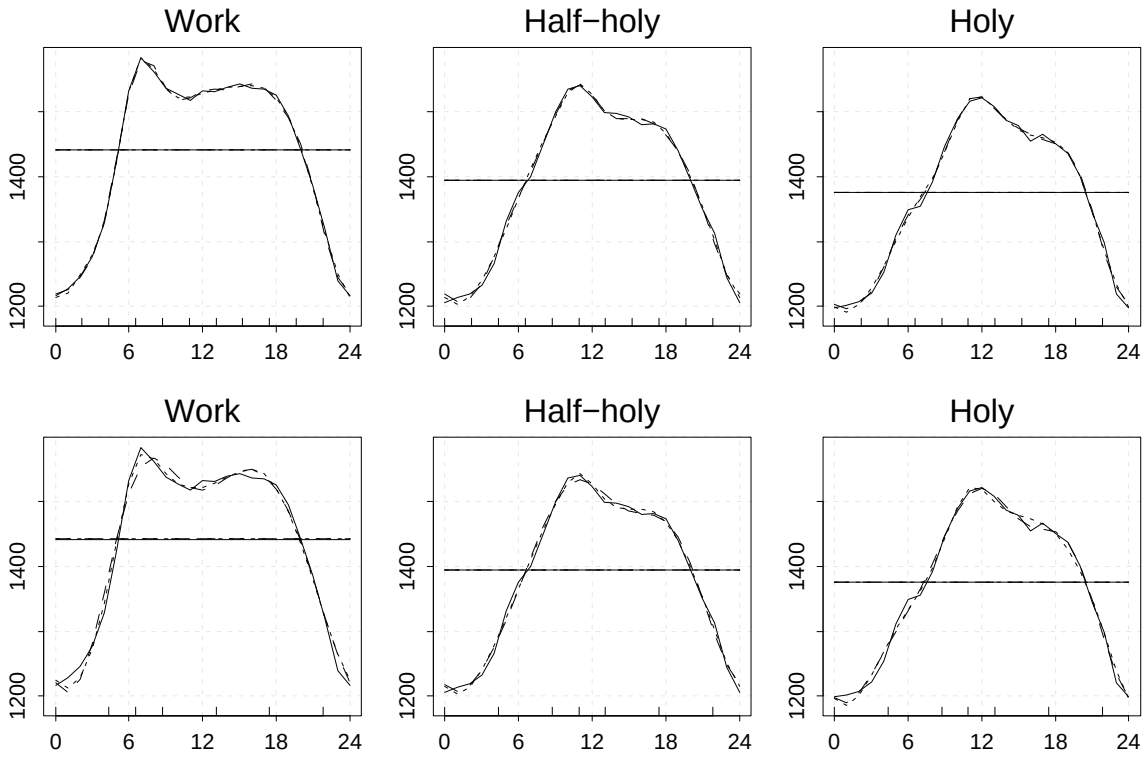

Figure 5.25: Estimated diurnal profiles (GJ/h) using different parametrizations. The solid lines indicates the estimates obtained when one free parameter is used for each combination of time and type of day. The dotted lines are results obtained using a periodic cubic B-spline basis and the dashed lines are results obtained using a harmonic expansion. The top row of plots corresponds to 43 degrees of freedom (11 internal knots / 5'th order harmonic expansion). The bottom row of plots corresponds to 31 degrees of freedom (7 internal knots / 3'rd order harmonic expansion). The internal knots are indicated by rugs on the time-axis. The horizontal lines shows the overall mean of the estimates.

Chapter 6

Predicting the heat consumption

In this chapter adaptive estimation as described in Section 3.4 will be applied to the entire data set. The investigation will be based on the same type of models as considered in Section 5.9, i.e.

$$\begin{aligned} Q_t &= \mu(h_t^{24}, \Upsilon_t) + a_{20}H_2(q)R_t \\ &+ a_{111}H_1(q)W_t + a_{120}H_1(q)T_{a,t} + a_{121}H_1(q)W_tH_1(q)T_{a,t} \\ &+ a_{100}H_1(q)R_t + a_{101}H_1(q)W_tH_1(q)R_t \\ &+ a_{211,0}W_t + a_{211,1}W_{t-1} + a_{220,0}T_{a,t} + a_{220,1}T_{a,t-1} + e_t, \end{aligned} \tag{6.1}$$

where the filters (5.11) and (5.12) are used, i.e.

$$H_1(q) = \frac{0.066}{1 - 0.934q^{-1}}, \tag{6.2}$$

and

$$H_2(q) = \frac{-0.350 + 0.612q^{-1} - 0.226q^{-2}}{1 - 1.703q^{-1} + 0.739q^{-2}}. \tag{6.3}$$

Q_t denotes the heat consumption at time t, W_t is the wind speed, $T_{a,t}$ is the air temperature, R_t is the solar radiation on a square pillar as described in Chapter 2, and $a_{\cdot}$ are the parameters. Above the structure of the diurnal variation $\mu(h_t^{24}, \Upsilon_t)$ is not specified. Here some of the same parametrizations as described in Section 5.9 will be investigated. Specificly the fifth order harmonic expansion, the periodic cubic B-spline basis with 11 equidistant internal knots, and the B-spline basis with seven knots derived from the former, cf. Figure 5.25. In Section 5.9 the estimated diurnal profiles for the day types "working", "half-holy", and "holy" are shown for the parametrizations mentioned above.

The adaptive estimates are calculated using an S-PLUS library written by Torben Skov Nielsen, Dept. of Math. Mod., Technical University of Denmark. In case of missing values neither of the updates (3.20) nor (3.21) on page 26 are performed.

6.1 Adaptive RLS estimates

In Appendix B starting on page 117 Figures B.1, B.2, and B.3 shows the cumulative squared prediction errors for the horizons 24, 48, and 72 hours. The predictions are calculated based on adaptive recursive least squares estimates in (6.1) with the three type of days mentioned above and under the assumption of known future climate. In case of missing values (only occurring in Q_t) the prediction errors are treated as zero in the plots.

It is seen that the three parametrizations of the diurnal variation result in quite similar results, but the choice of the forgetting factor is very important. Especially, for the 24 hour predictions and the low forgetting factor (0.99) the B-spline basis with seven internal knots are slightly inferior for data in 1996. However, the curves are practically parallel and therefore the predictions are equally good when considering this period separately. Hence, it is only during a short period of time in the end of January, 1996 that the B-spline basis with seven internal knots performs poorer than the two other parametrizations.

In conclusion there do not seem to be any benefit in using the special designed B-spline basis with only seven internal knots. It is therefore preferred to use the periodic cubic B-spline basis with 11 equidistant internal knots since this will not imply any restrictions on, e.g. the location of the morning peak load period estimated by the model. Alternatively, the fifth order harmonic expansion could be used. In Figure 6.1 the equivalent kernel corresponding to 12:00 of the two bases is shown assuming that data from one day is available. The equivalent kernel is the weights the observations receive in construction of the particular fitted value, i.e. $\hat{y}_{12} = \sum_{h=0}^{23} w_{12,h} y_h$, where y_h is observations of the dependent variable, $\hat{y}_{12}$ is the fitted value at 12:00, and $w_{12,h}$ are the weights. Although, the B-spline basis is preferable in that it dies out somewhat faster than the harmonic expansion, the two approaches have quite similar equivalent kernels.

Above each day of the year has been classified into one of the groups “working”, “half-holy”, and “holy”. Here the first group contains approximately five times as many observations as the two other groups. Since also the estimated diurnal profiles

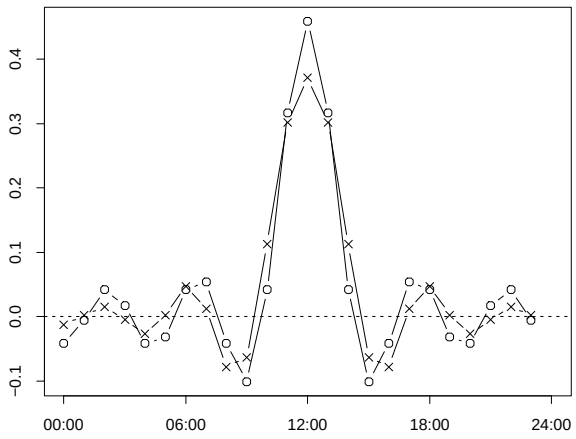

Figure 6.1: Equivalent kernel corresponding to 12:00 for a periodic cubic B-spline basis with 11 internal equidistant knots (×) and a fifth order harmonic expansion (∘).

corresponding to days in the groups "half-holy" and "holy" are similar, cf. Figure 5.25 on page 72, it may be advantageous to neglect the difference between these types of days for on-line predictions. These aspects are investigated using Figures B.4, B.5, and B.6 which shows the cumulative squared prediction error for 24, 48, and 72 hour prediction horizons and for different values of the forgetting factor.

Again, the forgetting factor seems to be more important than details about the model structure. However, when the low forgetting factor of 0.99 is used the grouping of days into all three groups seems advantageous for horizons 48 and 72 hours. When the overall better forgetting factor of 0.995 is used the methods perform almost equally well for horizons 48 and 72 hour. For the 24 hour predictions it seems advantageous to neglect the difference between half-holy and holy days. However, again the difference in performance seems to be generated almost entirely in the end of January, 1996.

Consequently, we shall entirely neglect the difference between half-holy and holy days. However, if there should be more subjective reasons to use all three day groups the loss is quite small.

For the model consisting of (6.1), (6.2), and (6.3), where the diurnal variation is modelled by a periodic B-spline basis with 11 internal equidistant knots forgetting factors ranging from 0.99 to 0.999 in steps of 0.001 are investigated for prediction horizons $1, 2, 4, 6, \ldots, 70, 72$ hours. Figure 6.2 shows the root mean square (RMS) of the prediction errors from August 15, 1995 and onwards. The optimal forgetting factor (in the sense of the RMS values mentioned) spans the interval 0.99 to 0.995, but for most

horizons the optimum is not very well defined. For the one and two hour predictions the optimum is at the lower bound of the interval investigated. However, since this bound corresponds to $1/(1 - 0.99) = 100$ effective observations only, and since it may occasionally result in very large prediction errors, smaller forgetting factors will not be investigated.

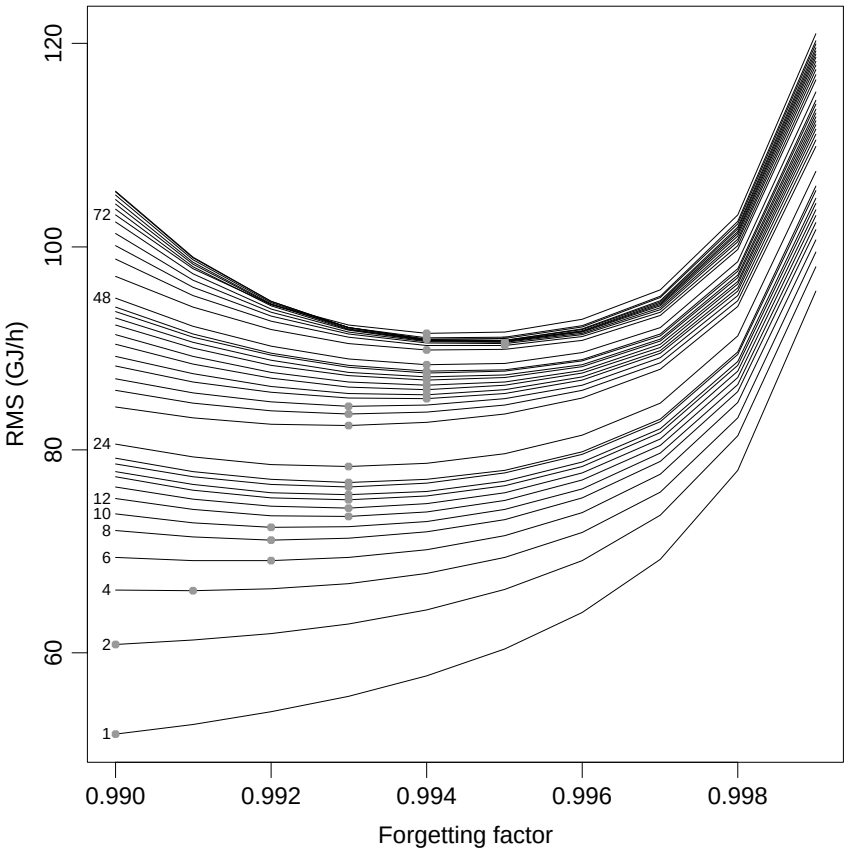

Figure 6.2: Root mean square (RMS) of prediction errors from August 15, 1995 and onwards versus forgetting factor. For selected horizons the actual horizon is indicated at the leftmost endpoint of the curve corresponding to that particular horizon. The minimum RMS for each horizon is indicated by a dot.

From the plots in Appendix B it is however clear that a forgetting factor constant over time is a sub-optimal solution. The small forgetting factors are good in periods in which the adaptive estimates should change quickly (spring and autumn), whereas larger forgetting factors are advantageous in more stable periods. These aspects are further considered in Section 6.3, starting on page 80.

For the "optimal" forgetting factor the root mean square of the prediction errors occurring in the same period as considered in Chapter 5 is 80.82, 87.79, and 88.49 GJ/h for the 24, 48, and 72 hour predictions, respectively.

6.2 Prediction errors from regression models

In Section 6.1 the correlation of the prediction errors from the adaptive estimates in the regression model is not used for generations of the actual predictions. Figure 6.3 shows the prediction error, the SACF, SPACF, and the SIACF for the 24 hour prediction. For $48h$ and $72h$ the picture is similar. From the plots it is seen that the prediction error seems to be well described by an AR(1) model, or a noise model as in (5.20) on page 62. However, here we will use an AR(1) model. Let $r_{t+k} = \widetilde{Q}_{t+k|t}$ be the k-step prediction error of the prediction of Q_{t+k} generated based on data available at time t. When predicting r_{t+k} only $r_t, r_{t-1}, \ldots$ are available. Hence if ϕ denotes the pole in the AR(1) model the prediction of r_{t+k} will be $\hat{r}_{t+k|t} = E[r_{t+k}|\{r_t, r_{t-1}, \ldots\}] = \phi^k r_t$. If $\phi \approx 0.9$, which seems plausible from the plots considered above, then $\phi^{24} \approx 0.08$, indicating that it is difficult to use the autocorrelation to improve the predictions.

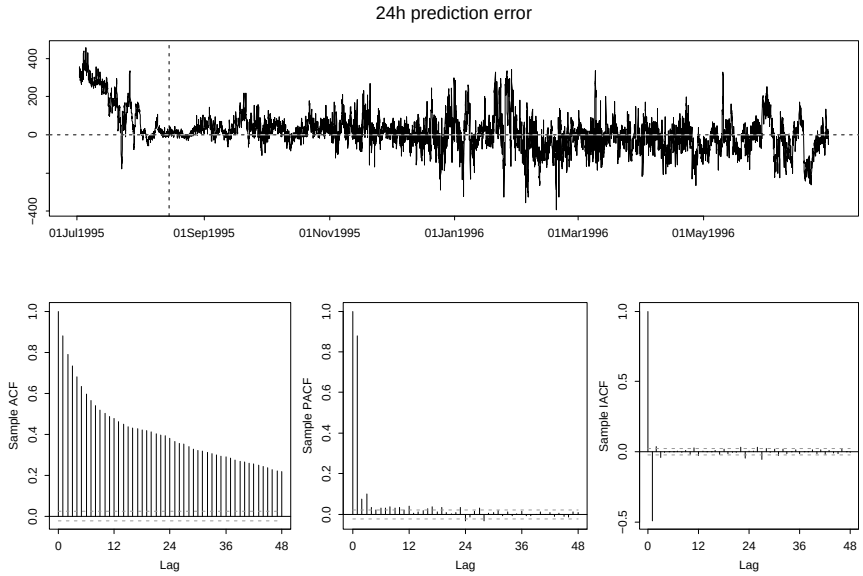

Figure 6.3: 24 hour prediction error and sample correlation functions of the prediction errors. Errors up to August 15, 1995 have not been used in the calculations.

To maximize the predictive ability of the procedure adaptive estimates turned for the

particular prediction horizon will be applied, i.e. for k-step predictions the model

$$r_{t+k} = \phi_k r_t + e_t, \tag{6.4}$$

will be used. Here e_t is the new prediction error and $\phi_k = \phi^k$, the difference being that the estimate is tuned to the specific horizon.

The regression model used is the same as considered in Figure 6.2 and the optimal forgetting factors indicated in this plot is used for adaptive estimation in the regression model. The predictions from the regression model, again assuming the future climate to be known, are then modified by adaptive estimation in (6.4). For (6.4) the forgetting factors $0.99, 0.991, \ldots, 0.999$, $0.9991, 0.9992, \ldots, 0.9999$, and $0.99991, 0.99992, \ldots,$ 0.99999 are investigated. Figure 6.4 shows the root mean square versus the forgetting factor. In general for the shorter prediction horizons the optimum is not very well defined and for the larger horizons forgetting factors outside the interval $[0.998, 0.9998]$ should be avoided. Furthermore, it seems reasonable to use a forgetting factor of 0.999 for all prediction horizons.

For the "optimal" forgetting factors corresponding to Figures 6.2 and 6.4 the root mean square of the prediction errors occurring in the same period as considered in Chapter 5 is 78.85, 86.55, and 87.48 GJ/h for the 24, 48, and 72 hour predictions, respectively.

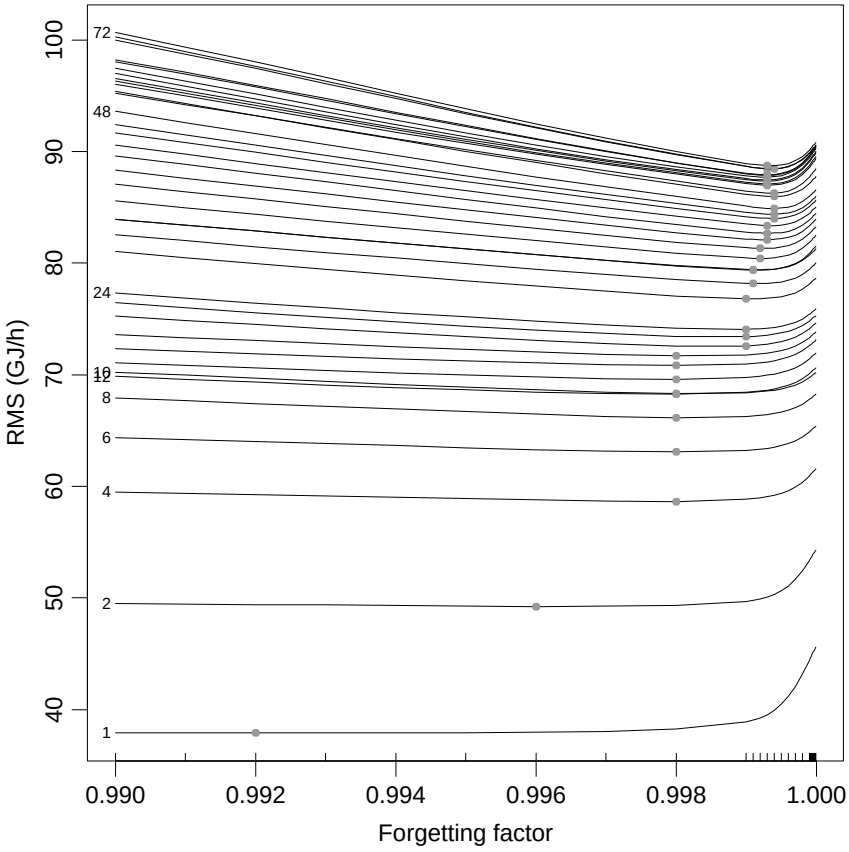

Figure 6.4: Root mean square (RMS) of prediction errors from August 15, 1995 and onwards versus forgetting factor used for adaptive estimation in (6.4), when the optimal forgetting factor from Figure 6.2 is used for adaptive estimation in the regression model. For selected horizons the actual horizon is indicated at the leftmost endpoint of the curve corresponding to that particular horizon. The minimum RMS for each horizon is indicated by a dot. The rugs on the x-axis indicate the forgetting factors used.

6.3 On-line selection of prediction method

As indicated by the previous sections the model described by (6.1), (6.2), (6.3), with a separate diurnal variation for working and non-working days and modelled by a periodic B-spline basis with 11 equidistant internal knots is appropriate for prediction when estimated adaptively. Furthermore, especially for the shorter horizons, it is important to correct the predictions from the regression model by an adaptively estimated autoregressive model as described in Section 6.2.

The quality of the predictions varies with the forgetting factor used for estimation in the regression model and the autoregressive model. Also, it is plausible that it will be beneficial to use different forgetting factors at different points in time, e.g. in the start of autumn when people start turning on the heating systems a relatively small forgetting factor is likely to be beneficial.

In this section ways of varying the forgetting factor is considered. Traditional approaches includes methods such as variable forgetting (Fortescue, Kershenbaum & Ydstie 1981) and selective forgetting (Parkum 1992). Selective forgetting is mainly concerned with the properties of the estimates and is complicated to implement. For these reasons it will not be considered further at this place. Variable forgetting modifies the forgetting factor used at each time step based on the last available prediction error. In this section methods related to variable forgetting will be considered. The methods are more computationally demanding, but allows for more general methods of measuring the current quality of the prediction method. Furthermore, the methods allows for the combination of predictions obtained by use of different models, or prediction methods in general.

Here the individual prediction methods are defined by the regression model mentioned in the beginning of this section, estimated adaptively with forgetting factors $0.990, 0.991, \ldots, 0.999$. The predictions based on each of these forgetting factors are then corrected using an autoregressive model of order one estimated adaptively with a forgetting factor of 0.999, cf. Section 6.2. However, the methods suggested applies to arbitrary prediction methods.

6.3.1 Definition of methods

Plots like e.g. Figure B.1 displays the performance over time for different methods of prediction. As described by Skouras & Dawid (1998) two prediction methods are equiv-

alent if a plot of the cumulative squared prediction error against time shows *parallel* curves. This also applies locally in time and from Figure B.1 it is e.g. evident that before December 15, 1995 the low forgetting factors of 0.99 and 0.995 are superior to 0.999 and from this date until ultimo January, 1995 the high forgetting factor is superior. Other measures of performance, e.g. plots of the cumulative absolute prediction errors, may also be applied.

However, for on-line applications the evaluation of prediction performance can only be based on past prediction errors for the simple (constant forgetting factor) prediction methods. Let $\tilde{Q}^{(i)}_{t|t-k} = Q_t - \hat{Q}^{(i)}_{t|t-k}$ denote the error observed at time t in predicting the heat consumption Q_t, k step before when using the prediction method indexed by i. The adaptive measure of performance used in this report is then found by exponential smoothing of the squared prediction errors, i.e.

$$MS^{(i,k)}_t = (1/n_{ms})(\tilde{Q}^{(i)}_{t|t-k})^2 + (1 - 1/n_{ms})MS^{(i,k)}_{t-1}, \tag{6.5}$$

where the n_{ms} is the effective number of hours behind the result. The recursions are initiated by setting $MS^{(i,k)}_1$ equal to the square of the first available prediction error.

Based on the adaptive measure of performance just defined two methods of generating the actual k-step predictions are suggested. The most simple approach is to use the predictor $\hat{Q}^{(i)}_{t+k|t}$ for which $MS^{(i,k)}_t$ attains a minimum over i. This method will be denoted *on-line selection of prediction method*. However, several prediction methods may have values of $MS^{(i,k)}_t$ near the minimum. Therefore a method denoted *on-line weighting of predictors* is suggested. With this method the actual prediction is generated as

$$\hat{Q}_{t+k|t} = \frac{\sum_i \hat{Q}^{(i)}_{t+k|t}/MS^{(i,k)}_t}{\sum_i 1/MS^{(i,k)}_t}, \tag{6.6}$$

i.e. as an weighted average of the individual predictions, with weights corresponding to the inverse of the most recent measure of performance. It is noted that by appropriate definition of weights both approaches can be described by an equation similar to (6.6). This also points towards a larger set of weighting schemes all based on an adaptive measure of performance.

6.3.2 Comparison of methods

Figures B.7, B.8, and B.9, starting on page 124, shows the cumulative squared 24, 48, and 72 hour prediction errors obtained using on-line selection of prediction method and n_{ms} in (6.5) corresponding to $3, 5, \ldots, 11$ days. In the figures the results obtained using

the constant forgetting factors are also displayed. For on-line weighting of predictors the results for 24, 48, and 72 hour predictions are depicted in Figures B.10, B.11, and B.12, respectively.

