KU LEUVEN

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

It should be made clear that this section about missing values is only applied during the data analysis. 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.

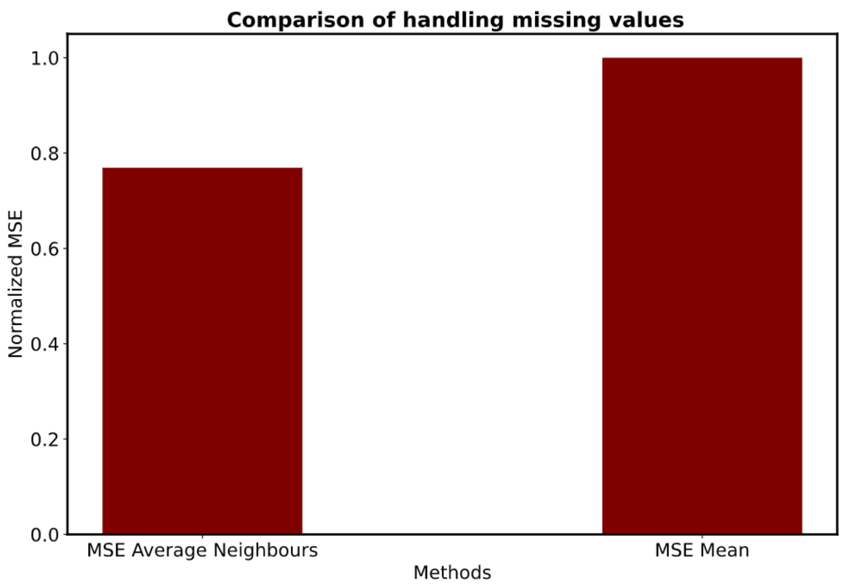


FIGURE 2.1: Resulting month of March after substitution of the missing values by the mean value of the measurements.

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.

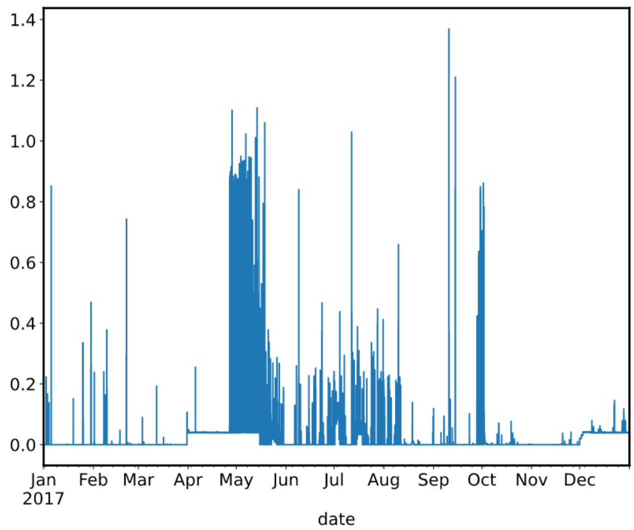


FIGURE 2.2: One of the 9 identified meters with multiple zero daily consumptions

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

it should be made clear that this normalization is only here used. Normalization is necessary because while absolute consumption differs, relative patterns of human behaviour are more similar [5]. 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 [5]. 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

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. Figure 2.3 shows all the maximum differences between the minimum and maximum weekly rolling averages. The red line on shows the cutoff and the smart meters above this line are defined as outliers and removed. The definition of an outlier that is used is the one and a half times the interquartile range. In total 256 smart meters remain.

