Smart meter consumption time-series forecasting

Ir. Stijn Staring

Thesis submitted for the degree of
Master of Science in Artificial
Intelligence, eg

Thesis supervisor:
Prof. dr. ir. Bart De Moor

Assessors:
Prof. dr. ir. Unknown
Prof. dr. ir. Unknown

Mentor:
Ir. Lola Botman

Academic year 2020 – 2021

Preface

I would like to thank everybody who kept me busy the last year, especially my promoter and my assistants. I would also like to thank the jury for reading the text. My sincere gratitude also goes to my wive and the rest of my family.

Ir. Stijn Staring

Contents

Abstract

The `abstract` environment contains a more extensive overview of the work. But it should be limited to one page.
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Abstract

In dit `abstract` environment wordt een al dan niet uitgebreide Nederlandse samenvatting van het werk gegeven. Wanneer de tekst voor een Nederlandstalige master in het Engels wordt geschreven, wordt hier normaal een uitgebreide samenvatting verwacht, bijvoorbeeld een tiental bladzijden.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

List of Figures and Tables

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

LoG	Laplacian-of-Gaussian
MSE	Mean Square error
PSNR	Peak Signal-to-Noise ratio

Symbols

42	"The Answer to the Ultimate Question of Life, the Universe, and Everything" according to [**?**]
c	Speed of light
E	Energy
m	Mass
π	The number pi

Chapter 1

Introduction

The first contains a general introduction to the work. The goals are defined and the modus operandi is explained.

1.1 Importance of topic

Customer is better informed what the bill is going to be at the end of the month/year. Energy producer can build a better trust with its customer by sending reliable bills. (Providing good service) Producent can better estimate the energy demand of the whole customer population. This will lead to cheaper electricity production because a better planning is possible where there is less need of the more flexible but more expensive electricity installations e.g. diesel engines.

1.2 Problem formulation and link with previous studies

Now going to forecast individual houses, not aggregated signals.

1.3 Thesis objective and structure

The goal of this thesis is to do short-term load forecasting for individual households. A forecast of the electrical load of a household for 24 hours.

Chapter 2

Data analysis

In this chapter details of the dataset are introduced and an analysis is performed. Things discussed about the dataset concern assessing missing data, removing zero days, normalizing the data and removing time-series with identified fundamental changes The analysis looks at the seasonality, influence of temperature, comparing weekdays with weekends, impact of holidays and the driving households characteristics. Finally the definition of a suitable baseline model is given, which will be used during the evaluation with more elaborate models in chapter 6.

2.1 Introduction to dataset

update pictures The data that is used in this thesis is made available for the IEEE-CIS technical challenge on energy prediction from smart data. It consists out of data from smart meters about the 1/2 hour granulated electricity consumption of 3248 households located in the United Kingdom in the year 2017. The definition of a household are all the people who occupy a single housing unit, regardless of their relationship to one another. Each smart meter collected thus a total of 17520 measurements that are performed by the the leading international energy provider, E.ON UK plc. Not all the 3248 smart meters consist of full data as can be seen in Figure A.1 in appendix A. It can be clearly seen that there are 12 steps in the amount of missing values. This is because the available data ranges from one month (only December) to a full year of data. This acknowledges that customers may have joined at different times during the year. Additionally, missing values are introduced due to errors in sending/receiving from smart meters.
Next to the electricity consumption of the different households, also information is available about the average, minimum and maximum temperature of the day on the location of the smart meter. This data is available at a daily resolution. Also, through voluntary surveys, incomplete information is collected about 2143 smart meters. This concerns e.g. dwelling type, number of occupants, number of bedrooms etc. Table A.1 displays all the attributes in appendix A.

Because of the additional information about the attributes that are summed up in

consumption.csv	
# households	3248
information	electric load
measurements	17520
granularity	$\frac{1}{2}$hour
timespan	year 2017
location	UK

weather.csv	
information	average temperature
	max temperature
	min temperature
granularity	daily

addInfo.csv	
# households	2143

Table 2.1: Table with information about the characteristics of the available datasets.

Table A.1, it can be better understood what kind of households are included in the consumption.csv. It is assumed that all the loads are measured form households of the type listed below and each household is made up of maximum four persons and has a maximum of five bedrooms. industrial loads or small businesses, a bakery for example, are not considered.

- flat
- bungalow
- detached house
- semi-detached house
- terraced house

2.2 Preprocessing

Following steps discuss the preprocessing done on the consumption time-series containing measurements for the entire year.

2.2.1 Missing data

As discussed above the consumption dataset contains additionally to the missing months also missing data due to sending/receiving errors of the smart meters. When this happens the data of the whole day is lost. It should be emphasized that a missing value should not always directly be seen as an error. It can be that the smart meter was put off because the inhabitants were on a holiday for example. The nan values then also gives information about the consumption behaviour, namely that it is possible that the inhabitants go on vacation and the electrical load will in this case normally correspond to a constant base load. However, the assumption is made that in the case of the "consumption.csv" missing data corresponds to a sending/receiving error of the smart meter. This assumption is valid because when full year data is assessed, the missing values always perfectly correspond to a day of missing values. It is therefore highly likely that the organizers of the competition manually deleted days in the consumption to increase the difficulty of the forecasting and to model sending/receiving errors of the smart meters. That the missing values correspond to sending/receiving errors is also stated in the data description of the competition.

Two methods to impute the missing values are compared. Method one substitutes the missing values of a time-serie by the mean of all the measurements done by the meter. Method two replaces the missing values by the mean consumption value of the same moment on the next and previous day. If the next or previous day is also missing, the closest known day is used. The resulting signals can be seen in Figure A.2 and Figure A.3 in appendix A.

In order to ascertain which method of the two performs the best, a reference dataset is needed in order to compare the estimated with the true values of the missing measurements. From the original dataset which contain 3248 meters it was found that for 181 meters the month March was given without missing data. These 181 complete signals of the month March are used as reference dataset. In order to create the test data in each of the 181 meter signals 7 random days of the month March were removed and estimated by the earlier two methods. The normalized mean square errors, MSE_{AN} and MSE_{mean} given by $\sum_{i=1}^{D} e_i^2$ and normalized by MSE_{mean} are given in Figure 2.1.

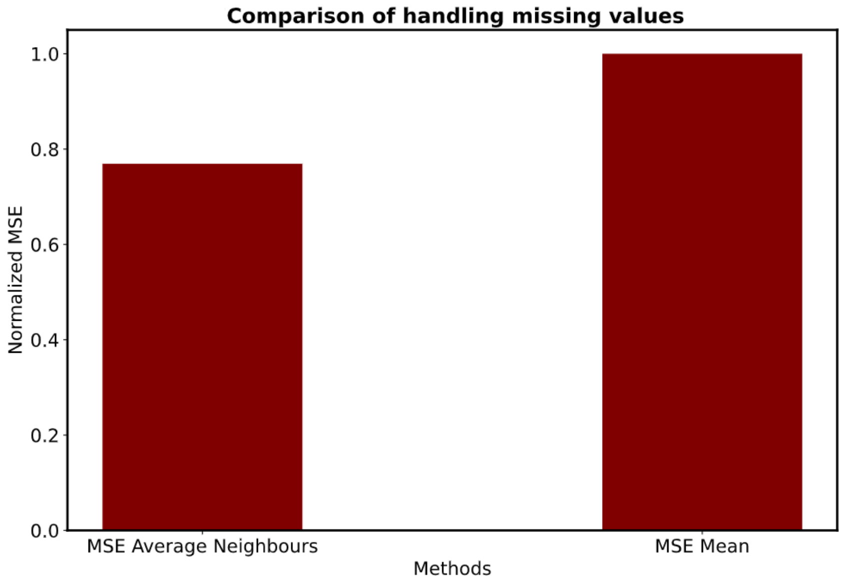

FIGURE 2.1: Resulting month of March after substitution of the missing values by the mean value of the measurements.

From Figure 2.1 it can be seen that using method 2 which estimates the missing values by the mean consumption value of the same moment on the next and previous day, outperforms method 1 which takes the mean of the signal. Therefore, all the missing values in the consumption dataset are estimated using method 2 with the only exception the first of January and thirty-one December. If one of these two days are missing, the method 1 is used because of the absence of two neighbouring days.

2.2.2 Zero days

When processing the consumption data, some untraditional meter measurements were identified. For example there were 9 meters that had multiple days with zero day consumption measurements. Because it is unlikely that a household produces exactly zero kWh on a day all these 9 meters were removed. The consumption time-serie of one of the meters is displayed in Figure 2.2 in appendix A.