For 24 hour predictions both methods are superior to the predictions obtained using constant forgetting factors and both quite insensible to the choice of n_{ms}. For the 48 hour predictions this is also true for the selection method, but the weighting method seems inferior in that the cumulative squared prediction errors are not consistently near the minimum over the predictions obtained using the constant forgetting factors. However, the curve corresponding to the weighted predictors has actual lower slope in the second half of the period than any of the other predictors and therefore it is superior in this period, but inferior during a short period in the beginning of 1996. Finally, for the 72 hour predictions the weighted predictors are mostly superior to the predictors obtained using constant forgetting factors. This is not true in the second half of the period for the selection method. For all prediction horizons it is evident that the weighting method is less sensible to the choice of n_{ms} than the selection method.

Figures 6.5, 6.6, and 6.7 compare cumulative squared 24, 48, and 72 hour prediction errors of both methods ($n_{ms} = 11 \times 24h$) together with the results obtained using the constant forgetting factor for which the final value of the curves in Figures B.7, B.8, and B.9 is lowest. Note that this comparison is biased towards selecting a constant forgetting factor as the best method of prediction. This is because the constant forgetting factor is selected based on the entire set of data.

It is seen that for the 24 and 48 hour predictions the selection method performs well in the difficult short period of time in the beginning of 1996. For data in 1996 the difference between the curves corresponding to selection and weighting decrease and overall the weighting method is superior to both the selection and constant forgetting factor methods for all prediction horizons. Since, it is rather insensible to the choice of n_{ms} and considering the bias in the comparison just mentioned it seems advantageous to use the weighting method for generating the predictions.

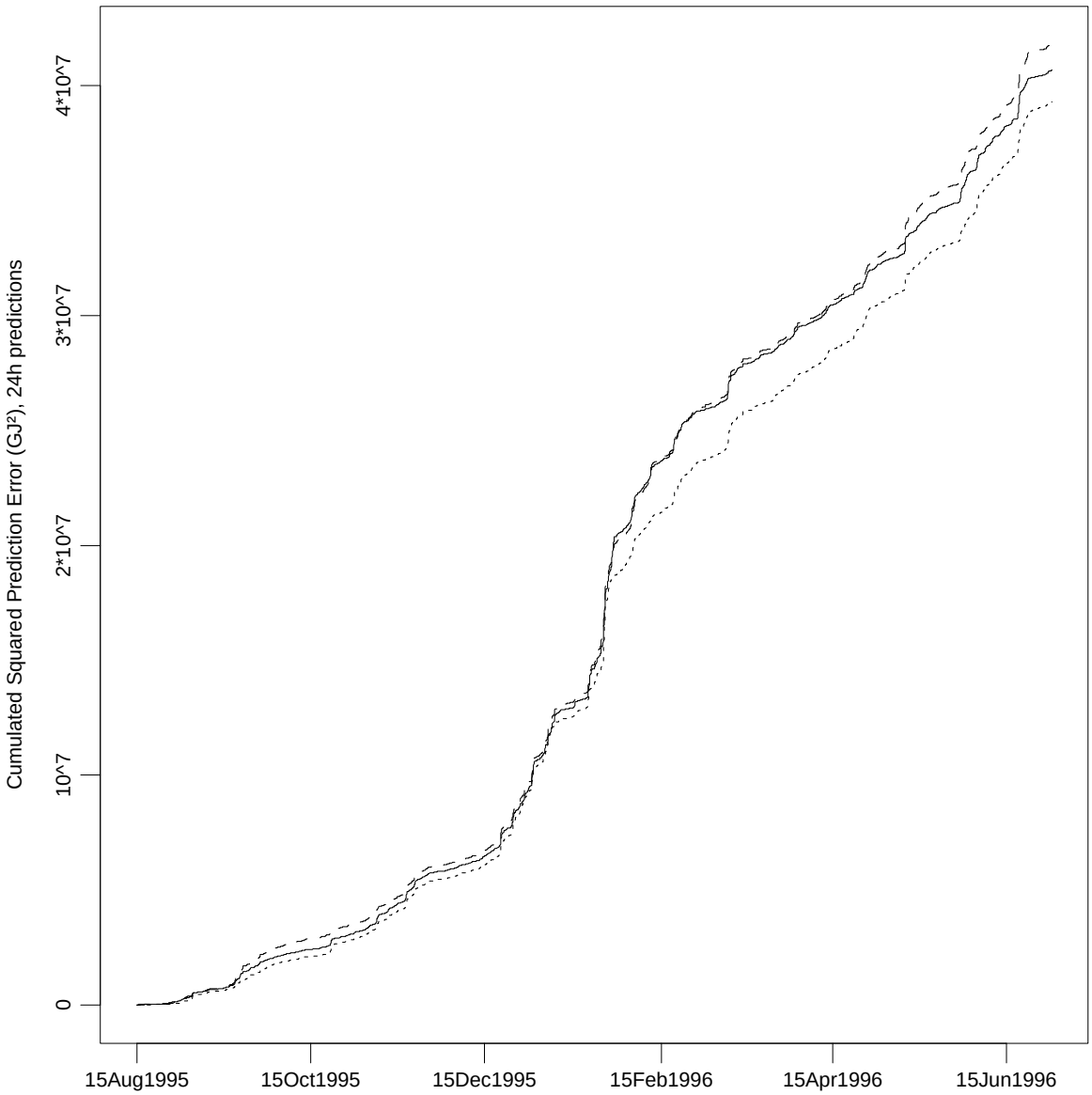

Figure 6.5: Cumulative squared 24 hour prediction errors for on-line weighting of prediction methods (solid), on-line selection of prediction method (dotted) and best constant forgetting factor (0.996, dashed). Prediction errors occurring before August 15, 1995 are disregarded.

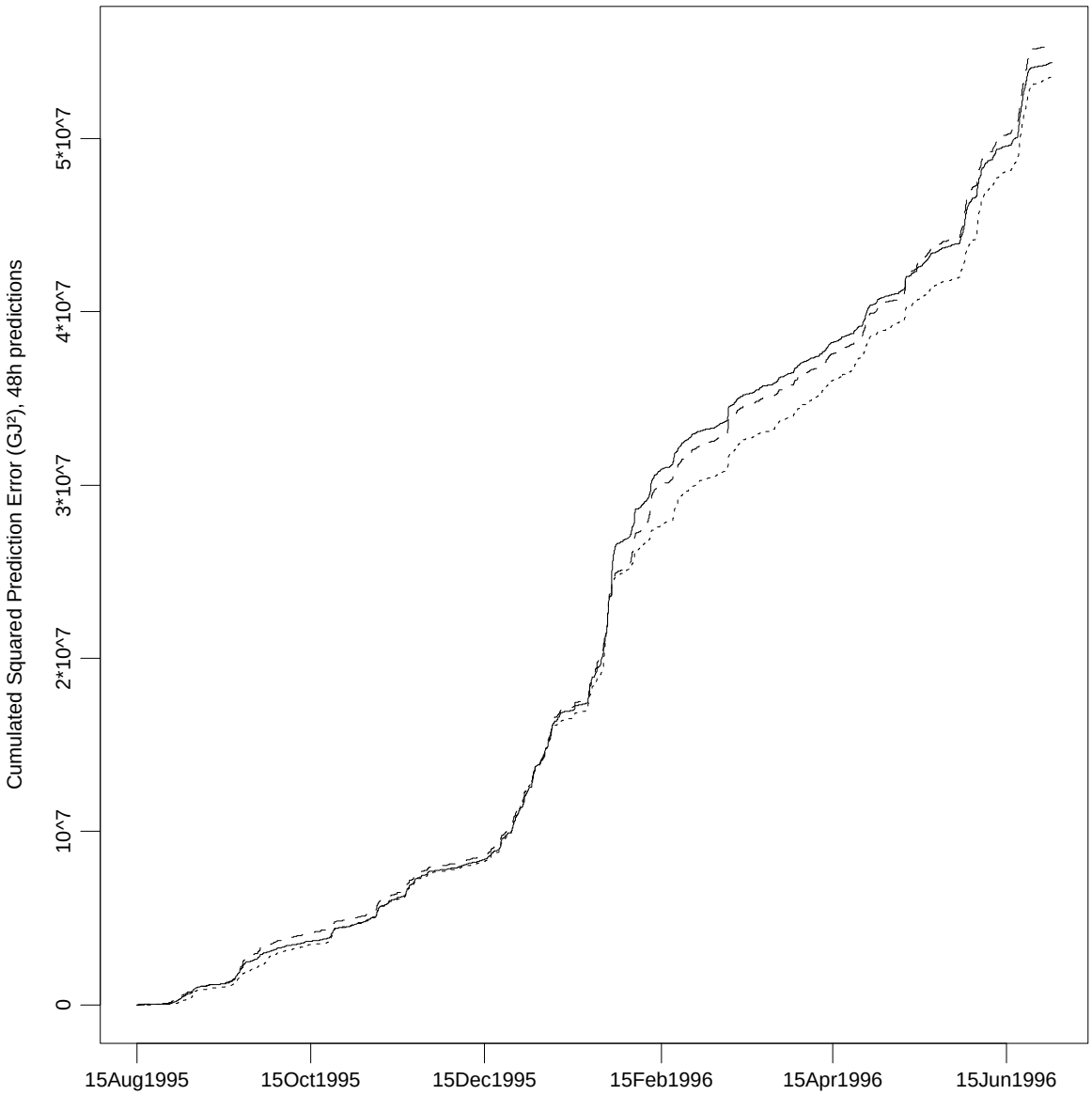

Figure 6.6: Cumulative squared 48 hour prediction errors for on-line weighting of prediction methods (solid), on-line selection of prediction method (dotted) and best constant forgetting factor (0.996, dashed). Prediction errors occurring before August 15, 1995 are disregarded.

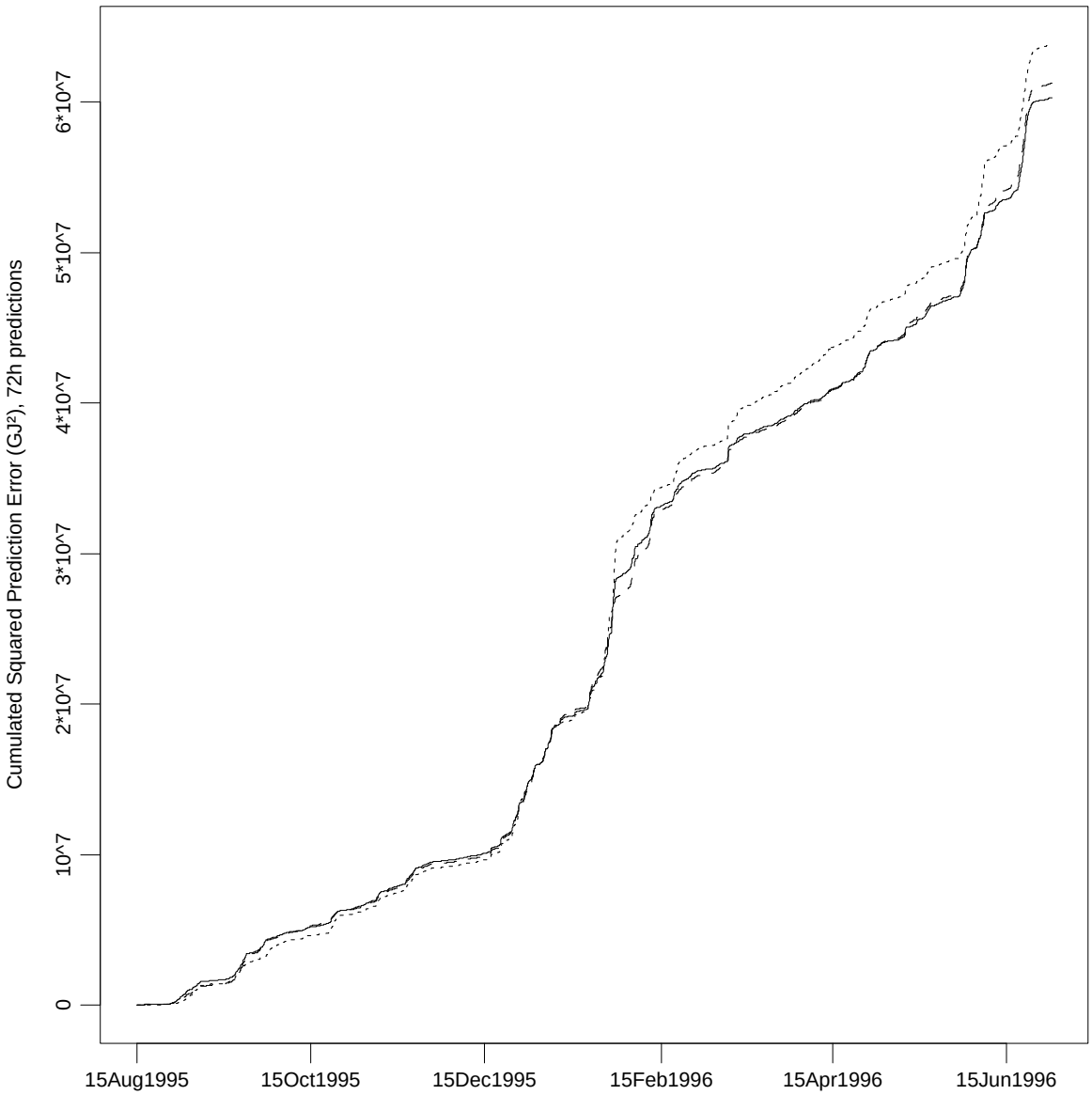

Figure 6.7: Cumulative squared 72 hour prediction errors for on-line weighting of prediction methods (solid), on-line selection of prediction method (dotted) and best constant forgetting factor (0.996, dashed). Prediction errors occurring before August 15, 1995 are disregarded.

6.4 Simplified use of climate data

In this section the effect on the prediction performance of each of (i) replacing the transformed solar radiation, cf. Chapter 2, with the direct measurement of global radiation, (ii) exclusion of solar radiation, (iii) exclusion of wind speed, and (iv) exclusion of both solar radiation and wind speed are assessed. In all cases the models are estimated adaptively using exponential forgetting with forgetting factors $0.99, 0.991, \ldots, 0.999$ and subsequently the weighting of prediction methods described in Section 6.3 are applied ($n_{ms} = 11 \times 24$). Each of these four methods of prediction are compared with the original method, i.e. the same weighting method as described above applied to the model defined by (6.1), (6.2), (6.3) and with a separate diurnal variation for working and non-working days and modelled by a periodic B-spline basis with 11 equidistant internal knots. Furthermore, in all cases, before the weighting is applied the predictions are corrected by an adaptively estimated autoregressive model using a forgetting factor of 0.999, as described in Section 6.2. In the analysis the future climate is assumed known.

For 24, 48, and 72 hour predictions the cumulative squared prediction errors are depicted in Figures 6.8, 6.9, and 6.10, respectively. It is seen that there is a slight loss in using the global radiation without transformation and for the 24 hour prediction horizon it performs marginally better. This is probably due to the relatively low forgetting factors used, whereby the annual variation of the orbit of the sun over the firmament can be corrected for by adapting the estimates related to the diurnal variation and to the global radiation.

Apparently the exclusion of the wind speed has a very significant effect on the prediction performance. However, a large fraction of the difference on the terminal sum of squared errors originates from a very short time interval in the end of January, 1996. An closer investigation of the data shows that during January 27, 1996 the wind speed drops to close to zero and remains so for a relatively long period. Furthermore, in a period up to the date mentioned the wind speed were relatively high. The prediction errors are large (up to 500 GJ/h) in the period January 28-30, 1996. These observations indicates that the part describing the dependence on wind speed in the original model is a good approximation in that it to some extend can handle the change in wind conditions which occurred on the date mentioned above. Consequently, when the wind speed are excluded from the model, the change in wind conditions has to be compensated for by adaption of the overall level of heat consumption.

During January 25-26, 1996 prediction errors of the same magnitude as during January 28-30, 1996 are present. Apparently, during January 25-26 the heat consumption is

higher than would be expected based on the observations of climate.

Overall excluding wind speed, solar radiation, or both will result in an increased slope on the plots of the cumulative squared prediction errors. Indicating that there is an advantage in using all climate variables, assuming that they can be predicted precisely enough. Note that when excluding the solar radiation the overall slope gets larger but in the end of January, 1996 the large prediction errors are absent. This is related to large recordings of solar radiation on January 28, 1996, which apparently is not reflected by the heat consumption.

In Table 6.1 the average slopes of the part of the curves in Figures 6.8, 6.9, and 6.10 starting on February 15, 1996 are listed. Except for the effect of curvature in the plots of the cumulative squared errors the slopes equal the mean squared prediction error. The slopes are found by fitting a strait line by use of least squares to the last part of the curves just mentioned. The same information is included for a number of other predictors; (a) all climate information excluded, (b) PRESS $24h$ predictor, (c) naive predictor, i.e. the prediction obtained by just using the last observed value, and (d) an other simple predictor obtained as the last observed value for a day of the same type. The PRESS predictor is model (8.6) in (Madsen, Nielsen & Søgaard 1996) with solar radiation excluded. Since here the supply temperature is irrelevant it is also excluded. The forgetting factor (0.99851) mentioned on page 118 of the reference just mentioned is used. Note that the naive predictor is only applicable because the prediction horizons are $24h$, $48h$, and $72h$.

The overall conclusion from table Table 6.1 is that the potential advantage from using meteorological forecasts is very large in that the slopes calculated with and without climate information differ approximately by a factor two. The so called simple predictor is inferior in performance to the remaining predictors. This is due to the very long actual prediction horizon used for this predictor, especially for Saturdays, but also in the beginning of the week.

Model	$24h$	$48h$	$72h$
Original	69.7^2	81.1^2	85.5^2
Global radiation	69.1^2	80.4^2	84.9^2
Solar radiation excl.	75.9^2	91.6^2	97.6^2
Wind speed excl.	75.3^2	86.2^2	89.4^2
Solar radiation and wind speed excl.	82.6^2	99.5^2	102.7^2
Climate information excl.	130.8^2	176.0^2	202.4^2
PRESS	129.0^2	–	–
Naive predictor	128.9^2	171.3^2	199.0^2
Simple predictor	184.3^2	218.9^2	238.0^2

Table 6.1: Average slopes in (GJ^2/h) of the part of the cumulative squared prediction error curves starting on February 15, 1996 for 24, 48, and 72 hour predictions. Note that for shorter horizons PRESS preforms much better than the naive predictor (Madsen et al. 1996, p. 118).

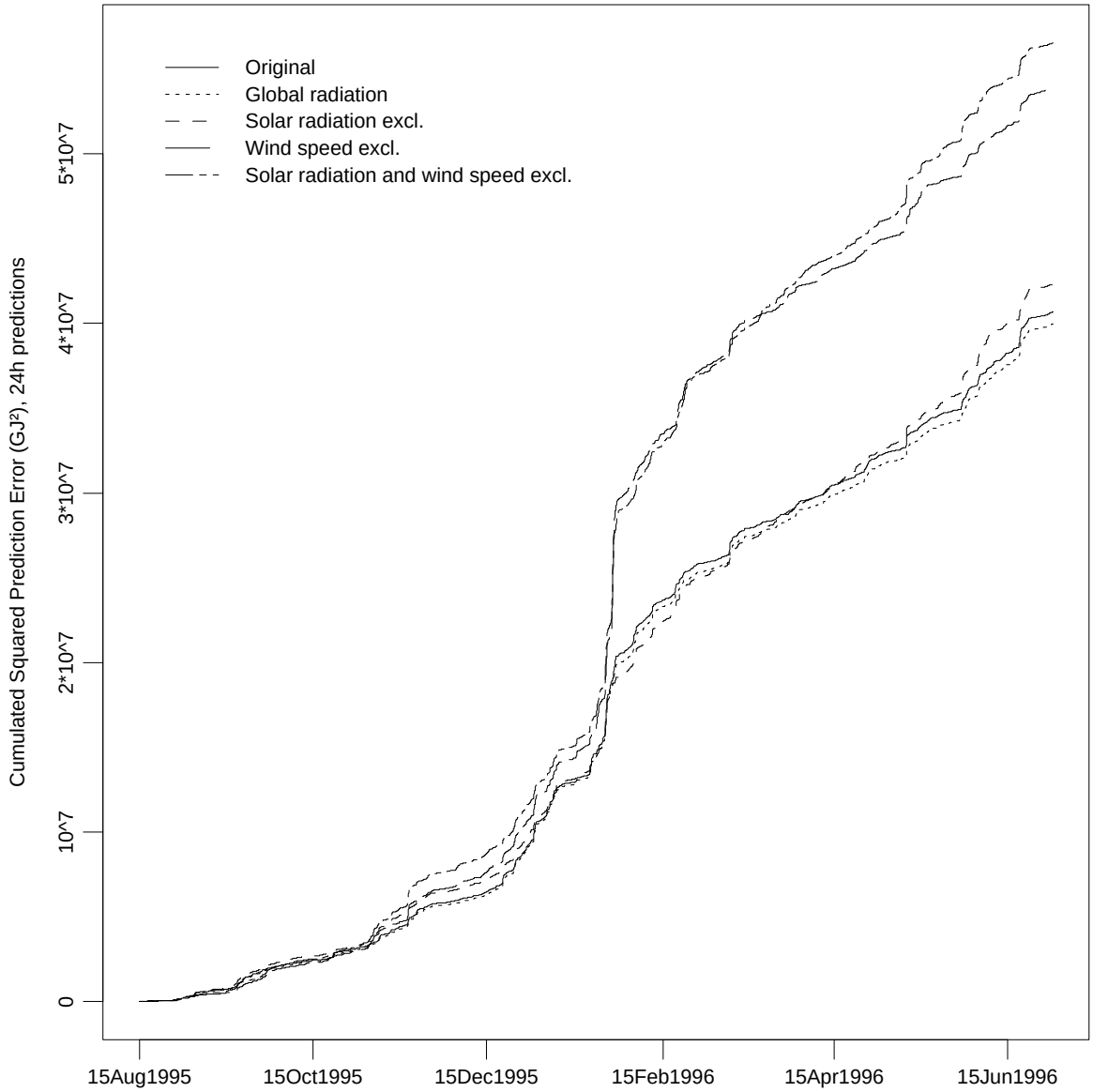

Figure 6.8: The effect of simplified use of climate data on the cumulative squared 24 hour prediction error, assuming known future climate. Prediction errors before August 15, 1995 are excluded.

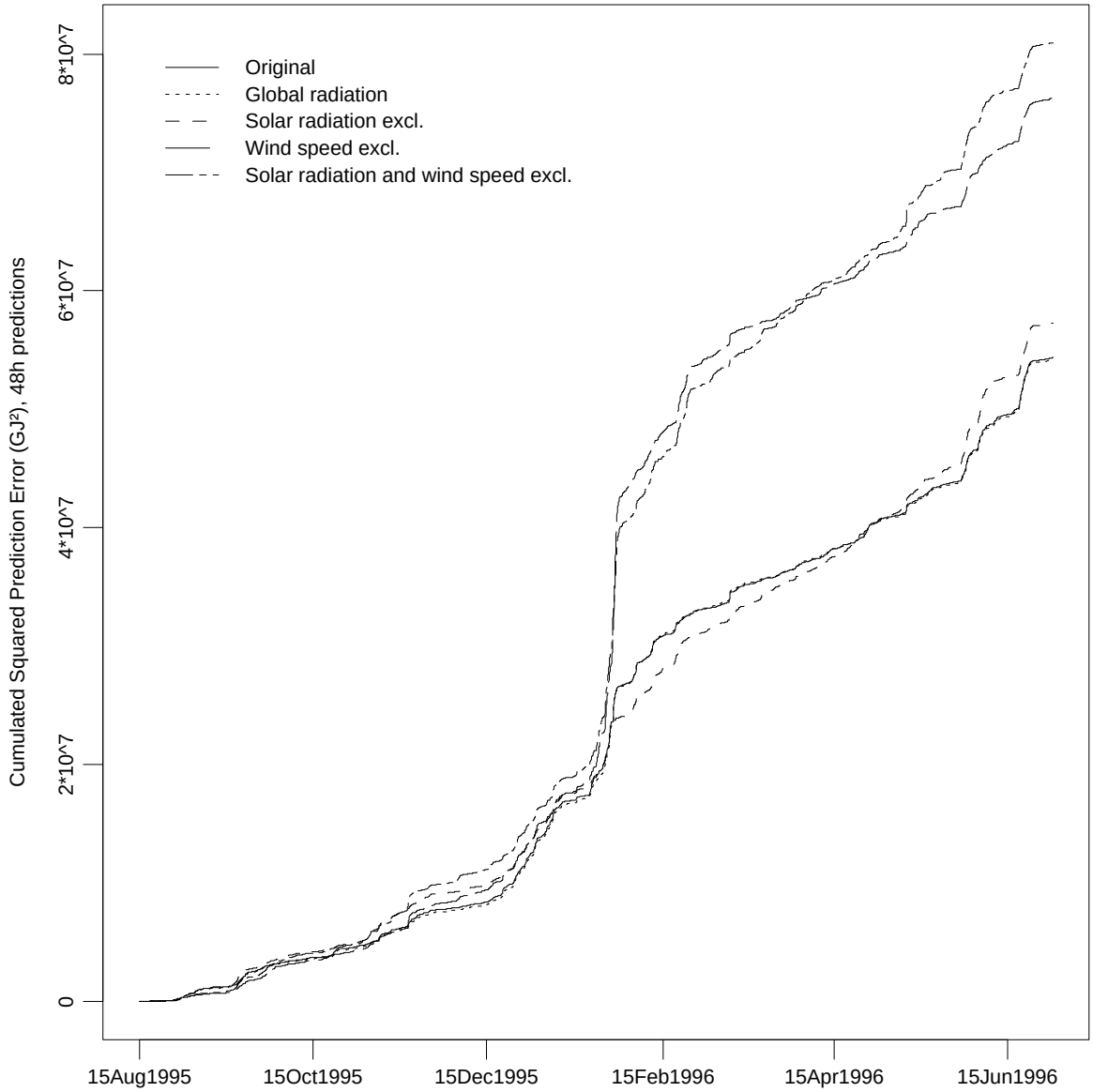

Figure 6.9: (The effect of simplified use of climate data on the cumulative squared 48 hour prediction error, assuming known future climate. Prediction errors before August 15, 1995 are excluded.

