DTU Compute
Department of Applied Mathematics and Computer Science

Statistical models for analysis of frequent readings of electricity, water and heat consumption from smart meters

In cooperation with SEAS-NVE

Anton Stockmarr (s164170)
Ida Riis Jensen (s161777)
Mikkel Laursen (s164199)

Kongens Lyngby 2019

Abstract

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Preface

This xxx thesis was prepared at the department of Applied Mathematics and Computer Science at the Technical University of Denmark in fulfillment of the requirements for acquiring a yyy degree in zzz.

Kongens Lyngby, May 24, 2019

Anton Stockmarr (s164170)
Ida Riis Jensen (s161777)
Mikkel Laursen (s164199)

Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Contents

CHAPTER 1

Introduction

1.1   Motivation

CHAPTER 2

Data

The data is provided by SEAS-NVE in three data sets. The house data consists of 71 `.csv`-files containing 8 attributes for each house which is antal data points in all. The second data set includes weather data containing 11,845 observations with 11 attributes. Furhtermore, the third data set is from Bygnings- og Boligregistret (BBR) and contain details for each of the houses e.g. total area, year of construction and type of house. Mangler muligvis lidt mere her.

The main focus of this section will be how data is prepared for the further analysis.

2.1 Original data

The original house and weather data include hourly observations from the period 31-12-2017 to 29-01-2019. The time period varies in the house data which will be taken into account when cleaning the data.

Table 2.1 below shows the attributes from the house data set.

Variable	Description
StartDateTime	Start time and date for measurements. Hourly values.
EndDateTime	End time and date for measurements.
Energy	Electricity consumption in *kWh*.
Flow	Amount of water passed through meter in $m^3/hour$.
Volume	in m^3.
TemperatureIn	Temp. of the water flowing into a house in Degrees/C.
TemperatureOut	Temp. of the water flowing out of a house in Degrees/C.
CoolingDegree	Difference between Temp.In and Temp.Out in Degrees/C.

Table 2.1: Attributes from the original house data..

The weather data set consists of the attributes seen in Table 2.2.

Variable	Description
StartDateTime	Start time and date for measurements. Hourly values.
Temperature	Temperature outside in Degrees/C.
WindSpeed	
WindDirection	
SunHour	
Condition	
UltravioletIndex	
MeanSeaLevelPressure	
DewPoint	
Humidity	
PrecipitationProbability	
IsHistoricalEstimated	

Table 2.2: Attributes from the original weather data..

StartDateTime and EndDateTime are always one hour apart. When there are missing observations the following the next StartDateTime is simply delayed. Energy is the measured energy consumption on the meter in the houses.

Noget med at vi også har BBR data.

2.2   Cleaning and preparation

In this section, it is described how the raw data is cleaned and prepared for the statistical analysis. Synes der mangler et eller andet her.

Both weather data and the house data are aggregated in order to convert hourly values into daily values since there are of interest when modelling i chapter 3. Loader en temporary data ind, som vi modificerer indtil vi putter den ind i vores endelige data. Data from 2017 in the house data are removed since data for the same period is missing in the weather data. The format for the attributes `StartDateTime` and `EndDateTime` is changed to d-m-Y H:min:sec. Likewise, the attribute `StartDateTime` in the weather data is converted to the same format as in the house data in order to merge the two data sets.

For nogle huse er der nogle hourly measurements der ikke er der. Der er huller i målingerne. Disse udfyldes med null, hvilket er bedre/lettere at arbejde med.

Attributen IsHistoricalEstimated ændres til logical, så vi kan compute med den.

Vi laver så temp. weather data så vi kan merge det med house data. Vi merger ikke al data, da mængden vil være en del større. Vi merger tmp weather data på house data i model processen.

In the house data there are some measurements missing and it can therefore be difficult to do modelling for the houses in question. To avoid these difficulties, a so

called "Data Checking" function has been made in order to check whether several constraints for the data are fulfilled. There most be a certain number of observations and the amount of missing data should not exceed a certain fraction of the data. Vi tilføjer en binær attribute for hver ferie, og endnu en for weekender. De forskellige ferier vi tager med er christmas break, winter break, spring break, autumn break

2.2.1   The sun and the wind

A physical factor that could possibly affect the heat consumption is the sun. In raw data, the attributes `Condition`, `SunHour`, and `UltraVioletIndex` can be seen as explanatory variables for the sun. Instead, an attribute, `Radiation`, is added to calculate the solar radiation for a given day. This attribute is determined with use of the `R` function `calcSol` from the library `solaR`. The ultraviolet index is a measurement of the strength of ultraviolet radiation and since the attribute `Radiation` is more exact, `UltravioletIndex` is removed from the weather data set.

Another physical factor that might be of importance is the wind. There are data available for both the wind direction in degrees and the wind speed. When the data is aggregated into daily values, it is important to pay special attention to the wind attributes, since it is not logical to take the average of degree values. For example, the average wind direction of 359 degrees and 0 degrees is not 179.5 degrees. Instead the wind direction and wind speed are interpreted as polar coordinates in a coordinate system. They are converted to rectangular coordinates. Then they are aggregated from hourly values into daily values, and returned to polar coordinates. When the wind is aggregated this way, wind directions with high wind speeds are weighted higher than wind directions with low wind speeds. Also the problem with the periodicity of the wind direction is solved.