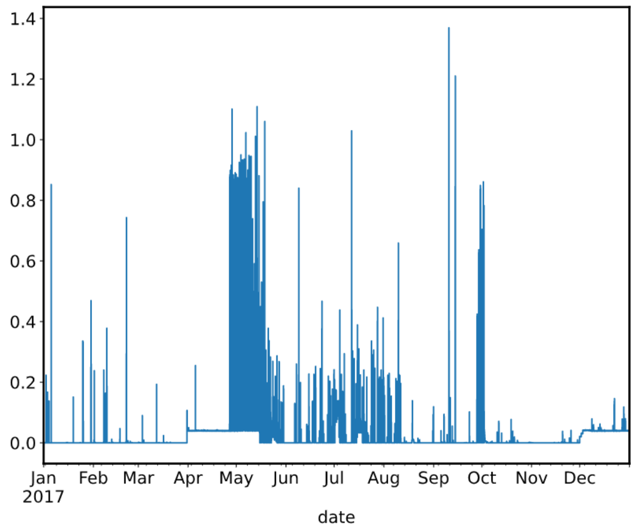

Figure 2.2: One of the 9 identified meters with multiple zero daily consumptions

Also, there has been looked if there were fundamental changes in the electricity consumption of certain meters. This is further discussed in section **??**.

2.2.3 Normalization of the data

Normalization is necessary because while absolute consumption differs, relative patterns of human behaviour are more similar [3]. The patterns in the human behaviour is what a forecasting model is trying to predict and normalization contributes by avoiding the disturbance of different magnitudes in which this human pattern may occur. Every individual household time-serie is normalized based on its yearly consumption as was done in [3]. The advantage of using the yearly consumption to normalize in comparison of the minimum and maximum values, is the robustness against measurements out shooters and every smart meter has a total consumption of one at the end of the year.

$$normalized \quad value = \frac{consumption_i}{\sum_{n=1}^{17520} consumption_i} \tag{2.1}$$

As discussed in section 2.3 the average is taken over all the normalized time-series to obtain a single signal.

2.2.4 Removing of fundamental changes in the consumption load

maybe have to remove this section. After normalization of all the individual time-series it is looked for fundamental changes in the consumption load due for example when an extra person lives in the house or when systems are installed that use a lot of electricity, during the year. An example of such a time-serie can be seen in Figure A.4 in appendix A. These changes are identified by looking at the maximum difference of the minimum and maximum rolling mean consumption over 7 days for each individual meter. If this difference can not anymore be explained by the dependency on the temperature and previous present appliances, it is assumed that a fundamental change in electricity consumption took place. It is desired that the mean consumption doesn't change much during the year. This is because the model that later will be used expects the household situation to be the same for the entire year and the time-series with a fundamental change can thus lead to a disturbance when it is kept in the training data. Figure 2.3 shows all the maximum differences between the minimum and maximum weekly rolling averages. The red line on 0.00125shows the cutoff and the smart meters above this line are removed. This cut off value chosen on sight. In total 239 smart meters remain. In Figure A.5 in appendix A the time-serie with the new maximum difference between the minimum and maximum weekly rolling averages is given.

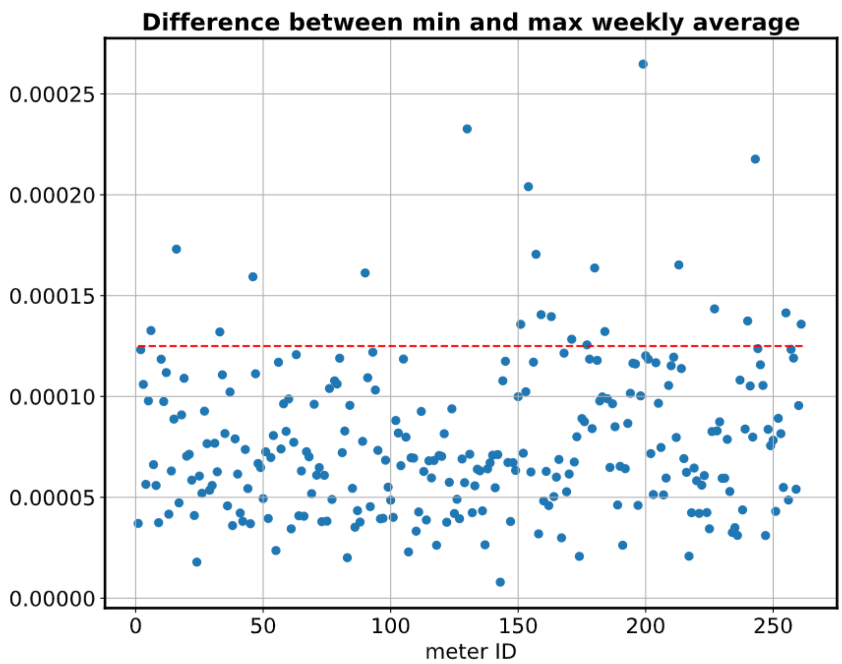

FIGURE 2.3: The maximum differences between the minimum and maximum weekly rolling averages for all the different time-series.

2.3 Analysis

Finally, the average is taken over all the remaining ... time-series to obtain a single signal. This is done to investigate the dependency of the smart meters on seasonality, temperature, weekends and holidays. At the end of this chapter a baseline forecasting will be discussed that will be used as null-hypothesis in chapter 6 to assess if the developed models lead to an improvement.

2.3.1 Seasonality

In this section the seasonality of the consumption data is discussed. In [2]it was concluded that all the forecasting algorithms that were considered, produced more accurate forecasts when they were combined with a preprocessing stage that extracted the seasonality before forecasting, compared to applying the same algorithms directly on raw data. The forecasting model is left with the task of modelling the deviation of the template consumption instead which is less challenging than performing a forecast out of the blue. These templates or filter are extracted from the consumption dataset by the use of equations 2.2 and 2.3. D and W gives respectively the number of days and weeks in the dataset. $\bar{y}_i$ and $\bar{y}_j$ gives the consumption of half an hour, averaged over respectively all days and weeks. **read paper again...**

$$\bar{y}_i = \frac{1}{D}\sum_{d=1}^{D} y_{di} i \in [1, 48] \tag{2.2}$$

$$\bar{y}_j = \frac{1}{W}\sum_{w=1}^{W} y_{wj} j \in [1, 336] \tag{2.3}$$

Figure A.6 shows the daily filter in appendix A.
Figure 2.4 shows the weekly filter.
In the daily and weekly filters there can clearly be seen a consumption peek after midnight. This is due to heat storage systems that use electricity in the hours of low tariff and that release heat during high electricity tariffs.

2.3.2 Influence of temperature

In following section the correlation between the temperature and the electricity consumption is discussed.

Pearson correlation

The Pearson correlation is a measurement of the linear dependency between two variables which is based on the covariance variable. A Pearson correlation values gives information concerning the magnitude of the association and the corresponding direction of it. A Pearson value of one and minus one give respectively a perfect positive and negative linear relation between the variables. A value of zero, corresponds

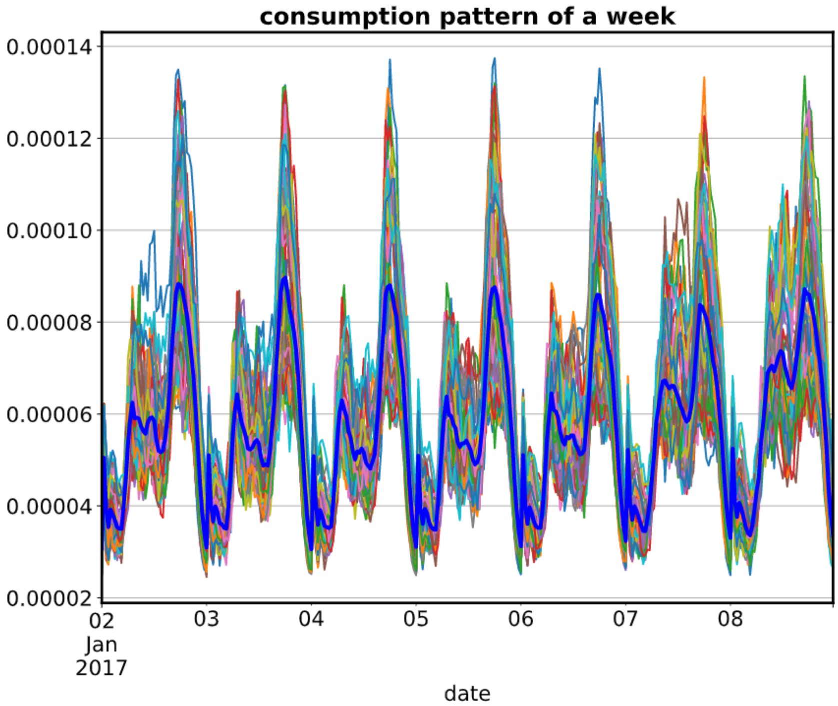

FIGURE 2.4: Figure that shows the seasonality of the electrical load during the week.

to independent behaviour. Following formula is used when calculating the Pearson correlation.

$$\rho_{X,Y} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y} \tag{2.4}$$

Assumptions concerning Pearson correlation are that samples used for the correlation should be independent, normal distributed and linear related to each other. Also, homoscedasticity is assumed. Homoscedasticity is important when performing linear regression and assumes that σ_x and σ_y are constant and not in function of each other. This final assumption is validated by making use of Figure 2.5.
This figure shows the classic cone-shaped pattern of heteroscedasticity. On days when it is warm there is overall similar human behaviour in lowering the electricity consumption. However, on colder days the variation in consumption is higher. Because the assumptions of the Pearson correlation are not fulfilled, care should be taken with its output.