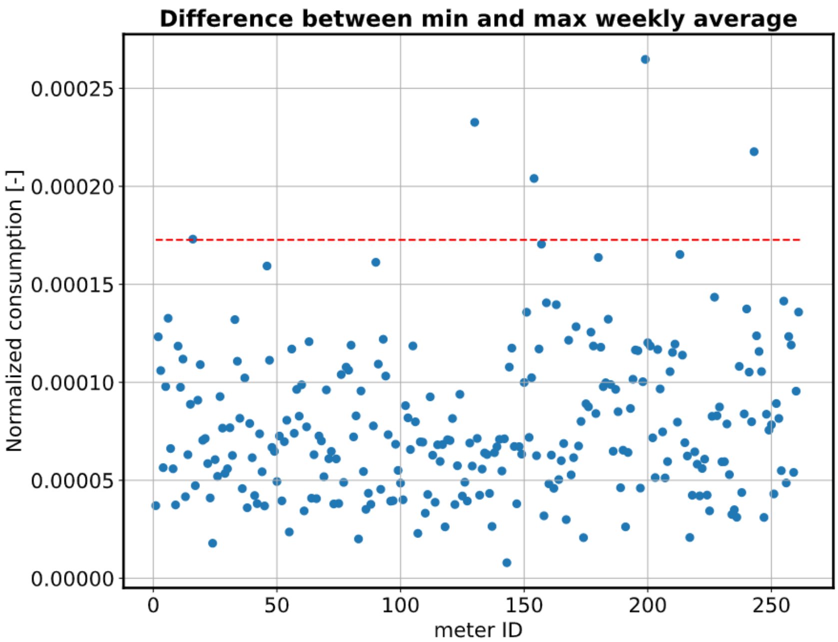


FIGURE 2.3: The maximum differences between the minimum and maximum weekly rolling averages for all the different time-series.

2.3 Analysis

Finally, the average is taken over all the remaining 256 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 from the template consumption instead of performing a forecast out of the blue. However in [2] they made forecasts of an aggregated signal which has a reasonably amount of regularity which is not the case for electrical consumption of individual households. These templates or filters 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.

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

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

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 Comparing weekdays with weekends

Weekdays vs weekends can be compared with the help of Figure 2.4. The reader is reminded that in order to get this graph, all the remaining household loads after preprocessing are averaged after which all the weeks are again averaged using equation 2.3. It can be seen that the consumption of the average business day is similar to a weekend day concerning the two main peaks during the day (7 am and 6 pm) and the sharp peak at midnight. However, it can be seen that the first peak during the day is higher and goes less down again during the weekend. This effect can be seen both during a Sunday and Saterday, but is most visible during a sunday. To proof previous statements similarity is measured by calculating the hourly difference of the 21 combinations that can be made of two different days. Figure **??** shows in blue

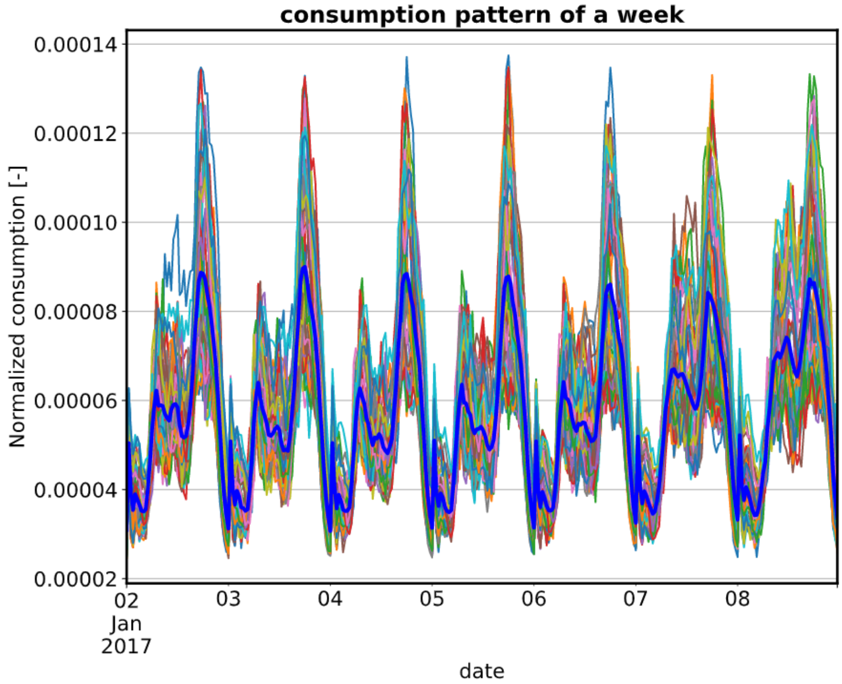


FIGURE 2.4: The seasonality of the electrical load during the week. The blue line shows the average week over all weeks in 2017.

and orange the error of combinations between business days or weekend days and in green the error of combinations between a business day and weekend day. The error value is calculated by summing the hourly errors between two days. It can be clearly seen that when a business day and weekend day are combined the error (green) is bigger and thus similarity smaller. Another thing that can be noticed is that the left cluster of green dots corresponds to a Saterday and the right to a Sunday. It can be noticed that Saterdays are more similar to a business day than a Sunday.

2.3.3 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 mitigate the temperature dependency. The resulting average holiday and business day is given in Figure 2.6. A holiday behaves similar to a weekend day with the first peak load going higher and goes less down over time. Figure 2.7 shows that a holiday behaves the most similar to a Sunday .

It can be seen that the consumption of a holiday behaves similar as a weekend day. Figure shows the average error between a holiday vs business day and a holiday vs weekend day. The error is calculated as discussed in section 2.3.2.