Exploratory Analysis

First part of the analysis is to explore the different attributes in the data in order to detect possible patterns or correlations. The exploratory analysis is also used to get an understanding of data and its behaviour. Hence, this chapter is about visualizing the different attributes focusing on their influence on the heat consumption. As the heat in each house is turned off in the summer period, data is segmented such that the summer period is excluded from the data used for modeling.

3.1   Examination of heat consumption

To get an overview of the heat consumption for each house, the daily average consumption for each house has been calculated and can be seen as a function of the time in figure 5.1.

Figure 5.1 shows the daily average consumption for all the houses and the daily consumption of two houses - one that follows the trend at one that deviates. It can be seen that the slopes around the summer months are close to 0. As mentioned, the data in focus in this project is where the heat is turned on, hence the period where the heat consumption is close to 0 needs to be removed. Exactly how this is done will be explained and discussed in the data segmentation section. All three plots show some unusual high data points around April 2018. This can be due to the fact that is was snowing in Denmark at that time Tilføj reference på det her.

The average of the attributes from the house data is examined through a scatterplot in order to find possible correlations. Figure 3.2 clearly shows that the consumption is close to 0 in the summer period. Pairs af gennemsnitlig house data - vi ser en masse sammenhænge mellem de forskellige attributer. Vi kan se at CoolingDegree skal være over 25, før at varmeforbruget stiger. CoolingDegree begynder at stige et stykke tid før flowet stiger, hvilket hænger godt sammen med at når man fx tænder en radiator så stiger CoolingDegree. De efterfølgende radiatorer man tænder øger volumnet.

The figure 3.3 shows the dependencies between the average consumption of the houses and the weather attributes.
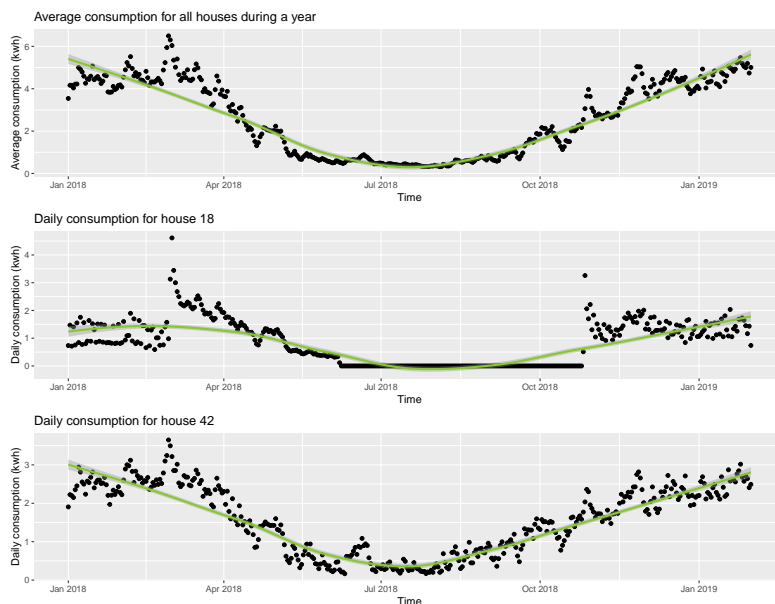
Figure 3.1: Daily consumption during a year (2018). The top plot shows the average consumption for all the houses. The plot in the middle shows an example of a house that follows the trend and the last plot shows a house that deviates from the trend.

It is already known that there is a dependency between the heat consumption and the time of year. During the summer period there is almost no consumption. The consumption in this period is probably mostly tap water. The next important thing is the relation between temperature and consumption. High temperatures tend to imply a higher consumption. And the reason why the consumption depends so clearly on the time of year can be assumed to that certain periods have similar temperature levels. It can also be seen that there is a correlation between dewpoint and consumption. This can be due to the correlation between dewpoint and temperature. Anton nævnte noget med SunHour og Ultravioletindex.

Figures 3.2 and 3.3 are used to investigate linear relationships which is desired when modeling. If a linear relation is not obtained this could give rise to a transformation on either the dependent or the independent variable. Jeg synes der mangler lidt her.