Applying the Pearson correlation on Figure 2.5 gives a correlation value of -0.87. This means there is a reasonable linearly decreasing relation.

Spearman correlation
Spearman correlation is a "Rank correlation". This means that the ordering of the consumption and temperature in a sample are each compared in their corresponding

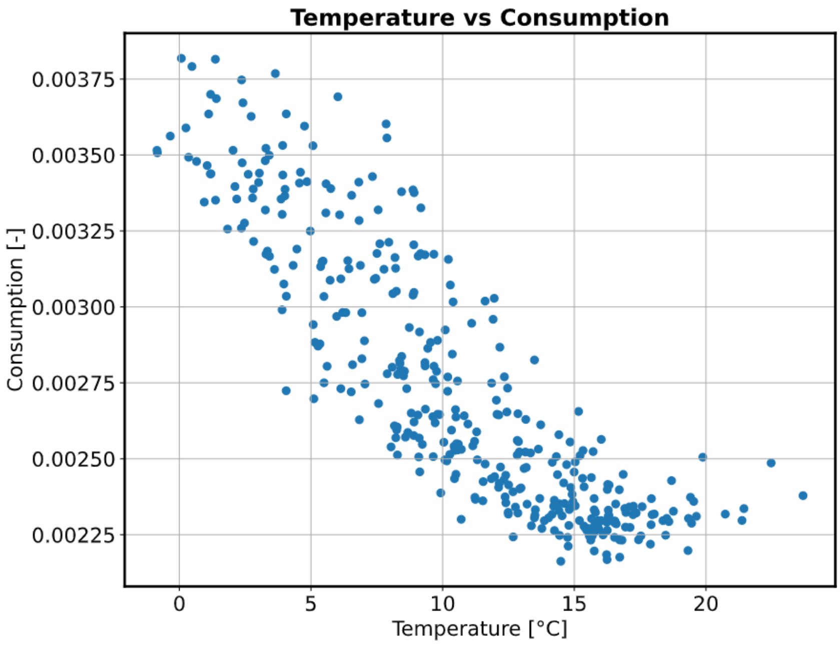

Figure 2.5: Relation between consumption and temperature.

array of measurements. When the ordering of both variables in a sample are similar, correlation is strong and positive. If the ordering is reversed, correlation is strong and negative. There is a perfect positive ordering if larger consumption always corresponds to a higher temperature. Notice that for a perfect ordering, no linear relation of the variables is necessary. The Spearman correlation coefficient is calculated using equation 2.4, but takes into account the rank of a variable in all the measurements of this variable instead of the measurement value itself.

In order to use the spearman correlation data has to be ordinal, which means that it can be ordered. The spearman correlation gives information about the monotonicity relation between the variables. $\rho = 1$ corresponds to a monotonically increasing relation.

Applying the Spearman correlation gives a correlation value of -0.89, which means there is a reasonable negative monotonicity relation.

Kendal correlation The "Kendal correlation" is also a rank based correlation. Here it is looked at the pairs of observation that are concordant, discordant or neither. A correlation coefficient close to one occurs when both variables have the same ranking and similar a coefficient close to minus one occurs when rankings in one variable are the reverse of the other. Equation 2.5 gives the equation to calculate the "Kendal correlation coefficient".

$$\tau = \frac{n^+ - n^-}{\sqrt{(n^+ + n^- + n^x)(n^+ + n^- + n^y)}} \tag{2.5}$$

- n^+ is the number of concordant pairs
- n^- is the number of discordant pairs
- n^x is the number of ties only in x
- n^y is the number of ties only in y
- concordant $\rightarrow (x_i > x_j)$ and $(y_i > y_j)$ or $(x_i < x_j)$ and $(y_i < y_j)$
- discordant $\rightarrow (x_i > x_j)$ and $(y_i < y_j)$ or $(x_i < x_j)$ and $(y_i > y_j)$
- neither $\rightarrow (x_i = x_j)$ or $(y_i = y_j)$
- if both $(x_i = x_j)$ and $(y_i = y_j) \rightarrow$ not included in either n^x or n^y

Applying the Kendal correlation gives a correlation value of -0.67, which means there is a reasonable negative monotonicity relation.

2.3.3 Comparing weekdays with weekends

Weekdays vs weekends can be compared with the help of Figure 2.4. It can be seen that the consumption of the average weekday is very similar to a weekend day.

2.3.4 Impact of holidays

In order to look at the impact of a holiday, all the holidays of the English and welsh holiday calendar are identified for the year 2017. For each of the 8 holidays a corresponding business day is selected with an as close as possible average temperature of the day. This is done to remove the temperature dependency. The resulting average holiday and business day with is given in Figure 2.6. It can be seen that the consumption of a holiday is reasonably similar to business day. **This is maybe not so important and can be left out...**

2.3.5 Identification of driving attributes

In this section the influence of the extra knowledge about the kind of household where the smart meter is located, is investigated. This is not done by using a single averaged signal as was the case in the previous analysis sections. Now, every meter with additional information is considered. In Figure 3.1 the monthly consumption of the month December in function of dwelling type is shown. The month December is chosen, because this month is known for every smart meter. Missing values of

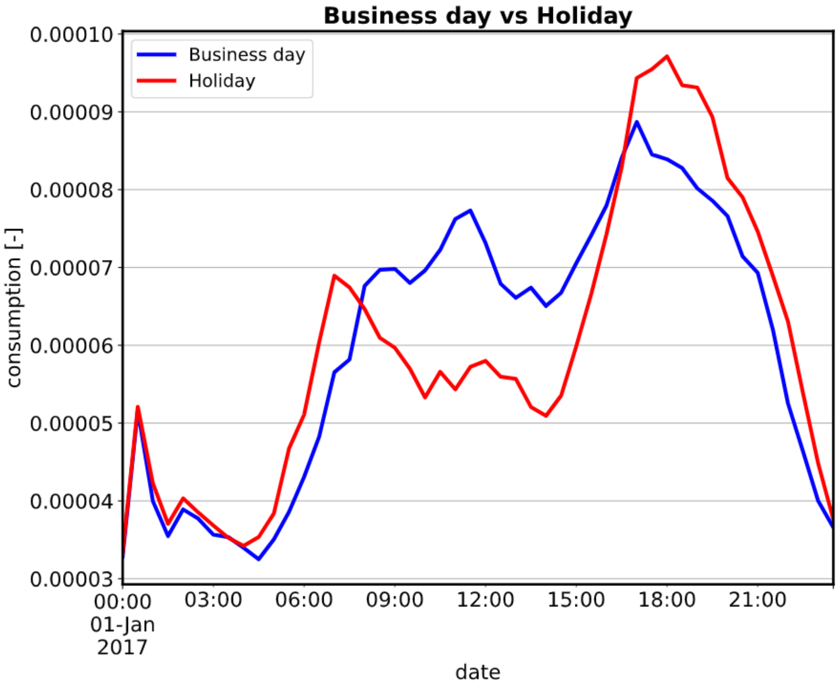

Figure 2.6: Figure with the comparison between holidays and business days.