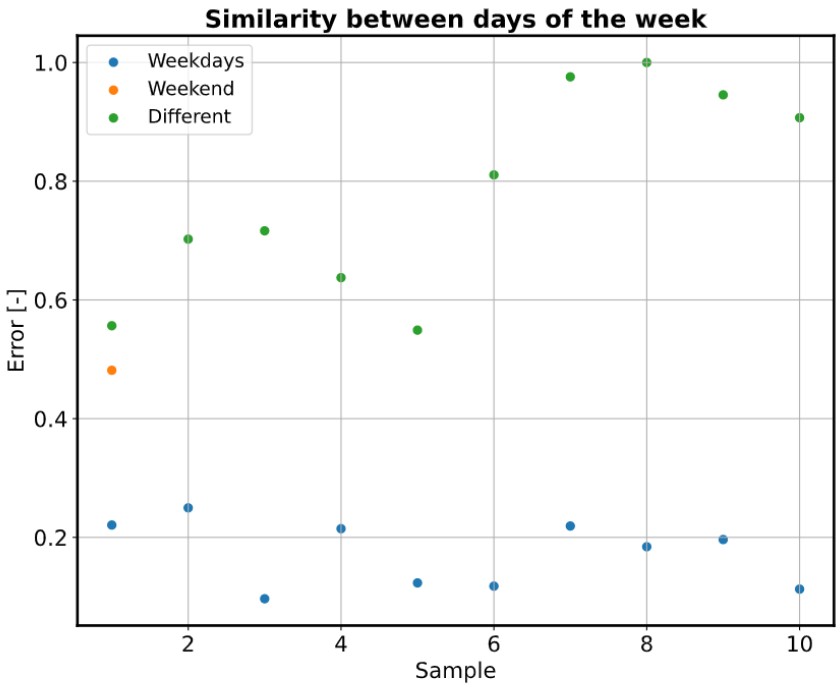


FIGURE 2.5: Error between different pairs of weekdays.

2.3.4 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 value 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 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 drawn, come in pairs, follow homoscedasticity and there are no outliers. Outliers are especially undesirable when there are not a lot of samples. The variables should be normal distributed, linear related to each other and be continuous.

The samples used for the correlation are generated by calculating the daily consumptions matched with the daily average temperature. In this case the above assumptions

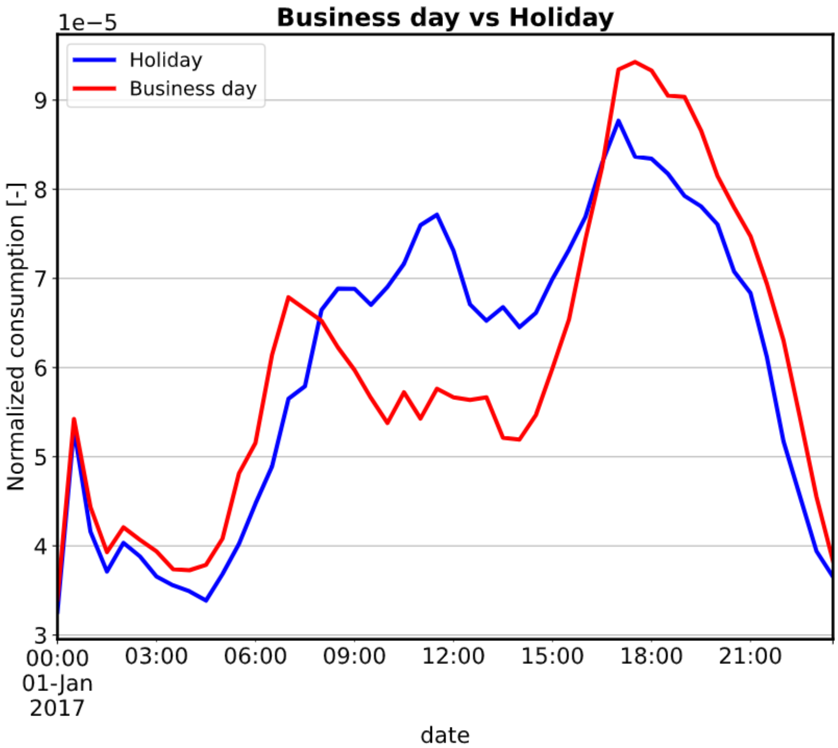


FIGURE 2.6: Figure with the comparison between holidays and business days.