3.1.1 BBR data

Presumably, the BBR data has influence on the heat consumption in particular the total area and year of construction. Mangler lidt her.
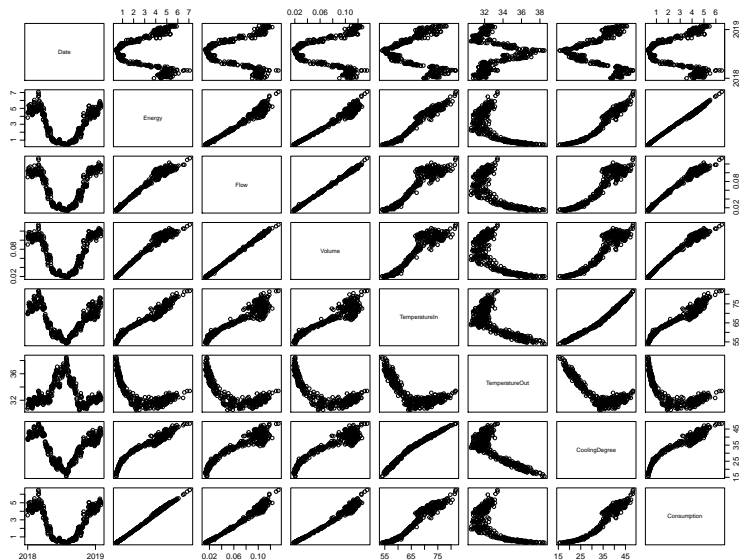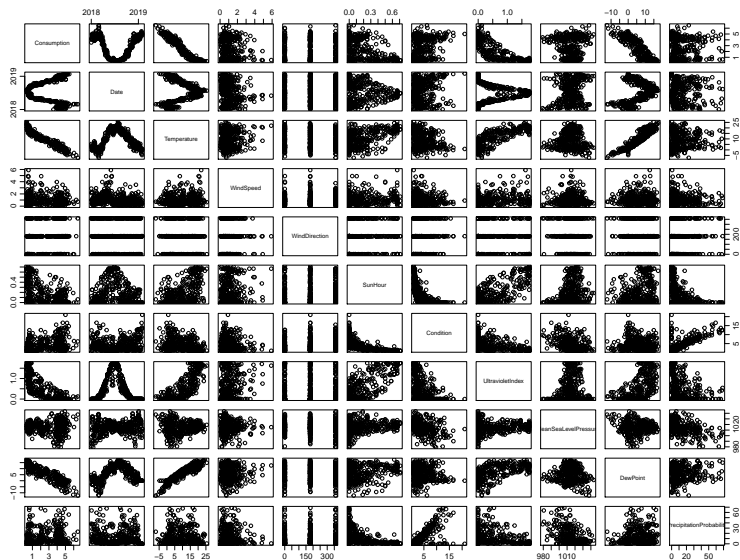
Figure 3.2: .

Figure 3.3: .

The average of the heat consumption for each house is found/determined for the winter period. By dividing the average consumption with the total area of the house the consumption pr. m^2 is calculated. Figure 3.4 shows the year of construction and the consumption for each of the houses. The year of construction is here determined by either the year of construction or the year of the latest reconstruction of a house.

Figure 3.4 clearly shows that the later a house is constructed (or reconstructed), the better is the insulation of the house as the consumption decreases with the year of construction. Furthermore, there is a clear outlier in the figure which has a remarkable high consumption pr. m^2. When looking up the house in the BBR data, it is seen that the outlier is an apartment of 61 m^2 build in 1920.

3.2   Data segmentation

Since one of the focuses of this paper is to estimate how much energy a house uses for heating depending on different outside temperatures, it is important to distinguish between when the house is actually being heated, and when the water is just being used for tap water consumption. If the inhabitants are not home for a longer period, there will probably be low consumption, even though it might be cold outside. This does not necessarily mean that the house is well isolated. And if there is consumption
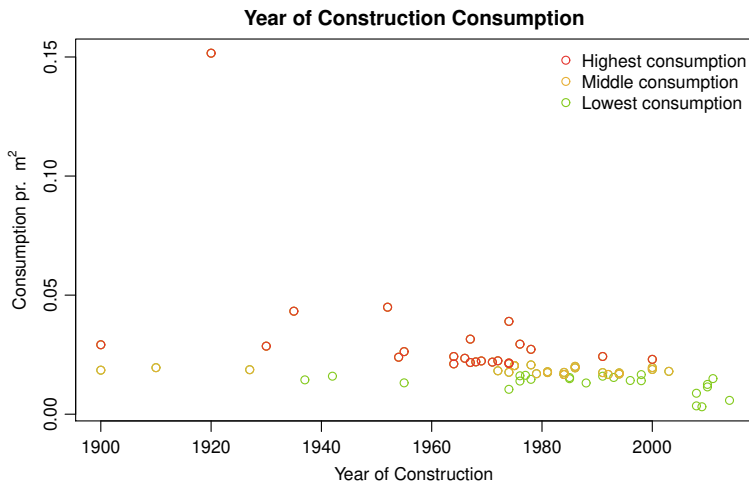

Figure 3.4: Plot showing the year of construction and the average consumption pr. m^2 for each house. It is clearly seen that there is a tendency that the later a house is built or reconstructed, the better is the insulation of the house.

in warm periods, it is likely to be tap water consumption, and not heating The data can be seen as part of two different distributions. One where the heating is turned off, and one where it is turned on. In this section different approaches will be examined on how to distinguish between the two distributions. The goal is to find some temperature, where it can be assumed that all data points below it belongs to the distribution with heating turned on. Two approaches will be described below, together with their pros and cons.

3.2.1   Segmentation by piece-wise optimization

The first approach is to make a linear regression on the data with two segments. A breakpoint α is found, such that the SSE is as small as possible. The second segment is restricted to being constant. This way the breakpoint illustrates when the consumption goes from being linearly dependent on the temperature, to having a constant value. This method was tested on every available house, where a new breakpoint was found for each house.

Figure 3.5 shows the regression for two different houses. On both houses the line fits rather well with the low-temperature data points. But it is not very accurate around the breakpoint. The house on the left shows very clearly, that the assumption that all points below the breakpoint belong to the distribution without heating, is not accurate. Even though this approach can easily take out a lot of data where there is clearly no heating, it will in many cases set the breakpoint too high. The "tail" of the low consumption distribution might still be included, causing a bias in the model, and some variation that is not accounted for. The method is also not very robust. Depending on how the points are spread out, the breakpoint is sometimes as high as 20 degrees, which is not desirable.