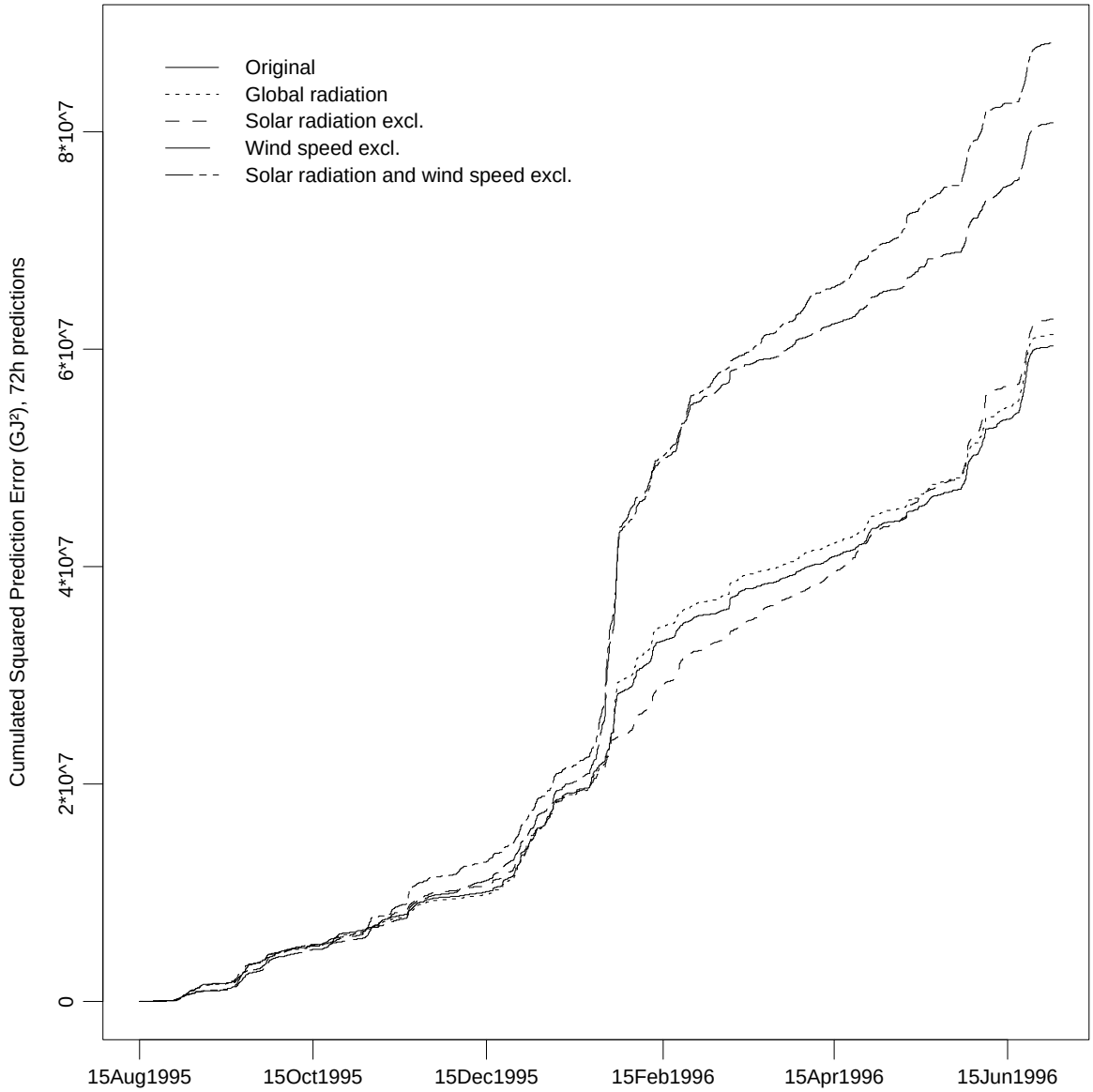

Figure 6.10: The effect of simplified use of climate data on the cumulative squared 72 hour prediction error, assuming known future climate. Prediction errors before August 15, 1995 are excluded.

6.5 Uncertainty on meteorological forecasts

For application of the prediction methods considered so far meteorological forecasts of the climate variables must be available. These forecasts will be associated with uncertainty. In (Nielsen & Amstrup 1998) these uncertainties are addressed. Here the standard deviations of the prediction errors for land stations from DMI-HIRLAM-D (D05) will be used. This particular meteorological model were chosen since the standard deviation do not show fluctuation with the prediction horizon and since it is superior to DMI-HIRLAM-G. However, the conclusions obtained in this section are not sensitive to the particular choice of meteorological forecast method. Figure 6.11 shows the evolution of the standard deviation of the prediction error with the length of the forecast horizon. Note that for prediction horizons below 12 hours the standard deviations of the prediction errors are approximately constant. Therefore it should be possible to improve the short term predictions of the climate variables by correcting the meteorological forecasts using statistical methods and on-line measurements of climate variables. A simple example of this is that for the climate data used in this report predicting the value one hour ahead as the most recent recorded value results in a standard deviation of 0.75^oC and $0.49m/s$ for the temperature and wind speed, respectively. These values are well below the minimum values for the meteorological forecasts.

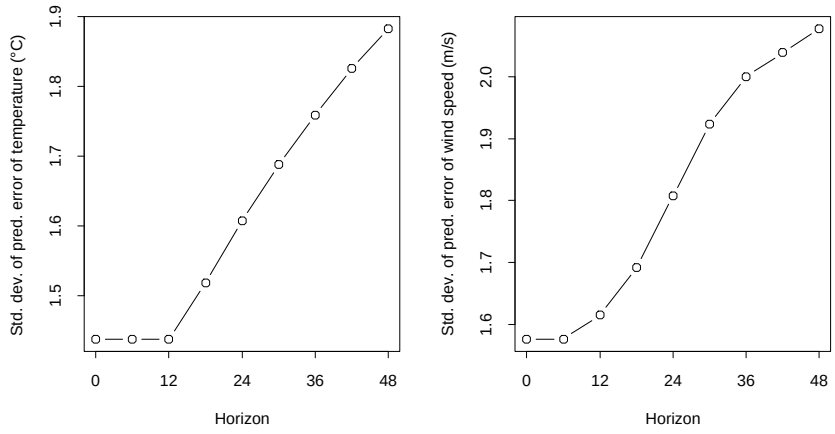

Figure 6.11: Standard deviation of the prediction error for prediction horizons 0-48 hour for the temperature 2m above ground level and the wind speed 10m above ground level.

When using meteorological forecasts it seems appropriate to use the best possible description (model) of the relations between the physical quantities, i.e. the climate variable and the heat consumption. Following this approach we would estimate the

parameters of the model adaptively based on measurements of heat consumption and climate variables as it is done in the previous sections and just use the meteorological forecasts to generate the predictions of the heat consumption. In the following this method will be called the *simple method* of including meteorological forecasts. One obvious problem with this method is that if the mean of the difference between the meteorological forecasts and the observations of climate is non-zero, constant or slowly changing over time, then this systematic bias will be reflected in the predictions of the heat consumption. However, this type of error can easily be corrected for by adaptive correction of the predictions as described in Section 6.2. A more serious problem with the simple method is that it will assign weight to the climate variables even if these are predicted so poorly that they can be considered unknown.

Jonsson (1994) considered these aspects for linear regression and compared different types of estimates with respect to the resulting prediction performance. He showed that it is often best to use the uncertain predictions of the independent variables (meteorological forecasts in this case) as data when estimating the parameters of the model. In this way the influence of the meteorological forecasts are included taking into account a combination of their uncertainty and their relative importance. Note that in this case the estimates of the model parameters are biased. For this reason care should be taken when interpreting components of the model. For instance the estimates can not be used for investigating the dependence of heat consumption on the air temperature.

Applying the concept in the present context implies that the adaptive estimates must be calculated separately for each prediction horizon. If the method described in Section 6.3 is used ten adaptive estimates are generated for each prediction horizon, resulting in 720 adaptive estimates when predictions up to 72 hours are to be generated. In the following this method will be called the *advanced method* of including meteorological forecasts. Note that this method does not pose a computational problem since in (Nielsen, Madsen & Nielsen 1999) approximately 1000 sets of adaptive estimations are run in parallel, and this is able to run 300 times real time on a 450 MHz Pentium III PC running Linux. Although the models in (Nielsen, Madsen & Nielsen 1999) contain fewer parameters than the models considered here the sampling interval is only half an hour opposed to one hour here.

Since, in the models used, the climate variables are filtered (see e.g. page 73) the correlation of the prediction errors will be important. However, only the standard deviation of the prediction errors for up to 48 hour horizons, cf. Figure 6.11, are known to us. Therefore, simulations are performed for different values of the correlation. Actually, the correlation has to be split in at least three types:

1. Correlation between prediction errors of meteorological forecasts originating from the same point in time and related to the same climate variable.

2. Correlation between prediction errors of meteorological forecasts sharing the same prediction horizon and related to the same climate variable.

3. Correlation between prediction errors of meteorological forecasts sharing the same prediction horizon and originating from the same point in time, but related to different climate variables.

To simplify the simulations a model in which both wind speed and global radiation is excluded is used. Therefore item 3 above can be disregarded.

Let $\tilde{T}_{a,t|s};\ t > s$ denote the prediction error of the meteorological forecast issued at time s of the air temperature at time t. A matrix of these prediction errors are then simulated as described in Appendix C and has the following structure:

$$\begin{bmatrix} \tilde{T}_{a,1|0} & \tilde{T}_{a,2|0} & \dots & \tilde{T}_{a,k|0} \\ \tilde{T}_{a,2|1} & \tilde{T}_{a,3|1} & \dots & \tilde{T}_{a,k+1|1} \\ \vdots & \vdots & & \vdots \\ \tilde{T}_{a,t+1|t} & \tilde{T}_{a,t+2|t} & \dots & \tilde{T}_{a,t+k|t} \\ \vdots & \vdots & & \vdots \\ \tilde{T}_{a,N-k+1|N-k} & \tilde{T}_{a,N-k+2|N-k} & \dots & \tilde{T}_{a,N|N-k} \end{bmatrix}, \tag{6.7}$$

where k is the prediction horizon considered (in hours) and the data used in this work is indexed as $1, 2, \dots, N$. Element $(1, 1)$ corresponds to the first data value and element $(N - k, N)$ to the last data value. The prediction errors are simulated such that

- they are normally distributed with zero mean,
- $\sqrt{Var[\tilde{T}_{a,t+k|t}]}$ follows the curve for the temperature $2m$ above ground level as shown in Figure 6.11 on page 92,
- the correlation between neighbour prediction errors in a row of (6.7) is constant ρ_r,
- the correlation between $\tilde{T}_{a,t+j|t}$ and $\tilde{T}_{a,t+i|t}$ depends on $|i-j|$ only and follows the auto correlation function of an AR(1) process with the pole in ρ_r, and
- each column in (6.7) is an AR(1) process with the pole in ρ_c.

Let $\hat{T}_{a,t+k|t} = T_{a,t+k} - \tilde{T}_{a,t+k|t}$ denote the simulated k-step meteorological forecast. According to the model $T^{(f)}_{a,t+k} = H_1(q)T_{a,t+k}$ must be used for predicting the heat consumption at time $t+k$. However, using meteorological forecasts, the best we can do is to use

$$\hat{T}^{(f)}_{a,t+k} = \sum_{s=0}^{k-1} h_s \hat{T}_{a,t+k-s|t} + \sum_{s=k}^{\infty} h_s T_{a,t+k-s}, \tag{6.8}$$

where $h_s; s \geq 0$ is the impulse response of $H_1(q)$. For simulations it is possible to calculate these values as

$$\hat{T}^{(f)}_{a,t+k} = T^{(f)}_{a,t+k} - \sum_{s=0}^{k-1} h_s \tilde{T}_{a,t+k-s|t}. \tag{6.9}$$

Since appropriate values of ρ_r and ρ_c are unknown to us different values will be investigated for both the simple and advanced method of including meteorological forecasts. Furthermore, the results will be compared for both the case where the predictions are adjusted by an autoregressive model as described in Section 6.2, and the case where this adjustment is not performed. On all plots of cumulative squared prediction errors a curve corresponding to known future air temperature will be included for reference. Also for all these plots the range on the axis corresponding to the cumulative squared prediction errors is fixed to the interval $[0, 2 \times 10^8]$ GJ^2. The investigations are performed for 24 hour predictions of the heat consumption only and the parameters of the model is estimated adaptively using exponential smoothing with a forgetting factor of 0.995, since for this particular data set this method performs almost as well as the more advanced method of averaging predictors, cf. Section 6.3.

When ρ_r and ρ_c are both close to one then, except for a constant factor between columns, all values in (6.7) will be of similar size. This will act as a bias compared with the observed air temperature and the simple method will result in large prediction errors due to this bias, but it will be possible to remove a large fraction of this bias if the predictions are adjusted as mentioned above. Therefore, when the predictions are adjusted the simple and advanced method will perform similarly. When ρ_r and ρ_c gets smaller the effect of the adjustment will be smaller and the advanced method will outperform the simple method. These aspects are illustrated for the case $\rho_r = \rho_c$ in Figure 6.12. From the figure it is seen that when the correlations are moderately high (0.9) adjustment of the predictions has a fair impact on both the simple and advanced method, but the simple method can not be "saved" by the adjustment. When the correlations are very high (0.999) the simple method performs very poorly without adjustment. With adjustment the methods perform similarly. The reasons for these observations are outlined above. In all cases the advanced method outperforms the simple method of including meteorological forecasts.

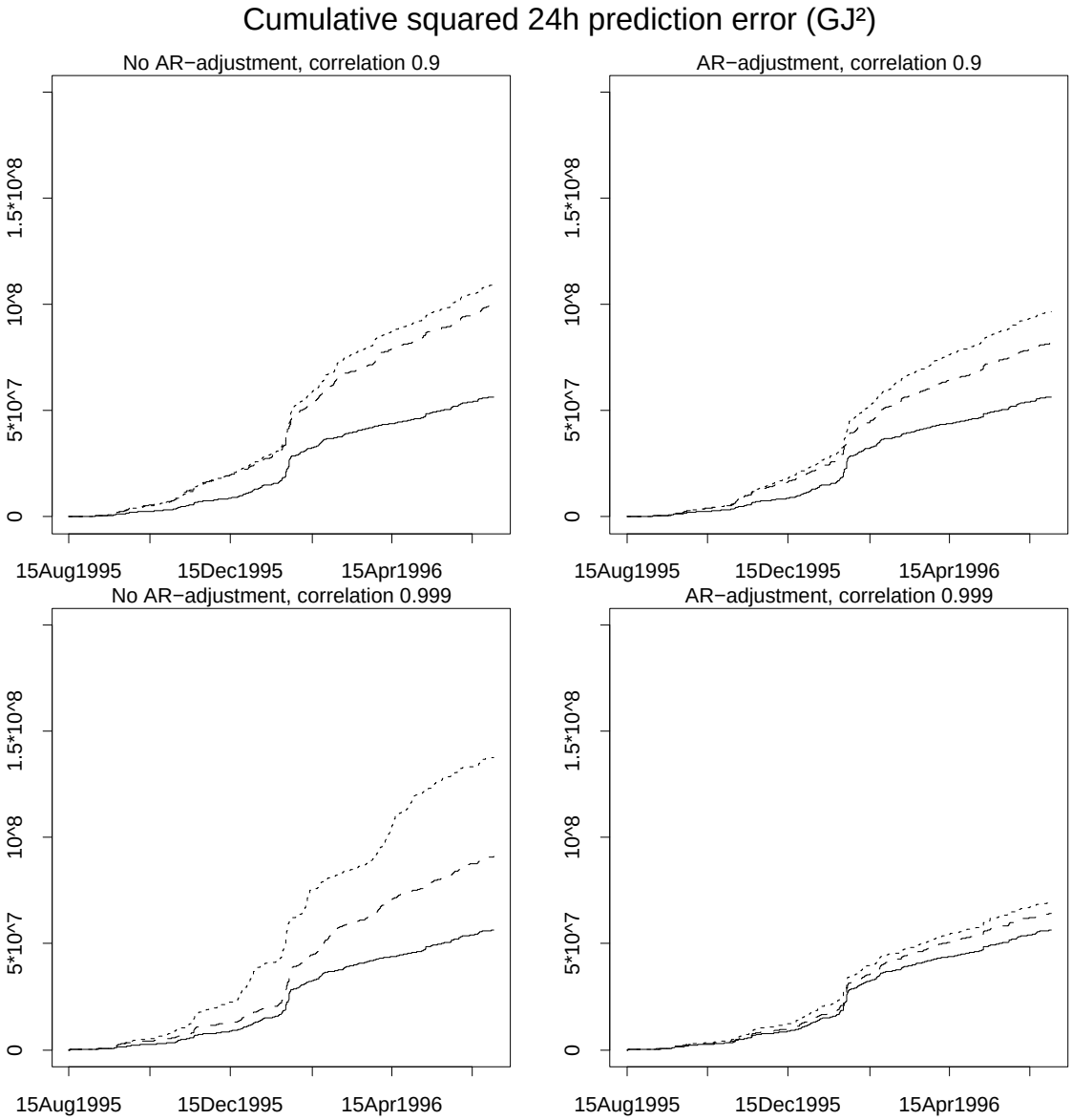

Figure 6.12: Cumulative squared 24 hour prediction error when $\rho_c = \rho_r$ for adjusted and non-adjusted predictions for reference (solid), simple (dotted), and advanced (dashed) methods. Prediction errors before August 15, 1995 are excluded.

In Figures 6.13 and 6.14 the cumulative squared prediction errors for the four combinations of $\rho_r = 0, 0.95$ and $\rho_c = 0, 0.95$ are shown for the simple and advanced method, respectively. It is seen that the performance of the simple method is very sensitive to the values of ρ_r and ρ_c, whereas the advanced method is not affected very much by the values of the correlations. Overall, the advanced method of including meteorological forecasts clearly outperforms the simple method.

Finally, for the advanced method Figure 6.15 shows the root mean square (RMS) of the prediction errors on August 15, 1995 and onwards for simulations in which ρ_c and ρ_r is varied over $0.05, 0.15, \ldots, 0.95$, i.e. 100 simulations in all. It seems that especially when the correlations are high there is some variation from one simulation to the other. In order to more clearly reveal the main features of the dependence of RMS on ρ_c and ρ_r the surface is smoothed using local linear regression with a 30% nearest neighbour bandwidth. It is seen that the value of RMS increase with approximately the same amount for both types of correlation, but the increase in RMS with increasing ρ_c and ρ_r seems to be larger than expected had the effects been additive.

From the results presented in this section it is concluded that the advanced method of including meteorological forecasts, i.e. using the meteorological forecasts when adaptively estimating the parameters of the model, should be applied. The method is not very sensitive to different correlation structures of the meteorological forecasts. As it is expected from theoretical considerations (Jonsson 1994) the advanced method is at least as good as the simple method of including meteorological forecasts, i.e. using the observations when adaptively estimating the parameters of the model, and in many cases the simple method is severely outperformed by the advanced method.

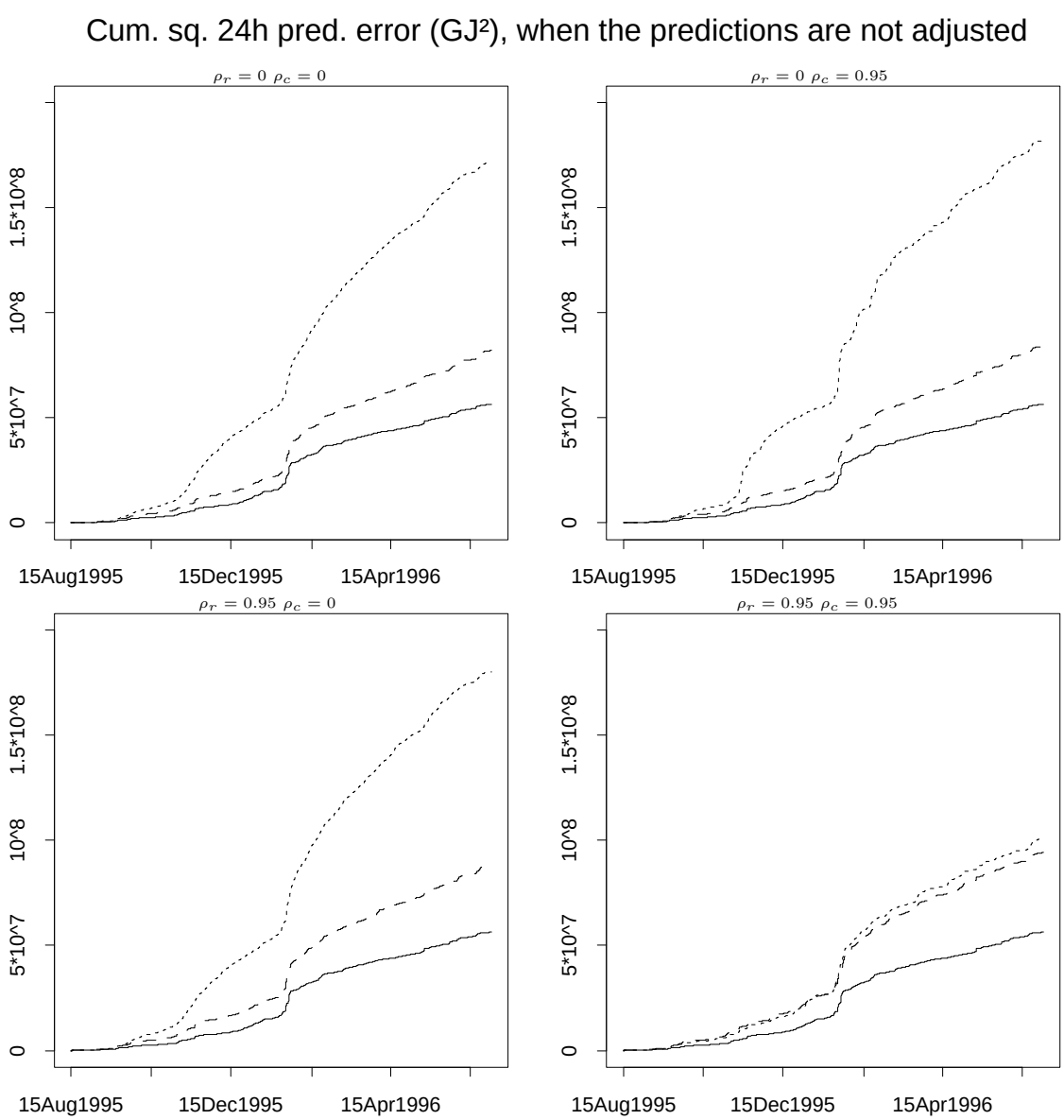

Figure 6.13: Cumulative squared 24 hour prediction error for unadjusted predictions for reference (solid), simple (dotted), and advanced (dashed) methods. Prediction errors before August 15, 1995 are excluded.