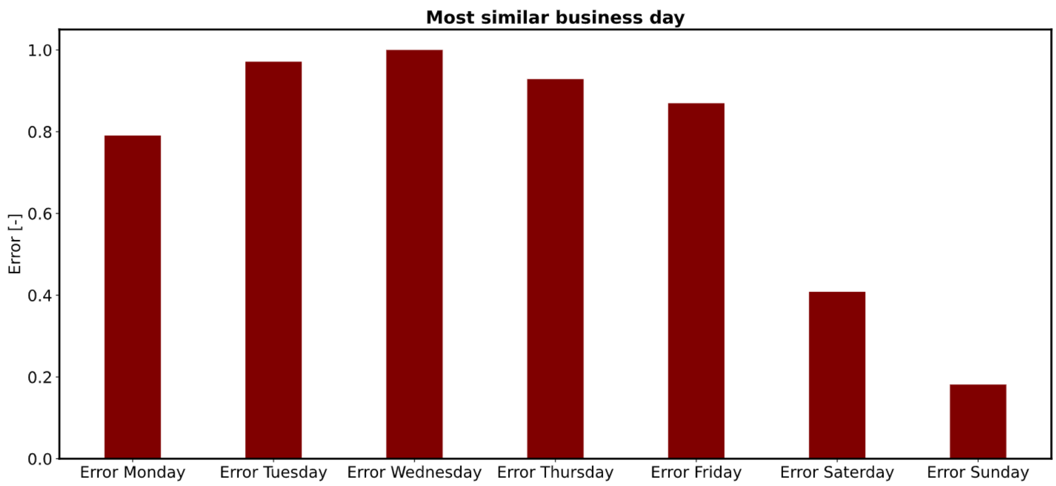


FIGURE 2.7: Error between a holiday and other days of the week.

are thus not valid. Homoscedasticity is important when performing linear regression and assumes that σ_x and σ_y are constant. This assumption is validated by making use of Figure 2.8.

This figure shows the classic cone-shaped pattern of heteroscedasticity. On days

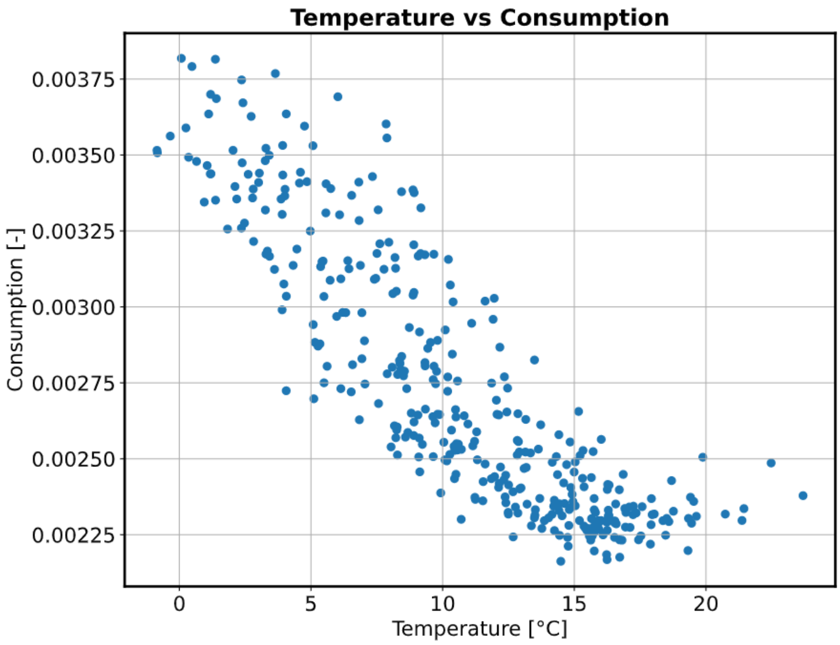


Figure 2.8: Relation between normalized daily consumption and daily temperature.

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, which means that homoscedasticity is not fulfilled. Because the assumptions of the Pearson correlation are not fulfilled, care should be taken with its output.

Applying the Pearson correlation on Figure 2.8 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 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 good negative monotone relation. This means if the temperature is higher, consumption is likely to be lower. Identically, if the temperature is lower it is likely that the consumption will be higher.

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.