3.2.2   Segmentation by significant deviations

In the second approach, the data points are examined from high temperatures to low. First, all data points from above 20 degrees are assumed to belong to the distribution without heating. If a data point is more than two standard deviations above from the mean of this distribution, it is assumed to belong to the distribution with heating. Now the data points are divided by temperature into one degree intervals. For each interval, starting from above and moving down, all data points in that interval are examined. The last interval where at least 20% of the data points are less than two standard deviations away, is chosen as the breakpoint of that house. An example of the approach is seen on figure 3.6. On the left the data points are plotted with standard deviations on the y-axis. The red line highlights the two standard devations. On the right there is a plot showing how many of the data points that are outside the interval. Here, the red line shows the 80% that determine the breakpoint. The orange line shows the breakpoint.
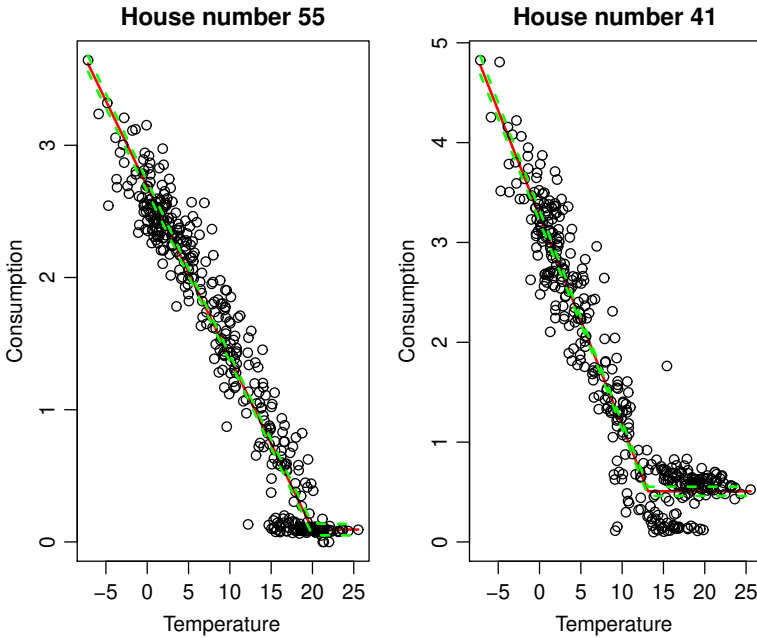
Figure 3.5: Piece-wise optimization of the consumption. The red line is the regression line and the green line is the confidence interval..

This model is more robust than the first. It is more selective, and provides a good way to set the breakpoint on the correct side of the mentioned "tail" that may occur at temperatures both with and without heating. When comparing figure 3.6 to figure 3.5, one can see that this method sets the breakpoint a bit lower, removing more points without heating. If the consumption data behaves badly, and chunks of datapoints are low enough to be within the two standard deviation, then a lot of data can potentially be removed, and there might be too little data left.

Until now the focus has been to find a breakpoint for every individual house. But it might be preferable to have a single breakpoint all houses. This way the segmentation becomes more robust to houses with unforseen heat consumption. Figure 3.7 shows a histogram of the breakpoint values for every house in the data set. The global breakpoint should be in the low end of the scale. It is better to remove data points that could have been used, than to include too many points that belong to a different distribution with a different variation, which could make the assumptions of the model worse. It would not be good to choose the minimum breakpoint, since that would be very vulnerable. A single house with a very low breakpoint might make the model bad for all the other houses. So the breakpoint that is chosen is the first quantile. As it is shown on the figure, this is 12 degrees. All models in the following sections will

only be considering data where the temperature less than or equal to 12 degrees.

3.3 Multicollinearity

Multicollinearity occurs when two or more explanatory variables are highly correlated. In linear regression, multicollinearity ... Multicollinearity can be investigated by calculating the correlation using the function `cor()` in `R`.

Figure 3.3 clearly shows that there is a high correlation between `Temperature` and `Dewpoint`. The exact correlation between the two attributes is calculated at 0.936, hence it is decided to remove `Dewpoint`. Furthermore, it is assumed that `Radiation` is a replacement for the attributes describing the sun, namely `Condition` and `SunHour`. This is the basis for expecting a correlation between the radiation and the sun attributes. Figure 3.8 shows a plot of the correlation matrix between the abovementioned attributes. There is a high correlation between `Radiation` and `SunHour` at 0.955, thus `SunHour` is removed from the weather data set.

The complete data set used for modeling in chapter 4 can be seen in table 3.1.

Variable	Description
Date	End time and date for measurements. Hourly values.
Temperature	Temperature outside in Degrees/C.
WindSpeed	
WindDirection	
Condition	
MeanSeaLevelPressure	Avg. atmospheric pressure at mean sea level in mbar.
PrecipitationProbability	Measure of the probability that precipitation will occur.
Observation	The number of observations for each day for each house.
Consumption	CoolingDegree times Volume from House data
Holiday	A categorical attribute with 6 levels: Working day, Weekend, Autumn break, Christmas break, Winter break and Spring break.

