Adaptive Load Forecasting

Philip Anton de Saint-Aubain

Kongens Lyngby 2011

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

Abstract

The purpose of this thesis is to contribute to the research in forecasting energy consumption in residential houses. The work is motivated by the Danish iPower project, which deals with investigation of possibilities for replacing fossil fuel with renewable energy. Renewable energy in Denmark is mostly based on wind power which is a highly fluctuating energy source and it is difficult to conserve. Energy consumption is also varying but independent of supply to the power plant. The fact that energy supply and energy consumption is not synchronized could be handled with a methodology that facilitates using the energy when present. The present work provides an adaptive method to get detailed knowledge of the energy consumption in residential houses. The method will be a contribution to forecasting energy consumption and to the development of Smart Grid technology.

The approach taken is to reveal the details in the heating consumption in residential houses by developing mathematical models for the heat load. Based on district heating consumption data from four houses in a small area in Denmark and data from a nearby meteorological station, models are developed for separating the heating signals into different components. One of the models is able to split the overall consumption into heating consumption and hot water consumption. The heating consumption is further separated into parts explained by diurnal variation and variation explained by changes in outdoor temperature and the amount of solar radiation present. The method is adaptive to changes in the consumption due to variation in the daily routine of the inhabitants.

The results are obtained by using mathematical modeling, statistics and time series analysis. For separating the hot water consumption and heating Low Pass Filters and advanced Kernel Smoothing techniques are used. The Kernel

Smoother is extended to contain robust estimation and polynomial shape kernels. The further separation of the heating consumption is done with Kalman Filter techniques for signal separation.

Keywords
Mathematical Modeling, Statistics, Time Series Analyses, Signal Separation, iPower, Smart Grid, Low Pass Filter, Kernel Smoother, Polynomial Kernel, Robust Estimation, Kalman Filter

Resumé

Formålet med dette eksamensprojekt er at bidrage til, hvordan man kan forudsige energiforbrug i parcelhuse. Arbejdet er motiveret af det danske iPower projekt, som har til formål at undersøge mulighederne for at erstatte fossilt brændstof med vedvarende energi. Vedvarende energi er i Danmark fortrinsvis baseret på vindenergi, som er en ustabil energikilde og vanskelig at lagre. Desuden varierer energiforbruget uafhængigt af forsyningen til kraftværket. Det faktum, at energiforsyning og energiforbrug ikke er synkroniseret, kan håndteres ved at energien bliver forbrugt når den er til stede. Dette projektarbejde giver en metode til at få detaljeret viden om energiforbruget i parcelhuse. Metoden vil være et bidrag til at kunne forudsige energiforbrug og til udvikling af Smart Grid teknologi.

I dette projekt er der udført en detaljeret undersøgelse af varmeforbruget i parcelhuse ved udvikling af matematiske modeller. Baseret på varmeforbrugsdata fra et fjernvarmeværk og klimadata fra en nærliggende vejrstation, er der udviklet modeller, der adskiller varmeforbruget i forskellige komponenter. En af modellerne giver mulighed for at adskille det samlede varmeforbrug i forbrug til opvarmning og forbrug til varmt vand. Varmeforbruget er yderligere opdelt i komponenter, der kan forklares ved rutinemæssig daglig variation og de komponenter af varmeforbruget, der kan beskrives ved ændringer i udendørstemperaturen og den aktuelle solstråling. Modellen er adaptiv med hensyn til ændringer i daglige rutiner.

Resultaterne er opnået ved at bruge matematisk modellering, statistik og tidsrækkeanalyse. Til at adskille varmtvandsforbruget og varmeforbruget er der brugt et Lavpasfilter og avanceret Kernel Smoothing. Kernel Smoothing er udvidet til at indeholde robust estimation og polynomisk kernel. Den efterfølgende adskillelse af varmeforbrugets elementer er udført med Kalman Filter for Signal Separation.

Keywords
Matematisk Modellering, Statistik, Tidsrækkeanalyse, Signal Separation, iPower, Smart Grid, Lavpasfilter, Kernel Smoother, Polynomisk Kernel, Robust Estimation, Kalman Filter

Preface

This thesis was prepared at Department of Informatics and Mathematical Modelling at the Technical University of Denmark in partial fulfillment of the requirements for acquiring the Master degree in engineering. The thesis was written during the period 1 February to 31 July 2011. The workload of six months is equivalent to 30 ECTS points. The thesis is linked to the Danish iPower project and is carried out in collaboration with the Danish company Grundfos and Enfor. All the described methods are implemented in the statistical software R.

Kongens Lyngby, July 2011

Philip Anton de Saint-Aubain

Acknowledgements

I am grateful to Professor Henrik Madsen (DTU) for encouragement, guidance and support from the initial to the final level of this project. Ph.D. Peder Bacher (DTU) has given much work in support and discussions which has been truly helpful. The necessary expert knowledge on modeling of the district heating has been given from Assoc. Prof. Henrik Aalborg Nielsen (Enfor) to whom I am deeply grateful. Discussions with Mr. Hakon Børsting (Grundfos) concerning the practical use of the results of this investigation have been most valuable. Thanks to Bengt Perers (DTU) for sharing his expert knowledge in solar radiation. Sønderborg Fjernvarme has collected the data and Enfor has made the data available for the project.

Contents

Chapter 1

Introduction

Burning of fossil fuels is considered the greatest contributor to human generated greenhouse gas emission. In 2009 the governments of EU confirmed a goal to reduce the greenhouse gases by 80%-95% in 2050 relative to 1990. In the decades to come global growth will require significantly greater amounts of energy. Therefore, it is expected that price of the fossil fuels, which today supply the majority of society's demand for energy, will increase. In 2008, 80% of the Danish energy consumption came from fossil fuel. Based on the concern of greenhouse gas emission and the increasing prices of fossil fuel the Danish Climate Commission has set the goal to reduce the 80% to 0% before the year 2050 [5]. The work of the Climate Commission has 40 recommendations, one of these is that about 40% of the energy supply should be generated by wind turbines. Wind turbines produce electricity which is more difficult and costly to conserve than oil, coal and gas. Further, the energy supplied by wind is extremely fluctuating, depending on when the wind is blowing and the speed of the wind. As the demand for energy is independent of the power supply but dependent of season and time of day, the energy is not necessarily available when energy is needed. This problem must be solved before wind energy can be accepted as a main energy source. To meet the problem of very irregular power supply the Climate Commission recommends two different approaches: one is to combine wind energy with other energy systems, the other is to make electricity consumption more flexible, which means to use electricity when it is generated and to reduce energy consumption when the demand exceeds the supply. To make energy consumption flexible and more suitable for the fluctuations of the renewable energy, rather than letting the society claim supply at certain times, is a new and highly interesting way of looking at energy production. To be forced to use renewable energy while it is available does not necessarily reduce total energy consumption, but will reduce the consumption of fossil fuel. One of the methods of reducing supply problems during the peak consumption periods in the morning and the afternoon is to reduce any unnecessary energy consumption during these hours. If there are any possibilities to make buildings work as buffer systems it will be possible to use extra energy before and after the peak periods, thereby reducing the consumption during the peak periods. As heating of residential buildings is one of the most energy consuming areas in Denmark and the heating supply is almost entirely produced by fossil fuel, it would be important to reveal and investigate to which extent the houses can act as energy buffers when heating will be based on electricity. If the houses could act as energy buffers they would be much more resistant to energy fluctuations. This study is a contribution to reveal the possibility of flexible energy consumption in family houses.

1.1 Aim

The aim of this study is to contribute to the knowledge of how and to which extent residential buildings heat consumption can be an active part of an electricity network based on renewable energy sources. This is under the assumption that the heating of houses will be based on renewable electric energy, either through district heating companies or by electric radiators. In this study statistical methods are used in order to gain insight in how the buildings' heat consumption can be a part of a grid based on renewable energy. The statistical methods are used on actual data of heating consumption of individual houses. The data available to this study is measurements of the total energy consumption in individual houses, which include both hot water and central heating consumption. The consumption will be separated into different parts, explaining each their contribution of the total consumption. In this way much information is obtained from only one measurement device. This investigation will contribute with the necessary information to further analyze how the houses can actively be a part of the future grid.

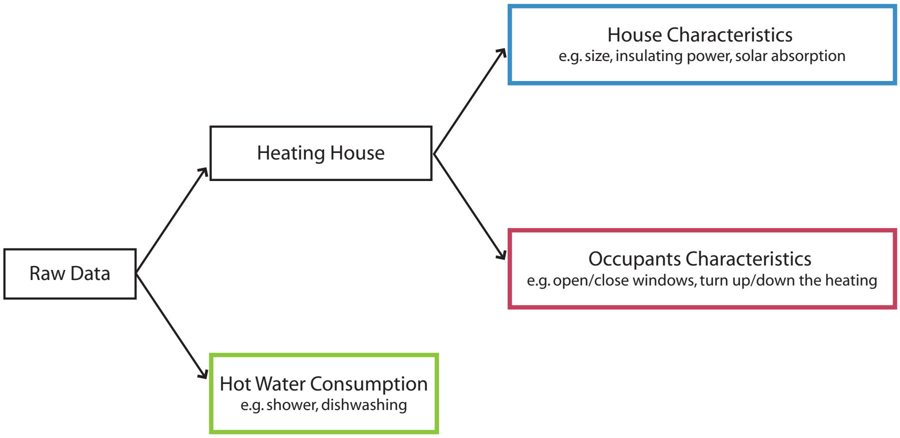

Figure 1.1: Diagram showing how the data will be separated.

1.2 Project Description

To analyze the overall consumption from the district heating company several methods are used to separate different parts of the consumption. Initially the energy is separated into the occupants heating consumption and their hot water use. Low Pass Filter and Kernel methodology is used to split the hot water use from the heating consumption. After this separation the heating consumption is further analyzed in order to reveal the characteristics of the houses and the occupants. This is done by separating the heating consumption into two parts, one part is the consumption influenced by the house characteristics and the other is the consumption influenced by the occupants' behavior. A Kalman Filter is used for this separation. The buildings characteristics are, for instance, influenced by, how well the building is insulated, the radiation to and from the house and how well the house is sealed. Furthermore, it is assumed that the outdoor temperature and solar radiation are exogenous factors which influence the heating consumption and will be used as input to the model. Thus, the part of the consumption which is explained by the house characteristics is the consumption which would be seen in a house with no inhabitants. It is assumed that some of the heating consumption is explained by the regular behavior of the occupants. The occupants' regular behavior is modeled by a diurnal variation. The part of the consumption, which is seen as daily variation, is the regular pattern of the occupants' behavior within a 24 hours period. The behavior could be opening and closing of windows and doors, manual changing of the radiators, use of electric devices which produce heat, together with the heat generated by the inhabitants themselves. As the behavior of inhabitants can change over time, this part is made adaptive, so it automatically detects changes in the behavior of the inhabitants and change the separation accordingly. The splitting procedure is illustrated in Figure 1.1.

1.3 Outline

Separating different kinds of energy consumption in data from a district heating company by statistical methods is described in this study. The content of the thesis is listed below.

Chapter Two	Description of the available data.
Chapter Three	Explaining methods for splitting heating and hot water consumption. The methods will be based on Low Pass Filters and Kernel Smoothers.
Chapter Four	Explaining method for separating heating consumption into different components using Kalman Filter. To investigate the preformance of the implemented Kalman Filter simulated data series were analyzed.
Chapter Five	Separating heating consumption into different components. Kalman Filter is used for data separation to distinguish house and occupants characteristics in the consumption.
Chapter Six	Discussion.
Chapter Seven	Conclusion.

Chapter 2

Data

The data available for this study is the heating consumption of individual residential buildings including central heating of the houses and hot water use. Sønderborg District Heating Company located in Southern Denmark delivered the data. The collection period was from 18^{th} November 2008 to 1^{st} of September 2010. Climate data are available from a nearby meteorological station.

Around 8000 households are connected to the district heating of which data from 56 households are available to this study. In order to start with a simplified investigation only four representative households were analyzed. The data was logged approximately every 10^{th} minute during the collection period. There were some technical failures during the data collecting, hence some of the data points are missing.

The measurement unit of the heating consumption is mega joule per hour [MJ/h]. The data points are time labeled Greenwich Mean Time [GMT]. To get from GMT to local Danish time you need to add one hour in the winter season (from last Sunday in October to last Sunday in March) and add two hours in the summer season.

A visual inspection of the data showed that the different houses have different characteristic. Some houses have a very regular consumption, while others showed a more irregular pattern. Four representative households were chosen in such a way that they cover the different characteristics. A few of the measured data contained some unrealistic high values, these outliers have been removed.

2.1 Houses

The four chosen houses are all single-family free-standing residential buildings. They are all build with a single floor and they do not have a basement. None of the chosen houses uses other heating devises than district heating, meaning no electrical heating, solar thermal collector, geothermal heat or any other kind of heat supply. The houses are located near to each other. An aerial photograph of Søenderborg with information about the four houses is shown in Figure 2.1.

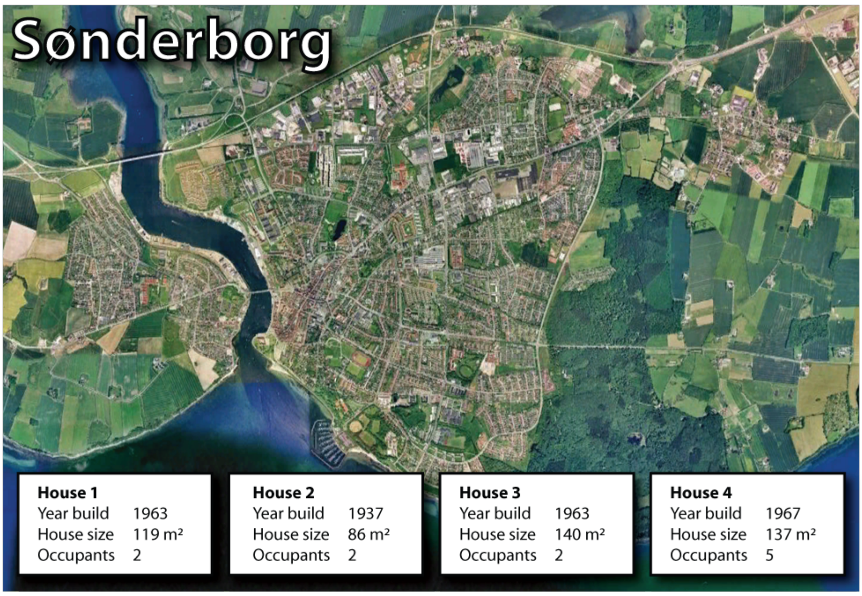

Figure 2.1: Aerial photograph of Sønderborg with house data.

The water from the district heating plant is used in the central heating systems of the houses. The clean water used for tap water is heated by the district water by a heat exchanger. Sometimes a hot water tank makes a buffer capacity. The consumption pattern could depend on whether a hot water tank is used or not. Using the heat exchanger without a hot water tank you expect spikes in the data every time the occupants are turning on the tab. When hot water tank is used the consumption will not be registered until the temperature in the tank is low enough for the thermostat to be active. It is not known whether the houses use a hot water tank. But Sønderborg District Heating Company recommends their customers not to use hot water tanks.