2.3.5 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 2.9 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 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 2.9 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

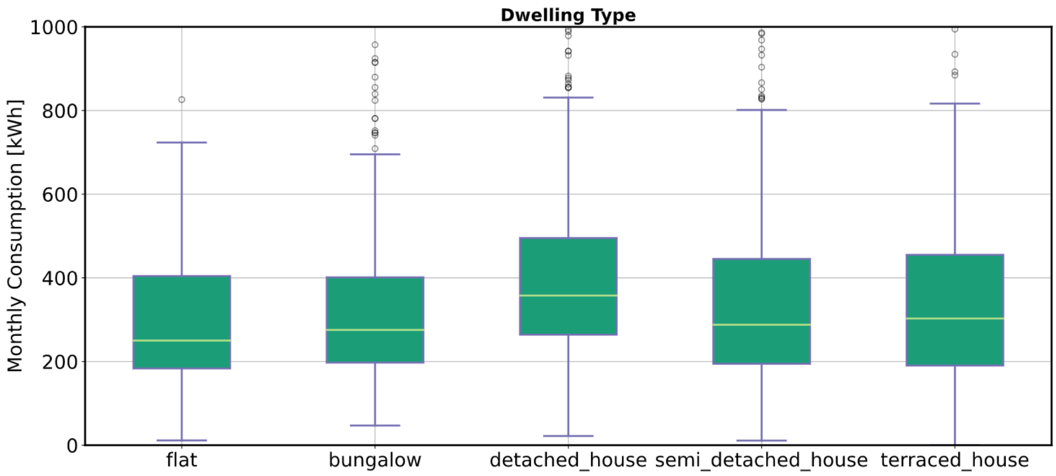


FIGURE 2.9: Figure with the comparison of the different dwelling types.

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. **Add concrete numbers!!**

- 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 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" are used. **See also paper TA2 –> aggregated vs individual forecasting.**

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 [6] 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 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 [8] 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

The simples configuration of deep networks are multilayer perceptrons and they are made up out of multiple fully connected layers of neurons. Figure 3.1 shows a MLP with one hidden layer.

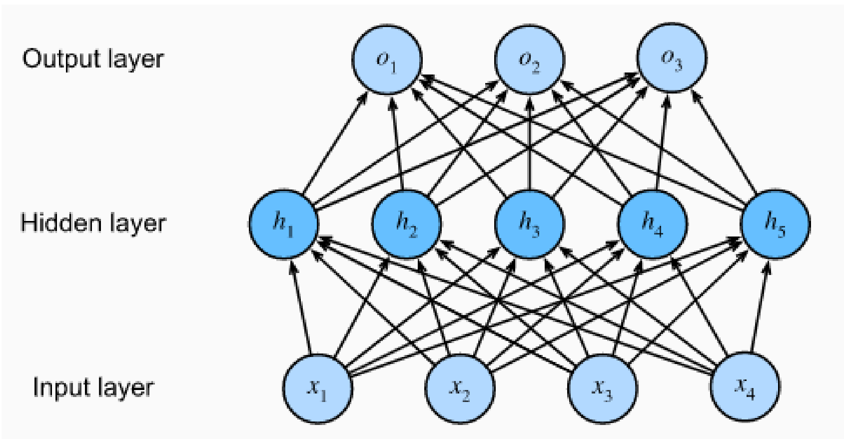


Figure 3.1: Figure of a MLP (source [10]).

All layers are connected to the next layer by the means of an affine function together with a non-linear activation function represented by sigma as shown by equation 3.1 with $\mathbf{L}^{(N)}$ the vector with outputs of the Nthe layer, $\mathbf{W}^(N)$ the Nth weight matrix and $\mathbf{b}^{(N)}$ the Nth bias.

$$\mathbf{L}^{N+1} = \sigma(\mathbf{W}^{(N)}\mathbf{L}^{N} + \mathbf{b}^{(N)}) \tag{3.1}$$

3.1.2 CNN

See oneNote

3.1.3 RNN

A recurrent Neural Network is a specialized neural network to better deal with sequential information. While traditional deep neural networks assume that inputs and outputs are independent of each other, the output of recurrent neural networks depend on the prior elements within the sequence. In order to take past information from previous inputs into account, a hidden variable h_t is used. By making use of this variable which makes a summary of the previous seen information, an exponential increase in the number of model parameters is avoided. Equation 3.2 shows how the previous hidden state and the current information are merged in the next hidden state with $\mathbf{X}^t \in \mathbb{R}^{d\times 1}$, $\mathbf{H}^t \in \mathbb{R}^{h\times 1}$, $\mathbf{W}_1 \in \mathbb{R}^{h\times d}$, $\mathbf{W}_2 \in \mathbb{R}^{h\times h}$ and $\mathbf{b} \in \mathbb{R}^{h\times 1}$.

$$\mathbf{H}^{t+1} = \sigma(\mathbf{W}_1\mathbf{X}^t + \mathbf{W}_2\mathbf{H}^t + \mathbf{b}) \tag{3.2}$$

The equation $\mathbf{X}^t$ corresponds to one example at time step t with dimensionality d. Also a deep RNN is possible, where multiple hidden state per time step are used.