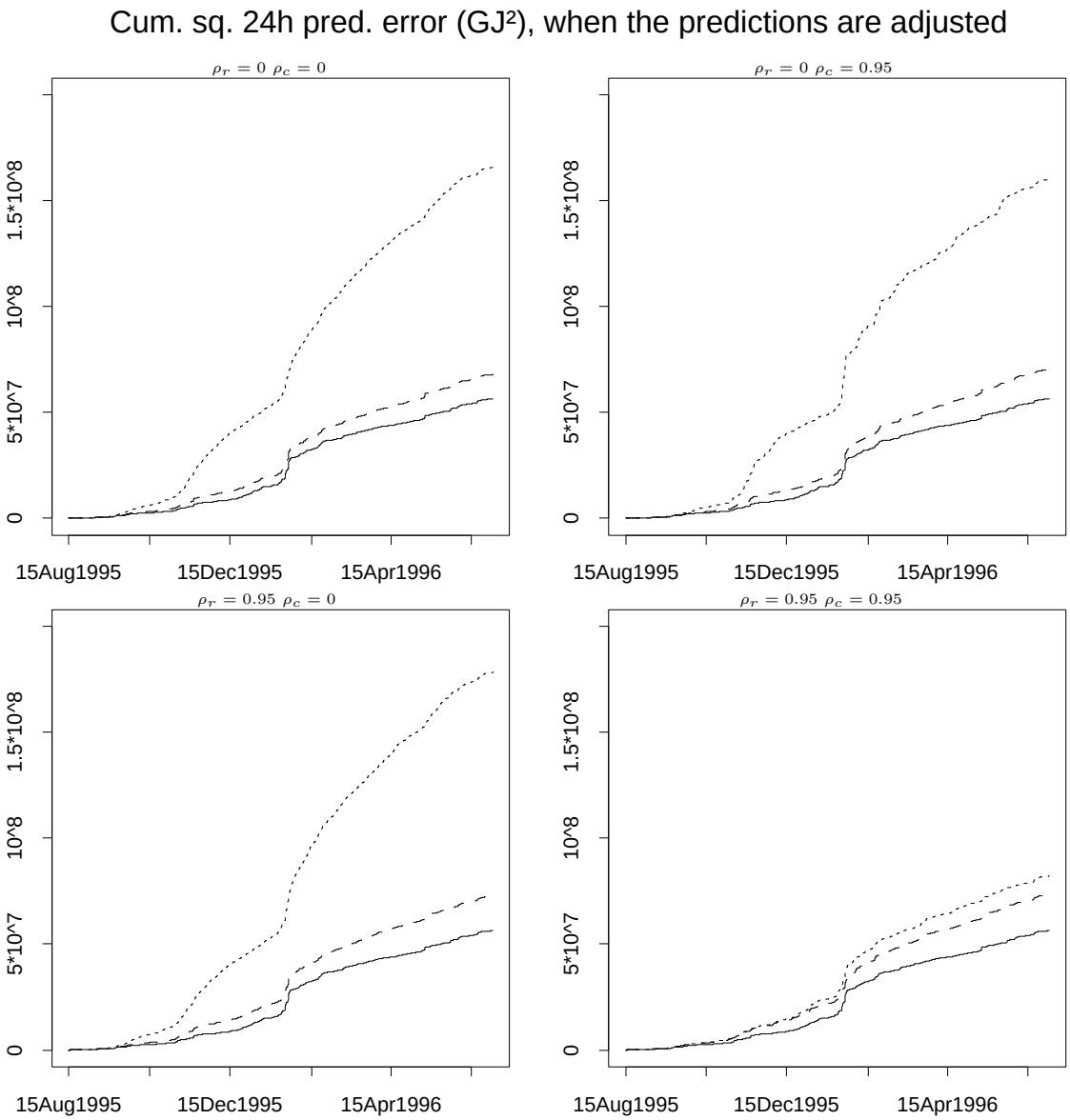

Figure 6.14: Cumulative squared 24 hour prediction error for adjusted predictions for reference (solid), simple (dotted), and advanced (dashed) methods. Prediction errors before August 15, 1995 are excluded.

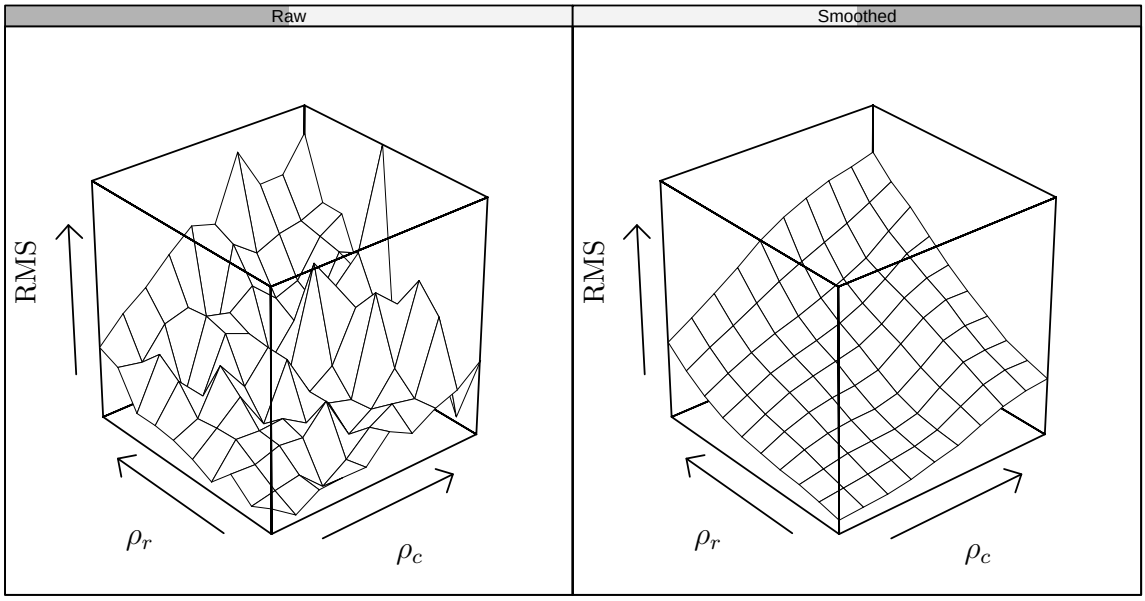

Figure 6.15: Raw and smoothed root mean square surface (24 hour predictions) for the advanced method with adjusted predictions. Prediction errors before August 15, 1995 are excluded. The range of the RMS-values is 93.6 to 102.1 GJ/h.

Chapter 7

Suggestions for implementation

Based on the work described in this report this chapter presents some suggestions for implementation of on-line prediction of the heat consumption. It will be assumed that meteorological forecasts of air temperature, wind speed, and global radiation are available on-line. Furthermore it is assumed that appropriate interpolation methods are implemented so that in the context considered here meteorological forecasts can be treated as arriving every hour with a one-hour resolution of the prediction horizon, see (Nielsen, Madsen & Nielsen 1999). The maximal prediction horizon which may be considered is determined by the maximal prediction horizon of the meteorological forecasts. Although the meteorological forecast may be uncertain for horizons larger than, say, three days the methods suggested in this report will still be able to adapt to slow changes in the overall level and diurnal variation and to the uncertainty of the meteorological forecasts. Hence, the methods will retain some predictive ability also when the precision of the meteorological forecasts is low. Preprocessing of on-line measurements of heat consumption, air temperature, wind speed and global radiation might be necessary to correct for errors which might otherwise deteriorate the quality of the predictions of heat consumption, cf. the plot of the uncorrected data on heat consumption in Figure A.1 on page 114. It will also be assumed that such methods for on-line correction of measurements are available.

The system considered in this report is the transmission system of Vestegnens Kraftvarmeselskab I/S (VEKS), which supplies heat to local district heating networks. The hourly measurements of heat consumption are the heat supplied to the local district heating networks during the preceding hour. For many networks supplying heat directly to the end-consumers the measurements of heat consumption will be the heat supplied from the district heating plant(s) to the network. If the supply temperature

from the plant varies over time it may be necessary to model the storage of heat and the heat loss in the network. In this case the model used here must be extended to allow for this. This may simply be done as in PRESS (Madsen & Nielsen 1997, Bøhm 1994). However, because of the adaptive estimation, if the supply temperature varies slowly, the model needs not to be extended. For instance, the effective No. of observations behind the estimates when a forgetting factor of 0.994 is used corresponds to one week.

Two types of implementations are suggested. In the so called *advanced implementation* on-line weighting of predictors as defined in Section 6.3 is used and in the *simple implementation* the constant forgetting factor method is applied. For both implementations the advanced method of including meteorological forecasts, as defined in Section 6.5, is used. In the models used in this report the climate variables are filtered to model the dynamics of the buildings. When using meteorological forecasts this cause some problems for on-line systems. This last aspect is considered in Section 7.1, the two types of implementations are considered in Sections 7.2 and 7.3. Finally, in Section 7.4 it is considered how the meteorological forecasts can be improved explicitly or implicitly by use of on-line measurements of the local climate or just by use of the measurements of the heat consumption.

7.1 On-line filtering using meteorological forecasts

Equation (6.8) on page 95 shows how meteorological forecasts of the air temperature for horizons $1, \ldots, k$ are used to generate a prediction of the filtered climate variable at time $t + k$ given information available at time t. The same principle applies for the other filtered climate variables in the model. To recapitulate, in the model a given climate variable x_t is filtered trough a transfer function $H(q)$ and enters the model as $x_t^{(f)} = H(q)x_t = \sum_{s=0}^{\infty} h_s x_{t-s}$ for prediction of the heat consumption at time $t+k$ given information available at time t the meteorological forecasts $\hat{x}_{t+k-s|t};\ s = 0, 1, \ldots, k-1$ is used to generate a prediction of $x_{t+k}^{(f)}$ as in

$$\hat{x}_{t+k|t}^{(f)} = \sum_{s=0}^{k-1} h_s \hat{x}_{t+k-s|t} + \sum_{s=k}^{\infty} h_s x_{t+k-s}. \tag{7.1}$$

This equation contains an infinite sum and can not be implemented directly. However, the largest pole of any of the filters used in the model selected in this report is 0.934. Therefore it is feasible to truncate the infinite sum, store e.g. one week of climate observations, and rerun the filter for every time step and prediction horizon. Depending on the structure of the filter it might be possible to derive an analytical result to the

infinite sum. The approach suggested here is not dependent on the structure of the filter.

7.2 Advanced implementation

It is suggested to use a model as defined by (6.1), (6.2), and (6.3) on page 73. The diurnal variation (including the level) $\mu(h_t^{24}, \Upsilon_t)$ should be modelled using a periodic cubic spline basis with 11 equidistant internal knots as described in Section 5.9, alternatively a harmonic expansion of order five could be used. A separate diurnal variation should be estimated for working and non-working days.

For each prediction horizon the advanced method of including meteorological forecasts as described in Section 6.5 should then be applied together with adaptive recursive estimation and exponential forgetting using forgetting factors $0.99, 0.991, \ldots, 0.999$. With a maximal prediction horizon of 72 hours this amounts to 720 sets of adaptive estimates. Each of the corresponding predictions should then be corrected using a k-step formulation of an AR(1) model estimated adaptively using a forgetting factor of 0.999, cf. Section 6.2. Finally, for each prediction horizon the method called on-line weighting of predictors in Section 6.3 should be used to generate the actual predictions. For the weighting of predictors the sensitivity to n_{ms} in (6.5) on page 81 seems to be low. In this report a value corresponding to 11 days ($n_{ms} = 264$) is found to be a good choice.

7.3 Simple implementation

To simplify the implementation the on-line weighting of predictors as suggested in the previous section can be left out of the implementation. Furthermore, the diurnal variation could be modelled using harmonic expansions instead of the periodic cubic B-spline basis.

With this approach the forgetting factor used for each prediction horizon must be selected. From Figure 6.2 on page 76 it can be deduced that the same forgetting factor can not be used for all horizons. However, in the figure just mentioned the predictions have not been adjusted using an AR(1) model and hence it cannot be used to select appropriate values of the forgetting factors. Figure 7.1 shows the result when the

predictions are adjusted as outlined in the previous section. From this plot it is seen that a forgetting factor of 0.995 or 0.996 is appropriate. However, the uncertainty of the meteorological forecasts might, to some extent, affect the choice of forgetting factor.

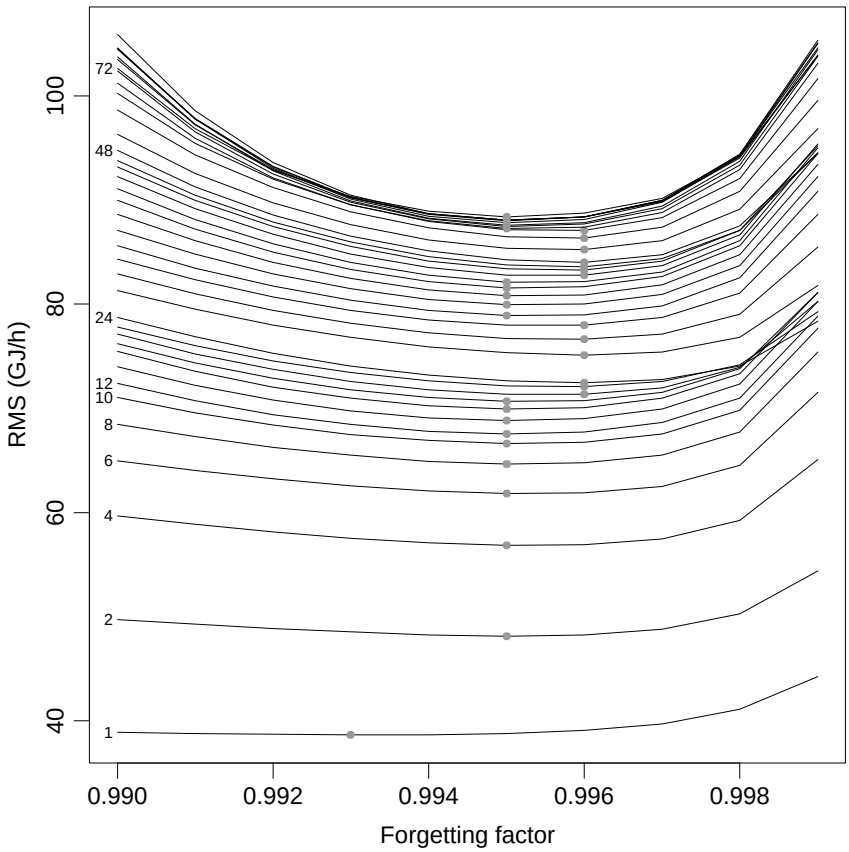

Figure 7.1: Root mean square (RMS) of prediction errors from August 15, 1995 and onwards versus forgetting factor when the predictions are adjusted using an AR(1) model. For selected horizons the actual horizon is indicated at the leftmost endpoint of the curve corresponding to that particular horizon. The minimum RMS for each horizon is indicated by a dot.

7.4 Correction of meteorological forecasts

As noted in Section 6.5 it might be appropriate to adjust the meteorological forecasts for the short prediction horizons. There is at least two ways to do this (i) explicitly by building prediction models / methods based on observations of the climate variables and recordings of meteorological forecasts, and (ii) implicitly by altering the structure of the model(s) used for prediction of the heat consumption. The explicit method will not be considered further at this place since it requires information not available to us. Furthermore, it will be quite time consuming.

Basicly, the implicit method can be thought of as including the structure of the models used in the explicit method into the model of the heat consumption (6.1). However, here a far more simple approach will be suggested. Let $\hat{x}_{t+k|t}^{(0,k-1)} = \sum_{s=0}^{k-1} h_s \hat{x}_{t+k-s|t}$ and $x_{t+k}^{(k,\infty)} = \sum_{s=k}^{\infty} h_s x_{t+k-s}$ whereby $\hat{x}_{t+k|t}^{(f)}$ in (7.1) can be written $\hat{x}_{t+k|t}^{(0,k-1)} + x_{t+k}^{(k,\infty)}$. In model (6.1) it is then suggested to replace each term containing a filtered climate variable with two terms, one corresponding to $\hat{x}_{t+k|t}^{(0,k-1)}$ and one corresponding to $x_{t+k}^{(k,\infty)}$. For instance

$$a_{111}\hat{W}_{t+k|t}^{(f)} = a_{111}(\sum_{s=0}^{k-1} h_s \hat{W}_{t+k-s|t} + \sum_{s=k}^{\infty} h_s W_{t+k-s})$$

is replaced by

$$a_{111,0}\hat{W}_{t+k|t}^{(0,k-1)} + a_{111,k} W_{t+k}^{(k,\infty)}.$$

For terms in the original model containing products of filtered climate variables the model can still be formulated so that it is linear in the parameters. For short horizons k where the meteorological forecasts of e.g. the wind speed should be adjusted by observations of the wind speed, the implicit method will be able to correct for this by placing more weight on $W_{t+k}^{(k,\infty)}$ than on $\hat{W}_{t+k|t}^{(0,k-1)}$. By the use of this method the number of parameters in the model will be increased by nine.

The method just outlined can easily be implemented. However, it must be noted that the performance of it has not been investigated.

Chapter 8

Conclusion and discussion

In this report methods for on-line prediction of the heat load hour by hour in a district heating system are considered for prediction horizons up to 72 hours. It is assumed that meteorological forecasts are available on-line.[1] Theoretical relations known from the theory of heat transfer have been used to select an initial model structure and data on heat consumption and climate (temperature, wind speed, and global radiation) is applied in combination with statistical methods to from an actual mathematical model of the heat consumption. Finally, based on this model adaptive methods for prediction are investigated and approaches to implementation are described.

Actual meteorological forecasts have not been available for the work described in this report. For this reason it is not possible exactly to quantify the expected size of the prediction errors of an on-line application based on this work. Therefore we shall only consider the quality when the climate is assumed to be known and hence indicate the potential for good meteorological forecasts. This is indicated by the plots of the cumulative squared prediction errors as shown in Figures 6.5, 6.6, and 6.7, starting on page 83. Using the weighting of predictors as suggested for implementation and the global radiation as a measure of the solar radiation, the mean absolute relative prediction errors is listed by month in Table 8.1. Also results obtained for the naive predictor and the predictor obtained by excluding climate information, cf. Section 6.4 and Table 6.1 on page 88, is included. Not surprisingly, the method developed is more superior during during cold periods than during warmer. Data for July 1995 is not included since the adaptive estimation used do not settle until about August 15, 1995.

[1] Such a service has recently been introduced by the Danish Meteorological Institute under the term "SAFE-Energy" (Petersen & Hilden 1999), where the forecasts are transmitted via the Internet.

	1995					1996					
	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
24h predictions											
Dev. meth.	6.5	7.6	4.4	3.5	3.3	4.5	3.9	3.1	6.2	7.2	14.3
Naive	5.2	9.5	9.4	11.4	6.8	6.4	7.9	5.9	11.2	14.2	14.9
No climate	6.9	12.0	9.6	11.7	6.4	6.5	8.3	5.7	13.2	14.1	16.7
48h predictions											
Dev. meth.	8.6	9.4	5.2	3.8	3.7	5.5	4.2	3.2	7.2	8.6	18.8
Naive	7.5	12.7	14.1	16.0	8.0	9.0	10.8	6.6	17.5	20.1	21.2
No climate	8.1	14.7	14.8	16.0	7.4	9.6	11.0	6.0	21.5	19.3	23.4
72h predictions											
Dev. meth.	11.2	10.8	6.2	3.8	3.7	5.5	4.4	3.4	7.5	8.8	21.6
Naive	9.2	16.1	17.0	17.0	8.8	10.8	13.1	7.1	21.5	23.3	25.8
No climate	8.9	16.8	17.0	17.1	8.7	11.7	12.9	6.6	26.1	21.3	28.4

Table 8.1: Mean relative prediction error in percent for the method developed assuming the future climate to be known, for the naive predictor, and when excluding climate information.

There are some differences between months. This is further revealed by Figure 8.1 in which the predictive ability of the method developed is summarized by week and by prediction horizon. It is clear that relative prediction errors tend to increase with decreasing heat consumption, whereas the actual prediction errors are large during the winter months. In June 1995 large errors occur, this is due to the drop in heat consumption during the first part of the month. This drop can apparently not be explained by the climate.

The system considered is Vestegnens Kraftvarmeselskab I/S (VEKS) and the variable predicted is the heat supplied from this network. Therefore, loss and storage of heat in the network is not an issue. However, often the variable to be predicted will be the heat supplied from the plant(s) to the district heating system. This variable will be influenced by the supply temperature and the predictions of it should be conditional on the future supply temperature. In such a situation the model used in this report must be extended with a term containing the supply temperature; which can be done as in (Madsen & Nielsen 1997).

Theoretical results such as those presented by Jonsson (1994) indicate that, since meteorological forecasts will contain some degree of uncertainty, the actual meteorological forecasts and not the observations of climate should be used when adaptively fitting the model used for prediction. Although the theoretical results are based on a more

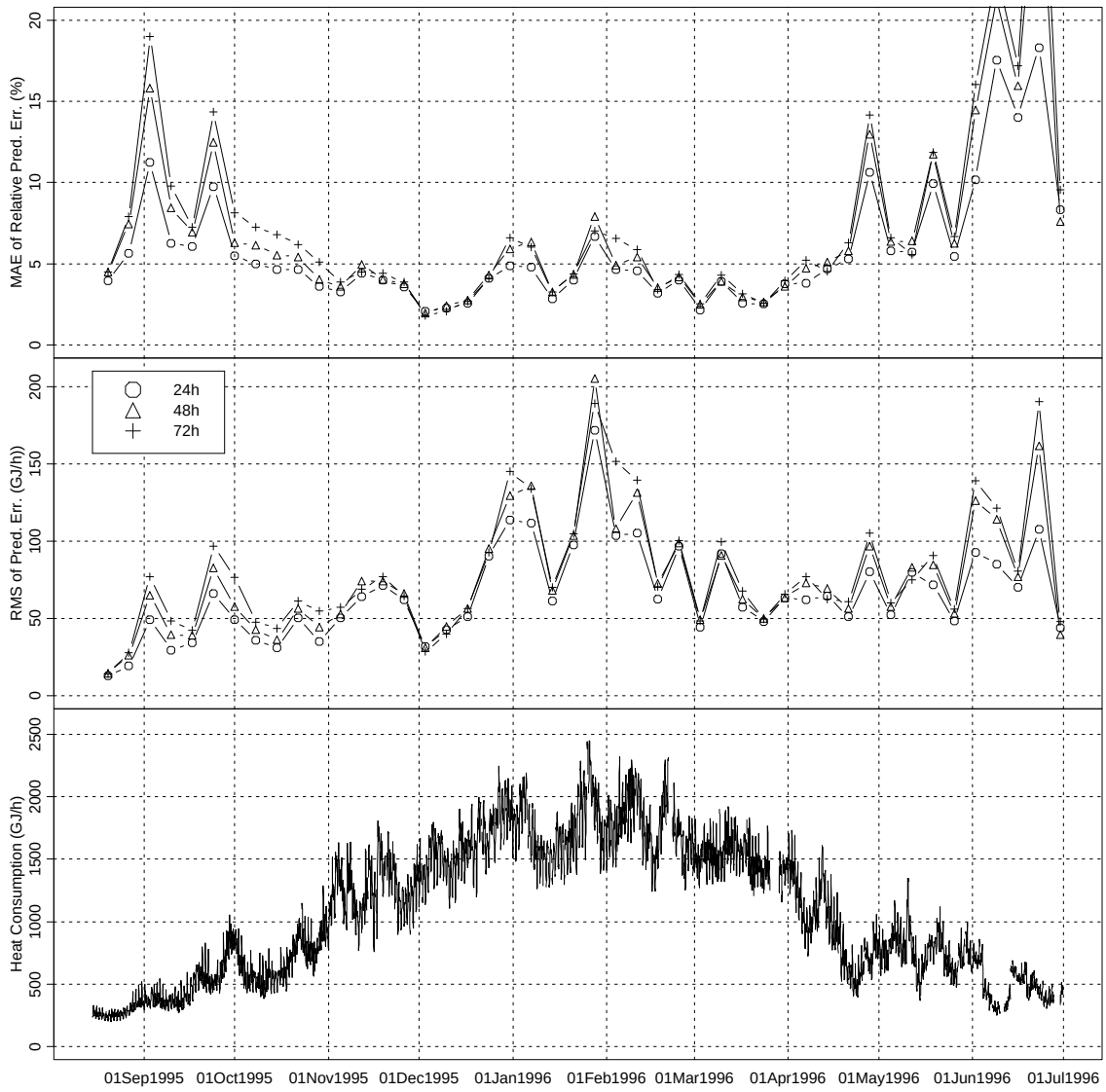

Figure 8.1: Mean relative prediction error by week (top), root mean squared prediction error by week (middle), and heat consumption (bottom). Values for dates before August 15, 1995 are excluded.

simple setup than used in this report it is plausible that the results also apply here. Simulations have been used to verify that using the actual meteorological forecasts is advantageous compared to using the observations. Depending on the characteristics of the meteorological forecasts, which are unknown to us, the use of observations may be severely inferior to the use of the actual meteorological forecasts.