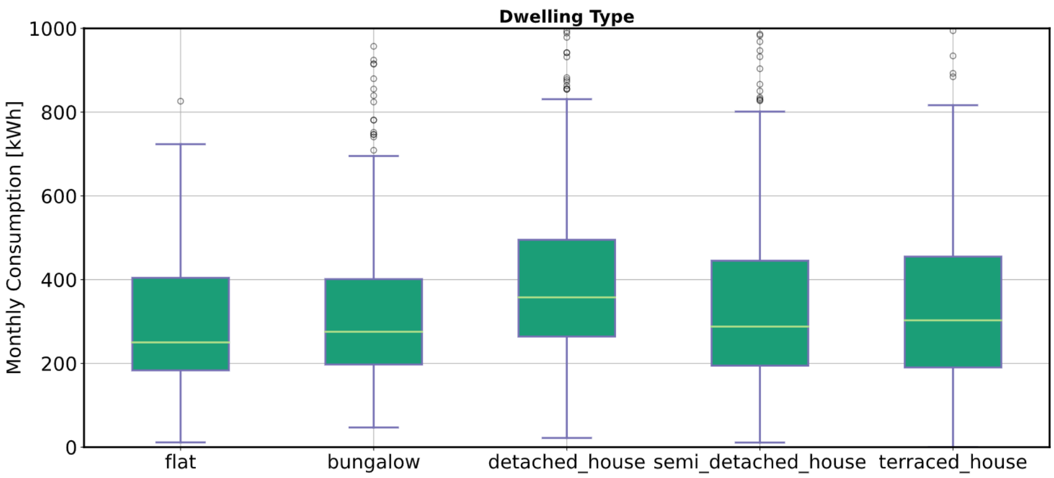

Figure 2.7: Figure with the comparison of the different dwelling types.

the smart meters are substituted by method two, as discussed in section 2.2.1. The amount of meters used for every visualization can be seen in Table A.1.

Similar as was done in Figure 3.1 is also done for the other characteristics of the smart meters. The conclusions are listed below. As can be seen in Table A.1, some characteristics have not much data or the data is not much distribute over the different options of a characteristic. If this is the case, no reliable conclusions could be drawn.

- There is a lot of variance in the monthly consumption of a detached house, but it has mostly a higher consumption than other dwelling types
- A “real” house (detached, semi-detached or terraced) tends to have higher monthly consumptions than a flat or bungalow.
- The order of monthly consumption according to the mean and median values: Flat < Bungalow < Semi-detached < Terraced < Detached
- More occupants means more monthly consumption
- More rooms in the house means more monthly consumption
- Almost all houses use gas as heating fuel
- Almost all houses use gas as hot water fuel
- The age of the boiler has no clear effect on the monthly consumption
- The vast majority of the lofts are insulated
- The majority of walls are insulated
- The vast majority heats till a temperature between 18 and 20 degrees
- The majority of people has an efficient lighting percentage between 75% and 100%

2.4 Baseline model

A naive baseline 24 hour forecasting model that is chosen is using the last 24 hour as prediction.

2.5 Conclusion

The final section of the chapter gives an overview of the important results of this chapter. This implies that the introductory chapter and the concluding chapter don’t need a conclusion.

Chapter 3

State of the art short-term residential load forecasting techniques

Forecasting the electrical load of the different individual households has a couple of challenges. There should be dealt with the missing values, as discussed in section 2.2.1. Also, the different time-series are influenced by exogenous factors as weather conditions and the day of the year. The dependency on exogenous variables can be a very non-linear relation and can have different effects on different households. For example depending on a house has solar panels, the consumption could be altered much. Only three indications of the temperature are given on a daily basis. Some additional information is know of certain households, but this data is very incomplete. Next, the individual load series have a high volatility and uncertainty with respect to a load signal on transmission level which shows more consistent seasonality and straight forward dependency on weather and calendar variables. This is because the contingency of the individual load data is mitigated due to averaging out of the uncertainty. Ofcourse, the obvious disadvantage is that only forecasts on this aggregated level can be made which is not the goal of our investigation.
To tackle the high non-linearity that is inherent to residential load forecasting in literature often “Neural Networks” to deal with this.

3.1 Introduction to Neural Networks

A standard multilayer feedforward neuralnetwork with locally bounded piecewise continuous activation function can approximate any continuous function to any degree of accuracy if and only if the network’s activation function is not a polynomial, as stated by **Leshno et al** in **1993**. This theorem proofs that a “universal approximator” exists for continuous functions, but it lacks the recipe to construct it. In [4] it is shown that a feedforward network with a single layer is enough to approximate any function by a specified accuracy if the hidden layer has the possibility to add an unlimited amount of hidden neurons in its layer. It is discussed that when a function is discontinuous, which means that it makes sudden, sharps jumps, it is not possible to approximate the function by any prescribed accuracy. However, in practise a continuous approximation is often good enough.

Neural networks are suitable of learning very non-linear mappings between inputs and outputs. The difference between “Deep Neural Networks” and “Shallow Neural Networks” is the amount of layers of neurons are used inside the network. These layers of neurons, that are not inputs or outputs are called “hidden neurons”. Because a “Deep Neural Network” has a a hierarchical layout of the different hidden layers, it not only learning features from the non-linear combinations of inputs, but uses other layers to learn features of combinations of features learned in lower hidden layers. This is possible because higher hidden layers get the outputs of lower hidden layers as input. As discussed in [5] due to this characteristic, deep learning is suitable to learn multiple uncertainties with differing sharing levels over different households e.g. the amount of sunshine. However, because of the higher expressiveness (and often the amount of the to learn parameters), a “Deep Neural Network” with respect to a “Shallow Neural Network”, suffers more of overfitting as is discussed in section 3.1.4.

3.1.1 MLP

figure, activation function, equations

3.1.2 CNN

3.1.3 RNN

3.1.4 Problems

Vanishing gradient problem –> short memory. Overfitting. Explosion of the gradient possible. tanh “Deep Recurrent Neural Network” –> more parameters –> more overfitting

Overfitting

Early stopping –> can be seen as regularization Training with regularization. Dropout regularization Pruning

Vanishing gradient

LTSM –> vanishing gradient problem is not solved! Mittigated. GRU Attention model Transer model

3.2 Short-Term residential electrical load forecasting

Classical ways to deal with uncertainty.

In this paper [5] a novel pooling-based deep recurrent neural network is proposed which collects a group of customers load profiles into a pool of training inputs. Pooling of households historical loads to serve as input of the “Deep Recurrent Neural Network”, is proposed to increase the data volume and diversity of load forecasting, which mitigates the effect of overfitting present in a DRNN. Also, due to the pooling of different households during training the DRNN is able to learn common uncertainties. From the pool of inputs every epoch a randomly chosen batch of load signals are fed to the network. LSTM is applied to mittigate the short term memory of the RNN. Additionally, there is been made use of early stopping to further avoid overfitting. To implement early stopping there has been looked at the “MSE” for k iterations, obtained by cross-validation. When the variance of this sequence gets smaller than a specified variable, training stops. When the training ends, performance is tested on each household by using the learned network to perform a feed-forward prediction of the electrical load.

An overview of the different steps that were done during the proposed method are: data cleaning and preprocessing → data pooling → data sampling → data training → benchmarking.

Performance of the proposed method was finally evaluated based on:

1. performance of the proposed method with respect to Vanilla RNN, SVR and DRNN (without pooling)
2. the effect of the neural network depth and pooling

The proposed DRNN with pooling outperforms all other four methods based on following three metrics:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(\hat{y}_t - y_t)^2}{N}} \tag{3.1}$$

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{3.2}$$

$$MAE = \frac{\sum_{t=1}^{N}|\hat{y}_t - y_t|}{N} \tag{3.3}$$

The effect of the depth of the DRNN and the pooling method is depicted in Figure **??**. It can be seen that without the pooling method the DRNN only benefits from extra layers till three are used. This is because from that point, overfitting will reduce the generalization capacity of the DRNN. With the pooling technique, extra layers stays beneficial. It can thus be concluded that introducing extra hidden layers is a good choice to model the non-linear realations, but this can only be done efficiently when overfitting is mittigated by the use of a pooling strategy.The RNN with pooling used for benchmarking consisted out of five layers and thirty hidden units in each layer.

3.3 Tables

Tables are used to present data neatly arranged. A table is normally not a spreadsheet! Compare Table **??** en Table **??**: which table do you prefer?