2.2 Heating Consumption

The energy consumption in the whole period for the four houses is shown in Figure 2.2. A spring and winter period of two weeks is shown in Figure 2.3. Figure 2.2 visualize how the energy consumptions changes during the year, from a high consumption in the winter to a low consumption in the summer time. Figure 2.3 visualizes the difference in consumption between spring and winter.

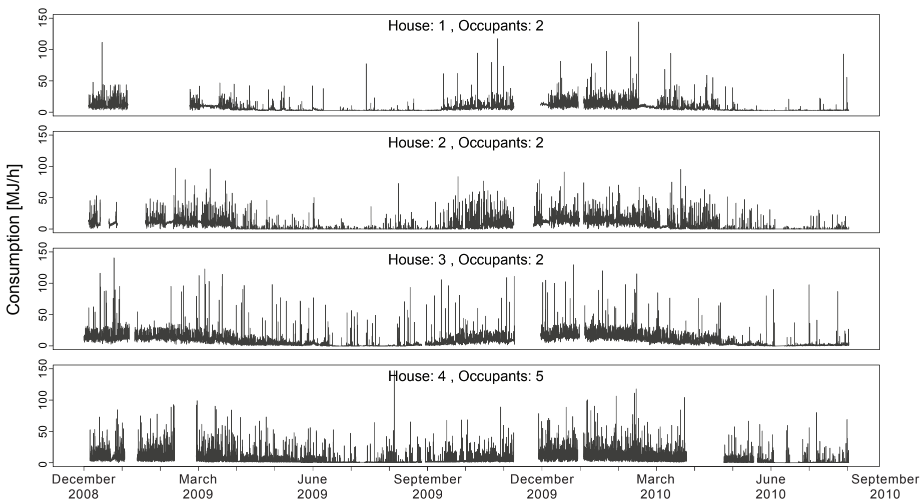

Figure 2.2: Consumption data for the whole sampling period.

Looking at the plots in Figure 2.2 and Figure 2.3, reveals that some slow varying changes and some fast varying changes are present in the data. The slow varying changes are mainly changes over the seasons or over day and night. The fast varying changes are seen as spikes in the plots and they are most probably due to the hot water usages. During the summer season the weather in Denmark is warm enough to turn off the radiators. Therefore, in the summer and spring period it should only be the hot water usages that are seen in the total consumptions. It is also worth to notice that the spikes are in most cases more irregular than the slow varying changes.

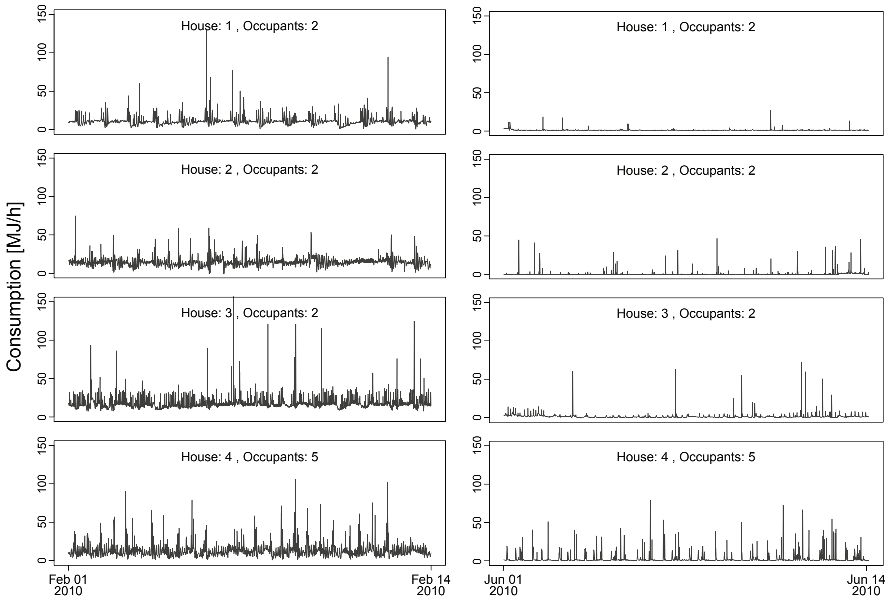

Figure 2.3: Consumption in 14 days of winter and 14 days of spring.

2.3 Climate Measurements

Climate measurements are available from a nearby meteorological station. The measured variables are wind speed $[m/s]$, air temperature [°C] and illuminance $[lux]$. The latter is a measure of the amount of visible light per area on a horizontal plane. Illuminance cannot be directly converted to power units, because it is weighted according to the wavelength of human brightness perception. But anyhow, the outdoor illuminance is expected to be highly correlated with the solar radiation. Thus the illuminance variable will be used as an indication of how much solar radiation is present at a given point of time. The measurements are given as 10 minute averages. Plots of the climate measurements are shown in Figure 2.4.

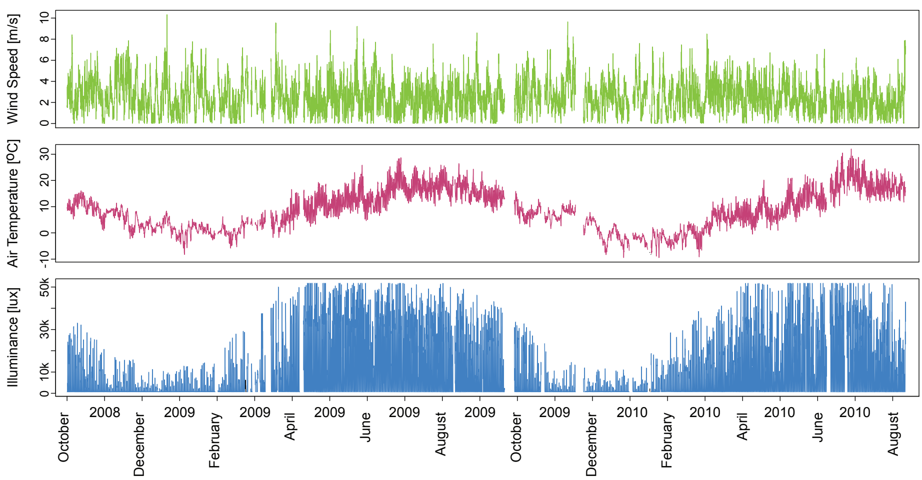

Figure 2.4: The measured climate variables.

Chapter 3

Splitting Hot Water and Heating Consumption

This chapter describes several methods for splitting the data into hot water consumption and heating consumption. When looking at plots of raw data it is apparent that data consist of two different kinds of fluctuations. One of the signals seems to have slow varying trends, while the other part looks like irregular spikes. It is expected that these two kinds of fluctuations describe heating consumption and hot water consumption, respectively. Since the measurements are 10 minutes average values, the high spikes are not expected to be a result of heating, as the heating systems hardly respond up and down in this high level in such short time. Another indication that the high spikes represent hot water consumption is that during summer time, where the heating is turned off, all consumption is seen as spikes. A Low Pass Filter method and a Kernel Smoother method will be introduced in order to separate the two different kinds of fluctuations.

To test the two methods data from house no. 1 for a period of March 2010 is chosen. Figure 3.1 shows the raw data of this period. Some of the spikes are as high as 160MJ/h and have been cut off by the frame in order to make the lower variations visible. The figure shows that in a two weeks period from Friday 12$^{\text{th}}$ until Friday 26$^{\text{th}}$ there are no spikes and the consumption has very little variation during this period. It is therefore assumed that the inhabitants were on holidays and left the house during these two weeks. This theory is supported by the observation that a similar period of low consumption without spikes is also seen at the same time in the previous year. Using a holiday period should reveal if the methods are successful in splitting the data, as no hot water consumption is expected during the holiday period contrary to the periods before and after.

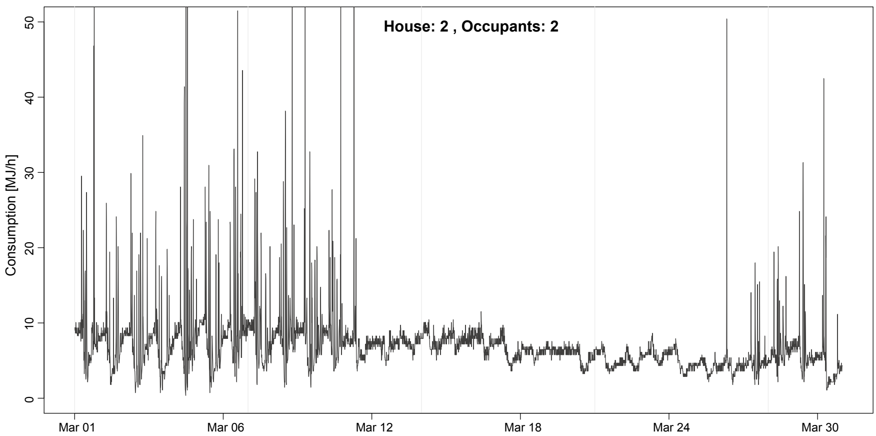

Figure 3.1: Raw data for house no. 1 during March 2010.

3.1 Low Pass Filter

To get a first impression of how to separate the data in the heating consumption and the hot water usage a Low Pass Filter is implemented. A Low Pass Filter is designed to dampen fast variation while keeping the slow long term trends. If the heating consumption and hot water usage operate in different frequencies, it would be possible to use this method to separate the consumption types. The heating consumption is assumed to change with the outdoor temperature and therefore to change slowly during the day and during the year. Contrary to this, the hot water usage is assumed to behave as short term fluctuations. If these assumptions hold, the Low Pass Filter is expected to deliver good results for splitting the raw data into hot water and heating consumption.

An ideal low pass filter cuts off all frequencies larger than a certain threshold value. Unfortunately it is not possible construct such a filter without having signals of infinite extent in time. Therefore, an approximate ideal low pass filter is used. The function `FIR1` from R is used in this work. `FIR1` is an implementation of a finite impulse response filter.

The result from the Low Pass Filter are shown in Figure 3.2. The heating consumption in the figure is calculated by the Low Pass Filter. The hot water consumption is calculated by subtracting the Low Pass Filter from the raw data. The uppermost plot shows the original data. The middle plot shows the hot water consumption. The lower plot shows the heating consumption. Hence, the sum of the two lower plots is equivalent to the uppermost plot.

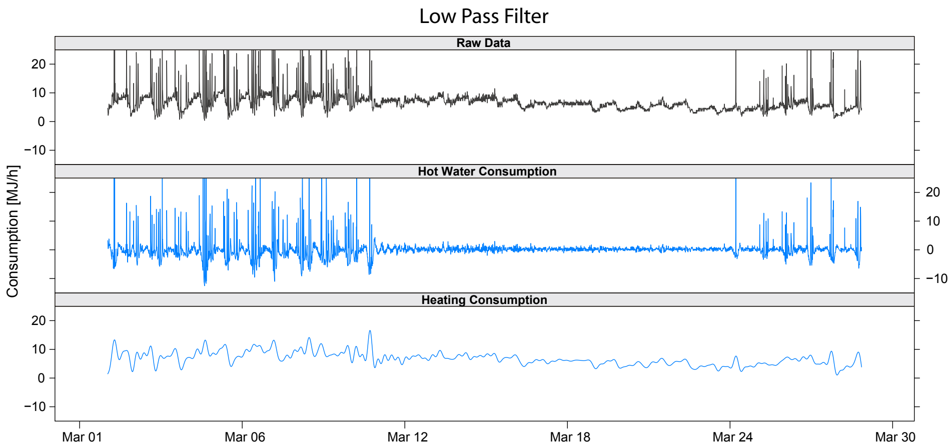

Figure 3.2: Splitting for house no. 1 using Low Pass Filter.

The Low Pass Filter seems to separate low and high frequencies as expected. But the method, however, shows some drawbacks, like negative values of hot water consumption, which are not possible. It also shows hot water consumption during the holiday which was not expected. One reasonable explanation could be that the thermostat of the radiators is turning on and off when the indoor temperature reaches a given level. This could give rise to high frequency signals not caused by hot water consumption. Therefore, the assumption that the heating only behaves as slow variation is not verified. Thus another method to separate the hot water and heating consumption is needed.

3.2 Kernel Smoothing

In this section it is presented how the splitting of heat consumption is carried out with a Kernel Smoother. The theory is that spikes significantly higher than the Kernel Smoother estimate will represent hot water consumption. It is therefore suggested that the occurrence of spikes represent hot water consumption. Thus it would be possible to separate the raw data into hot water consumption and heating consumption by isolating the spikes.

A Kernel Smoother is a method to estimate the underlying function of some given noisy measurements. Kernel estimation is a nonparametric estimation technique, where there is no assumption about the shape of the true function nor any parameters to estimate. The formula for the estimation is

$$\hat{g}(x) = \frac{\sum_{s=1}^{N} Y_s k\{\frac{x-X_s}{h}\}}{\sum_{s=1}^{N} k\{\frac{x-X_s}{h}\}} \tag{3.1}$$

where $\hat{g}(x)$ is the kernel estimate for a given x. N is the number of observations, X_t and Y_t are the x and y-value of the t^{th} observation in the time series, h is a chosen bandwidth parameter. Thus the Kernel Smoother is a Local Weighted Average around the point x. The function $k(\cdot)$ is the kernel, which determines how the weight should be put on the neighboring data points. The Gaussian kernel $k(u) = \frac{1}{2\pi}\exp\{-\frac{u^2}{2}\}$ is chosen. The bandwidth h is a smoothing parameter determining the width of the kernel used. As $h \to \infty$ the estimate will go toward the overall mean value $\hat{g}(x) = \bar{Y}$. Thus for large values of h the kernel estimate will be a bias at high curvature places in the data series. As $h \to 0$ the kernel estimate will just be the actual data point and thus there will be no bias, but instead a large variance. By empirically investigations $h = 3000$ is chosen as a reasonable value.

The Kernel Smoother estimate is used to split hot water and heating consumption. The hot water consumption is assumed to be the spikes definded as values above "1.25 $\cdot$ kernel estimate". It seems reasonable to use percentage value of kernel estimate instead of fixed value to separate, because both the heating fluctuations and the hot water spikes are larger in the cold period and smaller in warm periods. The heating consumption is measured by subtracting hot water from the raw data. The result is shown in Figure 3.3.

Heating consumption is not expected to have short-lived spikes, but such spikes are seen in Figure 3.3. The cause is that the kernel estimate is too affected by the very high spikes implying that even high spikes is below the "1.25 $\cdot$ kernel estimate". In order to improve the kernel method at separating the spikes, two different expansions of the method are investigated. The first approach is based on robust estimation techniques, while the second method uses local weighted polynomial instead of Local Weighted Average. It is necessary to show that the Kernel Smoother is equivalent to the Local Least Square method before the two methods are introduced, as this is the basis for both methods.