The prediction method contains some filters of which the parameters are not estimated adaptively. These filters are related to the dynamic response of buildings on climate variables. To be able to estimate the parameters of the filters adaptively by use of recursive least squares the model need to be reformulated and written in linear form. Actually, this is not entirely possible due to the product between e.g. filtered wind speed and filtered air temperature. Furthermore, the number of parameters will increase by a factor three and therefore the approach seems inappropriate. Fortunately, the filters are related to the construction of the buildings and therefore they are expected to be applicable to other geographical locations also. The main aspects of the dynamics found in this study are similar to the dynamics found in a study on electricity consumption (Nielsen & Madsen 1997). This further verifies that the filters are appropriate to district heating systems supplying heat to typical Danish houses. For other systems and countries it might be appropriate to change the filters. Based on data, this can be done as described in Chapter 5. Note that for the model structure with the explanatory variables available, i.e. air temperature, wind speed, and solar radiation, an appropriate value of the forgetting factor is approximately 0.995, corresponding to 200 hours of effective observations. This relative low forgetting factor is probably a consequence of the Danish coastal climate. Therefore the prediction method should only be applied in e.g. Central Europe after some investigations and possibly modifications. Furthermore, the prediction errors of the regression part of the model, as shown on e.g. Figure 6.3 (page 77), exhibit slow and irregular variations. This indicates that other explanatory variables than the ones used here might be appropriate. This could be other climate variables such as precipitation or it could be variables related to the operation of the district heating network. The last explanation implies that events occur where the heat consumption is not equal to the heat requested by the consumers. According to VEKS this might actually happen in case of pressure drops. One effect which is not considered in the previous part of the report is the heat loss from the DH-pipes in the networks connected to the VEKS transmission system. This is investigated in Appendix D. Overall it seems to be slightly advantageous to take this aspect into account, except for warm periods.

The method of prediction developed in this report may be applicable to other systems which supply energy for heating. It is quite plausible that the principles can be used as a basis for forming a method for prediction of natural gas consumption. If the pressure in the system varies this requires that a term taking this into account is introduced.

Furthermore, it is possible that the diurnal variation has to be modelled using more parameters than used here, because the peak around 18:00 might be more pronounced for natural gas consumption than for district heating.

To model the correlation of the noise an autoregressive model of order one (AR(1)) is used here. Although an adaptively estimated AR(1) model is expected to account for most of the correlation, the stochastic properties of the prediction errors of the actual meteorological forecasts might result in other models being appropriate. An analysis based on actual meteorological forecasts is required to clarify these aspects.

In the report it is demonstrated that non- and semi-parametric methods are valuable tools for investigating possible non-linearities. In the present case this is the effect of wind speed on the convection heat coefficient corresponding to the outside of the walls. It is concluded, however, that a linear dependence is appropriate for prediction purposes since improvement in the *in-sample* prediction performance is marginal when a general non-linearity is allowed for. However, it has not been investigated if the data contains support for the hypothesis of a non-linear dependence, i.e. the statistical significance of the non-linearity is not addressed.

Modelling a system by use of theoretical knowledge about the system to be modelled together with measurements performed on the system is sometimes called *gray box* modelling. The extremes of this approach is *white box* modelling in which no data is used and *black box* modelling in which no theoretical knowledge about the system is used. For the work described in this report the gray box approach seems to have been successful in that, as shown in Chapter 5, a model with good physical interpretation and residuals resembling white noise is obtained. However, the residuals have non-constant variance. This should be investigated further and possibly modelled. The model just mentioned contains a description of the correlation of the noise. Assuming the innovations to be Gaussian white noise, the noise and regression parameters of the model can be estimated simultaneously using maximum likelihood (ML) estimates. Disregarding the correlation of the noise the regression parameters can be estimated using least squares estimates; which in this case is called output error (OE) estimates. When these two sets of estimates are used to calculate the in-sample long term predictions, i.e. the predictions obtained when not conditioning on previous observations of the heat consumption, the OE estimates are very superior to the ML estimates in terms of root mean square (RMS) of the prediction errors (168.8 vs. 81.5 GJ/h). For one-step predictions the RMS when ML estimates are used is 38.5 GJ/h and the prediction errors based on the OE estimates can be corrected so that 46.7 GJ/h is obtained, cf. Section 5.7.

For prediction purposes the model parameters are estimated adaptively using recursive

least squares with exponential forgetting. As mentioned above this is only possible since some filters in the model are fixed. It might be considered to use the recursive prediction error method (Ljung 1987) for estimating all parameters of the model. However, since the filters are related to the heating of the buildings, during the summer period most of the information about the filters will be weighted down and must be learned again during the first part of the heating season. Therefore it is doubtful that the application of the recursive prediction error method will be successful. Furthermore, on page 54 it is argued that a singularity may exist for the non-linear least squares problem and it is possible that this could be reached, especially, when all previous information is weighted down during the summer period. Unless, very refined precautions are taken this could result in a breakdown of an on-line system after a long period of time in which it had been operating without problems.

Some new methods for on-line selection or weighting of predictions generated by different methods are suggested and investigated. It is concluded that especially the on-line weighting perform well and it is quite insensitive to the tuning parameter.

Appendix A

Plots of data

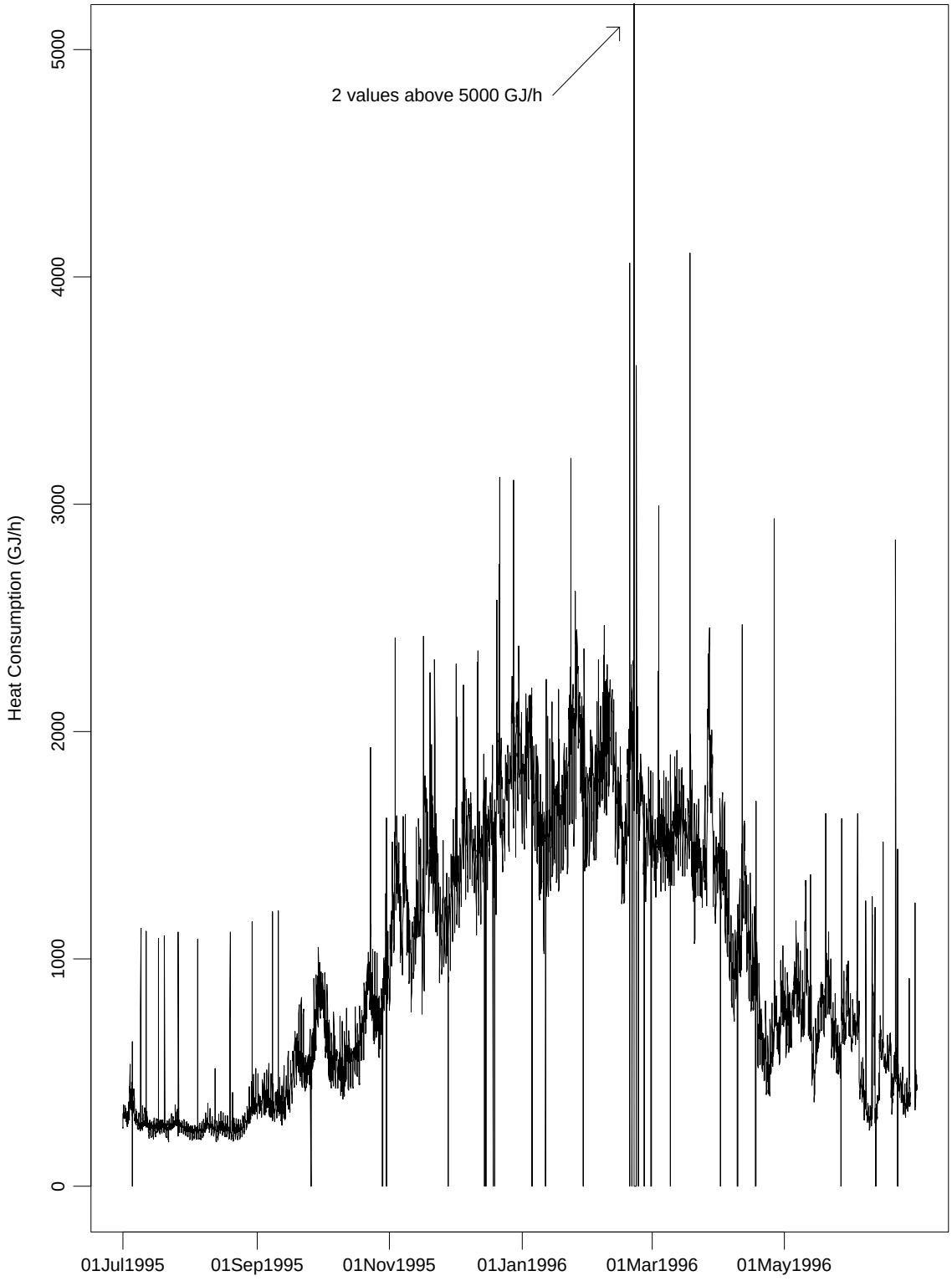

Figure A.1: Raw measurements of heat consumption.

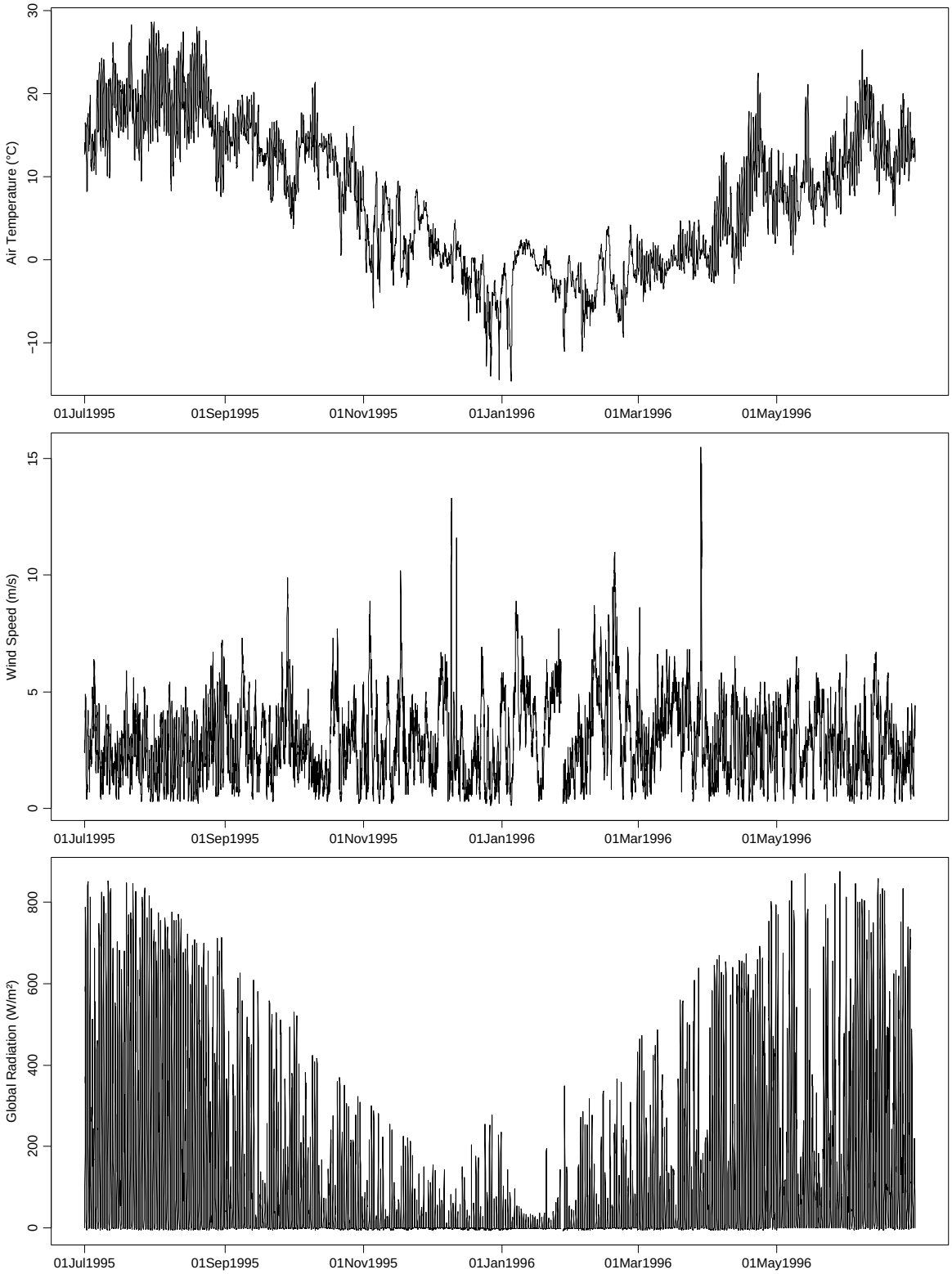

Figure A.2: Raw measurements of climate variables.

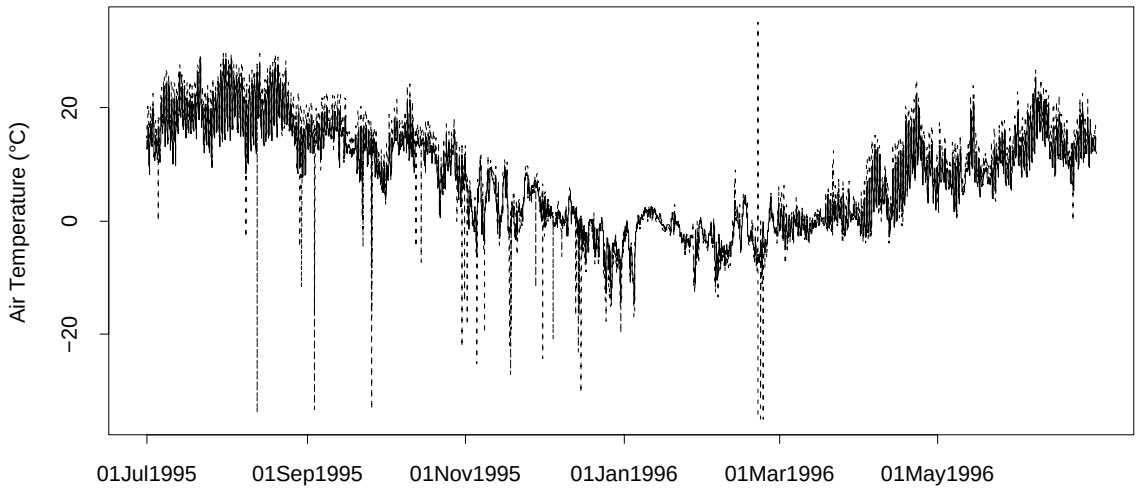

Figure A.3: Raw measurements of air temperature performed by The Royal Vet. and Agric. Univ. (solid) and VEKS (dotted).

Appendix B

Plots related to prediction

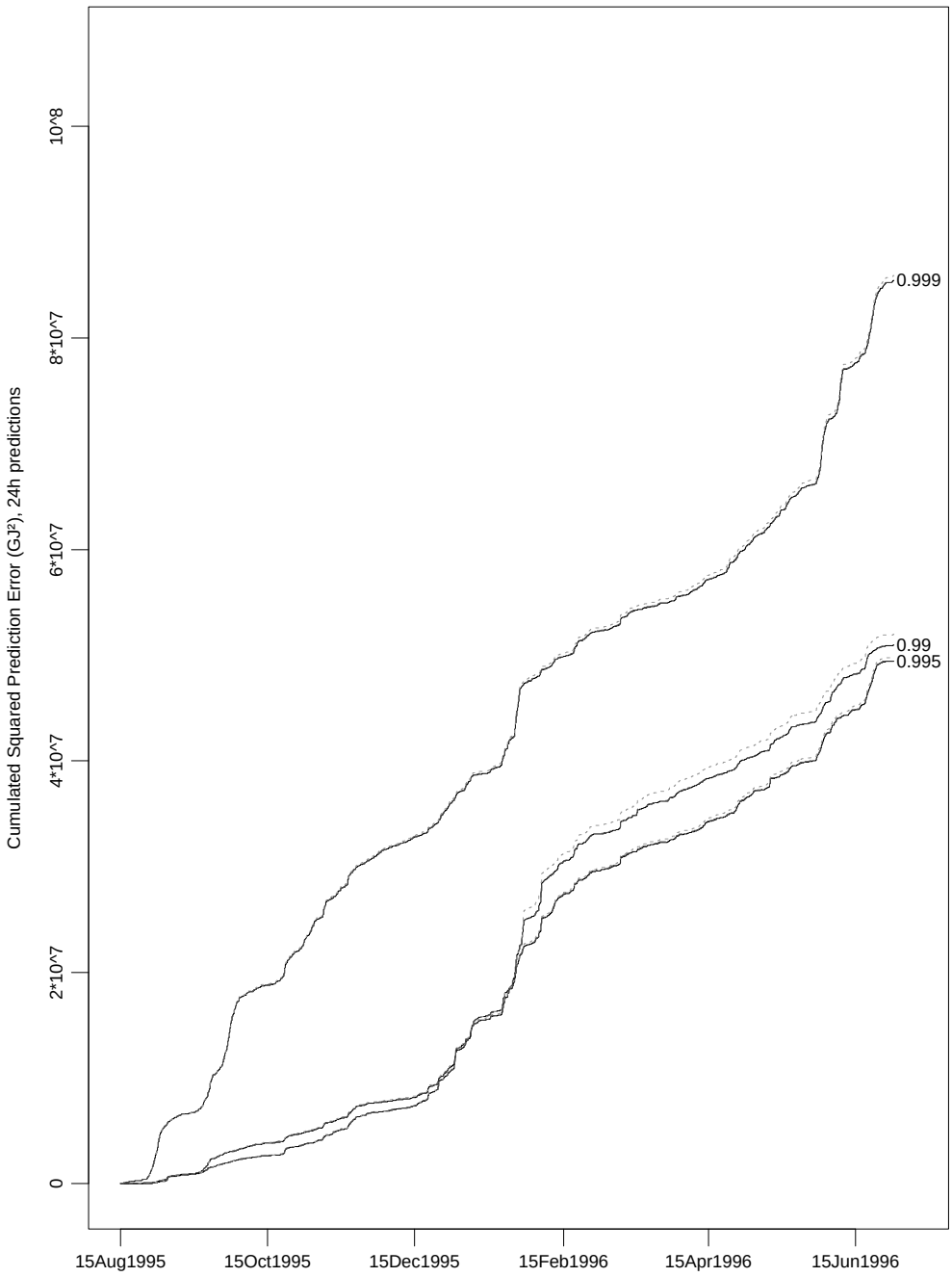

Figure B.1: Cumulative squared prediction error of 24 hour predictions based on adaptive RLS estimates in (6.1) for the forgetting factors 0.99, 0.995, and 0.999. Prediction errors occurring before August 15, 1995 are disregarded. Solid lines indicate the B-spline with 11 internal knots, dotted lines the B-spline with 7 internal (non-equidistant) knots, and dashed lines the fifth order harmonic expansion (hidden by the solid line). The forgetting factors are indicated at the end of the lines.

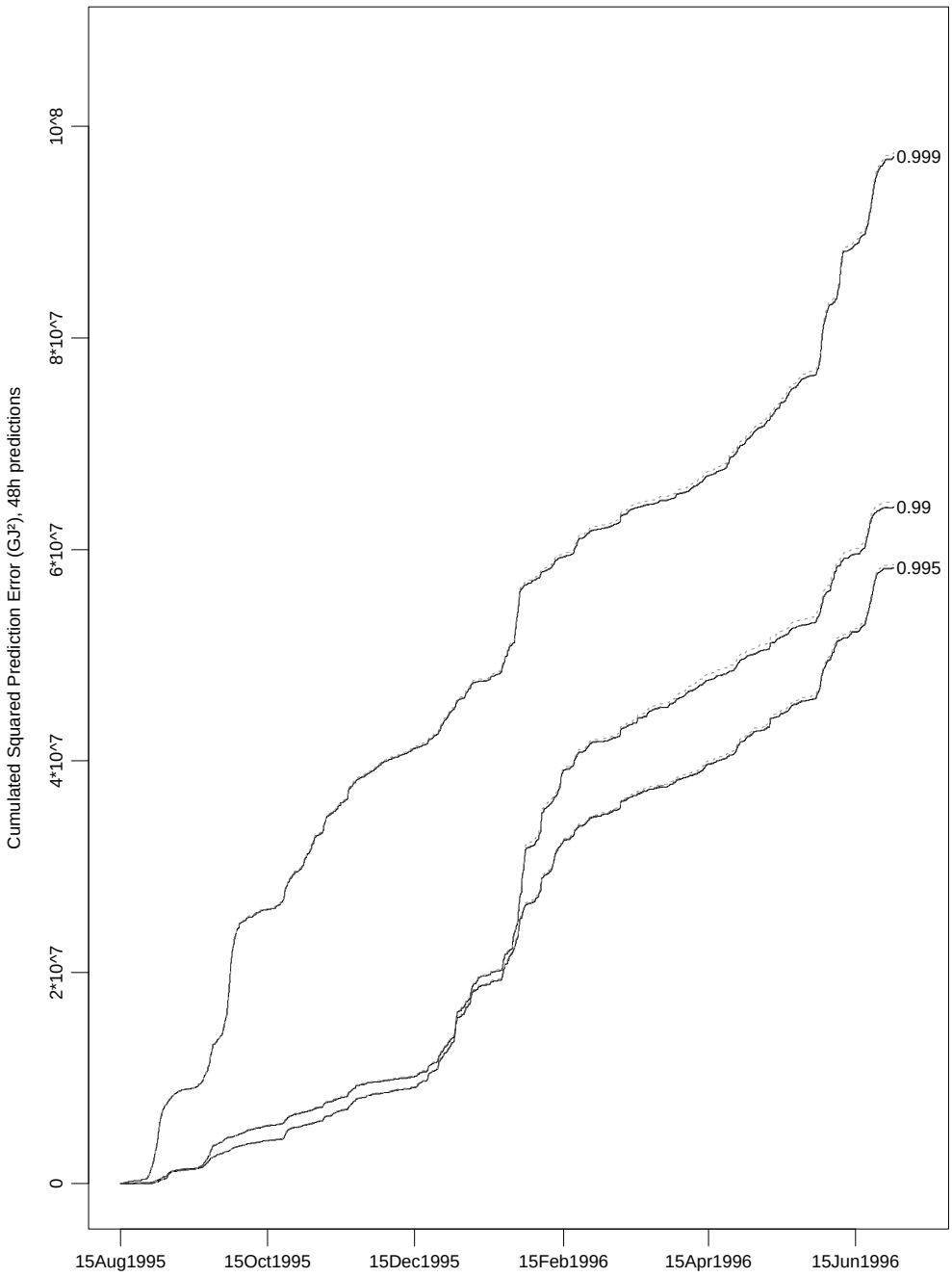

Figure B.2: Cumulative squared prediction error of 48 hour predictions based on adaptive RLS estimates in (6.1) for the forgetting factors 0.99, 0.995, and 0.999. Prediction errors occurring before August 15, 1995 are disregarded. Solid lines indicate the B-spline with 11 internal knots, dotted lines the B-spline with 7 internal (non-equidistant) knots, and dashed lines the fifth order harmonic expansion (hidden by the solid line). The forgetting factors are indicated at the end of the lines.