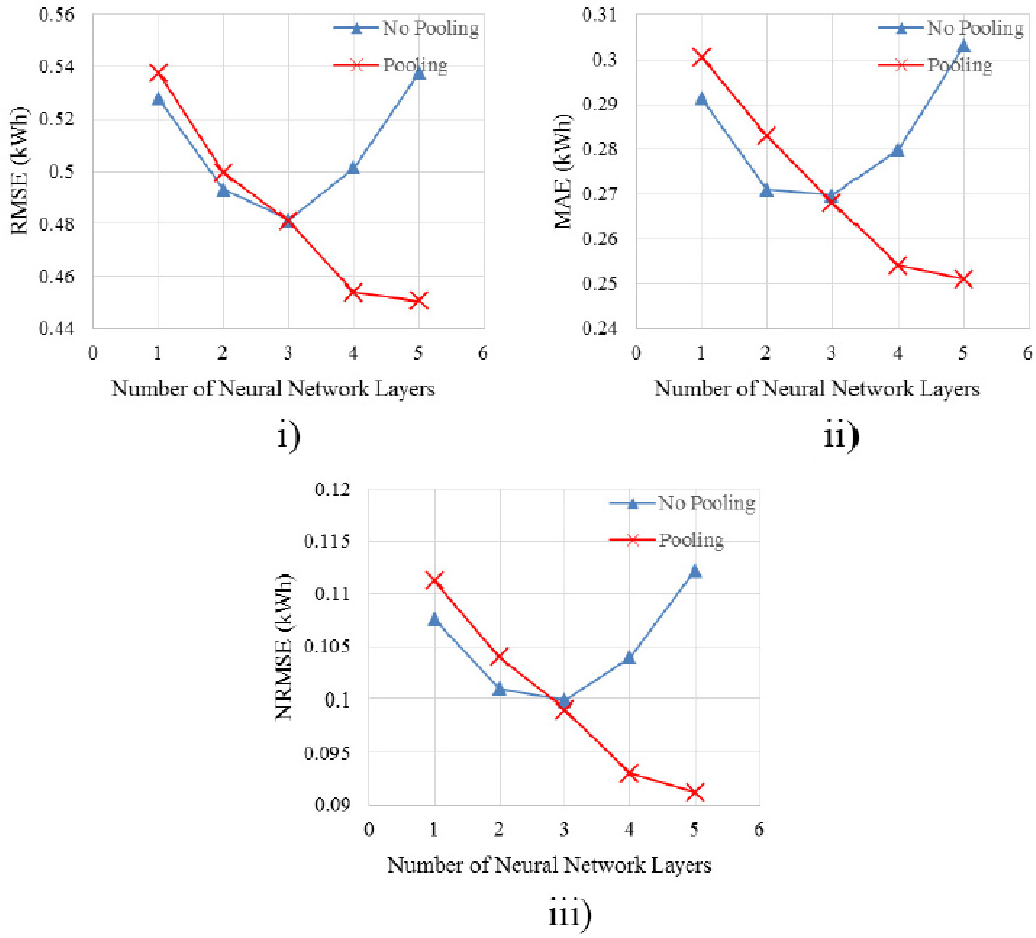

Figure 3.1: Influence of the number of layers and the pooling method.

3.4 Conclusion

The final section of the chapter gives an overview of the important results of this chapter. This implies that the introductory chapter and the concluding chapter don't need a conclusion.

Chapter 4

Clustering of the load profiles

Do a literature study about forecasting. What is the current state of the art methods to do forecasting.

4.1 The First Topic of this Chapter

4.1.1 Item 1

Sub-item 1

Nunc velit augue, scelerisque dignissim, lobortis et, aliquam in, risus. In eu eros. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Curabitur vulputate elit viverra augue. Mauris fringilla, tortor sit amet malesuada mollis, sapien mi dapibus odio, ac imperdiet ligula enim eget nisl. Quisque vitae pede a pede aliquet suscipit. Phasellus tellus pede, viverra vestibulum, gravida id, laoreet in, justo. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer commodo luctus lectus. Mauris justo. Duis varius eros. Sed quam. Cras lacus eros, rutrum eget, varius quis, convallis iaculis, velit. Mauris imperdiet, metus at tristique venenatis, purus neque pellentesque mauris, a ultrices elit lacus nec tortor. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent malesuada. Nam lacus lectus, auctor sit amet, malesuada vel, elementum eget, metus. Duis neque pede, facilisis eget, egestas elementum, nonummy id, neque.

Sub-item 2

Proin non sem. Donec nec erat. Proin libero. Aliquam viverra arcu. Donec vitae purus. Donec felis mi, semper id, scelerisque porta, sollicitudin sed, turpis. Nulla in urna. Integer varius wisi non elit. Etiam nec sem. Mauris consequat, risus nec congue condimentum, ligula ligula suscipit urna, vitae porta odio erat quis sapien. Proin luctus leo id erat. Etiam massa metus, accumsan pellentesque, sagittis sit amet, venenatis nec, mauris. Praesent urna eros, ornare nec, vulputate eget, cursus sed, justo. Phasellus nec lorem. Nullam ligula ligula, mollis sit amet, faucibus vel, eleifend ac, dui. Aliquam erat volutpat.

4.1.2 Item 2

Fusce vehicula, tortor et gravida porttitor, metus nibh congue lorem, ut tempus purus mauris a pede. Integer tincidunt orci sit amet turpis. Aenean a metus. Aliquam vestibulum lobortis felis. Donec gravida. Sed sed urna. Mauris et orci. Integer ultrices feugiat ligula. Sed dignissim nibh a massa. Donec orci dui, tempor sed, tincidunt nonummy, viverra sit amet, turpis. Quisque lobortis. Proin venenatis tortor nec wisi. Vestibulum placerat. In hac habitasse platea dictumst. Aliquam porta mi quis risus. Donec sagittis luctus diam. Nam ipsum elit, imperdiet vitae, faucibus nec, fringilla eget, leo. Etiam quis dolor in sapien porttitor imperdiet.

4.2 The Second Topic

Cras pretium. Nulla malesuada ipsum ut libero. Suspendisse gravida hendrerit tellus. Maecenas quis lacus. Morbi fringilla. Vestibulum odio turpis, tempor vitae, scelerisque a, dictum non, massa. Praesent erat felis, porta sit amet, condimentum sit amet, placerat et, turpis. Praesent placerat lacus a enim. Vestibulum non eros. Ut congue. Donec tristique varius tortor. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nam dictum dictum urna.
Phasellus vestibulum orci vel mauris. Fusce quam leo, adipiscing ac, pulvinar eget, molestie sit amet, erat. Sed diam. Suspendisse eros leo, tempus eget, dapibus sit amet, tempus eu, arcu. Vestibulum wisi metus, dapibus vel, luctus sit amet, condimentum quis, leo. Suspendisse molestie. Duis in ante. Ut sodales sem sit amet mauris. Suspendisse ornare pretium orci. Fusce tristique enim eget mi. Vestibulum eros elit, gravida ac, pharetra sed, lobortis in, massa. Proin at dolor. Duis accumsan accumsan pede. Nullam blandit elit in magna lacinia hendrerit. Ut nonummy luctus eros. Fusce eget tortor.
Ut sit amet magna. Cras a ligula eu urna dignissim viverra. Nullam tempor leo porta ipsum. Praesent purus. Nullam consequat. Mauris dictum sagittis dui. Vestibulum sollicitudin consectetuer wisi. In sit amet diam. Nullam malesuada pharetra risus. Proin lacus arcu, eleifend sed, vehicula at, congue sit amet, sem. Sed sagittis pede a nisl. Sed tincidunt odio a pede. Sed dui. Nam eu enim. Aliquam sagittis lacus eget libero. Pellentesque diam sem, sagittis molestie, tristique et, fermentum ornare, nibh. Nulla et tellus non felis imperdiet mattis. Aliquam erat volutpat.

4.3 Conclusion

Vestibulum sodales ipsum id augue. Integer ipsum pede, convallis sit amet, tristique vitae, tempor ut, nunc. Nam non ligula non lorem convallis hendrerit. Maecenas hendrerit. Sed magna odio, aliquam imperdiet, porta ac, aliquet eget, mi. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus.

Vestibulum nisl sem, dignissim vel, euismod quis, egestas ut, orci. Nunc vitae risus vel metus euismod laoreet. Cras sit amet neque a turpis lobortis auctor. Sed aliquam sem ac elit. Cras velit lectus, facilisis id, dictum sed, porta rutrum, nisl. Nam hendrerit ipsum sed augue. Nullam scelerisque hendrerit wisi. Vivamus egestas arcu sed purus. Ut ornare lectus sed eros. Suspendisse potenti. Mauris sollicitudin pede vel velit. In hac habitasse platea dictumst.

Suspendisse erat mauris, nonummy eget, pretium eget, consequat vel, justo. Pellentesque consectetuer erat sed lacus. Nullam egestas nulla ac dui. Donec cursus rhoncus ipsum. Nunc et sem eu magna egestas malesuada. Vivamus dictum massa at dolor. Morbi est nulla, faucibus ac, posuere in, interdum ut, sapien. Proin consectetuer pretium urna. Donec sit amet nibh nec purus dignissim mattis. Phasellus vehicula elit at lacus. Nulla facilisi. Cras ut arcu. Sed consectetuer. Integer tristique elit quis felis consectetuer eleifend. Cras et lectus.