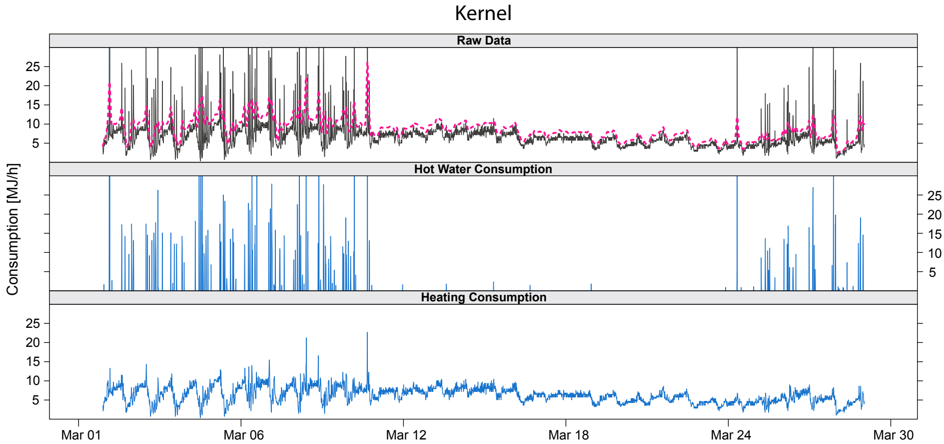

Figure 3.3: Splitting for house no. 1 using Kernel Smoother. The red dashed line is 1.25 · kernel estimate.

3.2.1 Least Square Parallel

In this section it is shown that the Kernel Smoother is equivalent to the Local Least Square method. First step is to rewrite the kernel equation (3.1) to

$$\hat{g}(x) = \frac{1}{N}\sum_{s=1}^{N} w_s(x) Y_s \tag{3.2}$$

where the weight is given by $w_s(x) = \frac{k\{x-X_s\}}{\frac{1}{N}\sum_{s=1}^{N} k\{x-X_s\}}$. The corresponding Least Square problem is

$$\arg\min_{\theta} \frac{1}{N}\sum_{s=1}^{N} w_s(x)\left(Y_s - \theta\right)^2 \tag{3.3}$$

The argument which minimizes the expression in (3.3) can be found analytically by differentiating the expression with respect to θ and equating it to zero. Differentiating it again reveals that the curvature is positive and hence the solution is a global minimum. Thus the Least Square solution to (3.3) is

$$\hat{\theta}_{LS} = \frac{\sum_{s=1}^{N} w_s(x) Y_s}{\sum_{s=1}^{N} w_s(x)} \tag{3.4}$$

Rearranging (3.4) gives $\sum_{s=1}^{N} w_s(x) Y_s = \hat{\theta}_{LS} \sum_{s=1}^{N} w_s(x)$. Plugging this into (3.2) and using the fact that $\frac{1}{N} \sum_{s=1}^{N} w_s(x) = 1$ gives

$$\hat{g}(x) = \frac{1}{N} \hat{\theta}_{LS} \sum_{s=1}^{N} w_s(x) = \hat{\theta}_{LS} \tag{3.5}$$

The equation states that the kernel estimate $\hat{g}(x)$ is equal to the least square solution $\hat{\theta}_{LS}$. Thus the Kernel Smoother is identical to a local Least Square problem.

3.2.2 Robust Estimation

When the Kernel Smoother was used the kernel estimate was too high during periods of spikes. In order to solve this issue the robust estimation is investigated. The idea behind robust estimation is to make the estimation method robust against large outliers. One way of doing this is to use Huber Estimation, which is a modification of the well known least square estimation. Many optimization methods try to minimize some function ρ of the residuals ε. The least square method tries to minimize the square of the residuals $\rho_{LS} = \varepsilon^2$, whereas the Huber estimation tries to minimize the Huber function

$$\rho_{\text{Huber}}(\varepsilon) = \begin{cases} \frac{1}{2\gamma}\varepsilon^2 & \text{if } |\varepsilon| \le \gamma \\ |\varepsilon| - \frac{1}{2}\gamma & \text{if } |\varepsilon| > \gamma \end{cases} \tag{3.6}$$

The Huber function is quadratic for small residuals and linear for residuals larger than γ. A plot of ρ_{Huber} and a scaled version of ρ_{LS}, together with their derivatives is shown in Figure 3.4. The derivatives are also known as the influence function.

Since the influence function is bounded, an outlier cannot cause significant displacement of the resulting estimate. The minimum of $\sum \rho(\varepsilon)$ satisfy the equation $\sum \rho'(\varepsilon) = 0$. From this it is seen that residuals larger than γ do not affect

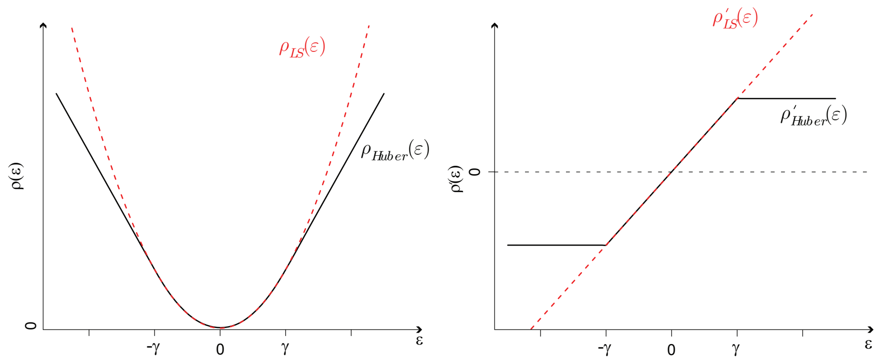

Figure 3.4: Left: Huber and a square function. Right: The derivatives also known as the influence function.

the estimate any more than if they were on the γ value. The parameter γ is a selected threshold for the Huber function, determining which residuals are large. By empirical investigations it is found that $\gamma = 3$ was a good choice in this case. `optimize` from R is used to find the solution to the minimization problem

$$\arg\min_{\theta} \frac{1}{N} \sum_{s=1}^{N} w_s(x) \rho_{\text{Huber}}(\varepsilon_s) \tag{3.7}$$

where $\varepsilon_s = Y_s - \theta$. The result is a generalized kernel estimation with a Huber function instead of Least Square function. The result is shown in Figure 3.5

The robust kernel appears to solve the problem that the original kernel estimate was too affected by the large spikes. As a consequence many of the spikes in the heating consumption are removed compared to the original kernel estimate. Therefore, the robust kernel appears to be more suitable for splitting raw data into hot water consumption and the heating consumption.

3.2.3 Polynomial Estimation

The previous versions of Kernel Smoother were based on Local Weighted Average. In this section the Kernel Smoother is based on an extended version of Local Weighted Average, namely local weighted polynomial. Above it was

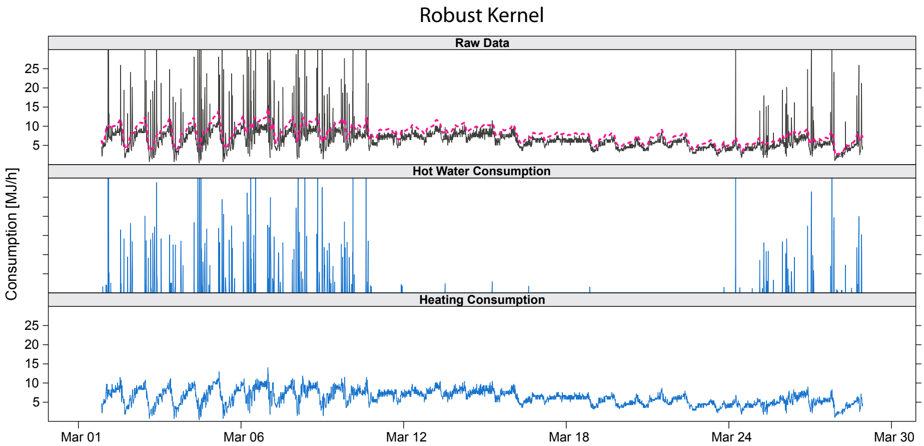

Figure 3.5: Splitting for house no. 1 using Robust Kernel Smoother. The red dashed line is 1.25 · kernel estimate.

shown that the Kernel Smoother could be written as a weighted Least Square problem of the form

$$\arg\min_{\theta} \frac{1}{N} \sum_{s=1}^{N} w_s(x)(Y_s - P_s)^2 \tag{3.8}$$

where $P_s = \theta$. Instead of only estimating one parameter θ, it is intended to estimate a polynomial of the form

$$P_s = \theta_0 + \theta_1(X_t - x) + \theta_2(X_t - x)^2 \tag{3.9}$$

The advantage is that a polynomial has an increased ability to follow local variation in the data. Normally kernel estimates have a tendency to get biased in high curvature places. This risk is reduced by using the polynomial. Thus the local kernel estimate will be given by $\hat{g} = \hat{\theta}_0$.

The result of splitting the raw data into hot water and heating consumption using the polynomial kernel estimation is shown in Figure 3.6. The polynomial kernel has improved the ability to remove the spikes from the heating consumption compared to the original Kernel Smoother method. However, it is weaker than the robust kernel. In the next section the robust kernel and polynomial fit is combined into one single method.

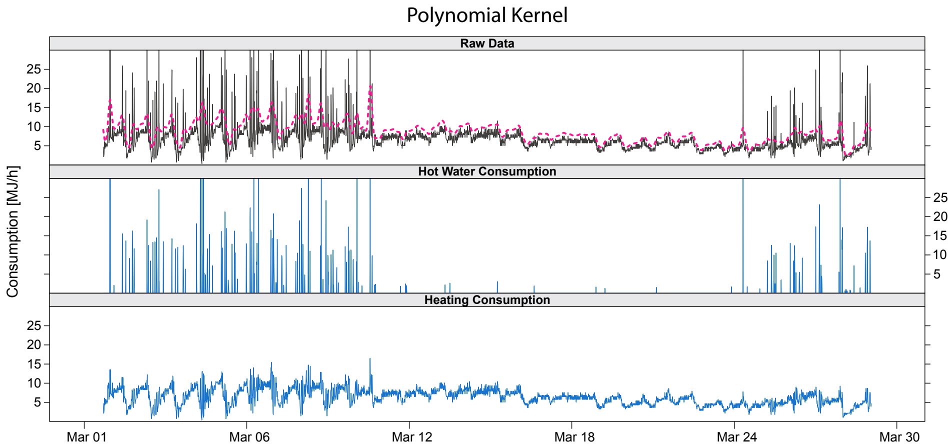

Figure 3.6: Splitting for house no. 1 using Polynomial Kernel Smoother. The red dashed line is 1.25 · kernel estimate.

3.2.4 Robust & Polynomial

The robust and polynomial estimation techniques can be combined into one method. The result is shown in Figure 3.7. Almost all spikes are separated from the raw data leaving the heat consumption without spikes. Under the assumption that the hot water consumption is the high spikes and heating consumption does not have spikes, this method is the most convincing obtained in this work. In conclusion, the combined Robust & Polynomial Kernel is chosen as the best method.

3.3 Results

The results of splitting the raw data from 2010 into hot water and heating consumption for the remaining three houses are shown in Figure 3.8, 3.9 and 3.10. The Robust & Polynomial Kernel Smoother, which was found as the best method for splitting the hot water and heating consumption, is used for splitting

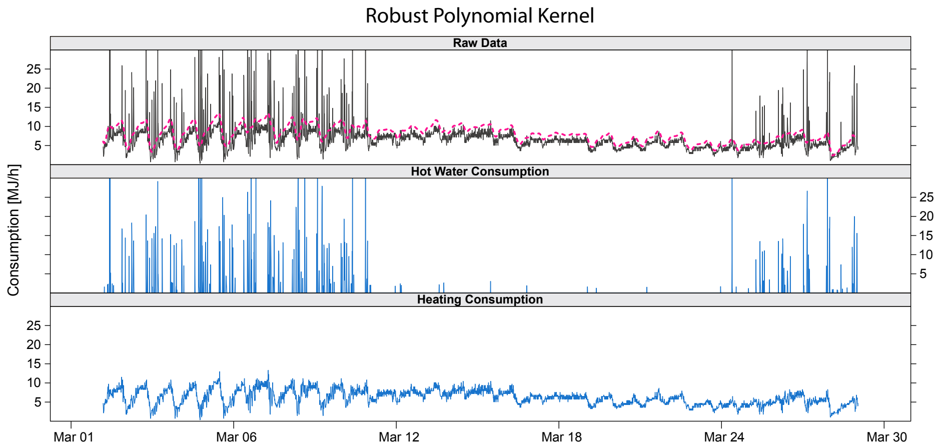

Figure 3.7: Splitting for house no. 1 using the Robust Polynomial Kernel Smoother. The red dashed line is $1.25 \cdot$ kernel estimate.

the data in the remaining houses. It is seen that house no. 4 has a much larger hot water consumption compared to the other houses. This was expected, as there are five occupants in house no. 4 compared to two occupants in each of the other houses.

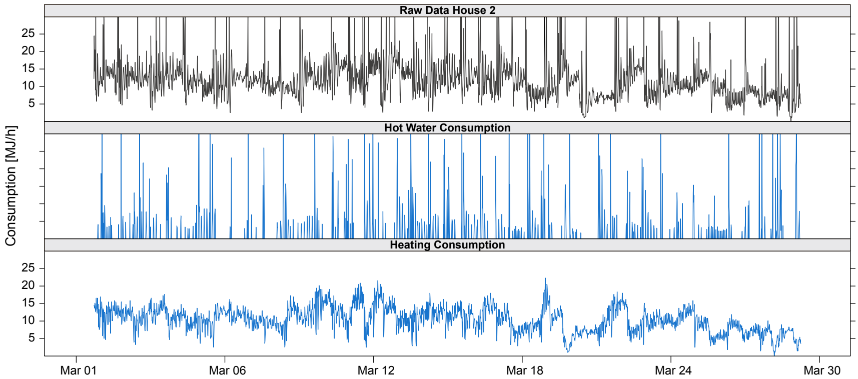

Figure 3.8: Splitting for house no. 2 using Robust Polynomial Kernel Smoother.

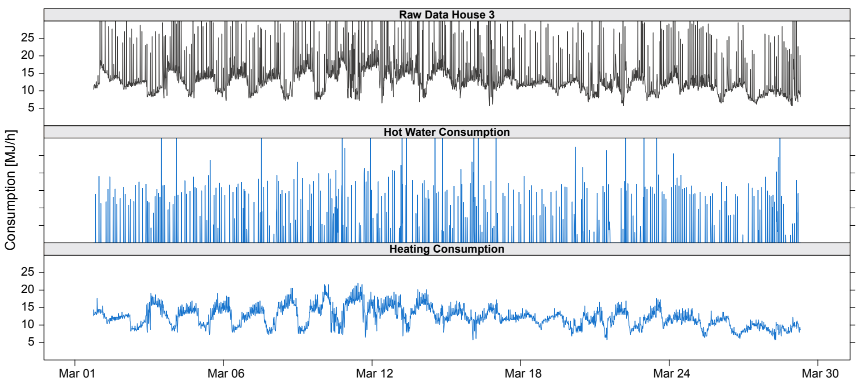

Figure 3.9: Splitting for house no. 3 using Robust Polynomial Kernel Smoother.