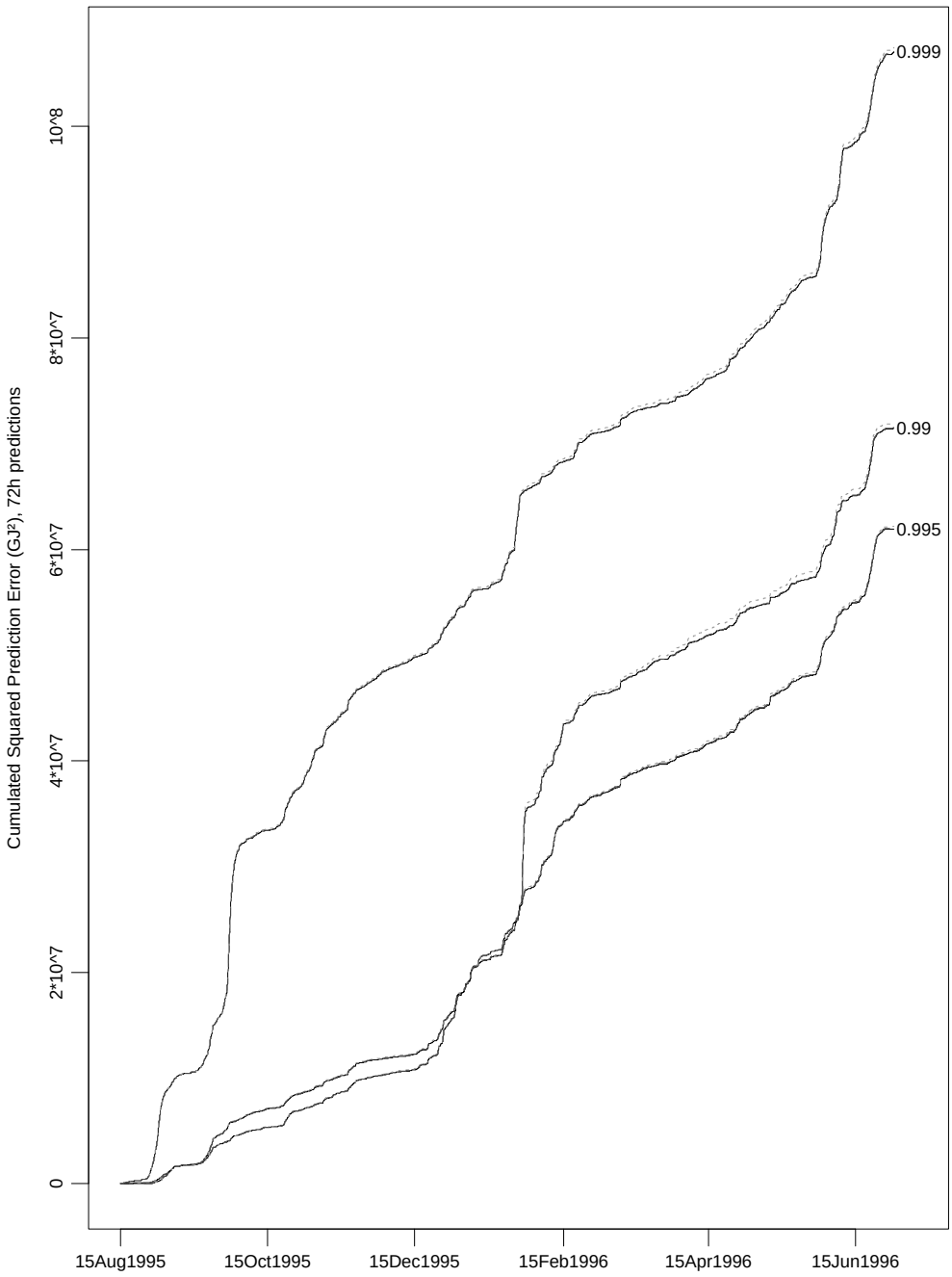

Figure B.3: Cumulative squared prediction error of 72 hour predictions based on adaptive RLS estimates in (6.1) for the forgetting factors 0.99, 0.995, and 0.999. Prediction errors occurring before August 15, 1995 are disregarded. Solid lines indicate the B-spline with 11 internal knots, dotted lines the B-spline with 7 internal (non-equidistant) knots, and dashed lines the fifth order harmonic expansion (hidden by the solid line). The forgetting factors are indicated at the end of the lines.

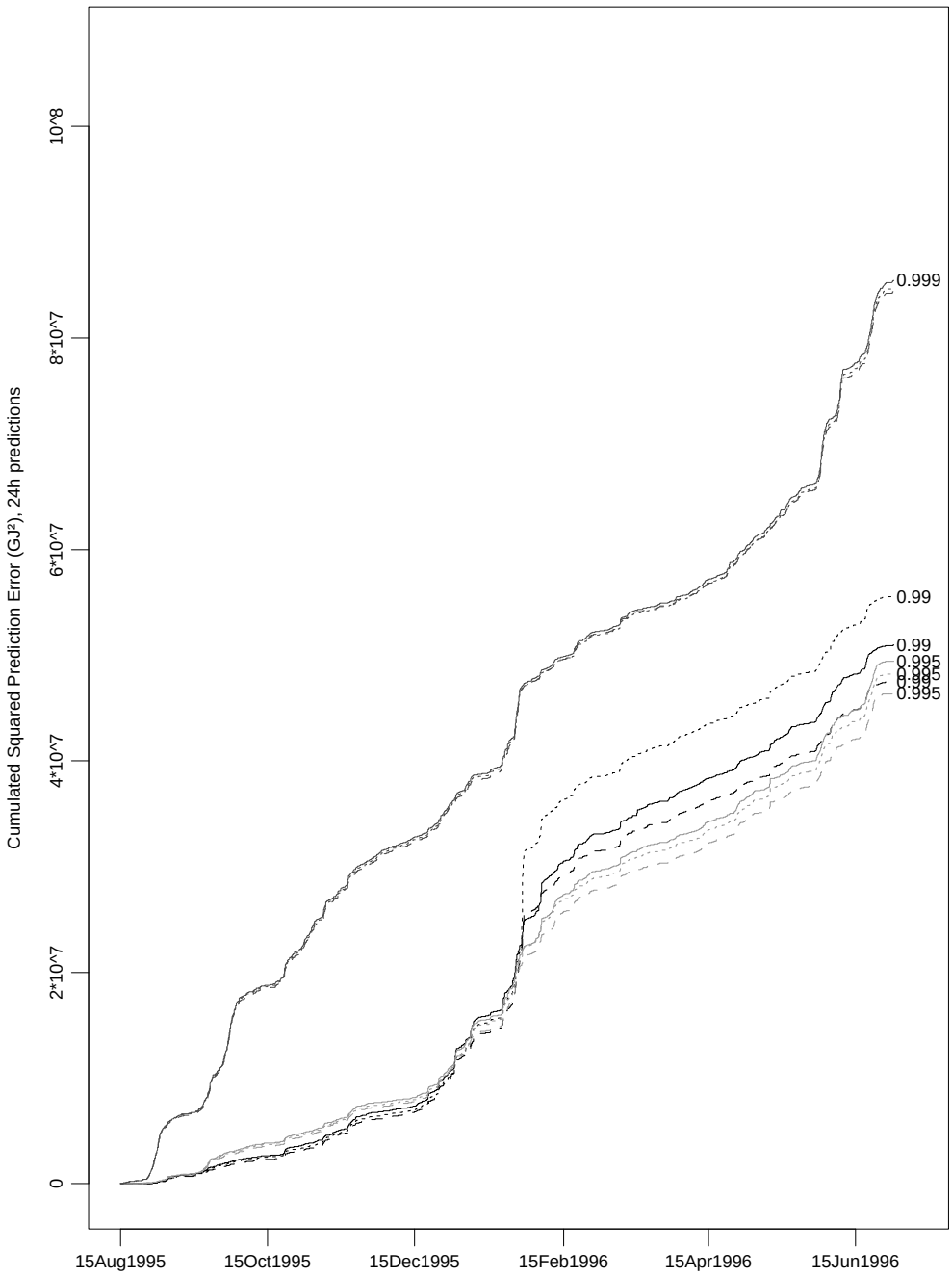

Figure B.4: Cumulative squared prediction error of 24 hour predictions based on adaptive RLS estimates in (6.1) for the forgetting factors 0.99, 0.995, and 0.999. Prediction errors occurring before August 15, 1995 are disregarded. The diurnal variation is modelled using a periodic cubic B-spline basis with 11 internal equidistant knots. Solid lines indicate that the days are grouped into working, half-holy, and holy days. Dashed lines indicate that the difference between half-holy and holy days are neglected entirely and dotted lines indicate that half-holy and holy days differ in mean value only. The forgetting factors are indicated at the end of the lines.

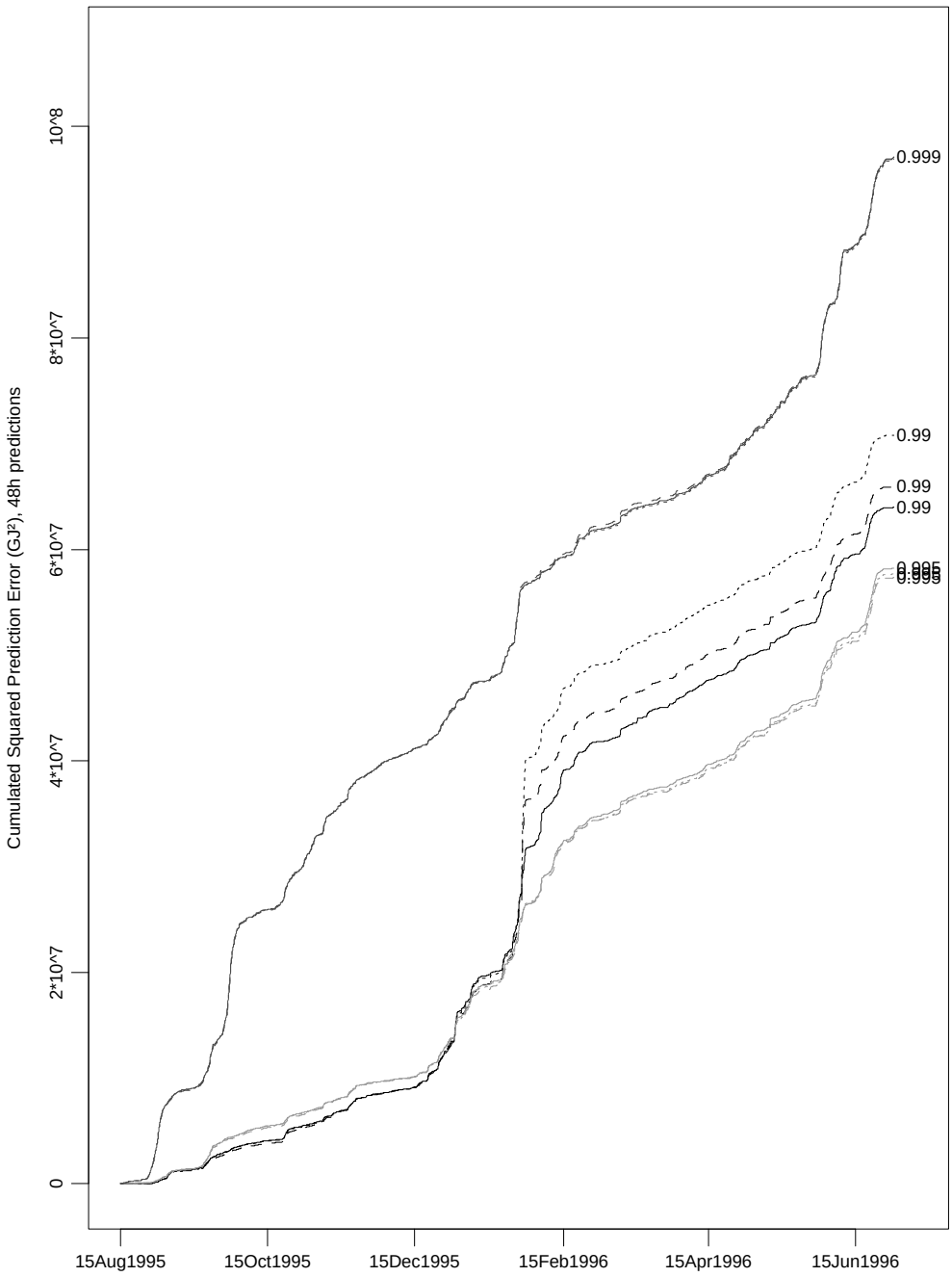

Figure B.5: Cumulative squared prediction error of 48 hour predictions based on adaptive RLS estimates in (6.1) for the forgetting factors 0.99, 0.995, and 0.999. Prediction errors occurring before August 15, 1995 are disregarded. The diurnal variation is modelled using a periodic cubic B-spline basis with 11 internal equidistant knots. Solid lines indicate that the days are grouped into working, half-holy, and holy days. Dashed lines indicate that the difference between half-holy and holy days are neglected entirely and dotted lines indicate that half-holy and holy days differ in mean value only. The forgetting factors are indicated at the end of the lines.

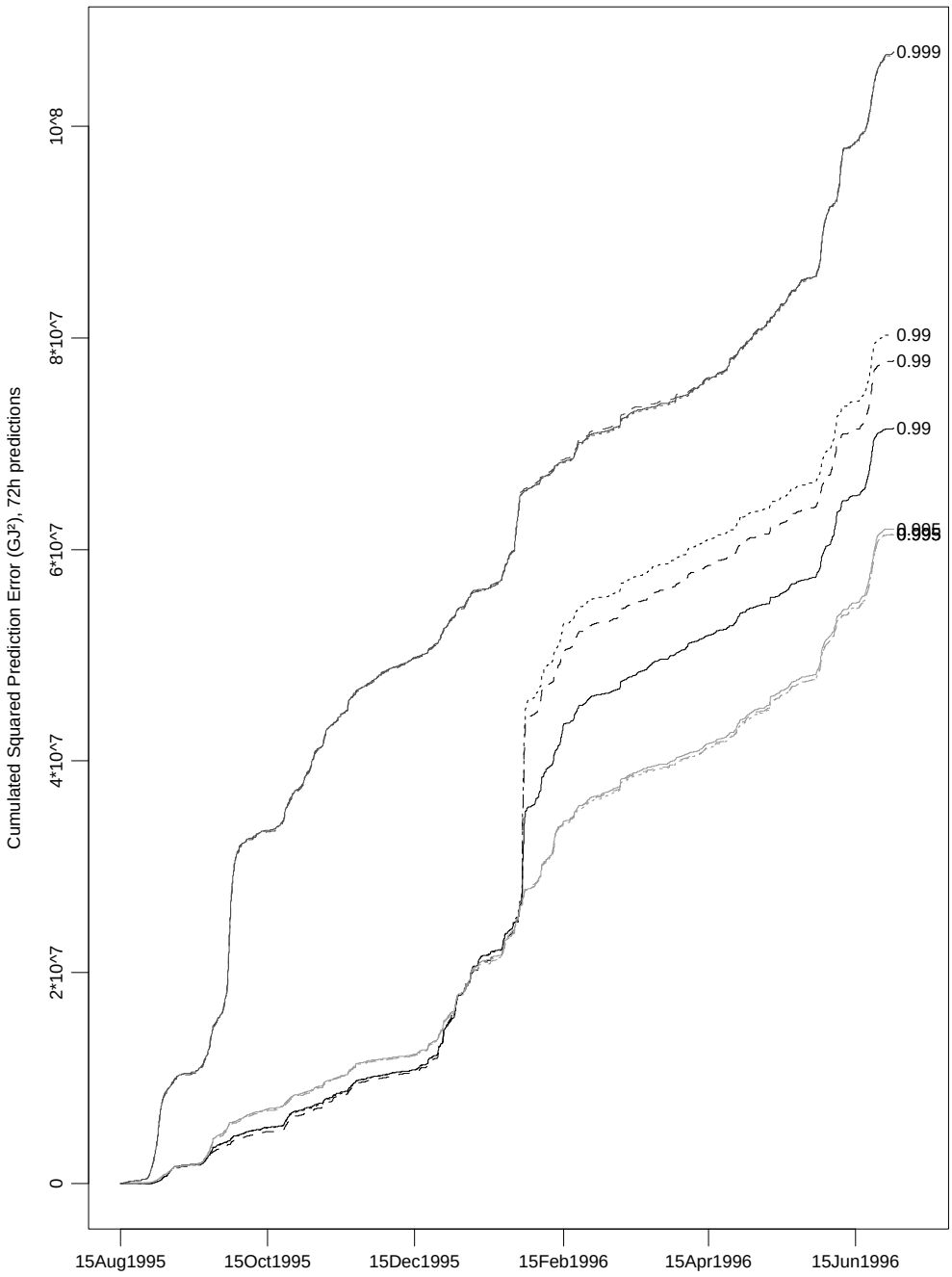

Figure B.6: Cumulative squared prediction error of 72 hour predictions based on adaptive RLS estimates in (6.1) for the forgetting factors 0.99, 0.995, and 0.999. Prediction errors occurring before August 15, 1995 are disregarded. The diurnal variation is modelled using a periodic cubic B-spline basis with 11 internal equidistant knots. Solid lines indicate that the days are grouped into working, half-holy, and holy days. Dashed lines indicate that the difference between half-holy and holy days are neglected entirely and dotted lines indicate that half-holy and holy days differ in mean value only. The forgetting factors are indicated at the end of the lines.

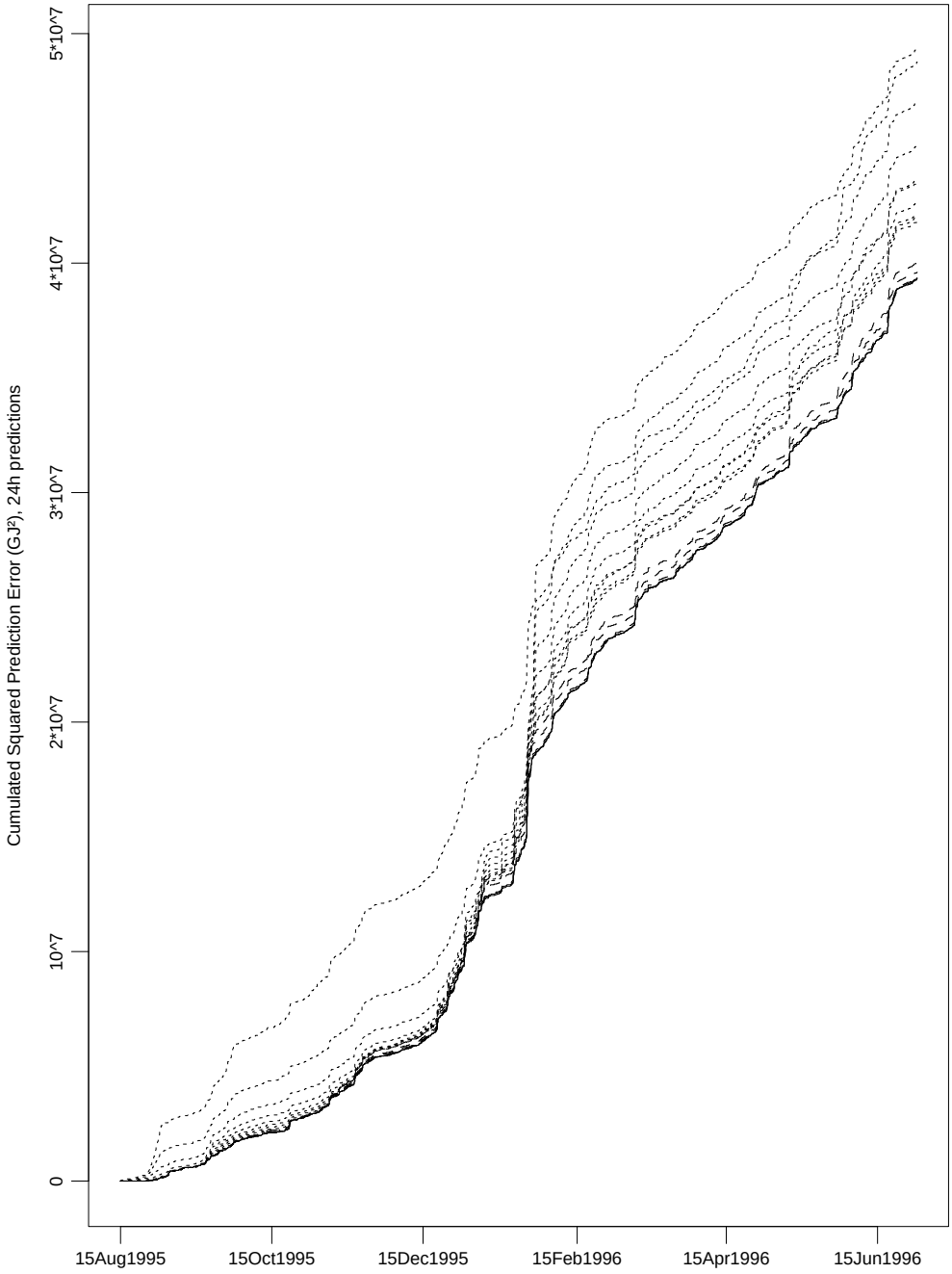

Figure B.7: On-line selection of 24 hour prediction method using n_{ms} in (6.5) on page 81 corresponding to 3, 5, and 7 days (dashed) and 9 and 11 days (solid). The dotted lines indicate the predictions obtained using constant forgetting factors. Prediction errors occurring before August 15, 1995 are disregarded.

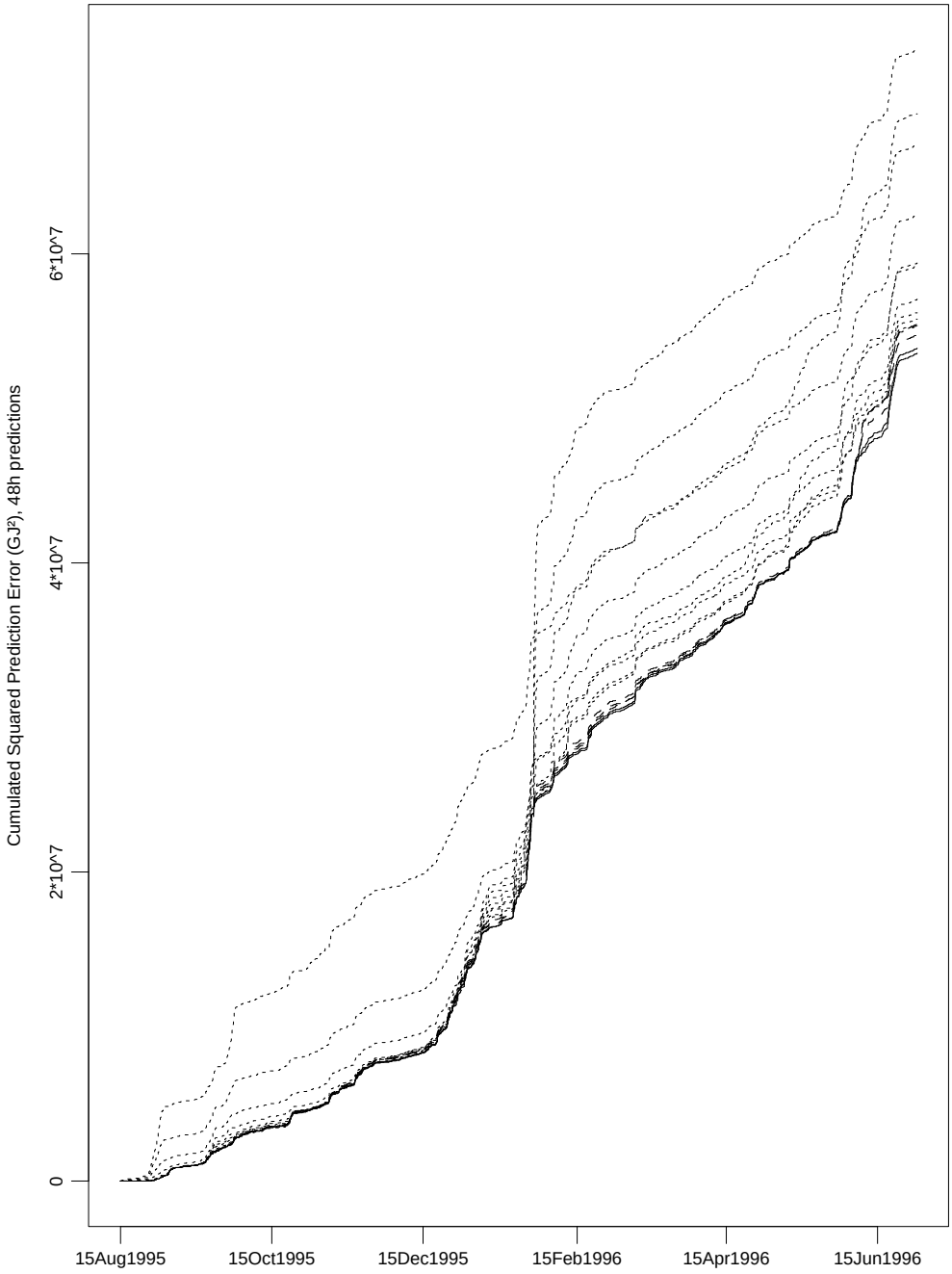

Figure B.8: On-line selection of 48 hour prediction method using n_{ms} in (6.5) on page 81 corresponding to 3, 5, and 7 days (dashed) and 9 and 11 days (solid). The dotted lines indicate the predictions obtained using constant forgetting factors. Prediction errors occurring before August 15, 1995 are disregarded.