Table 3.1: Attributes used for modeling.
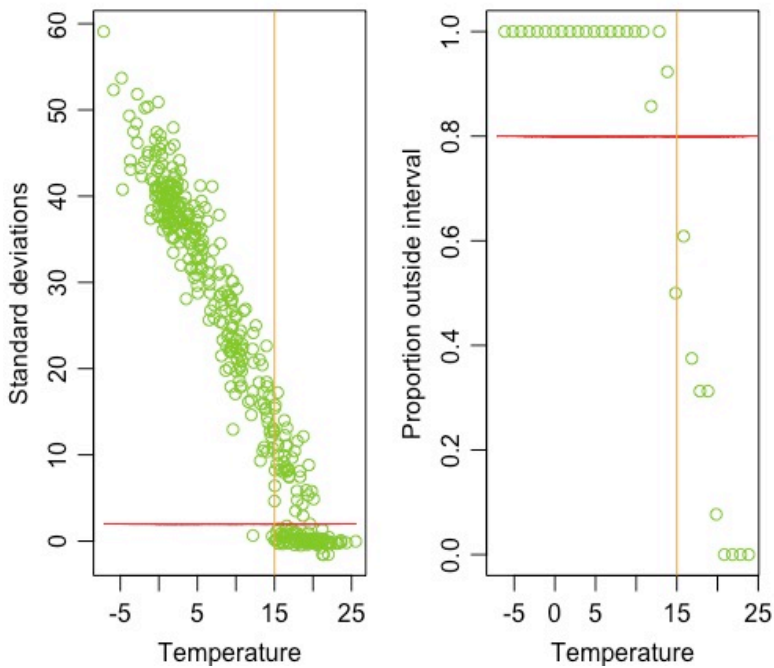
Figure 3.6: An illustration of how the breakpoint is found using segmentation by
significant deviations. On the left figure the line illustrates two standard
deviations from the high temperature distribution. The right figure
shows how many points are outside the two standard deviations. The
last point below 80% is the chosen breakpoint.
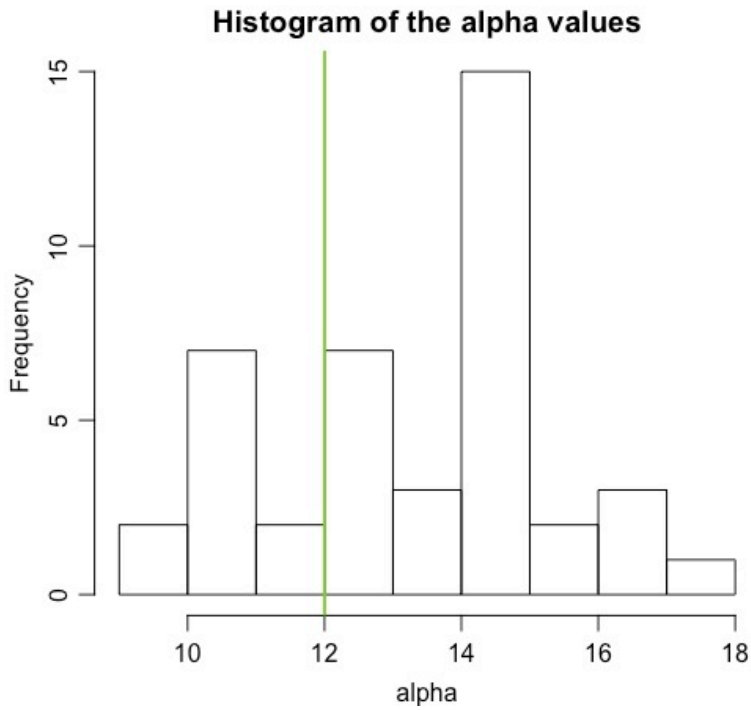
Figure 3.7: A histogram of the alpha values for every house in the third segmentation method. The first quantile is chosen as the overall breakpoint. It is 12 degrees, illustrated by the green line.
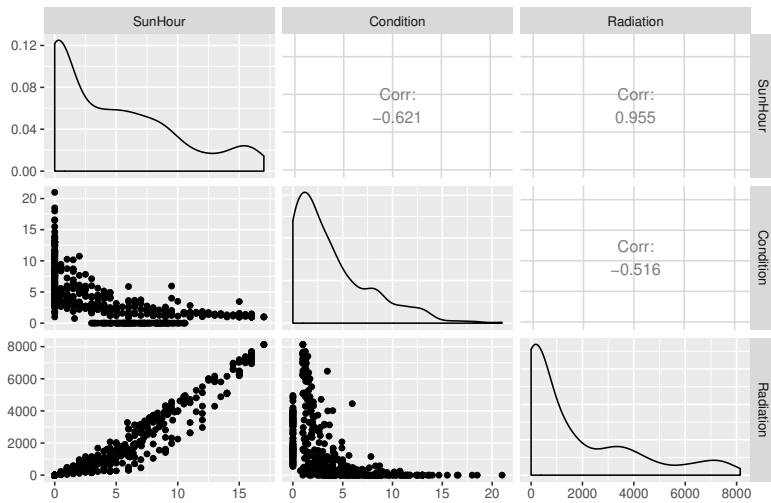
Figure 3.8: Scatterplot showing the correlations between the three attributes
`Condition`, `Radiation` and `SunHour`. It is clearly seen that the ra-
diation and the sun hour are highly correlated.

Statistical models

Now that data is cleaned and prepared, a statistical analysis consisting of data segmentation and linear regression models can be made. The purpose of the analysis is to detect which attributes affects the performance of a specific house.