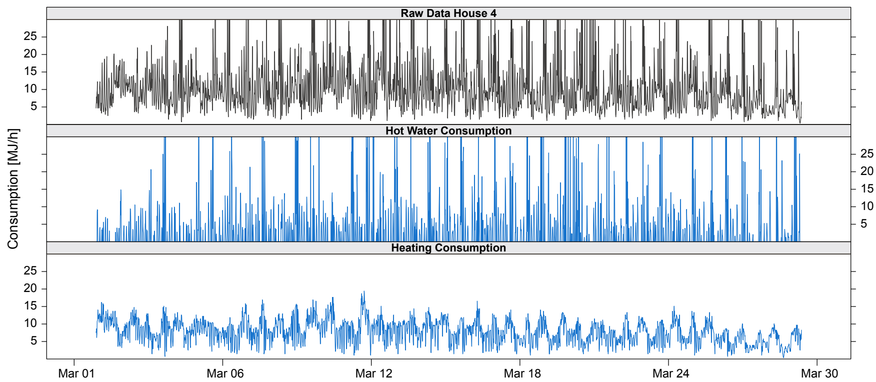

Figure 3.10: Splitting for house no.4 using Robust Polynomial Kernel Smoother.

Chapter 4

Kalman Filter

The heating consumption found in Chapter 3 will be separated into several components. It is expected that the heating consumption is influenced by several parameters such as the outdoor temperature, the solar radiation and daily variations in consumption pattern. Based on expectation of possible influences the heating consumption will be separated into components corresponding to different influences. Kalman Filter for Signal Separation, as described in [8], is used to separate the heating consumption. The meteorological measurement described in Chapter 2 will be a part of the input. In order to get experience with the Kalman Filter and the methods for modeling dynamics, some simulated data will be generated and analyzed. The advantage of using simulated data is that it is possible to test how well the Kalman Filter separates the data into the different components.

4.1 Kalman Filter for Signal Separation

A Kalman filter operates on a Linear Stochastic State Space Model in discrete-time. State Space Models are any model which includes an observation series Y_t and an unobservable state series $\boldsymbol{X}_t$. Furthermore, the model can include exogenous data inputs. The Stochastic State Space Models also contain two types of noise terms. State noise which allow states estimates to be subject to some uncertainties. Measurement noise which take into account the unavoidable noise in the measurements. The model consists of a state equation

$$\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{\Sigma}_1^* \boldsymbol{\varepsilon}_{1,t} \tag{4.1}$$

and an observational equation

$$Y_t = \boldsymbol{C}\boldsymbol{X}_t + \Sigma_2^* \varepsilon_{2,t} \tag{4.2}$$

Where $\boldsymbol{X}_t$ is the unobservable state vector, Y_t is the observations and $\boldsymbol{u}_t$ is the exogenous data input. The matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{\Sigma}_1^*$ and the scalar Σ_2^* are in principle known. Although in many practical problems, the matrices have to be estimated. As the structures of the matrices are modeled, only their elements need to be estimated. The noise terms $\boldsymbol{\varepsilon}_{1,t}$ and $\varepsilon_{2,t}$ are assumed independent normally distributed random variables. The covariance matrix of the system noise $\boldsymbol{\Sigma}_1^* \boldsymbol{\varepsilon}_{1,t}$ is given by $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1^* \ \boldsymbol{\Sigma}_1^{*T}$ and the variance of measurement noise $\Sigma_2^* \varepsilon_{2,t}$ is given by $\Sigma_2 = \Sigma_2^{*2}$. Calculating the covariance and variance in this way insures that the covariance matrix stays positive semi-definite and the variance stays positive. Since the model in equation (4.1) and (4.2) is a state space model, the Kalman Filter can be used to estimate the states $\boldsymbol{X_t}$. Kalman Filter is a recursive technique to find the optimal reconstruction and prediction of $\boldsymbol{X_t}$ given the observations Y_t and the exogenous data input $\boldsymbol{u}_t$. Thus the method only uses past observations to calculate the states. The recursive equations for the filter are shown below.

Prediction

$$\begin{aligned}
\widehat{Y}_{t+1|t} &= \boldsymbol{C}\left(\boldsymbol{A}\widehat{\boldsymbol{X}}_{t|t} + \boldsymbol{B}\boldsymbol{u}_{t+1}\right) \\
\widetilde{Y}_{t+1|t} &= Y_{t+1} - \widehat{Y}_{t+1|t} \\
\boldsymbol{\Sigma}^{xx}_{t+1|t} &= \boldsymbol{A}\boldsymbol{\Sigma}^{xx}_{t|t}\boldsymbol{A}^T + \boldsymbol{\Sigma}_1
\end{aligned}$$

Variance of Prediction Error

$$R_{t+1} = \boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T + \Sigma_2$$

Kalman Gain

$$\boldsymbol{K}_{t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T R_{t+1}^{-1}$$

Reconstruction

$$\begin{aligned}
\boldsymbol{\Sigma}^{xx}_{t+1|t+1} &= \boldsymbol{\Sigma}^{xx}_{t+1|t} - \boldsymbol{K}_{t+1}\boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{t+1|t} \\
\widehat{\boldsymbol{X}}_{t+1|t+1} &= \boldsymbol{A}\widehat{\boldsymbol{X}}_{t|t} + \boldsymbol{B}\boldsymbol{u}_{t+1} + \boldsymbol{K}_{t+1}\widetilde{Y}_{t+1|t}
\end{aligned}$$

The symbols of the equations are described in Table 4.1. The Kalman Filter is optimal in the way that it minimizes the expected squared prediction errors.

Symbols	Description
$\widehat{Y}_{t+1\|t}$	Measurement prediction
$\widetilde{Y}_{t+1\|t}$	One-step prediction error
R_{t+1}	Variance of prediction error
$\Sigma^{xx}_{t+1\|t}$	Covariance of predicted state
K_{t+1}	Kalman gain
$X_{t+1\|t+1}$	Reconstructed state
$\Sigma^{xx}_{t+1\|t+1}$	Reconstructed state covariance

Table 4.1: Description of symbols.

For a thorough discussion and explanation of the Kalman Filter see for instance [8]. To start the Kalman Filter initial values are needed.

$$\widehat{\boldsymbol{X}}_{0|0} = \boldsymbol{\mu}_0$$
$$\boldsymbol{\Sigma}^{xx}_{0|0} = \boldsymbol{V}_0$$

When working with real data, often one or more data points are missing due to failure of the measurement device. If the Kalman Filter reaches a missing data point it will use the prediction as an estimate of the reconstruction. As shown in the following equation

$$\widehat{\boldsymbol{X}}_{t+1|t+1} = \widehat{\boldsymbol{X}}_{t+1|t}$$
$$\boldsymbol{\Sigma}^{xx}_{t+1|t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t}$$

where $\widehat{\boldsymbol{X}}_{t+1|t} = \boldsymbol{A}\widehat{\boldsymbol{X}}_{t|t} + \boldsymbol{B}\boldsymbol{u}_{t+1}$. If the Kalman Filter reaches a series of missing observations the variance of prediction error R_t will increase for each step. This will continue until a new non-missing observation is reached. As mentioned, all elements of the matrices for the State Space Model $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{\Sigma}^*_1$, $\boldsymbol{C}$ and the scalar Σ^*_2, are not necessarily known, but the structures of the matrices needs to be chosen as part of the modeling. A Maximum Likelihood Estimation (MLE) technique will be used to estimate these elements. The MLE technique is built on the concept of calculating the probability of observing the measurement for a given set of parameters. When this probability is defined, standard optimization algorithms can be used to compute the set of parameters, which maximize the probability of observing the measurements. The Maximum Likelihood Estimation will be described in the following. Let $\mathcal{Y}_N = (Y_1, Y_1, \ldots, Y_N)$ denote all observations up to time N. The likelihood function for the unknown parameters $\boldsymbol{\theta}$ given the N observations $\mathcal{Y}_N$ is given by

$$\mathcal{L}(\boldsymbol{\theta}; \mathcal{Y}_{\mathcal{N}}) = f(\mathcal{Y}_N|\boldsymbol{\theta})$$

Where the function $f(\cdot)$ is the distribution function of the measurements given the parameter. By condition on Y_0 the function can be rewritten to a product of conditional densities

$$\mathcal{L}(\boldsymbol{\theta}; \mathcal{Y}_{\mathcal{N}}) = f(Y_N|\mathcal{Y}_{N-1}, \boldsymbol{\theta})f(Y_{N-1}|\mathcal{Y}_{N-2}, \boldsymbol{\theta}) \cdots f(Y_1|Y_0, \boldsymbol{\theta})f(Y_0|\boldsymbol{\theta}) \qquad (4.3)$$

Under the assumption, that the state and measurement noises are normally distributed, the conditional distribution of Y_{t+1} given all the past observations are given by

$$f(Y_{t+1}|\mathcal{Y}_t) = [2\pi R_{t+1}]^{-1/2} \exp\left[\frac{-\widetilde{Y}^2_{t+1|t}}{2R_{t+1}}\right]$$

Using this formula together with (4.3) gives

$$\mathcal{L}\left(\boldsymbol{\theta}; \mathcal{Y}_N\right) = f(Y_0|\boldsymbol{\theta}) \prod_{i=1}^{N} [2\pi R(i)]^{-1/2} \exp\left[\frac{-\widetilde{Y}(i)^2}{2R(i)}\right] \qquad (4.4)$$

Where $\widetilde{Y}(i)$ is the one-step prediction error for the i^{th} observation, and $R(i)$ is the variance of prediction error for the i^{th} observation. It is possible to optimize equation (4.4) in order to obtain a maximum likelihood estimate of the parameters. But to avoid numerical underflow it is preferred to work with the logarithmic transformation of the likelihood function given by

$$\log \mathcal{L}(\boldsymbol{\theta}; \mathcal{Y}_{\mathcal{N}}) = -\frac{1}{2} \sum_{i=1}^{N} \left[\log R(i) + \frac{\widetilde{Y}(i)^2}{R(i)}\right] - \frac{N}{2} \log(2\pi) + \log(f(Y_0|\boldsymbol{\theta}))$$

Assuming that $f(Y_0|\boldsymbol{\theta})$ is independent of the parameters the part "$-\frac{N}{2}\log(2\pi) + \log(f(Y_0|\boldsymbol{\theta}))$" is just a constant. Since the argument which maximizes the likelihood function will not be changed by a constant value, the constant value can eliminated from the equation. Then the loglikelihood function is defined as

$$\ell\left(\boldsymbol{\theta}; \mathcal{Y}_N\right) = -\frac{1}{2} \sum_{i=1}^{N} \left[\log R(i) + \frac{\widetilde{Y}(i)^2}{R(i)}\right]$$

Maximum likelihood estimate of $\boldsymbol{\theta}$ is found as the argument which maximizes the loglikelihood function $\ell\left(\boldsymbol{\theta}; \mathcal{Y}_N\right)$

$$\hat{\boldsymbol{\theta}} = \arg\left\{\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \mathcal{Y}_N)\right\}$$

4.2 Square Root Filter

The Kalman Filter and the Maximum Likelihood Estimate as described were implemented in R. However, it was not possible to get an estimate because $\boldsymbol{\Sigma}^{xx}_{t+1|t}$ becomes singular which makes R_{t+1} negative. If $\boldsymbol{\Sigma}^{xx}_{t+1|t}$ was positive-semidefinite $\boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T \geq 0$ and R_{t+1} would never become negative, according to the formulas for R_{t+1}. Thus it is not theoretically possible to have a variance R_{t+1} that is negative or a covariance $\boldsymbol{\Sigma}^{xx}_{t+1|t}$ that is singular, so the problem must be due to numerical instabilities. To solve the problem a modification of the implementation is made to ensure that $\boldsymbol{\Sigma}^{xx}_{t+1|t}$ stays positive-semidefinite. The modified implementation is known as Square Root Filter [4]. The Square Root Filter takes advantage of the fact that positive semidefinite matrices have a corresponding triangular matrix square root. When the correlation matrix is kept in the square root form, it will always retain the positive diagonal elements and its symmetric form and thus be positive semidefinte. In the remaining part of the project a standard implementation of Square Root Filter from R is used. In the following the Square Root version of the Kalman Filter is just referred to as the Kalman Filter.

4.3 Simulated Test

In order to test the performance of the above described methods a simulation is made. The test on simulated data will reveal the stability and make a realistic expectation of the unavoidable uncertainty of the method. The result of the simulation test make a beneficial experience before the method is used on real consumption data. The data series used for the simulation test is generated from a chosen model. The parameters are then estimated using MLE. Finally the separation is performed using the Kalman Filter.

The simulation model in the example is taken from [8]. Data is simulated from an additive model of the form

$$Y_t = T_t + S_t + V_t + \epsilon_t \tag{4.5}$$

where

$T_t = T_{t-1} + w_{1t}$
$S_t = -\sum_{i=1}^{6} S_{t-i} + w_{2t}$
$V_t = \phi_1 V_{t-1} + w_{3t}.$

T_t is a random walk, S_t is a seasonal component and V_t is a stochastic $AR(1)$ model. ϵ_t, w_{1t}, w_{2t} and w_{3t} are independent normally distributed random variables. In the simulated model the variances and the parameter for the AR model are chosen to

$\mathbb{V}[w_{1t}] = \sigma_T^2 = 10$, $\mathbb{V}[w_{2t}] = \sigma_S^2 = 0.1$, $\mathbb{V}[w_{3t}] = \sigma_V^2 = 5$, $\mathbb{V}[\epsilon_t] = \sigma_\epsilon^2 = 1$, $\phi_1 = 0.8$

5000 data points where simulated from the model, where the last 500 simulations are shown in Figure 4.1. The uppermost part of the plot is the sum of the four bottom simulations. Only the uppermost part Y_t will be used in the separation test.

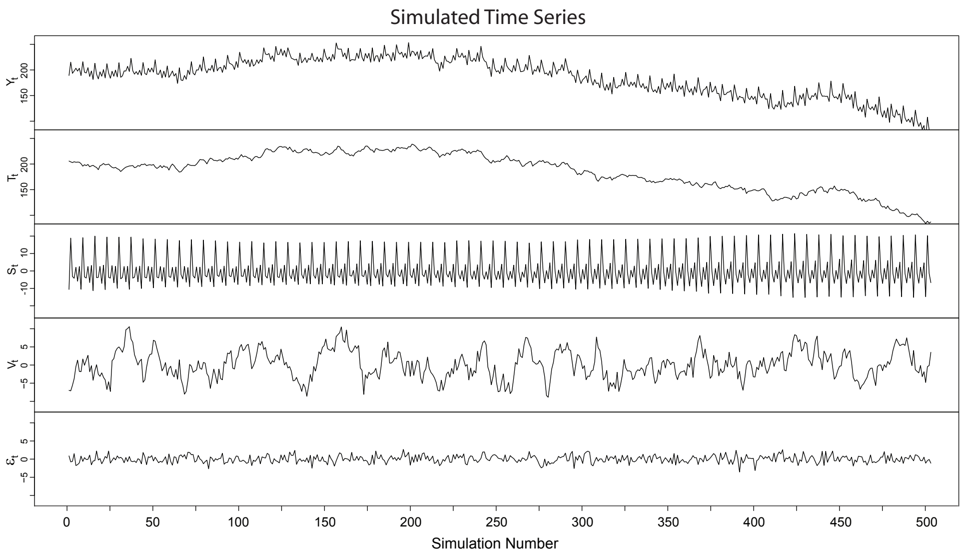

Figure 4.1: Simulated components.