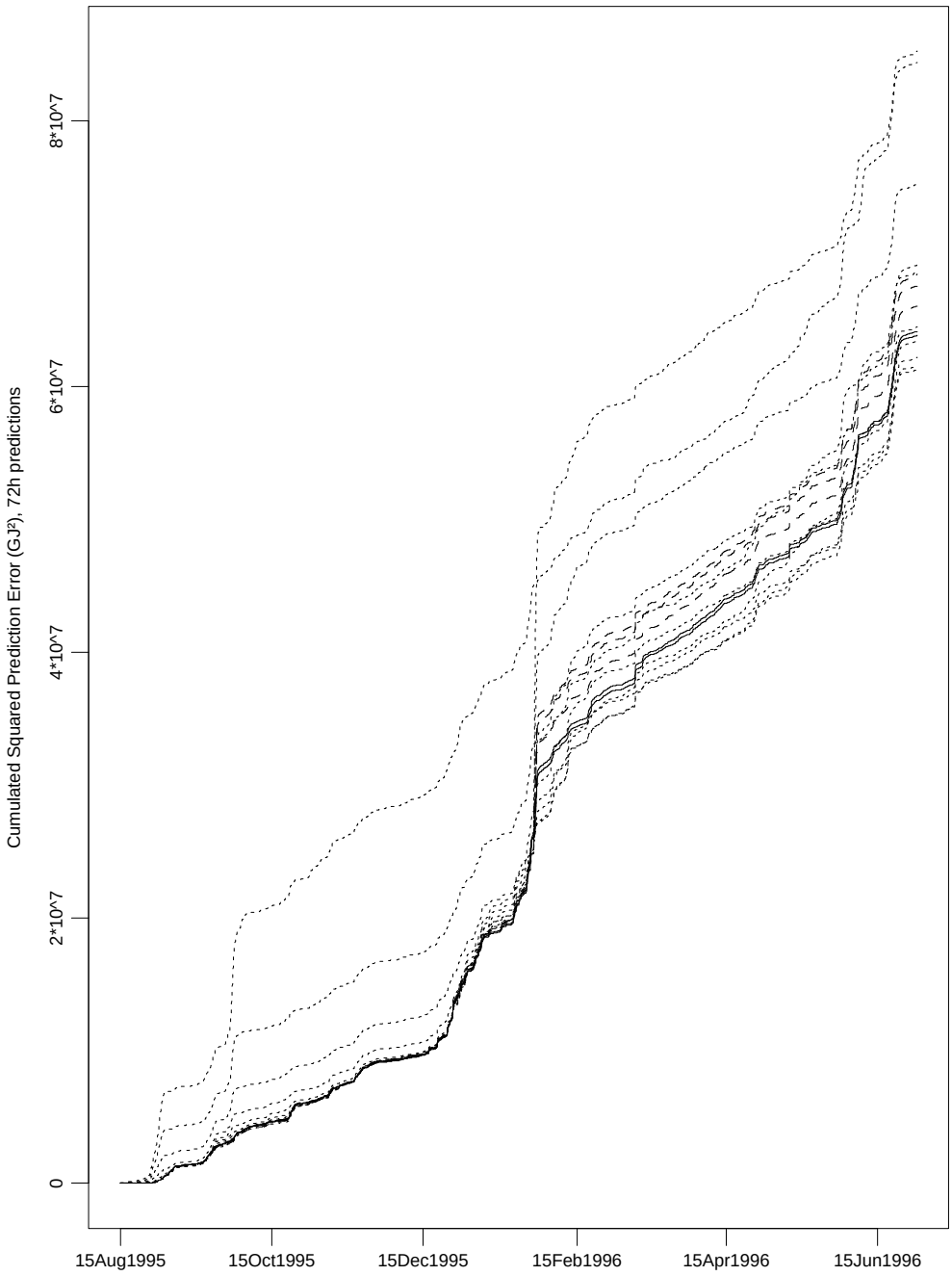

Figure B.9: On-line selection of 72 hour prediction method using n_{ms} in (6.5) on page 81 corresponding to 3, 5, and 7 days (dashed) and 9 and 11 days (solid). The dotted lines indicate the predictions obtained using constant forgetting factors. Prediction errors occurring before August 15, 1995 are disregarded.

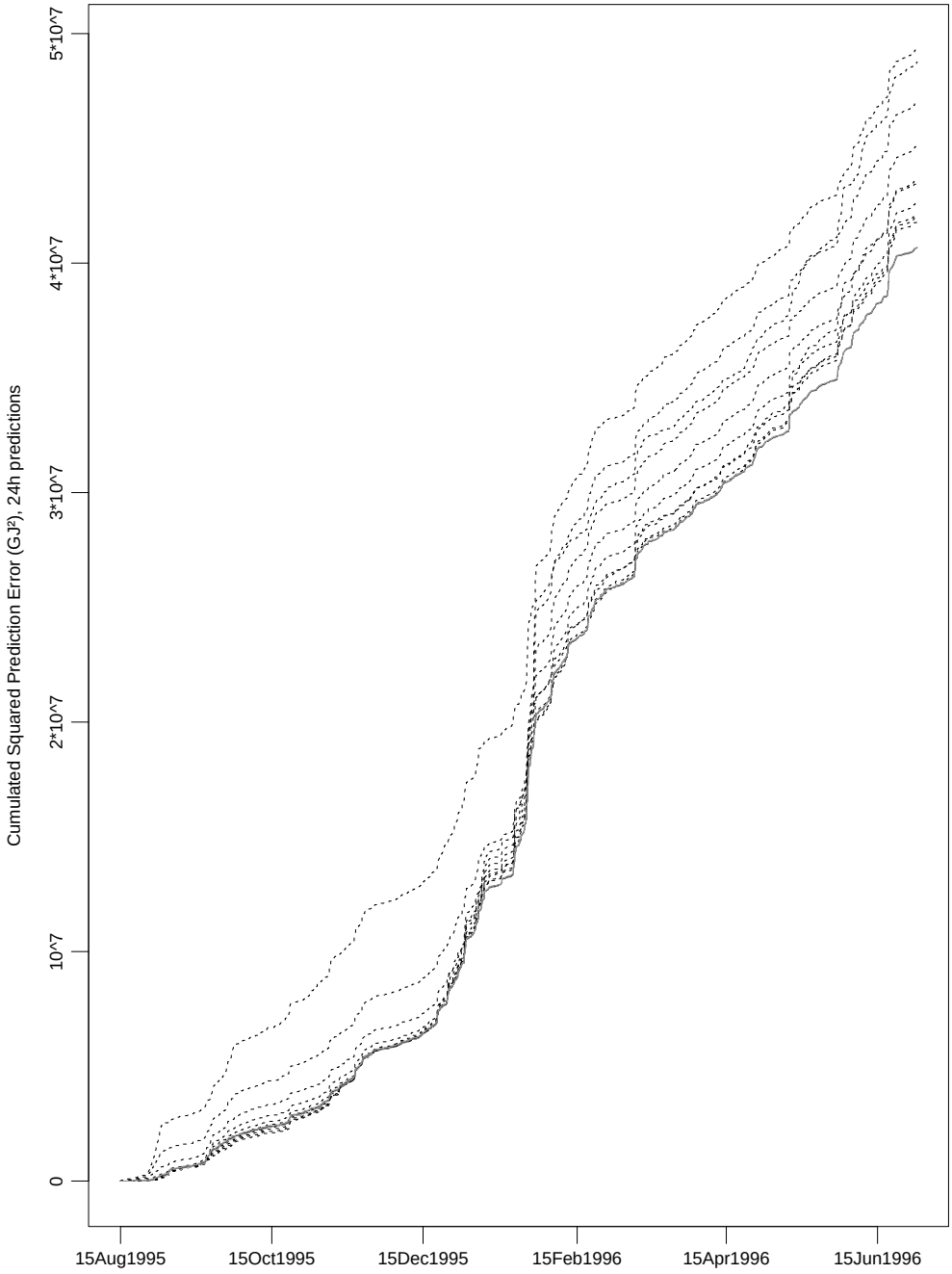

Figure B.10: On-line weighting of 24 hour predictors using n_{ms} in (6.5) on page 81 corresponding to 3, 5, and 7 days (dashed) and 9 and 11 days (solid). The dotted lines indicate the predictions obtained using constant forgetting factors. Prediction errors occurring before August 15, 1995 are disregarded.

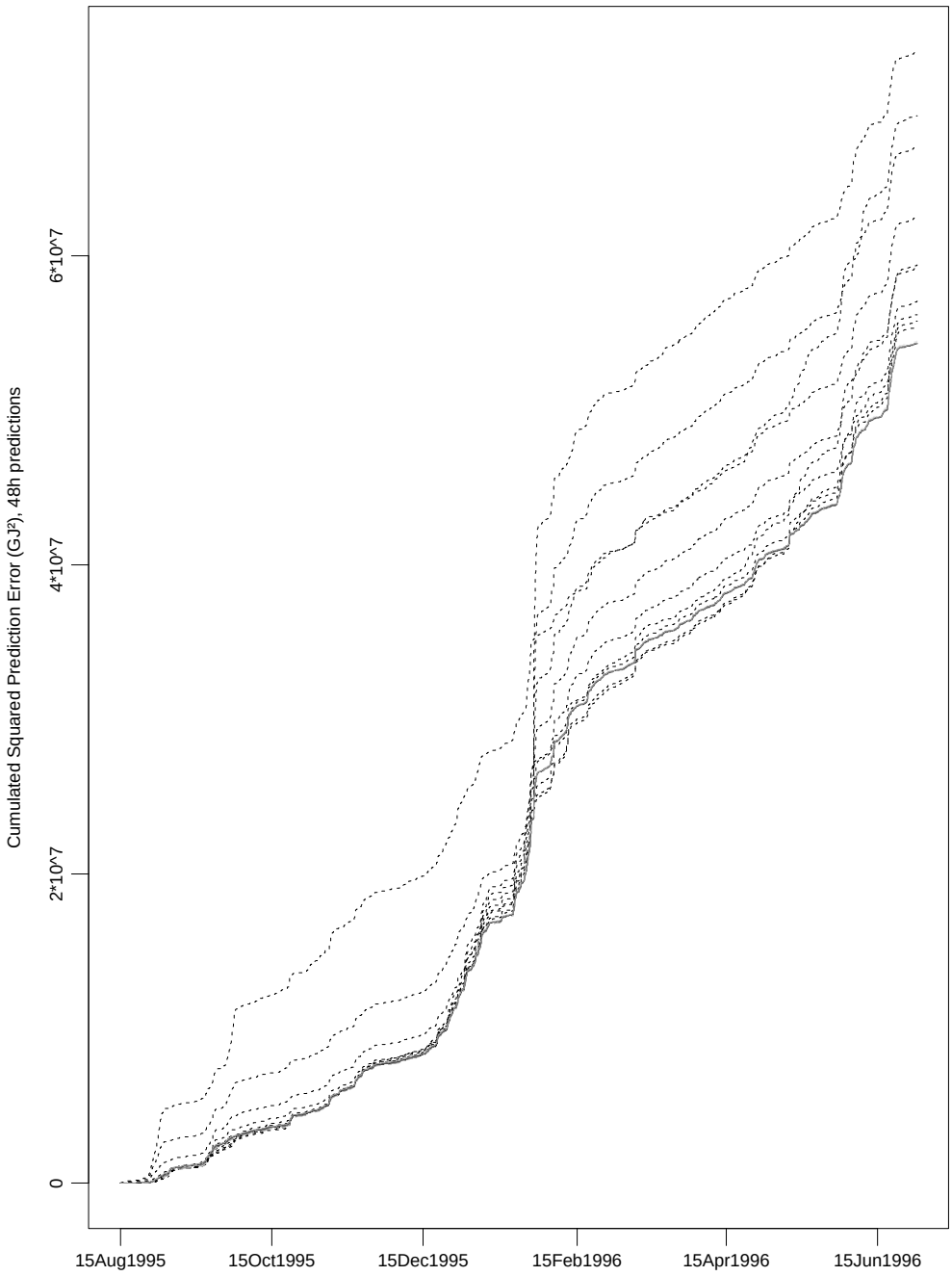

Figure B.11: On-line weighting of 48 hour predictors using n_{ms} in (6.5) on page 81 corresponding to 3, 5, and 7 days (dashed) and 9 and 11 days (solid). The dotted lines indicate the predictions obtained using constant forgetting factors. Prediction errors occurring before August 15, 1995 are disregarded.

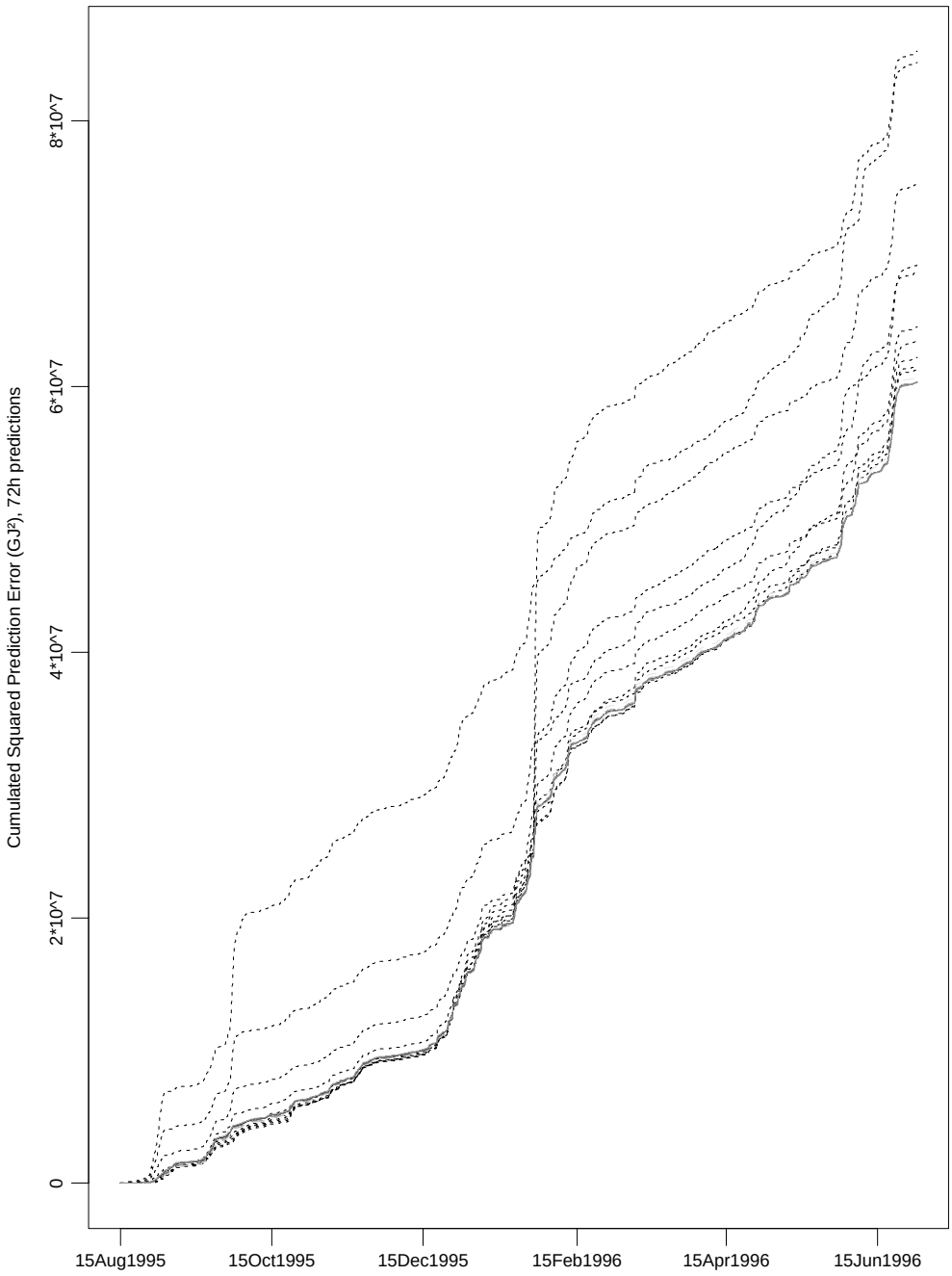

Figure B.12: On-line weighting of 72 hour predictors using n_{ms} in (6.5) on page 81 corresponding to 3, 5, and 7 days (dashed) and 9 and 11 days (solid). The dotted lines indicate the predictions obtained using constant forgetting factors. Prediction errors occurring before August 15, 1995 are disregarded.

Appendix C

Simulation of prediction errors

In this appendix it is described how correlated prediction errors of meteorological forecasts of one climate variable are simulated. These simulated values are used to simulate meteorological forecasts of the air temperature in Section 6.5.

Let $\tilde{x}_{t|s};\ t > s$ denote the prediction error of the meteorological forecast issued at time s of the climate variable x at time t. The simulated prediction errors are arranged in a matrix with the following structure:

$$\begin{bmatrix} \vdots & \vdots & & \vdots \\ \tilde{x}_{t+1|t} & \tilde{x}_{t+2|t} & \dots & \tilde{x}_{t+k|t} \\ \vdots & \vdots & & \vdots \end{bmatrix}, \tag{C.1}$$

where k is the prediction horizon considered, e.g. in hours. The prediction errors are simulated such that

- they are normally distributed with zero mean,
- $Var[\tilde{x}_{t+j|t}]$ is a predefined value σ_j^2 for all $j = 1, \dots, k$, cf. Figure 6.11 on page 92,
- the correlation between neighbour prediction errors in a row of (C.1) is constant ρ_r,
- the correlation between $\tilde{x}_{t+j|t}$ and $\tilde{x}_{t+i|t}$ depends on $|i - j|$ only and follows the auto correlation function of an AR(1) process with the pole in ρ_r, and
- each column in (C.1) is an AR(1) process with the pole in ρ_c.

A matrix like (C.1) with the properties mentioned can be simulated by the following steps:

1. Simulate k independent time series $\{e_{j,t}\};\ j = 1, \ldots, k$, of appropriate length, as Gaussian zero mean and unit variance AR(1) processes with a pole in ρ_c. The unit variance is obtained if the innovation variance is $1 - \rho_c^2$.

2. The first column of (C.1) is hereafter generated as

$$\tilde{x}_{t+1|t} = \sigma_1 e_{1,t} \tag{C.2}$$

3. The remaining columns of (C.1) are then generated as

$$\tilde{x}_{t+j|t} = \sigma_j \left[\frac{\rho_r}{\sigma_{j-1}} \tilde{x}_{t+j-1|t} + e_{j,t}\sqrt{1-\rho_r^2} \right]. \tag{C.3}$$

In the following it will be shown that this procedure results in simulated prediction errors with the properties stated above.

Normally distributed with zero mean: For $\tilde{x}_{t+1|t}$ this follows directly from (C.2) since $e_{1,t}$ has the same properties. Since $\tilde{x}_{t+2|t}$ is a linear combination of Gaussian variables it is itself a Gaussian variable. This argument can be repeated for the remaining values of j, establishing the normality of $\tilde{x}_{t+j|t}$ for all j. From (C.3) it follows that

$$E[\tilde{x}_{t+j|t}] = \sigma_j \left[\frac{\rho_r}{\sigma_{j-1}} E[\tilde{x}_{t+j-1|t}] + E[e_{j,t}]\sqrt{1-\rho_r^2} \right].$$

Since $E[e_{j,t}] = 0$ for all j and since $E[\tilde{x}_{t+1|t}] = 0$ it then follows that $E[\tilde{x}_{t+j|t}] = 0$ for all j.

Variance of $\tilde{x}_{t+j|t} = \sigma_j^2$: From (C.2) it follows that $V[\tilde{x}_{t+1|t}] = \sigma_1^2$ and from (C.3) and $V[e_{j,t}] = 1$ we get

$$V[\tilde{x}_{t+j|t}] = \sigma_j^2 \left[\frac{\rho_r^2}{\sigma_{j-1}^2} V[\tilde{x}_{t+j-1|t}] + (1-\rho_r^2) \right].$$

Setting $j = 2$ it follows that $V[\tilde{x}_{t+2|t}] = \sigma_2^2$ and hereafter the argument can be repeated to confirm that $V[\tilde{x}_{t+j|t}] = \sigma_j^2;\ j = 1, \ldots, k$.

Correlation between $\tilde{x}_{t+j|t}$ and $\tilde{x}_{t+i|t}$: The correlation of the prediction errors in the same row of (C.1) can be written as

$$Cor[\tilde{x}_{t+j|t},\ \tilde{x}_{t+j+l|t}] = \frac{Cov[\tilde{x}_{t+j|t},\ \tilde{x}_{t+j+l|t}]}{\sigma_j \sigma_{j+l}},$$

and since correlation and covariance is invariant to the order of the arguments it suffices to consider $j = 1, \ldots, k-1$ and $l = 1, \ldots, k-j$. Using (C.3) we obtain:

$$Cov[\tilde{x}_{t+j|t},\ \tilde{x}_{t+j+l|t}] = Cov\left[\tilde{x}_{t+j|t},\ \frac{\sigma_{j+l}}{\sigma_{j+l-1}}\rho_r \tilde{x}_{t+j+l-1|t} + \sigma_{j+l} e_{j+l,t}\sqrt{1-\rho_r^2}\right].$$

Since $e_{j+l,t}$ and $\tilde{x}_{t+j|t}$ is independent it follows that:

$$Cov[\tilde{x}_{t+j|t},\ \tilde{x}_{t+j+l|t}] = \frac{\sigma_{j+l}}{\sigma_{j+l-1}}\rho_r Cov[\tilde{x}_{t+j|t},\ \tilde{x}_{t+j+l-1|t}],$$

and if these calculations are repeated l times:

$$Cov[\tilde{x}_{t+j|t},\ \tilde{x}_{t+j+l|t}] = \frac{\sigma_{j+l}}{\sigma_j}\rho_r^l Cov[\tilde{x}_{t+j|t},\ \tilde{x}_{t+j|t}] = \sigma_{j+l}\sigma_j \rho_r^l.$$

Since correlation is invariant to the order of the arguments we have:

$$Cor[\tilde{x}_{t+j|t},\ \tilde{x}_{t+i|t}] = \rho_r^{|i-j|}, \tag{C.4}$$

which is equivalent to the autocorrelation function of an AR(1) process with a pole in ρ_r and the correlation between neighbour hood prediction errors in a row of (C.1) is constant ρ_r.

Auto correlation of the columns in (C.1): Since each column of (C.1) is generated as an scalar multiplied with an AR(1) process with the pole in ρ_c it follows directly that each column has the same properties.

Appendix D

Heat Loss From DH-pipes

At a presentation meeting held at the Department of Mathematical Modelling, Technical University of Denmark on March 16, 2000 the issue of heat loss from the pipes in the local district heating systems connected to the transmission system was raised. This aspect is considered in this appendix. The following persons participated in the meeting:

Anders N. Andersen, Energi- og Miljødata
Bente Andersen, Farum Fjernvarme
Egon Erlandsen, Frederiksberg Varmeværk
Helge Faurby, Danmarks Meteorologiske Institut
Henrik R. Hansen, Vestegnens Kraftvarmeselskab
Jens Chr. Hansen, Sønderborg Kraftvarmeværk
Jens H. Hansen, Rambøll
Klaus W. Hansen, ELSAM
Dorthe R. Jensen, Centralkommunernes Transmissionsselskab
Gert Jensen, Frederiksberg Varmeværk
Oluf Jørgensen, Esbjerg Kommune
Peter Laursen, ABB Energi & Industri
Henrik Madsen, Institut for Matematisk Modellering, DTU
Steffen Moe, Sønderborg Fjernvarme
Henrik Aa. Nielsen, Institut for Matematisk Modellering, DTU
Torben S. Nielsen, Institut for Matematisk Modellering, DTU
Hans Ravn, Elkraft
Jan Strømvig, Odense Fjernvarme
Jesper Thiesen, Danmarks Meteorologiske Institut

D.1 Introduction

In this report methods for predicting the heat consumption in district heating systems by use of meteorological forecasts are considered. The underlying models are developed using the gray box approach and the most important effects such as heat transfer via walls, windows, and ventilation are considered together with effects related to the overall behaviour of humans, see Chapter 4. The data used is the hourly heat consumption delivered from the transmission system of Vestegnens Kraftvarmeselskab I/S, Albertslund to the local distributors (DH-networks). The heat loss in the local networks is not considered in the gray box approach. This appendix evaluates the predictive ability obtained by considering this last mentioned effect. The evaluation is performed assuming the future climate to be known. Since the DH-pipes are affected only by a very low-pass filtered version of the climate, which will be quite easy to forecast, the benefit, if any, will in practice be larger than indicated by the analysis presented here.