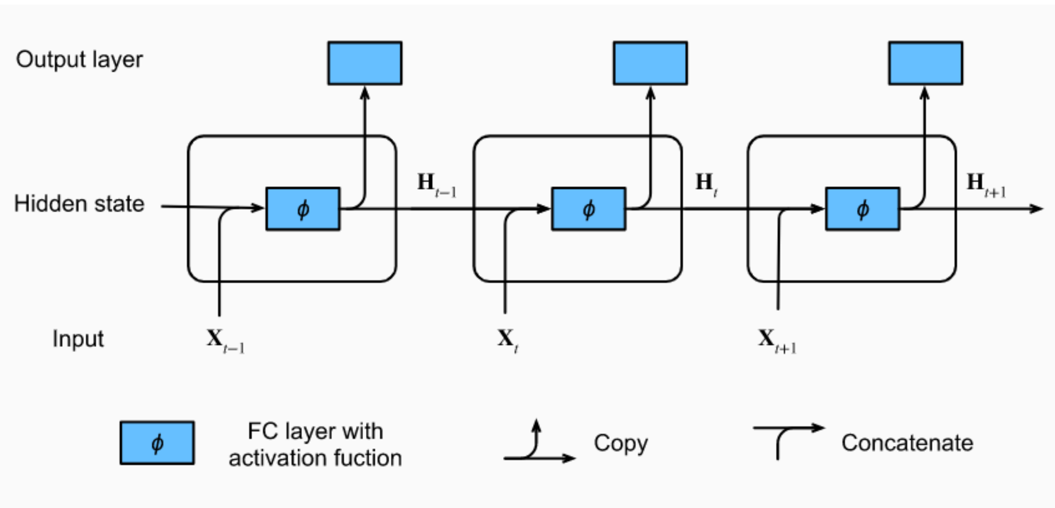


FIGURE 3.2: Figure of the logical flow of a vanilla RNN with a hidden state (source [10]).

3.1.4 Difficulties & Solutions of neural networks

Neural Networks have a high expressiveness but comes at the cost of overfitting and a vanishing gradient. when the NN is learning from training data, every epoch the error between the input and output of the training exemples is reduced. In the beginning the generalization error reduces simultaneously with the generalization error. The generalization error is the error that the model makes on data that is not in the training set. However, on a certain point during the training the generalization error increases while the training error still decreases. This means that the model is no longer learning “intelligent” general rules and patterns in the data, but is just remembering the training data and will therefore not apply in general. This is often the case in a model with high expressiveness because the model is less pushed to make generalizations and has the ability to just to remember the training data. Solutions to overfitting can be regularization which includes the parameter norms as a cost in the objective function. Typical choices for resembling the size of a parameter are the L_1 and the L_2 norms. Other methods that can be used are: early stopping, dropout and pruning.

It should also be noted that the gradient can increase very much, which in literature is called gradient explosion. The solution strategy for this is applying gradient clipping.

The second problem is the vanishing gradient problem which originates because while using the backpropagation algorithm to calculate the gradient which is used in different update methods of the weights, the gradient is calculated at the end of the NN and propagated back using every time the previous calculated gradient values who exponentially decrease. Therefore at the first layers of the network, the gradient has become so small that the weights are almost not updated anymore. In a RNN setting this corresponds to having a short term memory which means that initial inputs that were presented to the NN are being forgotten. Mitigation strategies often proposed in literature are LSTM and GRU. Both techniques have in common that they can learn which data in the sequence is important and should be retained and which information can be thrown away. It is important to state that LSTM and GRU are not solving the vanishing gradient problem as explained in [9]. The gradient is still exponentially decreasing, but the effect is less pronounced as can be seen for LSTM in Figure 3.3. τ gives the number of epochs.
Can put further explanation in attachment –> see assignament ANN

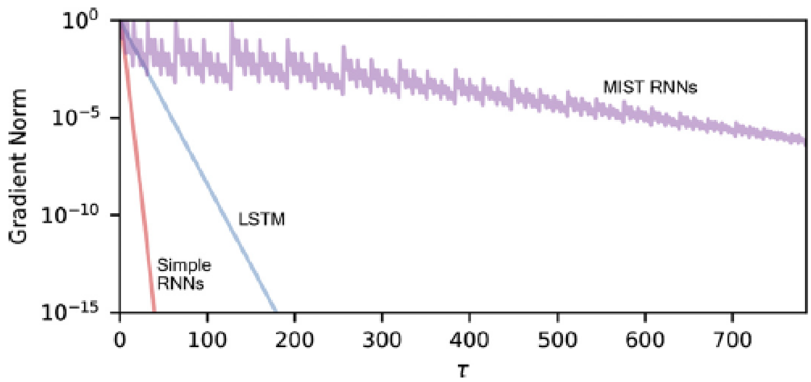


Figure 3.3: Exponential decrease of the gradient size of a simple RNN (red) or a LSTM (blue) (source [9]).