Before the Kalman Filter and MLE can be used to estimate the parameters and separate the components, the model need to be rewritten into a state space model of the form described in equation (4.1) and (4.2). Since T_t, S_t and V_t are independent the matrices can be written as

$$A = \begin{pmatrix} A_1 & \mathbf{0} & 0 \\ \mathbf{0} & \boldsymbol{A}_2 & \mathbf{0} \\ 0 & \mathbf{0} & A_3 \end{pmatrix}$$

$$B = 0$$

$$\boldsymbol{\Sigma}_1^* = \begin{pmatrix} Sig_1 & \mathbf{0} & 0 \\ \mathbf{0} & \boldsymbol{Sig}_2 & \mathbf{0} \\ 0 & \mathbf{0} & Sig_3 \end{pmatrix}$$

$$C = \begin{pmatrix} C_1 & \boldsymbol{C_2} & C_3 \end{pmatrix}$$

$$\Sigma_2^* = \sigma_\epsilon$$

B is set to zero because there are no exogenous data inputs in the test simulation. The components of the matrices are

$$A_1 = 1, \quad \boldsymbol{A}_2 = \begin{pmatrix} -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad A_3 = \phi_1$$

$$Sig_1 = \sigma_T, \quad \boldsymbol{Sig}_2 = \begin{pmatrix} \sigma_S \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad Sig_3 = \sigma_V$$

$$C_1 = 1, \quad \boldsymbol{C}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}^T, \quad C_3 = 1$$

The corresponding state vector is

$$\boldsymbol{X}_t = \begin{pmatrix} T_t \\ S_t \\ S_{t-1} \\ \vdots \\ S_{t-5} \\ V_t \end{pmatrix}$$

These state space matrices and the state vector together with the equations (4.1) and (4.2) describe the exactly same system as (4.5). Now the Kalman Filter and the MLE will be used to estimate the parameters and to reconstruct T_t, S_t and V_t given only the State Space Model and the simulated Y_t measurements. The estimated parameters are shown in Table 4.2.

	σ_ϵ^2	σ_T^2	σ_S^2	σ_V^2	ϕ_1
Theoretical	1.00	10.00	0.10	5.00	0.80
Estimated	0.90	10.49	0.10	4.62	0.79

Table 4.2: Estimated parameters.

Using the estimated parameters the Kalman Filter separates the simulated data using the reconstructed state $X_{t+1|t+1}$. The reconstructed separations together with the corresponding 95% confidence interval are shown in Figure 4.2. It is

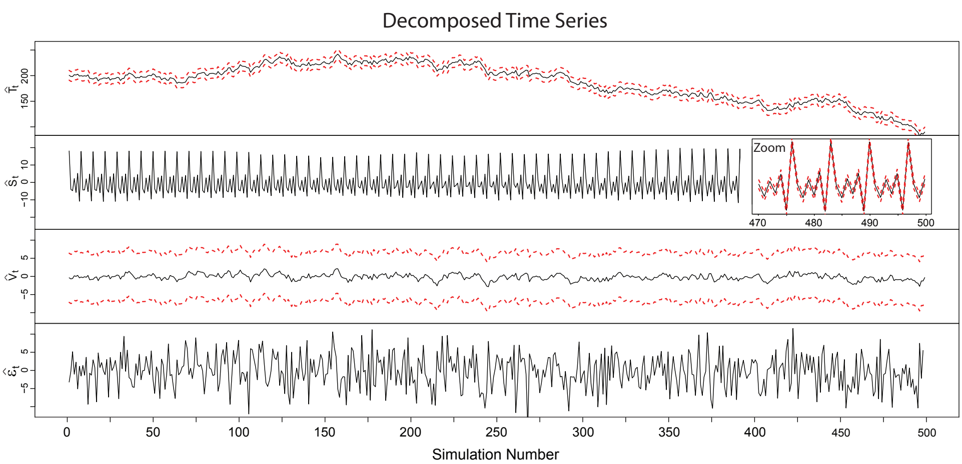

Figure 4.2: The extracted signals with 95% confidence interval.

seen that data separation is close to the simulated values. However, the confidence interval reveals that not all the components are equally well determined. $\hat{V}$ is not very well determined, but the true V_t lies mostly inside the confidence band.

Chapter 5

Kalman Filter for Signal Separation

In this section the Kalman Filter for Data Separation will be used on the heating consumption data, which was found with the Polynomial Robust Kernel in Chapter 3. A model will be developed and the results will be shown. The model describes how the heating consumption can be decomposed into independent components. Each component is the consumption explained by a certain factor. These factors are outdoor temperature, solar radiation and diurnal variation. The climatic measurements are included in the model as exogenous input.

The heating consumption data has been resampled from 10min values to hourly average values. 10 minutes variation is not useful when modeling building dynamics as short time variations is in praxis impossible to model. To avoid dealing with many missing data points, it is chosen to use measurements for a period of the early spring 2010, where only few data points are missing.

5.1 Model

According to [2] the change in indoor temperature can be modeled with this simple model

$$\frac{dT_{indoor}}{dt} = \frac{1}{RC}\left(T - T_{indoor}\right) + \frac{1}{C}\Phi_S + \frac{1}{C}\Phi_h$$

Where T_{indoor} and T is the indoor and outdoor temperature respectively. Φ_S is the amount of energy coming from the sun and Φ_h is the amount of energy coming from the heaters in the house. R is the thermal conductivity of the walls and C is the heat capacity of the house. By assuming that the indoor temperature is constant, the equation can be rearranged to

$$\Phi_h = -\frac{1}{R}T - \Phi_S + \frac{1}{R}T_{indoor} \tag{5.1}$$

It is assumed that there is a affine transformation between the energy from the sun Φ_S and the measurement of solar radiation S.

$$\Phi_S = c_1 S + c_2$$

Using this assumption together with the fact that R and T_{indoor} are unknown, equation (5.1) can be rewritten to

$$\Phi_h = UA \cdot T + gA \cdot S + Cons$$

where UA, gA and $Cons$ are unknown parameters. Thus the amount of energy from the heaters Φ_h can be separated to a part which depends on outdoor temperature, amount of solar radiation and a constant value. Where the $Cons$ value is the amount of energy used at $0°C$ and no solar radiation. This formula is only valid when the outdoor temperature and solar radiation are constant over time. However, the model have to handle changing climate variables, as changed in temperature and solar radiation will effect the heating consumption of the houses. Since the house is insulated and has a heating capacity, the outdoor temperature and solar radiation will not have immediately effect on the heating consumption. Some time will go before a change in temperature will give rise to a change in heating consumption. Therefore, it is not possible to just to replace T and S with the actually climate measurements T_t and S_t. To model the delayed effect from climate measurements a simple low pass filter is included in the model. The low pass filter is formulated as

$$\begin{aligned} T_{filter,t} &= \phi_T T_{filter,t-1} + (1-\phi_T)T_t \\ S_{filter,t} &= \phi_S S_{filter,t-1} + (1-\phi_S)S_t \end{aligned}$$

where T_t and S_t denotes the measured climate data at time t of outdoor temperature and solar radiation respectively. $T_{filter,t}$ and $S_{filter,t}$ is the filtered values at time t which will be used instead of T_t and S_t to describe the timedelay between change in climate to a change in heating consumption. ϕ_T and ϕ_S are parameters to be estimated between zero and one. The two estimated parameters will give useful information about the heating dynamics of the house, as they describe how fast the building will respond to changes in climate. For instance, if ϕ_T is close to zero the $T_{filter,t}$ will react fast to changes in outdoor temperature, and that could indicate that the house is not well insulated. On the other hand if ϕ_T is close to one the heating consumption will hardly change with the temperature. Replacing T and S with their filtered values in equation (5.1) gives

$$\Phi_{h,t} = UA \cdot T_{filter,t} + gA \cdot S_{filter,t} + Cons$$

Thus the measured heating consumption $\Phi_{h,t}$ at time t can be separated to a part which depends on outdoor temperature, amount of solar radiation and a constant value. Where UA and gA describes how the consumption changes with the filtered values of temperature and solar radiation.

Another factor that is expected to describe changes the heating consumption is the daily routines of the inhabitants. This daily variation will be modeled with a diurnal component given by

$$D_t = -\sum_{i=1}^{23} D_{t-i}$$

The daily variation gives information of the diurnal pattern in heating consumption. The sum of consumption explained by the daily variation component over any period of 24-hours will be zero. Thus, the daily variation gives no information of how much energy the inhabitants' daily routines contribute to the overall consumption. If, for instance, the inhabitants make a regular imbalance like opening a window at a certain time each day, the increased overall consumption will be captured in the $Cons$ part, while the variation of the consumption will be captured by daily variation. A problem arises when series of data points are missing. If the length of the series is not a dividable by 24 the diurnal component will lose track of the time, as a jump in one data point would not change the time of day by one hour. To eliminate the problem all series of missing data points are forced to have a length dividable by 24. This modification removes information from the data series, but makes the estimation of the diurnal component more robust.

To make the model able to handle noise in the states, state noise will be included as a part of the model. Noise will be included into the values of temperature and solar radiation and into the diurnal component. Including noise into these terms will make it possible for the states to change in other direction than described in the model. For instance, the state noise on the daily variation component will make the model able to adapt to changes in the inhabitants' daily routines. The influence from climate measurements on heating consumption will be able to change over time when noise is included to the model. The Kalman Filter can also handle measurement noise, so the noise from the measured heating consumption is also included as part of the model.

All the described components are included into the final state space model. Where the filtered temperature $T_{filter,t}$, filtered solar radiation $S_{filter,t}$, the constant $Cons$ and the daily variation elements D_t are considered as the unobservable state series. The heating consumption $\Phi_{h,t}$ are the measurements. The climate measurements are included as exogenous data inputs. The Kalman Filter will be used on the state space model to separate the heating consumption into these components. The model is rewritten in state space form, with the state equation

$$
\begin{pmatrix} T_{filter} \\ S_{filter} \\ Cons \\ D_0 \\ D_{-1} \\ \vdots \\ D_{-23} \end{pmatrix}_t =
\begin{pmatrix}
\phi_T & 0 & 0 & & & & & \\
0 & \phi_S & 0 & & & \mathbf{0} & & \\
0 & 0 & 0 & & & & & \\
 & & & -1 & -1 & \dots & -1 & -1 \\
 & & & 1 & 0 & \dots & 0 & 0 \\
 & \mathbf{0} & & 0 & 1 & \dots & 0 & 0 \\
 & & & \vdots & \vdots & \ddots & \vdots & \vdots \\
 & & & 0 & 0 & \dots & 1 & 0
\end{pmatrix}
\begin{pmatrix} T_{filter} \\ S_{filter} \\ Cons \\ D_0 \\ D_{-1} \\ \vdots \\ D_{-23} \end{pmatrix}_{t-1}
+ \begin{pmatrix} 1-\phi_T & 0 & 0 \\ 0 & 1-\phi_S & 0 \\ 0 & 0 & Cons \\ & \mathbf{0} & \end{pmatrix}
\begin{pmatrix} \boldsymbol{T} \\ \boldsymbol{S} \\ \mathbf{1} \end{pmatrix}
+ \begin{pmatrix}
\sigma_T & & & & & & \\
 & \sigma_S & & & \mathbf{0} & & \\
 & & 0 & & & & \\
 & & & \sigma_D & & & \\
 & & & & 0 & & \\
 & & \mathbf{0} & & & \ddots & \\
 & & & & & & 0
\end{pmatrix} \varepsilon_{1,t}
$$

and the observational equation

$$
Y = \begin{pmatrix} UA & gA & 1 & 1 & 0 & 0 & \dots & 0 \end{pmatrix} X_t + \sigma_\varepsilon \varepsilon_{2,t}
$$

The model parameters to be estimated are ϕ_T, ϕ_S, UA, gA and $Cons$, together with the standard deviations of the states σ_T, σ_S, σ_D and of the measurement noise σ_ϵ. The minimum solar elevation angle $\theta_{\min}$, which will be described in Chapter 5.1.1, also needs to be estimated.

In order to ensure that the estimated parameters make physical sense and improve the robustness of the estimation, some of the parameters are reparameterized. This makes it possible to keep the parameters within physical boundaries. To keep ϕ_T and ϕ_S between zero and one, a logistic transformation ($T(t) = \frac{1}{1+\exp(t)}$) has been applied.

It is assumed that the indoor temperature is constant, thus at any given time the amount of energy leaving the house should be equal to the amount of energy entering the house. This is an approximation as the indoor temperature may changes through time. However, since the measurements are resampled to hourly values, fluctuations in indoor temperature within the hour will not be recognized. Furthermore, regular diurnal changes in indoor temperature will be captured in the part explained by daily variation. Hence the approximation of constant indoor temperature is acceptable.

The initial conditions for the method are chosen as

$$\begin{pmatrix} T_{filter} \\ S_{filter} \\ Cons \\ D_0 \\ D_{-1} \\ \vdots \\ D_{-23} \end{pmatrix}_{0|0} = \begin{pmatrix} \frac{1}{4}\sum_{i=1}^{4} T_i \\ \frac{1}{4}\sum_{i=1}^{4} S_i \\ \text{mean}(\Phi_h) \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_{0|0}^{xx} = 10^8\, \boldsymbol{I}$$

The initial states for T_{filter} and S_{filter} is chosen to the mean value of the first four measurements for solar radiation and temperature respectively. The initial states for $Cons$ is the overall mean value for the heating consumption. These starting values should be close to the actual state and therefore give a good starting point for the method. Initial values for the covariance of predicted state $\boldsymbol{\Sigma}_{\mathbf{0|0}}^{\boldsymbol{xx}}$ is chosen such that the variance is high for all the states and there is no covariance between states. This is a reasonable starting value as no other information is available.

5.1.1 Data Transformation

Solar radiation influence the heating consumption of the house, as the house is heated when the sunbeams hit the roof, outer walls and get through the windows. It is assumed that the radiation through the windows has the dominating effect. As the windows usually are placed in vertical walls, it is only relevant to use the horizontal part of the radiation. It is the vertical radiation that is measured, as the measuring device lies in the horizontal plane. Hence data is transformed from vertical to horizontal part of the radiation, see Figure 5.1.

To transform the measured solar radiation S_{vertical} to the horizontal part $S_{\text{horizontal}}$ it is necessary to know the solar elevation angle θ_S. Since the location of the houses and the timestamps for each measurement is known it is possible to calculate θ_S for each data point.

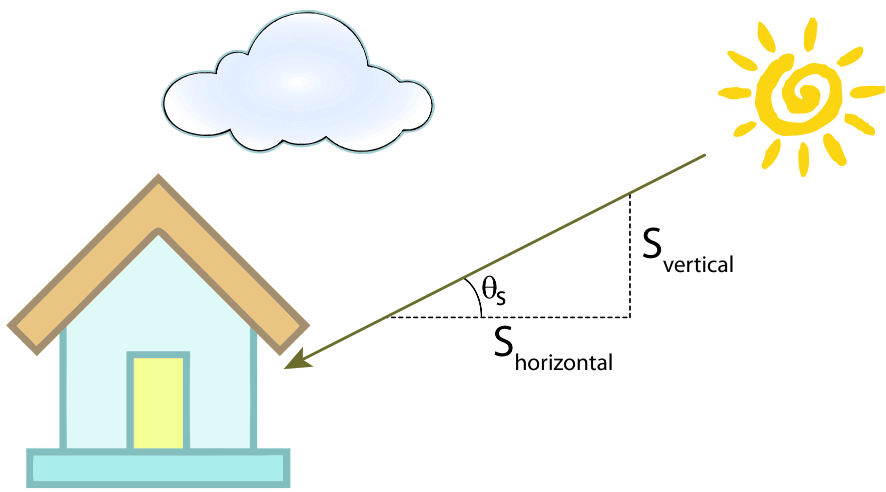

Figure 5.1: The solar radiation's vertical and horizontal part.