4.1 Linear regression

Linear regression is a method to model the relationship between a dependent variable and one or more independent variables where the unknown model parameters are estimated from the data. Mangler nok lidt her. With the dependent variable Y and the independent variables $x_1, \ldots, x_n$, the linear regression model is formulated as

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} + \varepsilon_i, \quad \varepsilon_i \; \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, n. \quad (4.1)$$

The variables ε_i are errors which are assumed to be white noise while also being i.i.d (independent and identically distributed). Equation (4.1) shows a multiple linear regression model as it contains more than one explanatory variable. In this section both a simple linear model and a multiple linear model has been fitted to data given in table 3.1.

As the best linear model Y_i is desired, the total deviation from the data has to be as small as possible. The least squares method given as

$$\mathrm{SSE} = \sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots \beta_p x_{i,p}))^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (4.2)$$

is chosen for estimating the model. The parameters β_j are optimized to minimize the sum of squared errors of prediction (SSE).

4.1.1 Model assumptions

When SSE is minimized the model needs to be validated by checking whether the underlying model assumptions are fulfilled.

1 Normality of residuals

2 Variance homogenity

3 Variance should be independent of location

4 Linear relationship between x_j and Y

If these assumptions are not met ...

The fitting of the regression models is carried out by using the method stepwise regression Bruger vi adjusted R-squared? which updates the model in each step. In each step it is considered whether a variable is added or subtracted from the set of explanatory variables based on specific criteria.

Both a simple linear and a multiple linear regression model will be implemented in order to detect which attributes affect the performance of a specific house. This will be done by interpreting the estimates of the relation between the different explanatory attributes and `Consumption`. As mentioned, the p-value of the estimates of the explanatory variables will be the main focus when investigating which attributes influence the performance.

4.2   Simple linear regression model

A simple linear regression model is fitted to each house with `Consumption` as a function of `Temperature`. Since it is expected that the temperature is the physical phenomenon with the greatest influence on the heat consumption, it is chosen as the independent variable. The models are performed by using the `lm()` function. The models will then be validated by examining whether the model assumptions in Chapter 4.1.1 are met.

Opskriv hvilken simpel lineær regressionsmodel, vi bruger.

4.2.1   Validation

Indsæt et eksempel der viser hvordan det skal se ud.
    Det interessante at kigge efter er de huse, hvis residuals har en mærkværdig opførsel eller den simple modellering. Som eksempel ses hus 18 i figur 4.1. Det ses tydeligt, at residualerne for modellen for dette specifikke hus er gakket. Q-Q plottet ligger ikke fint langs en ret linje. Residuals vs. fitted viser en underlig opførsel, som ikke er randomly scattered.

4.2.2   Results

Plot a estimater af beta coefficienter for den simple lineære model.
Overordnet kan den simple lineære regressionsmodel ikke beskrive trenden. Den antager, at temperaturen er den eneste faktor der påvirker husenes varmeforbrug.
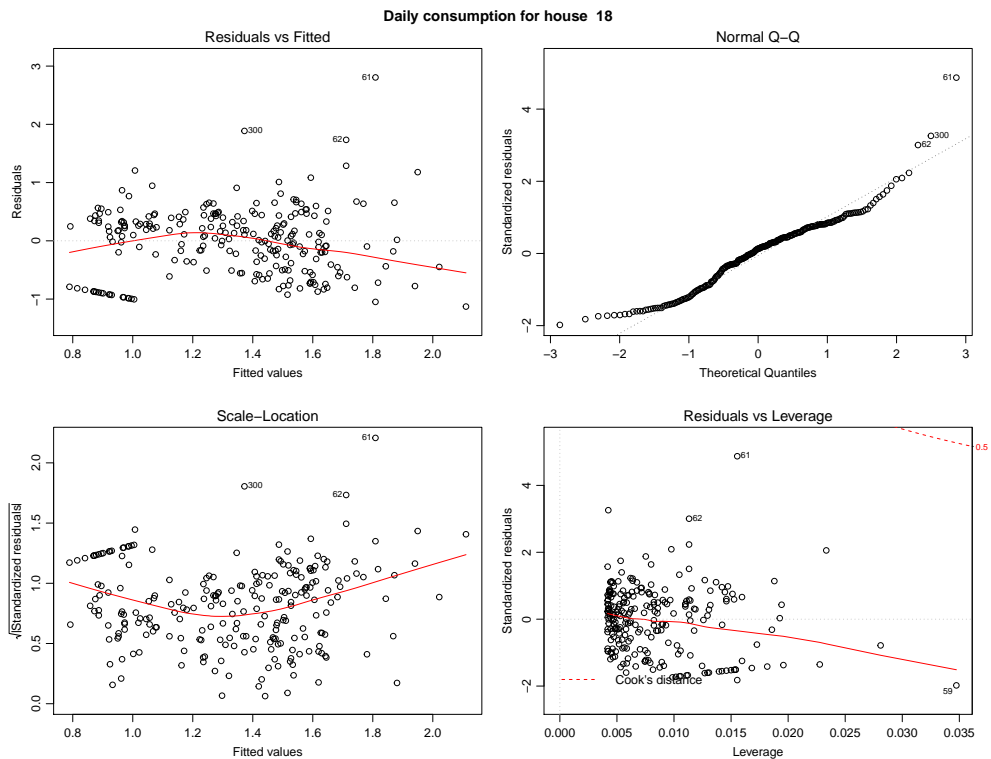
Figure 4.1: Residual plots of house 18 based on the simple linear regression model given in equation (**??**). The model assumptions of a linear regression model are not fulfilled for this specific house..