Ut congue malesuada justo. Curabitur congue, felis at hendrerit faucibus, mauris lacus porttitor pede, nec aliquam turpis diam feugiat arcu. Nullam rhoncus ipsum at risus. Vestibulum a dolor sed dolor fermentum vulputate. Sed nec ipsum dapibus urna bibendum lobortis. Vestibulum elit. Nam ligula arcu, volutpat eget, lacinia eu, lobortis ac, urna. Nam mollis ultrices nulla. Cras vulputate. Suspendisse at risus at metus pulvinar malesuada. Nullam lacus. Aliquam tempus magna. Aliquam ut purus. Proin tellus.

Chapter 5

Forecasting of time-series

Typical variables used in a forecasting model are: past electricity consumption loads, weather information, calendar information and error-correction terms [1].

5.1 The First Topic of this Chapter

5.1.1 Item 1

5.2 Conclusion

Chapter 6

Evaluating results

Morbi malesuada hendrerit dui. Nunc mauris leo, dapibus sit amet, vestibulum et, commodo id, est. Pellentesque purus. Pellentesque tristique, nunc ac pulvinar adipiscing, justo eros consequat lectus, sit amet posuere lectus neque vel augue. Cras consectetuer libero ac eros. Ut eget massa. Fusce sit amet enim eleifend sem dictum auctor. In eget risus luctus wisi convallis pulvinar. Vivamus sapien risus, tempor in, viverra in, aliquet pellentesque, eros. Aliquam euismod libero a sem.

6.1 The First Topic of this Chapter

6.1.1 Item 1

Sub-item 1

Nunc velit augue, scelerisque dignissim, lobortis et, aliquam in, risus. In eu eros. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Curabitur vulputate elit viverra augue. Mauris fringilla, tortor sit amet malesuada mollis, sapien mi dapibus odio, ac imperdiet ligula enim eget nisl. Quisque vitae pede a pede aliquet suscipit. Phasellus tellus pede, viverra vestibulum, gravida id, laoreet in, justo. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer commodo luctus lectus. Mauris justo. Duis varius eros. Sed quam. Cras lacus eros, rutrum eget, varius quis, convallis iaculis, velit. Mauris imperdiet, metus at tristique venenatis, purus neque pellentesque mauris, a ultrices elit lacus nec tortor. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent malesuada. Nam lacus lectus, auctor sit amet, malesuada vel, elementum eget, metus. Duis neque pede, facilisis eget, egestas elementum, nonummy id, neque.

Sub-item 2

Proin non sem. Donec nec erat. Proin libero. Aliquam viverra arcu. Donec vitae purus. Donec felis mi, semper id, scelerisque porta, sollicitudin sed, turpis. Nulla in urna. Integer varius wisi non elit. Etiam nec sem. Mauris consequat, risus nec congue condimentum, ligula ligula suscipit urna, vitae porta odio erat quis sapien. Proin luctus leo id erat. Etiam massa metus, accumsan pellentesque, sagittis sit amet, venenatis nec, mauris. Praesent urna eros, ornare nec, vulputate eget, cursus sed, justo. Phasellus nec lorem. Nullam ligula ligula, mollis sit amet, faucibus vel, eleifend ac, dui. Aliquam erat volutpat.

6.1.2 Item 2

Fusce vehicula, tortor et gravida porttitor, metus nibh congue lorem, ut tempus purus mauris a pede. Integer tincidunt orci sit amet turpis. Aenean a metus. Aliquam vestibulum lobortis felis. Donec gravida. Sed sed urna. Mauris et orci. Integer ultrices feugiat ligula. Sed dignissim nibh a massa. Donec orci dui, tempor sed, tincidunt nonummy, viverra sit amet, turpis. Quisque lobortis. Proin venenatis tortor nec wisi. Vestibulum placerat. In hac habitasse platea dictumst. Aliquam porta mi quis risus. Donec sagittis luctus diam. Nam ipsum elit, imperdiet vitae, faucibus nec, fringilla eget, leo. Etiam quis dolor in sapien porttitor imperdiet.

6.2 The Second Topic

Cras pretium. Nulla malesuada ipsum ut libero. Suspendisse gravida hendrerit tellus. Maecenas quis lacus. Morbi fringilla. Vestibulum odio turpis, tempor vitae, scelerisque a, dictum non, massa. Praesent erat felis, porta sit amet, condimentum sit amet, placerat et, turpis. Praesent placerat lacus a enim. Vestibulum non eros. Ut congue. Donec tristique varius tortor. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nam dictum dictum urna.

Phasellus vestibulum orci vel mauris. Fusce quam leo, adipiscing ac, pulvinar eget, molestie sit amet, erat. Sed diam. Suspendisse eros leo, tempus eget, dapibus sit amet, tempus eu, arcu. Vestibulum wisi metus, dapibus vel, luctus sit amet, condimentum quis, leo. Suspendisse molestie. Duis in ante. Ut sodales sem sit amet mauris. Suspendisse ornare pretium orci. Fusce tristique enim eget mi. Vestibulum eros elit, gravida ac, pharetra sed, lobortis in, massa. Proin at dolor. Duis accumsan accumsan pede. Nullam blandit elit in magna lacinia hendrerit. Ut nonummy luctus eros. Fusce eget tortor.

Ut sit amet magna. Cras a ligula eu urna dignissim viverra. Nullam tempor leo porta ipsum. Praesent purus. Nullam consequat. Mauris dictum sagittis dui. Vestibulum sollicitudin consectetuer wisi. In sit amet diam. Nullam malesuada pharetra risus. Proin lacus arcu, eleifend sed, vehicula at, congue sit amet, sem. Sed sagittis pede a nisl. Sed tincidunt odio a pede. Sed dui. Nam eu enim. Aliquam sagittis lacus eget libero. Pellentesque diam sem, sagittis molestie, tristique et, fermentum ornare, nibh. Nulla et tellus non felis imperdiet mattis. Aliquam erat volutpat.

6.3 Conclusion

Vestibulum sodales ipsum id augue. Integer ipsum pede, convallis sit amet, tristique vitae, tempor ut, nunc. Nam non ligula non lorem convallis hendrerit. Maecenas hendrerit. Sed magna odio, aliquam imperdiet, porta ac, aliquet eget, mi. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Vestibulum nisl sem, dignissim vel, euismod quis, egestas ut, orci. Nunc vitae risus vel metus euismod laoreet. Cras sit amet neque a turpis lobortis auctor. Sed aliquam sem ac elit. Cras velit lectus, facilisis id, dictum sed, porta rutrum, nisl. Nam hendrerit ipsum sed augue. Nullam scelerisque hendrerit wisi. Vivamus egestas arcu sed purus. Ut ornare lectus sed eros. Suspendisse potenti. Mauris sollicitudin pede vel velit. In hac habitasse platea dictumst.

Suspendisse erat mauris, nonummy eget, pretium eget, consequat vel, justo. Pellentesque consectetuer erat sed lacus. Nullam egestas nulla ac dui. Donec cursus rhoncus ipsum. Nunc et sem eu magna egestas malesuada. Vivamus dictum massa at dolor. Morbi est nulla, faucibus ac, posuere in, interdum ut, sapien. Proin consectetuer pretium urna. Donec sit amet nibh nec purus dignissim mattis. Phasellus vehicula elit at lacus. Nulla facilisi. Cras ut arcu. Sed consectetuer. Integer tristique elit quis felis consectetuer eleifend. Cras et lectus.

Ut congue malesuada justo. Curabitur congue, felis at hendrerit faucibus, mauris lacus porttitor pede, nec aliquam turpis diam feugiat arcu. Nullam rhoncus ipsum at risus. Vestibulum a dolor sed dolor fermentum vulputate. Sed nec ipsum dapibus urna bibendum lobortis. Vestibulum elit. Nam ligula arcu, volutpat eget, lacinia eu, lobortis ac, urna. Nam mollis ultrices nulla. Cras vulputate. Suspendisse at risus at metus pulvinar malesuada. Nullam lacus. Aliquam tempus magna. Aliquam ut purus. Proin tellus.

Chapter 7

Conclusion

The final chapter contains the overall conclusion. It also contains suggestions for future work and industrial applications.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Appendices

Appendix A

Introduction to the dataset

Appendices hold useful data which is not essential to understand the work done in the master's thesis. An example is a (program) source. An appendix can also have sections as well as figures and references[**?**].