According to [3], the solar elevation angle can be calculated with a good approximation by the following formula

$$\theta_S = \arcsin\left(\cos h \cos\delta \cos\Phi + \sin\delta \sin\Phi\right)$$

where the hour angle h and the sun declination δ can be calculated by

$$h = 15^\circ \cdot (\text{time} - \text{SolarNoon})$$

$$\delta = -23.44^\circ \cdot \cos\left[\frac{360^\circ}{365}(N + 10)\right]$$

where SolarNoon is the time when the sun is highest in the sky. In Sønderborg SolarNoon is approximately 11:30 GMT. N is the number of days since January 1. Φ is the local latitude in Sønderborg which, according to Google Earth, is roughly $54°55'$. When θ_S is known the horizontal part of the solar radiation can be calculated as

$$S_{\text{horizontal}} = \frac{S_{\text{vertical}}}{\tan \theta_S}$$

The tangent to an angle close to zero is close to zero. Implying that at sunrise and sunset, where the elevation angle is close to zero, $S_{\text{horizontal}} \to \infty$ if $S_{\text{vertical}} > 0$. Due to diffuse reflection and light from the surrounding city S_{vertical} will be larger than zero at sun rise and sun set. Furthermore, in residential areas no solar radiation will reach the windows of the houses until the solar elevation angle is above a certain minimum level, due to shadows from the neighboring houses. Therefore, it is decided to disregard solar radiation when the corresponding solar elevation angle is below a certain threshold. This threshold value is denoted by θ_{min} and is estimated, in radians, for each of the individual houses.

Before using the horizontal solar radiation in the Kalman Filter, the data will be transformed with the square root function. Empirical investigations show that the data need a transformation, because a high solar radiation is so powerful in influence on the heat consumption, that the model ignored low solar radiation. The effect of using the square root transformation is that on high radiation levels a certain raise in radiation has less effect than when the level of radiation is low. This is illustrated in Figure 5.2. The drawback of using the square root transformation of the solar radiation is that it has no physical foundation.

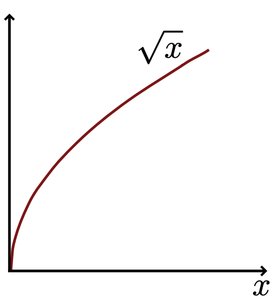

Figure 5.2: The square root function.

5.1.2 Result of Simple Model

The result of the Maximum Likelihood Estimation and the separation of the heating consumption for house no. 3 is shown in Table 5.1 and Figure 5.3. The separated heating consumption is calculated by the reconstructed states of the state space model. The result consists of six plots, the uppermost plot shows the heating consumption found in Chapter 3. Number two to five is the part of the heat consumption explained by their individual factors: temperature, solar radiation, constant value and daily variation. The plot in the bottom is the residuals. Hence, the sum of the five lower plots is equivalent to the uppermost plot. As seen from the result the parts can both be positive and negative. For instance, the part explained by solar radiation is negative because the heating consumption decreases with increasing solar radiation.

It is seen that the effect of solar radiation only have little changes between the days, which seems somewhat unrealistic. The fluctuations seem more dependent of the day-night rhythm that of the actual solar radiation. The daily variation, on the other hand, is more noisy than expected. In order to handle these issues and to improve the model another approach of transforming the solar radiation is described in the following.

Param	Estimate
ϕ_T	0.954
ϕ_S	0.786
UA	−0.639
gA	−0.0314
$Cons$	14.198
σ_T	12.600
σ_S	1.000
σ_D	8.021
σ_ε	19.409
$\theta_{\min}$	0.0428

Table 5.1: Estimated Parameters for house no. 3.

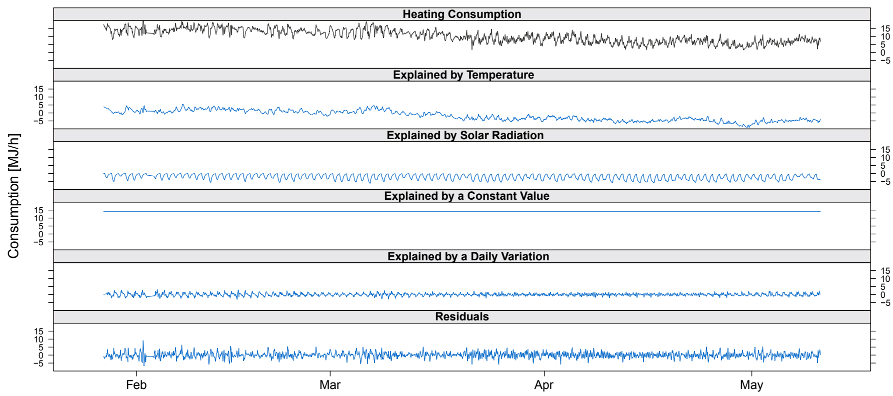

Figure 5.3: Separated heating consumption with simple model for house no. 3.

5.2 Improved Model

The way the solar radiation was handled in the model is not quite satisfying, as the solar radiation was being transformed with a function without any physical foundation. Another approach to handle the solar radiation is to use the physical interpretation of radiation in the atmosphere. This new approach uses the fact that the total solar radiation (global radiation) consists of direct beam and diffuse radiation. The advanced method splits the global radiation into diffuse and direct radiation and transforms them individually to horizontal radiation. The method also includes the reflection from the ground to the windows. According to [13] measurements of solar radiation can be spitted into direct beam and diffuse solar radiation with the following approximate method

$$G_0 = 1367 \cdot \left(1 + 0.033 \cdot \cos\left(360^\circ \frac{N}{365}\right)\right) \sin\theta_S$$

$$k_t = \frac{\text{Global}}{G_0}$$

$$R_{\text{Diffuse}} = R_{\text{Global}} \cdot (0.952 - 1.041 e^{-\exp(2.300 - 4.702 k_t)})$$

$$R_{\text{Direct}} = R_{\text{Global}} - R_{\text{Diffuse}}$$

where G_0 is the extraterrestrial radiation, which is the radiation outside the earth's atmosphere. k_t is the clearness index, which is the fraction of extraterrestrial radiation that hits the ground. θ_S can be calculated as in Chapter 5.1.1. N is the number of days since January 1. R_{Global} is the measured solar radiation. R_{Diffuse} and R_{Direct} is the diffuse and direct part of the global solar radiation respectively. In order for this method to work the unit of the solar radiation must be $watt/m^2$, but the measured solar radiation is in lux. Since lux is a measurement of visible light, it is not possible to directly convert lux to $watt/m^2$ without knowing the distribution of wavelengths of the measured light. To get an approximate conversion from lux to $watt/m^2$ the measured illuminance is divided by 68.3.

The calculated direct and diffuse part of the radiation is measured as the vertical part. Since it is suggested that dominating effect on the heating consumption is the radiation coming through the windows of the house, the radiation needs to be transformed. According to [12] the total radiation through the windows G_t is the sum of direct radiation G_{Direct} transformed to the horizontal part, a transformed diffuse radiation G_{Diffuse} and the ground reflection $G_{\text{Reflection}}$ on the windows

$$G_t = G_{\text{Direct}} + G_{\text{Diffuse}} + G_{\text{Reflection}}$$

where G_{Direct} is calculated by transformning the direct radiation to the horizontal part

$$G_{\text{Direct}} = \frac{\text{Direct}}{\tan \theta_S}$$

The diffuse radiation is further split into circumsolar and isotropy radiation, by the anisotropy index $A_i = \frac{\text{Direct}}{G_0}$. The horizontal part of circumsolar radiation can be calculated just as for the direct radiation, whereas the isotropy radiation is transformed by multiplying with a factor 0.5

$$G_{\text{Diffuse}} = A_i \cdot \frac{\text{Diffuse}}{\tan \theta_S} + (1 - A_i)\,\text{Diffuse} \cdot 0.5$$

The horizontal part of ground reflection is given by

$$G_{\text{Reflection}} = \rho_g \cdot 0.5 \cdot \text{Global}$$

where the ground reflection coefficient is chosen as $\rho_g = 0.2$.

The total radiation on the vertical windows G_t will be used as input for the Kalman Filter instead of the measured solar radiation. The estimated parameters of the improved model for house no. 3 are shown in Table 5.2 and the result is plotted in Figure 5.4. The advantage of the improved model is that the effect of solar radiation on heat consumption changes from day to day and seems to follow the actual solar radiation. The daily variation is more regular, which indicates that the improved model captures the pattern of daily variation. From the estimated parameters, it is also seen that the variances decrease for the improved model. Based on these results it is possible to conclude that the improved model has a better separation of the heating consumption explained by solar radiation and daily variation. In the following, the improved model will be validated and discussed.

Param	Estimate
ϕ_T	0.955
ϕ_S	0.602
UA	-0.981
gA	-0.00329
$Cons$	10.408
σ_T	9.034
σ_S	0.700
σ_D	0.466
σ_ε	0.795
$\theta_{\min}$	0.0805

Table 5.2: Estimated parameters.

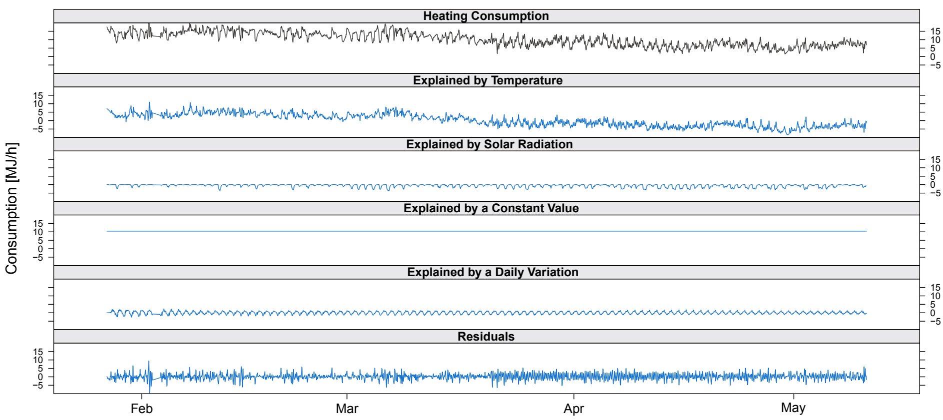

Figure 5.4: Separated heating consumption with improved model for house no.3

5.3 Model Validation

The results of separating the heating consumption for house no. 3 are validated by investigating the residuals. If the residuals are random white noise it is not possible to improve the model any further. If, on the other hand, there are any trends in the residuals, the model does not capture all of the information in the data. The residuals for house no. 3 are shown in Figure 5.5. Different statistical methods will in the following be used to investigate if there are any trends in the residuals.

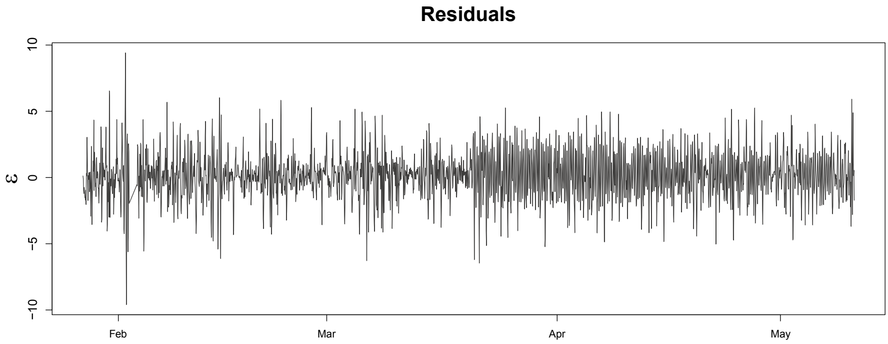

Figure 5.5: Residuals for house no. 3.

From theory it is know that if a process is white noise it has a flat power spectral density. The power spectral density of a data series can be estimated by the periodogram. Thus a cumulated periodogram of a white noise sequence is a straight line. It is possible to plot a cumulated periodogram together with a confidence band based on the Kolmogorov-Smirnov test. The confidence limits are constructed so that the whole cumulated periodogram lies within the limits with probability 0.95. The cumulated periodogram with 95% confidence interval of the residuals for house no. 3 is shown in Figure 5.6. As the cumulated periodogram is not entirely inside the confidence interval, it can be concluded that the residuals is not a white noise sequence.

The residuals are not entirely white noise therefore further investigation is needed. An autocorrelation function is plotted in Figure 5.7 and a periodogram is plotted in Figure 5.8. These plots may support an interpretation of the trend in the residuals.

The autocorrelation plot indicates an oscillation with a period of about 5-6 hours. The periodogram is a useful method for identifying the dominant frequencies in a time series. In the periodogram there are dominant changes around

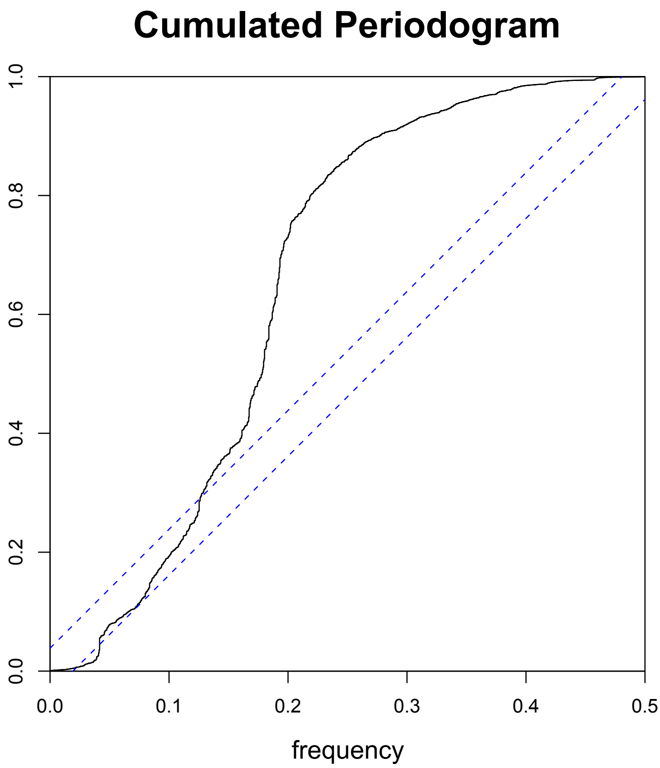

Figure 5.6: Cumulated periodogram with 95% confidence interval for house no.3

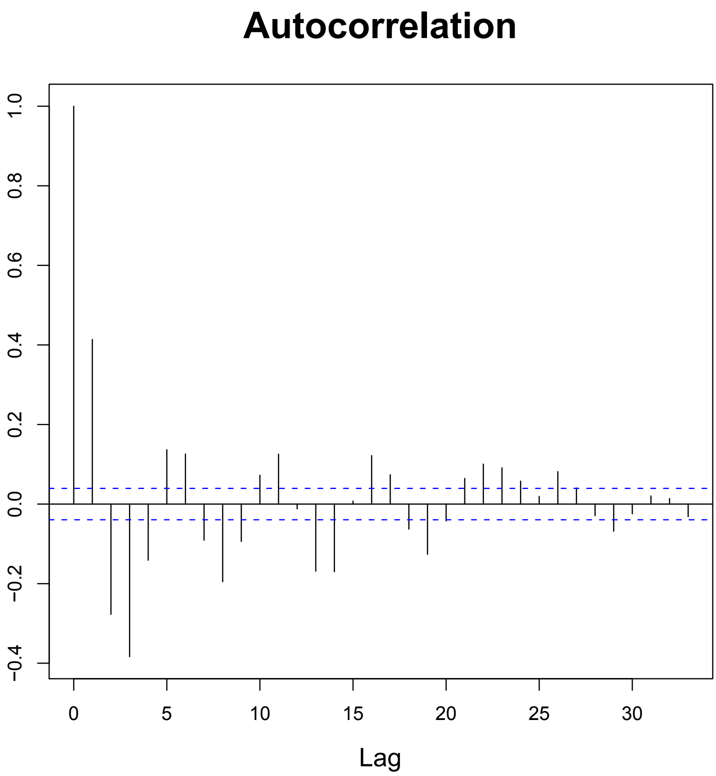

Figure 5.7: Autocorrelation function for house no. 3.