Men ved at undersøge hvorvidt model assumptions er opfyldt, så 'failer' modellen i de fleste tilfælde. Dette tyder på, at der findes flere faktorer, der påvirker varmeforbruget, hvilket selvfølgelig er forventet.

4.3   Multiple linear regression model

The linear regression model is extended to a multiple linear regression model as the inclusion of several independent variables is expected to improve the model. The simple model clearly showed that the heat consumption is affected by other physical factors than temperature. Hence, a full multiple linear regression model containing the attributes given in Table 3.1 is performed on the model data. Since `Condition` and `PrecipitationProbability` are not normalised, they are excluded from the

model. In addition, it is mentioned in Chapter 2 that the house data consists of house with observations for approximately a year and house with observations for approximately six months. Thus, the two distinct lengths of observations are modeled slighty different. There do not exist observations for winter break and spring break in the data containing the short houses. This lead to the following two multiple linear regression models: Indsæt modellerne.

The models show that the interactions between the attribute `Holiday` and the other attributes are chosen to be excluded. The reason is that `Holiday` is used to investigate how the consumption changes during holiday periods. The parameters will be denoted as follows: Intercept (I), Temperature (T), North (N), East (E), South (S), West (W), Mean Sea Level (MSL), Solar Radiation (SR), Winter Break (WB), Spring Break (SB), Autumn Break (AB), Christmas Break (CB), Weekend (WKND), the interaction between the temperature and the different wind directions (T:N, T:E, T:S, T:W).

4.3.1 Splines

In the multiple regression model, splines will be used to model the wind direction. It does not make much sence to include the wind direction as it is in the model. It is not useful to know how significant the wind direction is, if it is not connected to the wind speed and if it is not known which directions are important. By modeling the wind direction with splines, each spline will represent a specific general direction.

Modellere wind direction Lave en parameter om til flere vind retninger. 2. degree splines Knots Mellem retningerne Giver mere mening for brugeren

Vi vil gerne vægte vores vind i forskellige retninger i vores model, så derfor bruger vi splines til at modellere de forskellige retninger.

Tilføj billede af splines

4.3.2 Results

When performing the two models given in , without reduction, Et eller andet Table A.2. In addition, Table 4.1 and Table 4.2 are generated in order to determine which parameters are significant for the majority of the houses.

	I	T	N	E	S	W	MSL	SR	WB	SB	AB	CB	WKND	T:N	T:E	T:S	T:W
Sum of ***	5	41	0	18	2	24	6	22	3	3	1	5	4	0	1	7	9
Sum of **	5	1	1	9	5	12	4	10	6	2	0	2	0	1	1	9	9
Sum of *	6	1	5	7	7	2	2	2	3	6	5	3	8	2	3	5	11
Total	17	43	6	34	14	38	12	34	12	11	6	10	12	3	5	21	29

Table 4.1: The distribution of significant parameters from the multiple linear regression model for long houses.

	I	T	N	E	S	W	MSL	SR	AB	CB	WKND	T:N	T:E	T:S	T:W
Sum of ***	0	27	0	4	0	15	0	5	2	0	0	0	0	0	3
Sum of **	2	0	0	6	2	5	2	6	0	0	3	0	0	2	5
Sum of *	2	0	1	8	4	4	4	4	2	5	2	1	1	3	9
Total	4	27	1	18	6	24	6	15	4	5	5	1	1	6	17

Table 4.2: The distribution of significant parameters from the multiple linear regression model for short houses.

Hvis bare én af retningerne har over halvdelen signifikant skal alle retningerne med.

4.4   Regression model for comparing houses

Baseret på tabeller over signifikante parametre for både korte og lange huse, kan vi lave en general model for alle huse, der inkluderer temperaturen, splines og radiation.

4.4.1   Validation

4.4.2   Results

4.5   Comparison
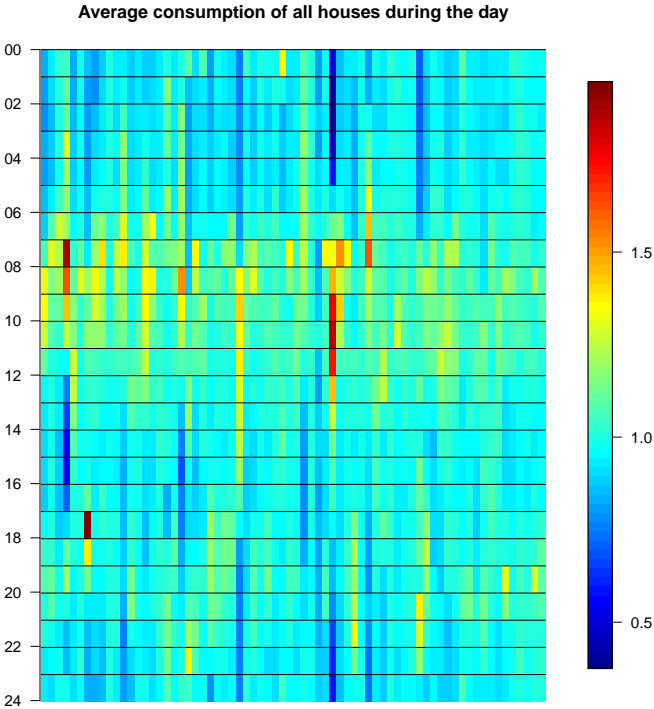
CHAPTER 5

Models on the Hourly Consumption

Average consumption of all houses during the day

Figure 5.1: .