A.1 Introduction to the dataset

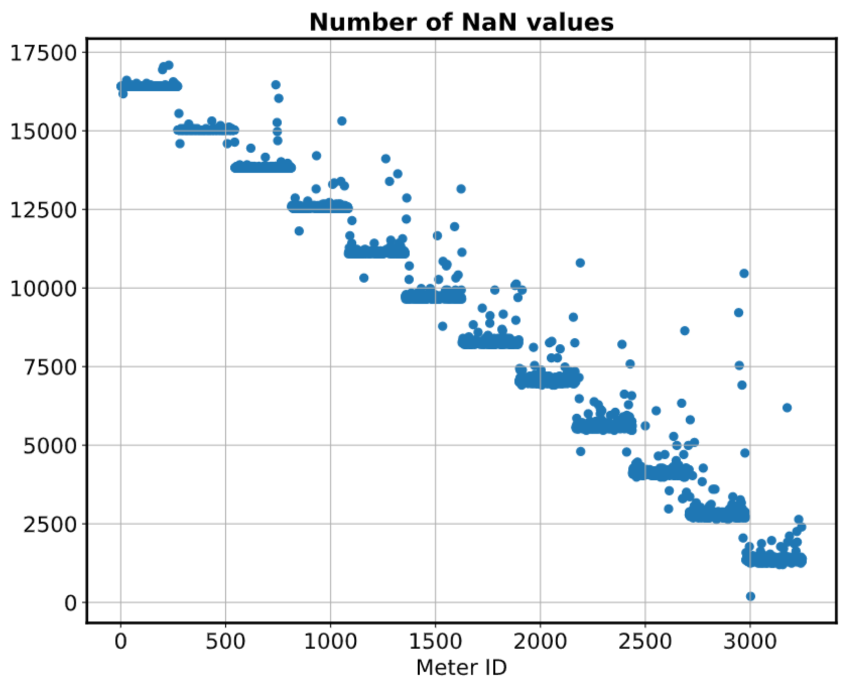

FIGURE A.1: The amount of NaN values in all the 3248 smart meters.

Attribute	**Filled places**
Dwelling type (5 cat.)	1702
# Occupants (max 4)	74
# Bedrooms (max 5)	1859
Heating fuel (4 cat.)	78
Hot water fuel (3 cat.)	76
Boiler age (2 cat.)	74
Loft insulation (2 cat.)	75
Wall insulation (5 cat.)	75
Heating temperature (4 cat.)	74
Efficient lighting percentage (4 cat.)	73
Dishwasher (0,1,2)	76
Freezer (0,1,2)	70
Fridge freezer (0,1,2)	70
Refrigerator (0,1,2)	73
Tumble Dryer (0,1,2)	76
Washing machine (0,1,2)	76
Game console (0,1,2,3)	72
Laptop (0,1,2,3,4)	70
Pc (0,1,2,3)	70
Router (0,1,2)	69
Set top box (0,1,2,3)	70
Tablet (0,1,2,3,4)	70
Tv (0,1,2,3,4)	75

Table A.1: Amount of response on the voluntary questionnaires.

A.2 Missing values

A.2.1 Fundamental change

A.3 Daily filter

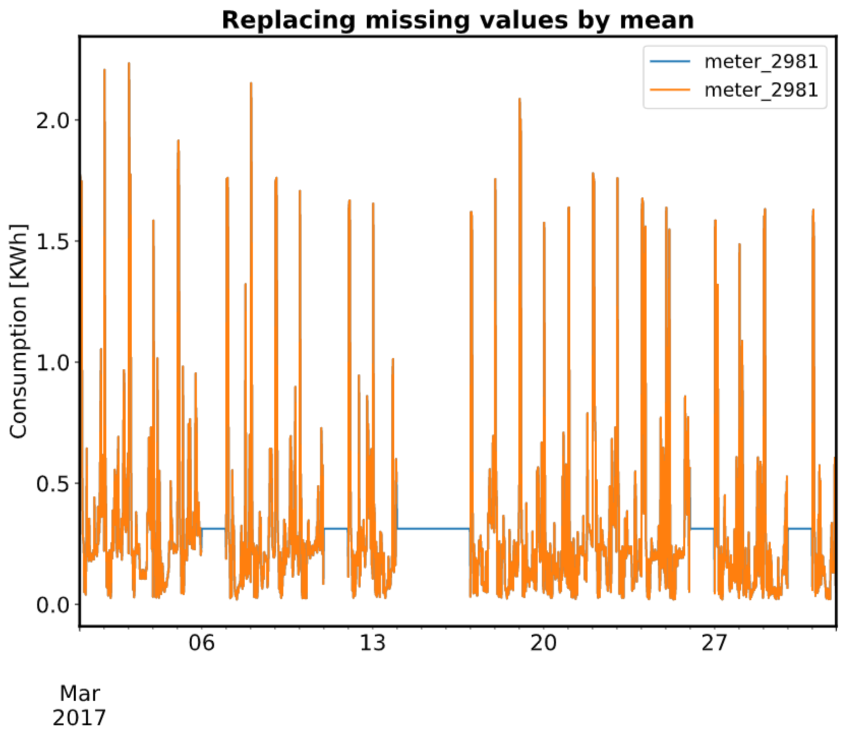

FIGURE A.2: Resulting month of March after substitution of the missing values by the mean value of the measurements.

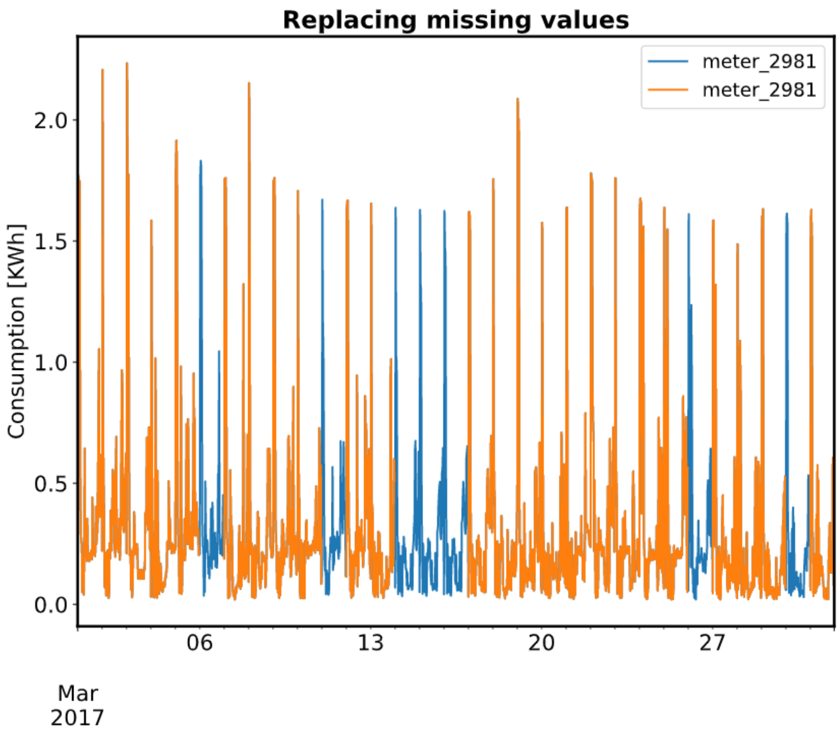

Figure A.3: Resulting month of March after substitution of the missing values by the mean value of the same moment on the next and previous day.

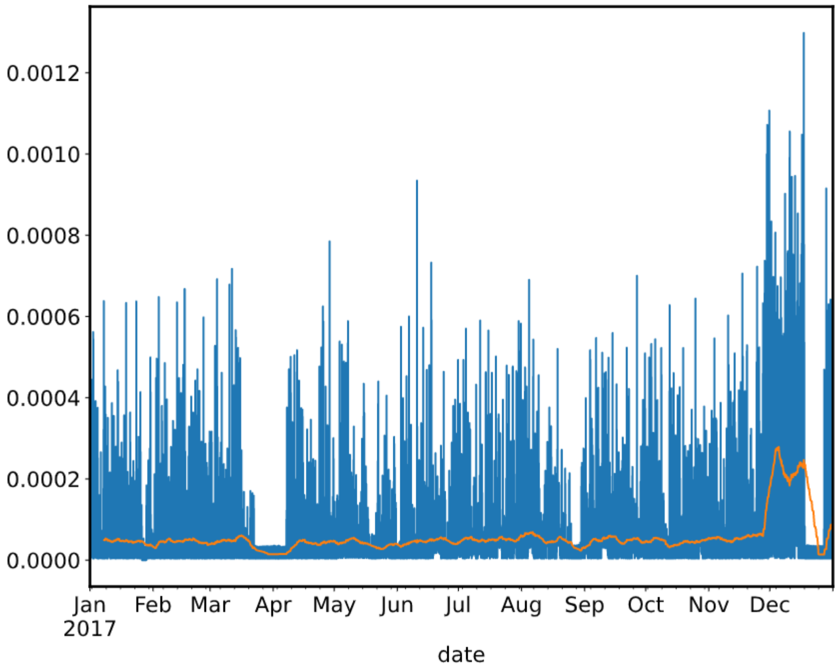

FIGURE A.4: The time-serie with the original maximum difference between the minimum and maximum weekly rolling averages.