LSTM
Next, a LSTM (long short term memory) neural network will be assessed to model the Santa Fe data set instead of the MLP regression neural network that was used in the previous section . A LSTM has as advantage over a standard RNN that I can better handle the vanishing gradient problem and LSTM is not suffering as much of a short term memory. Therefore, it can longer take important aspects of the presented time-serie into account when working on a new forecasting. LSTM makes use of three gates to learn which data of the sequence is important and should be remembered. In connection of these gates is a cell state which serves as a memory state and a connection to transport relative information throughout the different time steps. The forget gate decides what information should be kept or removed from the input and hidden state. The input gate updates the cell state and the output state decides what the next hidden state will be.
In order to train a LSTM neural network there are more parameters that have to be learned. There are now four different weight matrices. One form the forget gate, two from the input gate and one from the output gate. It can be noted that in order to train all these weights, this network is more data hungry than a simple RNN. When the lag value gets bigger, so do all the different weight matrices and the calculation load.

Add here a picture of the the LSTM and the corresponding formula's!!

3.2 Short-Term residential electrical load forecasting

Pooling paper Classical ways to deal with uncertainty.
Residential electrical load series have a high amount of volatility and uncertainty due to the contingency of the electrical consumption. Classical ways to deal with this are discussed in [8] and listed as follows:

1. Clustering to group similar houses based on historic load or exogenous consumption driving variables. Because the load or driving variables are similar in a cluster, the variance of uncertainty is also decreased. However, performance is very dependent of the dataset. **But the uncertainty on the whole is reduced –> on single household stays the same!!**

2. Aggregating the residential loads to cancel out the uncertainties. The aggregated signal will show more regular patterns which means that is easier to predict.The downside is that the aggregated forecast will do a poor job of serving as forecast for a household

3. A spectral analysis e.g. wavelet analysis, Fourier transforms and empirical mode decomposition aim at seperating a load serie into a regular pattern, an uncertain signal and noise. Because the amount of regularity is low in a residential load serie, this method is infeasible.

In this paper [8] a novel pooling-based deep recurrent neural network is proposed which collects load profiles of neighbouring houses into a pool of training inputs. Pooling of neighbouring 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. The idea is as quoted by [8] to use the interconnected spacial information to compensate insufficient temporal information.Thereby, the pool of data allows to learn the correlations between neighbouring households and the shared uncertainties coming from external factors e.g. temperature. Also, due to the pooling of different households during training the DRNN is able to learn common uncertainties. In paper [8] pools consisting of 10 households are used. 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 a test set of the last 30 days and consisting out of :

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:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(\hat{y}_t - y_t)^2}{N}} \tag{3.3}$$

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{3.4}$$

$$MAE = \frac{\sum_{t=1}^{N}|\hat{y}_t - y_t|}{N} \tag{3.5}$$

Actually LSTM network The amount of which the PDRNN outperformed the other methods can be seen in Table **??**. The effect of the depth of the DRNN and the pooling method is depicted in Figure 3.5. 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 relations, 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.

Network Architecture	*RMSE (kWh)*	*NRMSE (kWh)*	*MAE (kWh)*
ARIMA	0.5593	0.1132	0.2998
RNN	0.5280	0.1076	0.2913
SVR	0.5180	0.1048	0.2855
DRNN	0.4815	0.0974	0.2698
PDRNN	0.4505	0.0912	0.2510
Improvement from DRNN to PDRNN	6.45%		6.96%
Improvement from ARIMA to PDRNN	19.46%		16.28%

FIGURE 3.4: Results obtained in paper [8] using the PDRNN method.

GRU (Gated Reset Update) or LSTM (Long Short Term Memory) can be implemented. They are both enhancements of the vanilla RNN which suffers from a vanishing gradient which causes it to to behave without a long term memory. In practise to know which one works often both are tried [9]. Stochastic gradient descent means that the approximated gradient is calculated from a random subset of the available data instead from the entire dataset.