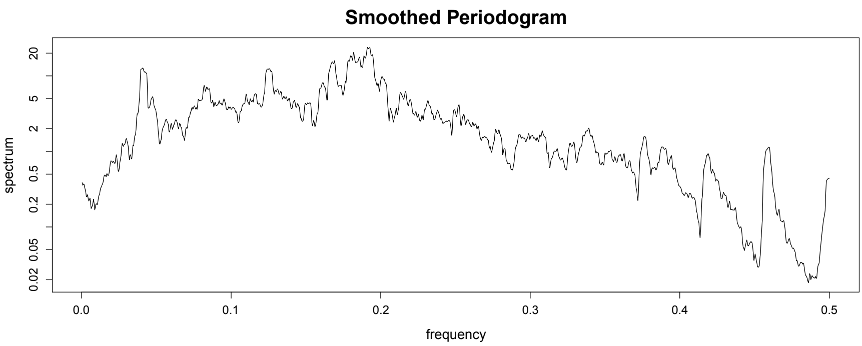

Figure 5.8: Periodogram of residuals for house no. 3.

frequency 0.18 corresponding to a period of 1/0.18 = 5.6 hours, which correspond to the result of the autocorrelation. A plausible explanation of the periodicity of 5.6hours could be the heating control of the thermostats. Heating controls work by turning on and off the radiators when the indoor temperature reaches a certain level. This control will give rise to a periodic heating consumption, where a period of 5.6hours is not unrealistic. There is also dominant frequency around 0.041, which corresponds to a period of around 24 hours. The trend seen with a period of 24 hours could be some daily variation not captured in the part explained by daily variation.

5.3.1 Solar

The residual for house no. 3 in Figure 5.4 shows a change of increasing intensity rather suddenly in the medio of March. A close look at the whole period shows that periods of small residuals alternate with periods of large residuals. When the residual is compared to the intensity of solar radiation, a connection is revealed. Periods with high solar intensity are simultaneous with periods of large residuals. This indicates that the solar radiation has a more pronounced influence on heating consumption than the model has revealed. Based on these observations an investigation of the solar radiation in a 24 hours period is analyzed.

The way the solar radiation is handled in the Kalman Filter may include some issues. Though the elevation of the sun is included in the model, the horizontal movement during the day is not. The sun's horizontal movement could influence the sun's effect on heat consumption. For instance, houses could have larger windows pointing in one direction than in others or there could be shadows from larger trees et cetera. An investigation is made to reveal if there is a correlation between the residual and the time of the day. Figure 5.9 shows a plot comparing the residual and the solar radiation for some sunny days during the year. It can be concluded that the residuals do not change significantly during the sunny days, thus there is no indication that the horizontal movement of the sun has a significant effect of heat consumption.

It is noticed that the measured solar radiation is not symmetrical around noon. It seems that in the morning the radiation is lower than in the afternoon. As the figure only shows days of full sunshine, the asymmetrical radiation could be due to fog in the morning, which is common in clear days. Another explanation could be that the measuring device has not been exactly in the horizontal plane.

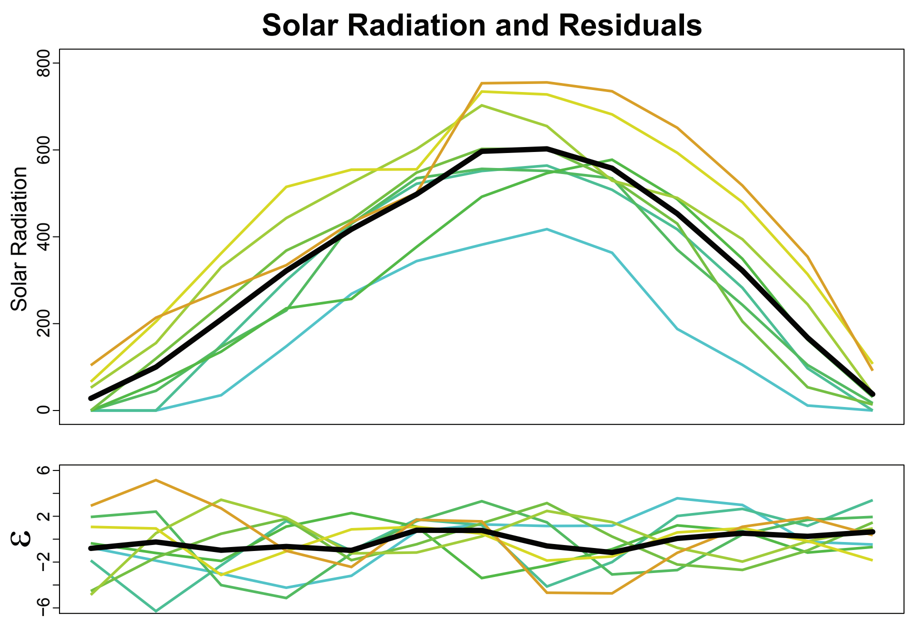

Figure 5.9: Uppermost: Measured solar radiation during the day for eight distinct days with full sun. Bottom: Residuals of the Kalman Filter for the same days. Black line: Mean value.

5.4 Results

The Kalman Filter was used to separate the heating consumption in all of the four houses. The results showed that the method has equally good separation of heating signals for all the houses and thus gives similar results. The result for house no. 3 is shown in details to represent the performance of the results. The results of the other houses are shown in the next section. The result-figures consists of six plots. The uppermost plot shows the heating consumption found in Chapter 3. Number two to five is the part of the heat consumption explained by their individual factor: temperature, solar radiation, constant value and daily variation. The plot in the bottom is the residuals. Hence, the sum of the five lower plots is equivalent to the uppermost plot. The result for house no. 3 is shown in Figure 5.4 and will be the basis for the presentation of the results.

5.4.1 Climate Variables

The relation between the heat consumption and the climate variables is shown in Figure 5.10 and 5.11. The red lines are rescaled outdoor temperature and a rescaled version of the horizontal part of solar radiation. The plots illustrate that the heat consumption explained by temperature decreases when the temperature increases and vice versa. The same effect is seen for the level of solar radiation. This fact is confirmed in Table 5.2, where gA and UA are negative.

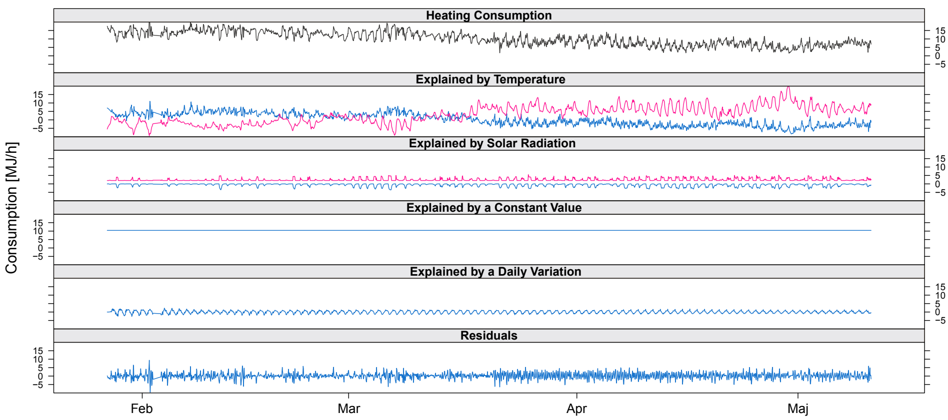

Figure 5.10: Result for house no. 3. Blue: Consumption, Red: Rescaled climate variables.

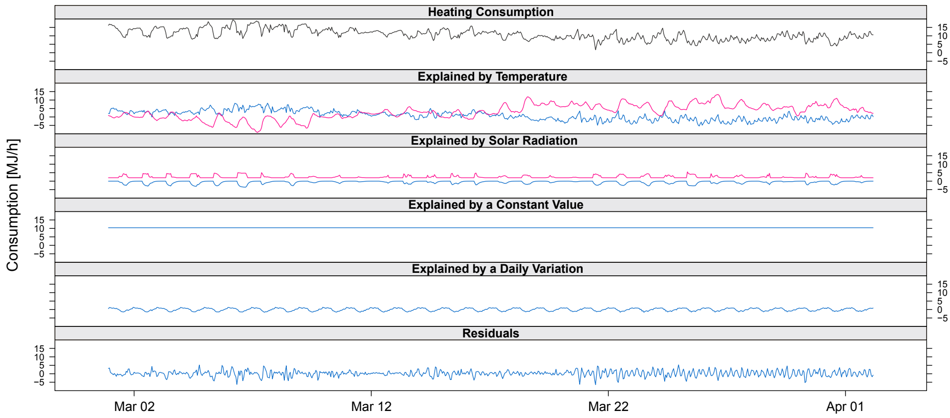

Figure 5.11: Result for house no. 3 during March. Blue: Consumption, Red: Rescaled climate variables.

From Figure 5.11 it is seen that changes in the climate dependent heating consumption respond much faster to changes in solar radiation than to outdoor temperature. A careful inspection of the plots indicates that a raise in heating consumption due to a drop in temperature is delayed, which means that the increased heating consumption happens at little after the actually temperature drop. The heating consumption is delayed because the indoor temperature does not drop in the same instance that the outdoor temperature drops. The delay in the response to outdoor temperature changes is due to the isulation and the heat capacity for the house. Contrary, the solar radiation has an instant effect on the indoor temperature and thus on the heating consumption. The estimate of ϕ_S is much smaller than ϕ_T, which confirms that the response to solar radiation is faster than the response to outdoor temperature. To clarify the difference between slow and fast changes in heating consumption, a detailed plot of the result is shown in Figure 5.12. In Box 1, the heating consumption shows fast changes, which is recognized as effect of the solar radiation part. In Box 2, the heating consumption shows slow changes, which is recognized as the part explained by temperature.

Table 5.2 shows that the minimum solar $\theta_{\min}$ is estimated to 0.0805rad which corresponds to 4.6°. This estimate seems reasonable as the neighborhood consist of single floor houses, which only cast shadows at each other when the sun's elevation angle is low.

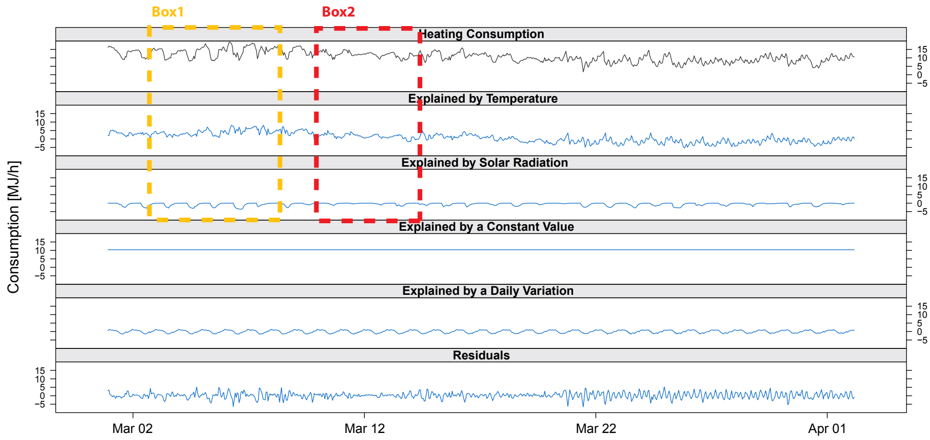

Figure 5.12: Result for house no. 3 during March.

5.4.2 Constant Value

The constant value is a measure of how much energy the house will use at $0°C$ and with no solar radiation. The estimation is $10.408MJ/h$ as shown in Table 5.2.

5.4.3 Daily Variation

Daily variation explains the part of the consumption which has a regular diurnal pattern and is not captured by the temperature or solar radiation. The regular pattern seen in the result could be caused by time dependent automatic temperature regulating devices. It could also be a response to the inhabitants' behavior, as people have tendency to be regular in their daily routines.

The plots show that the daily variation has a cycle of 24 hours, which confirm that this part is diurnal as it was intended. At the beginning of the estimation period, there are some irregularities in the daily variation. These irregularities are even more evident for house no. 2, see Figure 5.16. This phenomenon is a consequence of the Kalman filter method. The Kalman filter estimates recursively over time using incoming measurements, therefore a number of cycles must be completed before the result of the daily variation is stable.

The result for house no. 1 in Figure 5.15 shows a change in the daily variation starting in the middle of March. This change could be a consequence of the inhabitants' holyday, which started 12th March. The result shows that the method is adaptive to changes and further that human behavior influence the pattern of daily variation.

The different houses have very different daily variation patterns, which support the suggestion that human behavior influence the daily variation. House no. 4 has the largest fluctuations in daily variation, see Figure 5.4. This could be connected to the fact that house no. 4 has five inhabitants compared to two inhabitants in the other houses.

Figure 5.13 shows a detailed plot of the daily variation part for house no. 3 on a representative day. This single plot can be used for inspection as the nearby days have similar patterns in the daily variation. It is seen that heating consumption is lower during the day than during the night. This could be caused by an automatically device that lower the indoor temperature during the hours where the family is away from the house. An interesting observation is that the heating consumption decrease every day around 19:00GMT. An explanation of this decrease in heating consumption could be an increase in the indoor temperature due to heat generating devices. When the family returns to their home, they start several electrical devices such as electric light, television and cooking devices, which all generate heat.

Figure 5.14 shows the daily variation for house no. 4 on a representative day. During the night, the consumption is low. Also during the daytime the consumption is lowered. While, during morning and evening, the consumption is increased. This could indicate that the family uses an automatic device that lower the temperature in the night and when they are away from the house during the day.