APPENDIX A

Tables

A.1 Significance of parameters from multiple linear regression model

Index	I	T	N	E	S	W	MSL	SR	WB	SB	AB	CB	WKND	T:N	T:E	T:S	T:W
1		−***		+.		+***	+*			−***	−***					+.	−***
2	+**	−***		+**		+***		−*		−*		+*					−***
3		−***		+***		+***	+***	−***	+.			−*				+**	−*
4	−*	−***	+*		+***	+**		−***									
5		−***		+***		+***	+***	−***	+**	+***					+*	+***	−**
7	+***	−***	−**	+***	−**	+**	−***	+**	+*	+***					−***	+*	−**
11	+**	−***		+***		+***		−***	+***			+*				+**	
12		−***				+*							−*				
14	+*	−***		+***		+***		−**	+***	−*						+**	
18	+***	−*		+***	−*	+**	−**	+.				−.				+**	−***
21	+**	−***		+*		+***				+*							−***
22		−***			+***	+**		−***		−**		+.				−*	
23	+*	−***	−.	+*		+***		−***			−.					+***	−***
28		−***		+**		+.		−***		−*		−***	−***				
29	+.	−***		+***		+***		−**	−*	+.				+.		+*	−***
30	+.	−***		+***		+***		−***	+*	−*						+.	−.
31		−***		+***		+***	+**	−***		+.		−.		+**		+***	−***
32		−**			+*	+***		−**		−*					+*		−*
33		−***		+***				−**									
34	+.	−***		+***	+.	+***		−**			−*	+***			−*		*
36	+**	−***		+***		+**		−***								+**	
37		−***		+***	+**	+**											
38	−***	−***		+**		+***	+***	−***	+**	−.		−*		+.		+.	−***
40	+*	−***		+*	+**	+**		−***	−**		−*					+.	−*
41	+***	−***		+***		+***	−.	+.			−.					+.	−.
42		−***		+***		+***	+***	−***	+.				−***			+*	−*
44		−***	−*									+**				+**	
45	+**	−***			+*	+***		−.				+*			+*	+.	−**
46		−***		+*	+*	+**		−**							−*		−**
47		−***		+**	+*	+***	+.	−*	−.			+***	+***		−**		−***
48	+*	−***		+**		+**		−**	+**				−***				−**
49	+*	−***						−**	+*								
50		−***	−*	+***	−**	+*		−***	+**					+.		+***	
52	+.	−***	−*	+*	−*			−**	−***	−***		−***		+.		+***	
54		−***		+**		+**		−***				−.	−*			+**	−*
55		−***		+*		+***	+*	−***				−.	−*			+***	−*
56	+**	−***		+**		+***		−**	−.	−*						+*	−*
57	+.	−***		+***		+**											−*
58		−***	−.	+.		+***		−***				+**			+*	+***	−*
61	+***	−***		+**		+**	−***	+.	+**			−.			−.		
64		−***	−*	+***		+**		−***	+.				−*		−*	+**	−**
65		−***		+*		+***		−***								+*	−**
66		−***		+***	+*	+***		−***		−.	−.					+**	−**

Table A.1: Significance of parameters from the full multiple linear regression model performed on 'long' houses.Punktum betyder, at det er mellem 5 og 10%.

Index	I	T	N	E	S	W	MSL	SR	AB	CB	WKND	T:N	T:E	T:S	T:W
6		–***		+*		+**		–*	–.						
8	+**	–***		+*		+***		–*						+.	–*
9		–***	–.			+**	+**	–***	–*		–*	+.		+*	–**
10		–***		+***		+***	+.	–**			–**	+.			–*
13		–***		+*		+**	+*	–*			–.	+.		+**	–*
15		–***		+***		+**		–.		–.	–.				–.
16		–***		+*		+***	+.	–*							–**
17		–***		+**	+**	+***		–**			–.				–***
19		–***		+**	–.		+.	–.			–.		+.	+**	
20		–***			+*										
24		–***		+*		+***	+*								–***
25		–***		+**		+***			–.					+.	–*
26		–***		+.	+.	+***	+*	–**	–***						–*
27		–***		+*	+**	+***				–*					–*
35		–***		+**	+*	+***		–**		+*					–**
39	+**	–***				+*	–*	–.			–.				
43		–***		+**	+*	+***		–**		+*					–**
51		–***		+.		+***		–**			–*			+*	–*
53		–***		+*		+*									–.
59		–***	–.	+***	+*	+***		–.				+.		+.	–.
60		–***	–*	+**	+.	+***		–***				+*		+*	–***
62	+*	–***				+*		–.	–*						
63	+.	–***		+*							–**		+.		
67	+*	–***		+***		+***		–***	+.	+*			–*	+.	–**
68		–***				+*		–***	+*	–**				+*	–*
69		–***			+.	+**		–.	–***				+.		
70		–***		+.	+.	+***	+**	–***							–*

Table A.2: Significance of parameters from the full multiple linear regression model performed on 'short' houses. Punktum betyder, at det er mellem 5 og 10%.

APPENDIX **B**

Figures