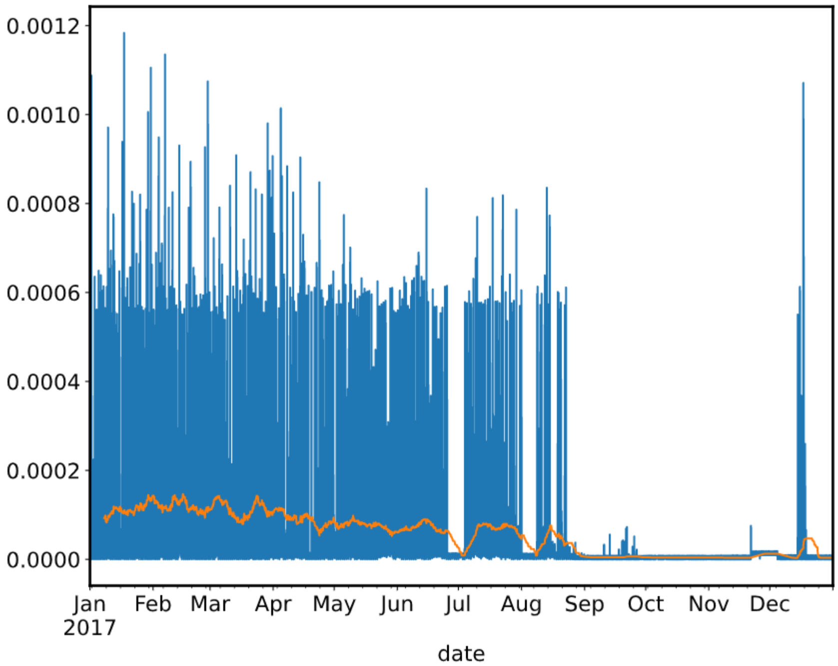

FIGURE A.5: The time-serie with the new maximum difference between the minimum and maximum weekly rolling averages.

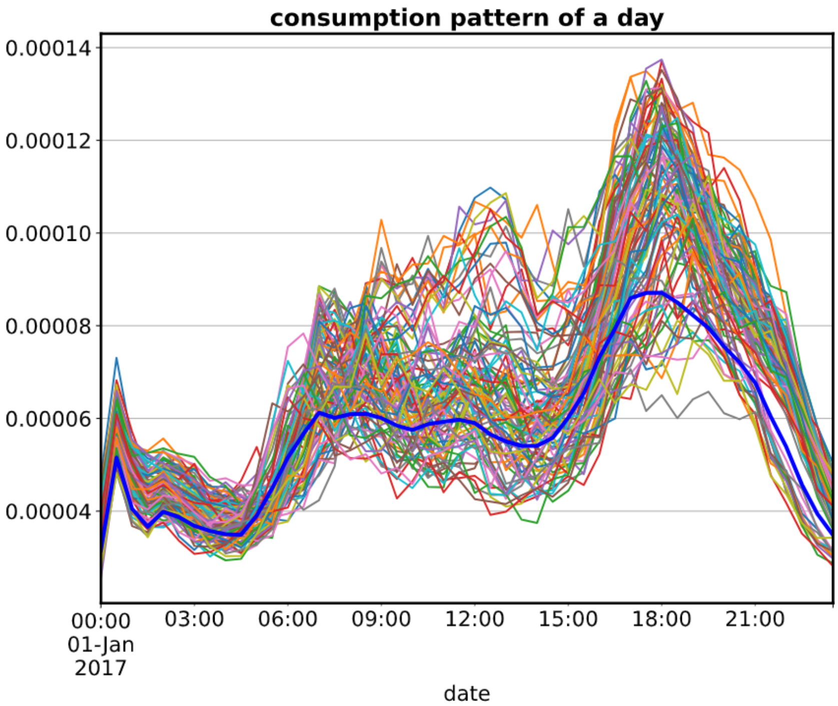

Figure A.6: Figure that shows the seasonality of the electrical load during the day.

Appendix B

Old things

B.0.1 Removing outliers

After the missing values are replaced by estimations, the outliers of the electricity consumption signals are identified. This is done by looking at the z-scores of the yearly consumptions. A z-score is calculated using equation **??** and assumes that the yearly consumptions are normally distributed around the average consumption. Consumptions that have a very low probability to occur are removed by imposing that $|z - score| < 3$.

$$z - score = \frac{x - \mu}{\sigma} \tag{B.1}$$

Figure B.1 gives the obtained z-values. It can be seen that 6 meters with an unlikely high or low consumption are removed.

B.0.2 Normalization of the data

Normalization is necessary because while absolute consumption differs, relative patterns of human behaviour are more similar [3]. The patterns in the human behaviour is what a forecasting model is trying to predict and normalization contributes by avoiding the disturbance of different magnitudes in which this human pattern may occur. Every individual household time-serie is normalized based on its maximum and minimum value according to equation B.2.

$$normalizedvalue = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{B.2}$$

As discussed in section **??** the average is taken over all the normalized time-series to obtain a single signal.**Ask if this is good??** Because the maximum is taken into account during the normalization, measurement out shooters have an influence on the normalization.

B.1 ARIMA

What is ARIMA. Assumptions of ARIMA...

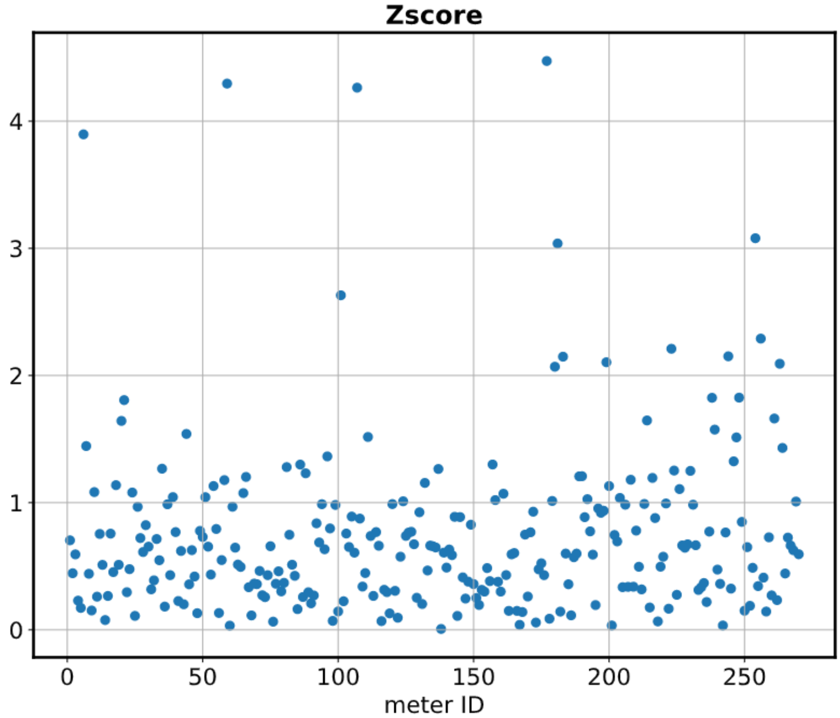

Figure B.1: Z-scores calculated from the yearly consumptions.

Stationarity
https://machinelearningmastery.com/remove-trends-seasonality-difference-transform-python/ When data is modelled it is assumed that the statistics of the data are consistent or stationary. This means the mean and standard deviation is not changing in time. However, because time series are often subdued to a trend or seasonality this assumption of stationarity is violated. In order to model not stationary observations by a stationary model as ARIMA, trends and seasonal effects should be removed. A way to check the stationarity of your observations, the “Dicky-Fuller test” can be used. A way to remove non-stationarity is by using “Difference Transform”. Here the trend and seasonality is subtracted from the observations leaving behind a stationary dataset.

Bibliography

[1] M. Espinoza, J. Suykens, R. Belmans, and B. De Moor. Electric Load Forecasting. *IEEE control systems magazine*, (October 2007):43–57, 2007.

[2] B. A. Hoverstad, A. Tidemann, H. Langseth, and P. Ozturk. Short-Term Load Forecasting With Seasonal Decomposition Using Evolution for Parameter Tuning. *IEEE Transactions on Smart Grid*, 6(4):1904–1913, 2015.

[3] J. Lago. A ratios and clustering based approach to forecast electricity consumption, 2020.

[4] M. A. Nielsen. Neural Networks and Deep Learning, 2015.

[5] H. Shi, M. Xu, and R. Li. Deep Learning for Household Load Forecasting-A Novel Pooling Deep RNN. *IEEE Transactions on Smart Grid*, 9(5):5271–5280, 2018.