Due to the low-pass filtering of the climate before reaching the DH-pipes and since the prediction methods used in the report are adaptive, it is doubtful if the direct modelling of the heat loss will improve the predictive ability. Since sufficiently low-pass filtered versions of the air temperature and the global radiation are almost linear dependent and since a sufficiently low-pass filtered version of the wind speed is near to constant only the air temperature will be used in this investigation.

D.2 Prediction Method

The same model as in Chapter 6 will be used, but with a low-pass filtered version of the air temperature $H_{soil}(q)T_{a,t}$ added.

$$\begin{aligned} Q_t &= \mu(h_t^{24}, \Upsilon_t) + a_{20}H_2(q)R_t \\ &+ a_{111}H_1(q)W_t + a_{120}H_1(q)T_{a,t} + a_{121}H_1(q)W_tH_1(q)T_{a,t} \\ &+ a_{100}H_1(q)R_t + a_{101}H_1(q)W_tH_1(q)R_t \\ &+ a_{211,0}W_t + a_{211,1}W_{t-1} + a_{220,0}T_{a,t} + a_{220,1}T_{a,t-1} \\ &+ a_{soil}H_{soil}(q)T_{a,t} + e_t, \end{aligned} \tag{D.1}$$

with the predefined filters

$$H_1(q) = \frac{0.066}{1 - 0.934q^{-1}}, \tag{D.2}$$

and

$$H_2(q) = \frac{-0.350 + 0.612q^{-1} - 0.226q^{-2}}{1 - 1.703q^{-1} + 0.739q^{-2}}. \tag{D.3}$$

In D.1 Q_t denotes the heat consumption at time t, W_t is the wind speed, $T_{a,t}$ is the air temperature, R_t is the solar radiation on a square pillar as described in Chapter 2, and $a.$ are the parameters. Above the structure of the diurnal variation $\mu(h_t^{24}, \Upsilon_t)$ is not specified. Here the periodic cubic B-spline basis with 11 equidistant internal knots will be used and groups of days corresponding to working and non-working days will be considered, cf. Chapter 6. The model will be estimated adaptively using a forgetting factor of 0.995, which in Chapter 7 is found to be a good overall forgetting factor.

The low-pass filter

$$H_{soil}(q) = \frac{1 - \phi_{soil}}{1 - \phi_{soil}q^{-1}} \tag{D.4}$$

corresponding to the heat loss from the DH-pipes is used and different values of the pole ϕ_{soil} is investigated below. It is decided to investigate values of the pole ranging from 0.97 to 0.999, the low value is chosen based on the largest pole (0.934) in (D.2) and (D.3). Ten values of ϕ_{soil} are investigated. These are chosen equidistant from $1/log(0.97)$ to $1/log(0.999)$ and then transformed back to the original scale. Figure D.1 shows the filtered values of the air temperature.

D.3 Results

The results are presented both for the case where the model (D.1) is used directly to generate the predictions and for the case where the predictions from the model are adjusted using a multi step AR(1) model estimated adaptively using a forgetting factor of 0.999, see Chapter 6. Results are presented for 72 hour predictions only.

Figures D.2 and D.3 show plots of the cumulative squared prediction errors for unadjusted and adjusted predictions, respectively. Prediction errors before August 15, 1995 are disregarded. The results corresponding to the original model, i.e. excluding $H_{soil}(q)$, are indicated by thick lines.

For unadjusted predictions it is clearly advantageous to include the filter. For the adjusted predictions the advantage is not obvious, but it is clear that using the filter is not advantageous in the very first and last part of the period. This may be because the consumers are not reacting on the climate at these time points and maybe models not including climate information should perhaps be used in these periods.

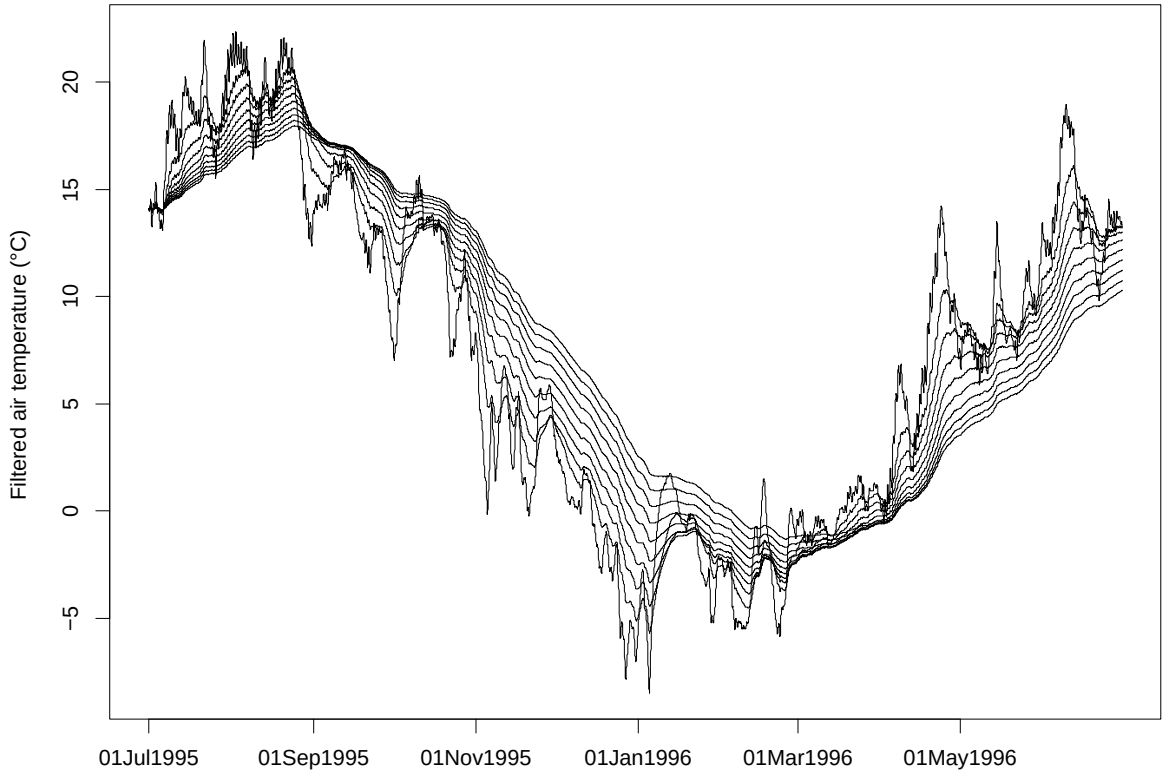

Figure D.1: Filtered values of the air temperature using ten different poles.

To summarize the information contained in the plots mentioned above the RMS of the prediction errors are calculated for the four last periods defined by the dates on the to horizontal axes of the plots. These periods corresponds to points in time in which a fairly uniform behaviour of the prediction methods are apparent. The results are displayed in Figure D.4.

Except for the last period (June, 1996) it is advantageous to include $H_{soil}(q)T_{a,t}$ in the model. There seem to be no good explanation for the strange curvature for the period December 16, 1995 to February 15, 1996 and this will not be considered further. For the period October 1 to December 15, 1995 the unadjusted predictions perform marginally better than the adjusted predictions for some values of ϕ_{soil}. The opposed (and expected) behaviour are observed for the period February 16 to May 31, 1996.

Overall, a pole in the interval 0.998 to 0.9985 seems to be advantageous together with adjusted predictions. However, for the summer period it should be considered dropping the climate information from the model. This may be done by use of the weighting of predictors in Chapter 6.

D.4 Conclusion

The benefit from including a low-pass filtered version of the air temperature as an additional explanatory variable in a model used for adaptive prediction of the consumption in district heating systems is investigated. Overall this seems to be beneficial using a first order filter with a pole in the interval from 0.998 to 0.9985. There is some indication that climate information should be excluded from the model during summer periods. One approach to this is to apply prediction methods both with and without climate information and generate the final prediction by the method called *on-line weighting of predictors*, cf. Chapter 6. In this appendix only one forgetting factor is investigated. As argued in Chapter 6 a constant forgetting factor is not optimal. This can also be handled by on-line weighting of predictors.

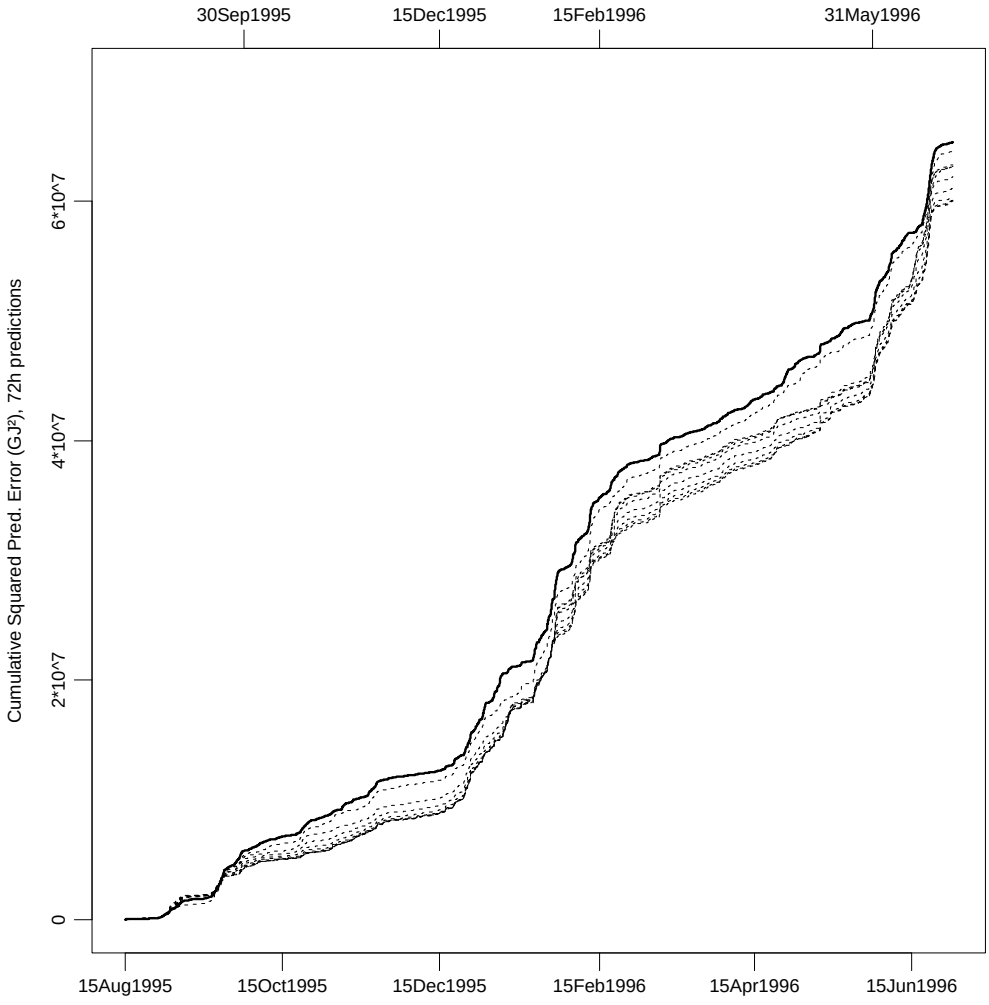

Figure D.2: Cumulative squared prediction errors for unadjusted 72 hours predictions.

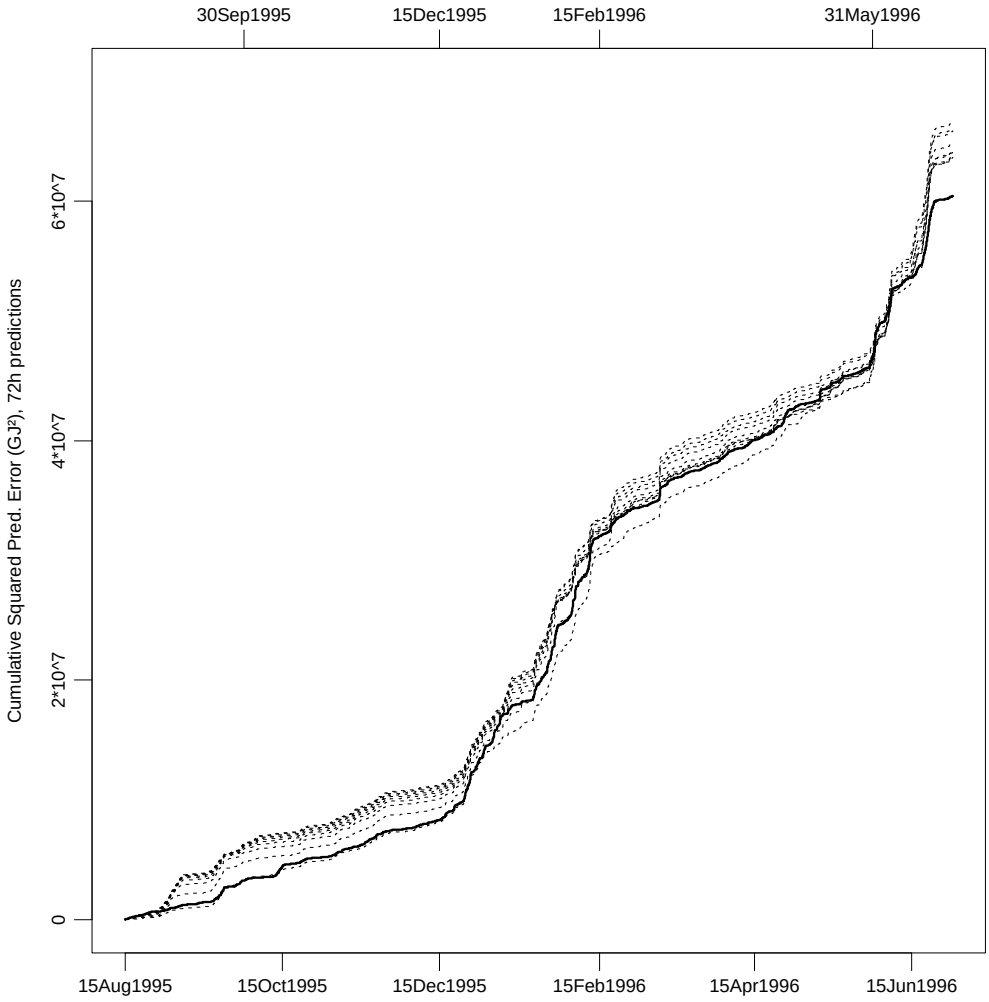

Figure D.3: Cumulative squared prediction errors for adjusted 72 hours predictions.

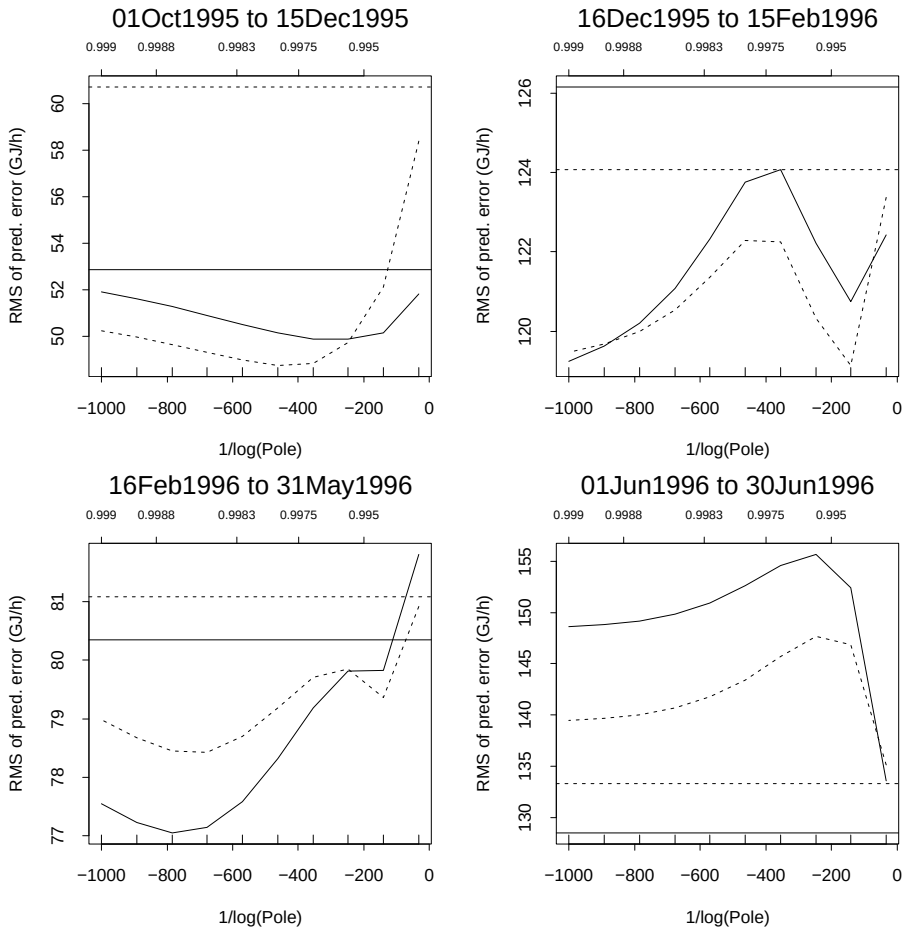

Figure D.4: RMS of 72h prediction errors for the periods indicated. Results corresponding to adjusted predictions are indicated by solid lines and dotted lines indicate results corresponding to unadjusted predictions. The horizontal lines indicate results obtained with the original model.

Bibliography

Abraham, B. & Ledolter, J. (1983), *Statistical Methods for Forecasting*, Wiley, New York.

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

Becker, R. A., Clark, L. A. & Lambert, D. (1994), 'Cave plots: A graphical technique for comparing time series', *Journal of Computational and Graphical Statistics* **3**, 277–283.

Bhansali, R. J. (1980), 'Autoregressive and window estimates of the inverse correlation function', *Biometrika* **67**, 551–566.

Bøhm, B., ed. (1994), *Optimum Operation of District Heating Systems*, Laboratory of Heating and Air Conditioning Technical University of Denmark, DK-2800 Lyngby, Denmark. EFP91/1323/91-0010, 292 pages.

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

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

Brockwell, P. J. & Davis, R. A. (1986), *Time Series: Theory and Methods*, Springer-Verlag, Berlin/New York.

Chatfield, C. (1984), *Analysis of Time Series: An Introduction*, third edn, Chapman & Hall, London/New York.

Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. (1990), 'STL: A seasonal-trend decomposition procedure based on loess', *Journal of Official Statistics* **6**, 3–33. (C/R: p33-73; Corr. to Comment: p343-348).

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

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

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

Dunsmuir, W. & Robinson, P. M. (1981), 'Estimation of time series models in the presence of missing data', *Journal of the American Statistical Association* **76**, 560–568.

Efron, B. & Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall, London/New York.

Fortescue, T., Kershenbaum, L. & Ydstie, B. (1981), 'Implementation of self-tuning regulators with variable forgetting factors', *Automatica* **17**(6), 831–835.

Gallant, A. R. (1987), *Nonlinear Statistical Models*, Wiley, New York.

Gay, D. M. (1983), 'Algorithm 611. subroutines for unconstrained minimization using a model / trust-region approach', *ACM Transactions on Mathematical Software* **9**, 503–524.

Graybill, F. A. (1976), *Theory and Application of the Linear Model*, Duxbury Press, North Scituate, MA.

Hansen, H. E., Kjerulf-Jensen, P. & Stampe, O., eds (1987), *Varme- og Klimateknik, Grundbog*, DANVAK ApS., DK-2800 Lyngby. ISBN: 87-982652-1-0.

Harvey, A. C. & Pierse, R. G. (1984), 'Estimating missing observations in economic time series', *Journal of the American Statistical Association* **79**, 125–131.

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

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

Incropera, F. P. & DeWitt, D. P. (1985), *Fundamentals of Heat and Mass Transfer*, second edn, John Wiley & Sons.

Øivin Holter, Ingebretsen, F. & Parr, H. (1979), *Fysikk og Energi Ressurser*, Universitetsforlaget, Oslo · Bergen · Tromsø. ISBN 82-00-01894-6.

Jensen, S. E. (1996), 'Agroclimate at Taastrup 1995-96. Agrohydrology and Bioclimatology', Department of Agricultural Sciences. The Royal Vet. and Agric. Univ., Copenhagen.

Jones, R. H. (1980), 'Maximum likelihood fitting of ARMA models to time series with missing observations', *Technometrics* **22**, 389–395.

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

Jørgensen, B. (1993), *The Theory of Linear Models*, Chapman & Hall, London/New York.

Kohn, R. & Ansley, C. F. (1985), 'Efficient estimation and prediction in time series regression models', *Biometrika* **72**, 694–697.

Kohn, R. & Ansley, C. F. (1986), 'Estimation, prediction, and interpolation for ARIMA models with missing data', *Journal of the American Statistical Association* **81**, 751–761.

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

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

Madsen, H. (1985), Statistically determined dynamical models for climate processes, PhD thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Madsen, H. (1995), 'Time series analysis (in danish: Tidsrækkeanalyse), 2nd edition', Department of Mathematical Modelling, Technical University of Denmark.

Madsen, H. & Nielsen, T. S. (1997), 'PRESS – PRognose- og EnergiStyrings System', Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby.

Madsen, H., Nielsen, T. S. & Søgaard, H. T. (1996), *Control of Supply Temperature*, Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark.

Miller, A. J. (1992), '[Algorithm AS 274] Least squares routines to supplement those of Gentleman', *Applied Statistics* **41**, 458–478. (Correction: 94V43 p678).

Moses, L. E. (1986), *Think and Explain With Statistics*, Addison-Wesley, Reading, MA/Menlo Park, CA.

Myers, R. H. (1986), *Classical and Modern Regression With Applications*, Duxbury Press, North Scituate, MA.

Nielsen, H. A. (1997), LFLM version 1.0, an S-PLUS / R library for locally weighted fitting of linear models, Technical Report 22/97, Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark.

Nielsen, H. A. & Madsen, H. (1997), *Development of methods for evaluation of electricity savings and load levelling Measures, Part 1: Aggregated Power Consumption Models for the Eastern Part of Denmark*, Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark in collaboration with NESA A/S, DK-2900 Hellerup, Denmark. EFP95/1753/95-0001, 122 pages.

Nielsen, H. A., Nielsen, T. S., Joensen, A. K., Madsen, H. & Holst, J. (1999), Tracking time-varying coefficient-functions, Technical Report 9/99, Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark.

Nielsen, N. W. & Amstrup, B. (1998), 'DMI-HIRLAM verification report for the third quarter of 1998', Danish Meteorological Institute, Copenhagen.

Nielsen, T. S., Madsen, H. & Nielsen, H. A. (1999), 'Using Meteorological Forecasts in On-line Predictions of Wind Power', Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark.

Parkum, J. E. (1992), Recursive identification of time-varying systems, PhD thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby, Denmark.

Petersen, N. K. & Hilden, A. (1999), 'SAFEtable Ver1.00 – beskrivelse', Danish Meteorological Institute, Copenhagen.

Priestley, M. B. (1981), *Univariate Series*, Vol. 1 of *Spectral Analysis and Time Series*, Academic, New York/London.

SAS Institute Inc. (1993), *SAS/ETS User's Guide, Version 6*, second edn, SAS Institute Inc., Cary, NC.

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

Skouras, K. & Dawid, A. P. (1998), 'On efficient point prediction systems', *Journal of the Royal Statistical Society, Series B, Methodological* **60**(4), 765–780.

Statistical Sciences (1995*a*), *S-PLUS Guide to Statistical & Mathematical Analysis, Version 3.3*, StatSci, a division of MathSoft, Inc., Seattle.

Statistical Sciences (1995*b*), *S-PLUS Guide to Statistical & Mathematical Analysis, Version 3.3*, StatSci, a division of MathSoft, Inc., Seattle.

Åström, K. (1970), *Introduction to stochastic control theory*, Academic Press, London.

Venables, W. N. & Ripley, B. D. (1997), *Modern Applied Statistics With S-Plus*, 2 edn, Springer-Verlag, Berlin/New York.

Ye, J. (1998), 'On measuring and correcting the effects of data mining and model selection', *Journal of the American Statistical Association* **93**(441), 120–131.