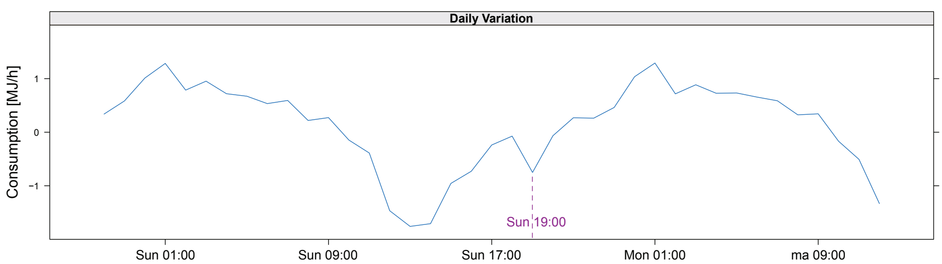

Figure 5.13: Part explained by daily variation for house no. 3 during 21-22 of February.

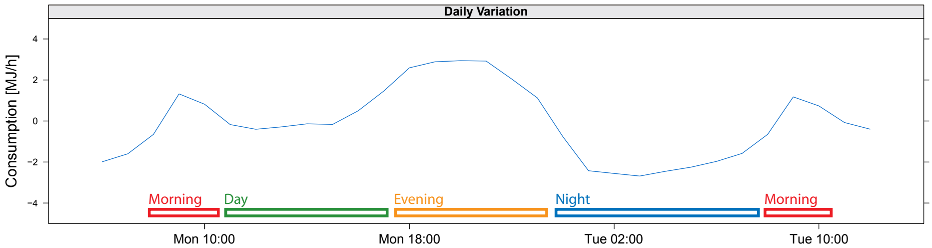

Figure 5.14: Part explained by daily variation for house no. 4 during 29-30 of March.

5.4.4 Residuals

As stated in Chapter 5.3, the residuals are not entirely white noise as some unexplained trend is present. The residual for house no. 3 shows a change of increasing intensity rather suddenly in the medio of March. A close look at the whole period shows that periods of small residuals alternate with periods of large residuals. When the residual is compared to the intensity of solar radiation, a connection is revealed. Periods with high solar intensity are simultaneous with periods of large residuals. This indicates that the solar radiation has a more pronounced influence on heating consumption than the model has revealed. The solar radiation has undoubtedly a very complex influence on residential houses. Only the solar radiation through the windows is considered, although radiation hitting the roof and walls also has an influence on the heating consumption.

5.4.5 The Other Houses

The parameter estimate of all the houses are listed in Table 5.3 and the separation results for the remaining three houses in a spring period of 2010 are shown in Figure 5.15, 5.16 and 5.17. The results of the remaining houses behave much like house no. 3, and will not be analyzed further.

	House 1	House 2	House 3	House 4
ϕ_T	0.944	0.923	0.955	0.873
ϕ_S	0.704	0.548	0.602	0.658
UA	−0.180	−0.638	−0.981	−0.501
gA	−0.00339	−0.00163	−0.00329	−0.00334
$Cons$	6.933	10.959	10.408	8.811
σ_T	10.033	6.504	9.034	2.820
σ_S	1.500	0.600	0.700	0.500
σ_D	0.323	0.101	0.466	0.124
σ_ϵ	1.926	3.669	0.795	2.291
θ_{min}	0.106	0.0724	0.0805	0.129

Table 5.3: Estimated parameters.

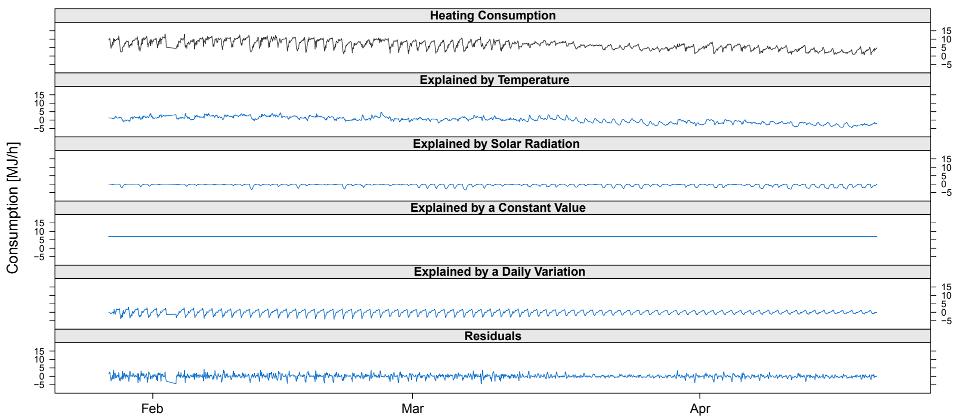

Figure 5.15: Result for house no. 1.

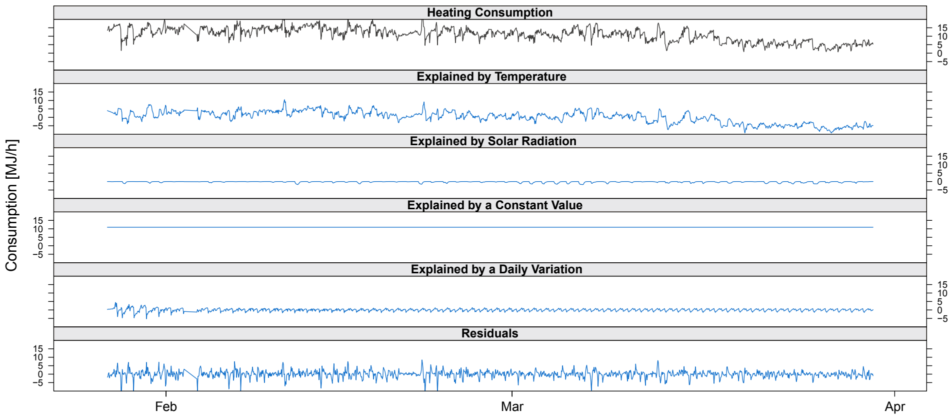

Figure 5.16: Result for house no. 2.

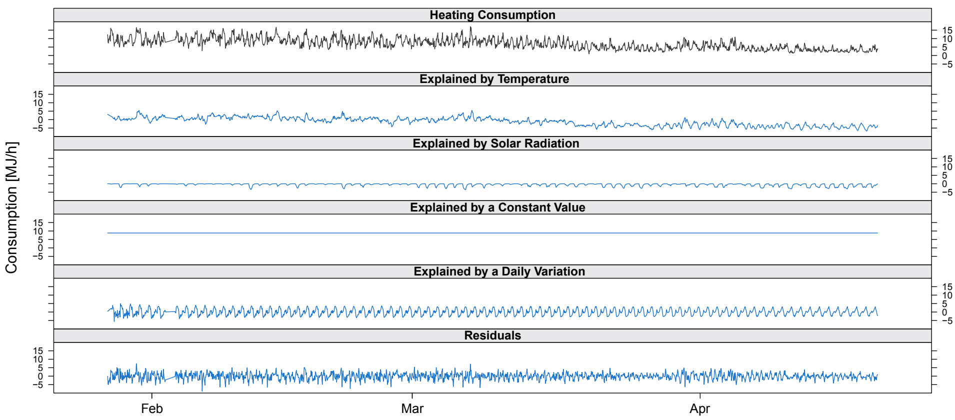

Figure 5.17: Result for house no. 4.

Chapter 6

Discussion

The statistical analyzes presented show that data from a single measurement device is sufficient for obtaining valuable information on heat consumption in residential houses. By the Robust & Polynomial Kernel Smoother, it is possible to split the hot water consumption and the heating consumption alone based on total consumption data. The quality of the splitting is convincing, as hot water consumption has a fluctuation very different from the heating pattern. The result is further confirmed by the fact, that no hot water consumption is seen during the inhabitants' holydays and the finding that the hot water use is much higher in the house with five inhabitants than in the houses with two inhabitants. The heating consumption was then analyzed using the Kalman Filter for Signal Separation and meteorological measurements of outdoor temperature and luminance. The result is a four-part split revealing the influence of outdoor temperature, solar radiation, constant value and the diurnal variation on the heating consumption. In addition, information on the heat dynamics of the house is revealed through the estimated parameters.

Conclusive interpretations of the result for the four houses in this investigation are not possible as too few houses are investigated. However, the useful information that can be extracted from the estimated parameters is worth investigating. The $Cons$ is the heat consumption at $0°C$ and no sun radiation. This parameter is independent of local weather and other external factors and is therefore useful to compare houses located in different areas. Though behavior of the inhabitants like the preferred indoor temperature can have effect on $Cons$. The parameter ϕ_T indicates how fast the heat consumption changes with the outdoor temperature, a low ϕ_T indicats a fast change. A high heat capacity or a high insulation level can influence ϕ_T positively. The parameter ϕ_S indicates how fast the heat consumption changes with solar radiation. The UA parameter is the responses of heat consumption to a change in outdoor temperature. When UA is very negative the response in heat consumption is high. The UA value is influenced by the insulation of the house. The gA is the responses of heat consumption to a change in solar radiation.

6.1 Limitations

There are elements of uncertainty, which are not straightforward to include in the model. One of them is irregular human behavior. Inhabitants use hot water or their use of electric devices will release energy to the house and may influence the heating consumption. However, many of these uncertainties are a part of the inhabitant's daily routines and will be captured in the daily variation. The conversion of solar radiation data from luminance in unit of lux to energy of unit $watt/m^2$, is another unavoidable uncertainty. The effect of solar radiation on heating consumption is complex and not all aspects are captured in the model. Wind speed is a factor, which could influence the heating consumption but not included in the model.

6.2 Further Work

Wind speed could influence the heating consumption, but the level of influence depends on the outdoor temperature, meaning that the model will be nonlinear. It is not possible to handle nonlinearities in the ordinary Kalman Filter. However, it is possible to use the Extended Kalman Filter for nonlinear models, even though it can be cumbersome to work with. Another difficulty of including the wind speed into the model is that the wind's influence on heating consumption is expected to be dependent of the wind direction. Thus, it should be possible to include wind speed in the model, but it may not be a simple task.

The solar radiation influence on heating consumption is a very complex mechanism. In the model presented, the solar radiation influence on the houses is simplified. Only the solar radiation through the windows is considered to influence the heating consumption. A more enhanced model would include the effect of radiation hitting the walls and roof of the house, although this effect will have much slower influence on the consumption. Furthermore, it is expected that a fraction of the sunbeams hitting the windows will be reflected and hence will not contribute to the heating consumption. This fraction will be dependent of the angle at which they hit the windows. If the sunbeams are perpendicular to the windows, only few beams will be reflected, whereas if the sunbeams are almost parallel to the windows the fraction of the reflected beams will be much higher. As with the wind speed and direction, the influence of solar radiation could be dependent of the position of the sun. Including these factors in the model may improve the overall result, but the transformation of solar data from illuminance to an energy unit will still be subject to a considerable approximation.

A possibility for improving the utility of the model is to make it able to predict the future. The Kalman Filter is intended for predictions, but the Kernel method for splitting the hot water and heating consumption is not currently able to make the splitting alone based on preceding measurements. It is possible to develop a method to split the hot water and heating consumption based on preceding measurements, but it will be at the expense of the quality of the splitting. In Chapter 5.3, it was revealed that the residuals were not entirely white noise. It seems that the dynamics of the thermostats was creating a systematic behavior of the residuals. This regulating mechanism by the thermostats could be modeled by an autoregressive (AR) model, as the AR model uses previous patterns to predict the future.

Another useful enhancement of the model would be to make the parameter estimate adaptive. In that case change in house environment or the behavior of the inhabitants would change the parameters accordingly.

6.3 Potential

Today, energy labeling of houses is based on experts' calculation from the insulation materials and subjective estimates. The developed model could make a cost efficient and objective evaluation of energy consumption, as the method is using the actual consumption for the individual house.

It is also poissible to expand the model so the inhabitants could track their real-time consumption online. This would make surveillance of the house efficient, and in cases where the consumption increases it would be easy and fast realize the problem. Further, it would be easy to locate the problem, whether it is hot water waste, diurnal behavior or some changes in the heat dynamics of the house. If the heating consumption increases fast when outdoor temperature decreases, the problem would be recognized by the model, and based on this the inhabitants can be warned.

Using weather forecasts together with the model makes it possible to predict the heating consumption of the houses. Short time scale predictions are valuable information for the district heating company, as they can schedule how much energy they will have to supply in the next couple of hours. Such short time predictions could make it possible to adjust the supply of energy to the system, and benefit from variations in the energy price. By this, the district heating company can act as a buffer in a Smart Grid network. The heat capacity of the residential houses can be a part of this buffer system, by smoothing out the variation in supply.

Chapter 7

Conclusion

The heating consumption measurements of four individual residential buildings including central heating and hot water use are analyzed in this study. Splitting the hot water consumption from the heating consumption was investigated by Low Pass filter. The result of this method was not convincing. Afterwards the Kernel Smoother method was tested, which gave promising results. Rewriting the Kernel Smoother to its least square parallel made it possible to expand the method to Robust and Polynomial Kernel. The combined Robust & Polynomials Kernel showed good results. Therefore, it is concluded that the method is suitable for splitting hot water use from heating consumption.

After splitting the hot water consumption and the central heating consumption, the Kalman Filter for Signal Separation was used together with meteorological measurements to analyze the heating consumption. The result is a four-part split revealing the influence of outdoor temperature, solar radiation, constant value and the diurnal variation on the central heating consumption. Further, the heat dynamics of the houses was revealed by estimating parameters of the model. Though the residual was not entirely white noise, indicating that not all trends have been captured in the model, the results showed that Kalman Filter for Signal Separation is suitable for analyze of the heating consumption in residential houses.

Bibliography

[1] *Analysis of energy consumption in single family houses.* 2010.

[2] Peder Bacher. Matematisk modellering infor klima og energi. 2011.

[3] Tian Pau Chang. The suns apparent position and the optimal tilt angle of a solar collector in the northern hemisphere. *Solar Energy 83*, 2009.

[4] Mohinder S. Grewal and Angus P. Andrews. *Kalman Filtering - Theory and Practice Using MATLAB.* Wiley, 2008.

[5] Klimakommissionen. Green energy - the road to a danish energy system without fossil fuels. Technical report, Danish commission on climate change policy, 2010.

[6] Tine Steen Larsen, Henrik Nellemose Knudsen, Anne Marie Kanstrup, Ellen Tove Christiansen, Kirsten Gram-Hanssen, Mette Mosgaard, Henrik Brohus, Per Heiselberg, and Jørgen Rose. Occupants influence on the energy consumption of danish domestic buildings. 2010.

[7] H. Madsen and J. Holst. Estimation of continuous-time models for the heat dynamics of a building. 1994.

[8] Henrik Madsen. *Time Series Analysis.* Chapman & Hall, 2008.

[9] Henrik Madsen and Jan Holst. *Modelling Non-Linear and Non-Stationary Time Series.* 2000.

[10] Kaj Madsen and Hans Bruun Nielsen. *Introduction to Optimization and Data Fitting.* 2010.

[11] Henrik Aalborg Nielsen, Stig Bousgaard, Peder Bacher, and Henrik Madsen. Analysis of energy consumption in single family houses. 2010.

[12] Bengt Perers. Description of program for calculation of annual solar collector energy output. 2010.

[13] J.A. Ruiz-Arias, H. Alsamamraa, J. Tovar-Pescadora, and D. Pozo-Vázqueza. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Conversion and Management*, 2010.