Short-term Residential load forecasting based on LSTM RNN paper

In [4] it is chosen for a LSTM approach to forecast the complex temporal consumption pattern which characterises a single household electricity load. It is discussed that the diversity in the aggregated level of the individual electrical loads, smooths the daily load profile. This has as effect that the aggregated electrical load time-serie becomes more predictable, while a single household electrical load is more dependent on the human behaviour of its residents. This is substantiated by making use of a density based clustering technique where it was shown that the different daily

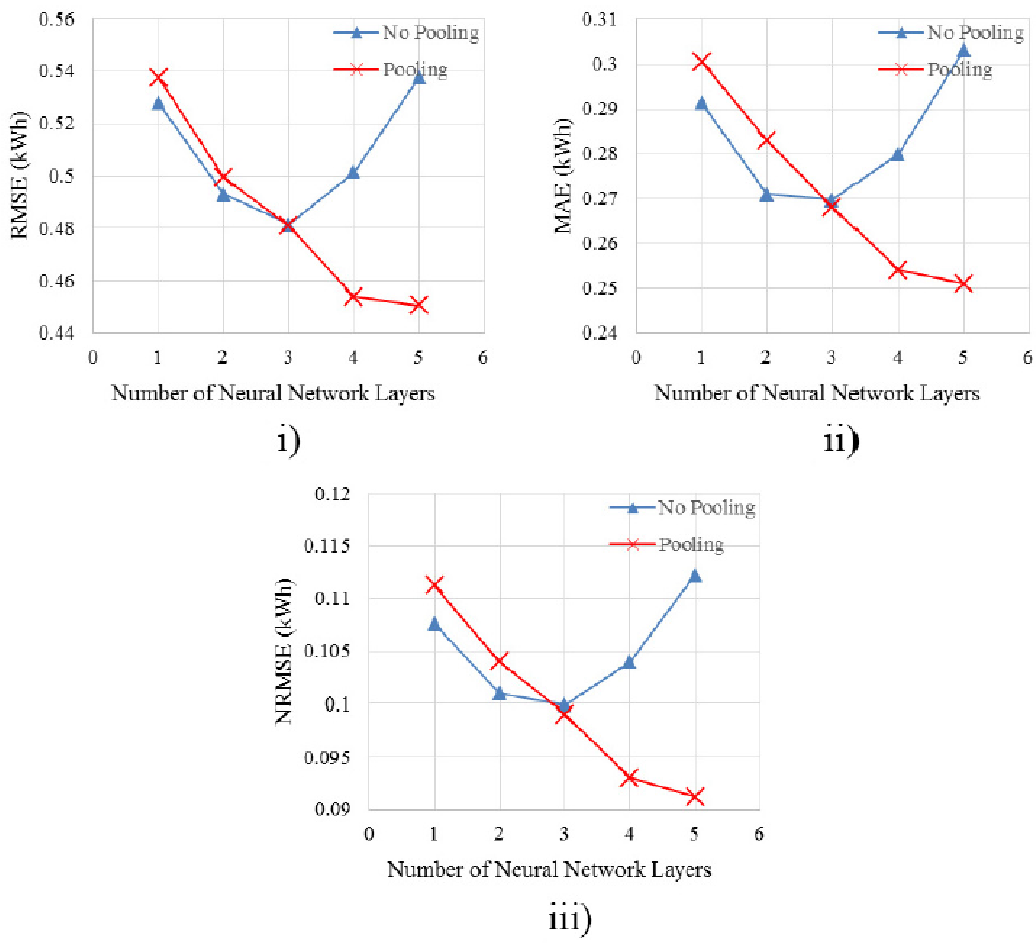


Figure 3.5: Influence of the number of layers and the pooling method used in [8].

consumptions of the aggregated signal could be described by one cluster and no outliers. An outlier means that a daily consumption could not be assigned to a cluster. On the other hand for individual time series the amount of outliers could range to over 80. To compare the consistency of different individual load signals the amount of outliers could therefore be used.
Because the residents daily routine is characterizing the household load so much, this is tried to be learned directly inside the LSTM RNN.
Inputs that are given to the LSTM are k past half hour load measurements, the time of when this measurements were taken, the day of the week of the measurements and if this day is a holiday or not. In table 3.6 the results are shown of the LSTM RNN method in comparison with other forecasting techniques. It can be noted that the proposed technique outperforms the rest based on the average performance of 29,808 individual forecasts of half an hour individual loads. Forecasting was performed on 69 different electrical loads coming from households in Australia. However, for individual load series forecasting the MAPE minimization is also remarkable when

considering its simplicity in comparison with LSTM. Next, it was concluded that learning methods that had good performance on aggregated time-series e.g. IS-HF and KNN, perform much worse when predicting individual loads.
Further, by making use of a regresion technique in function of the amount of outliers it is shown that LSTM and BPNN (Back-Propagation Neural Network) perform similar for, as previously discussed, consistent individual loads. The LSTM only starts to differentiate in performance when inconsistency grows. To conclude things that lack in [4] are practical useful forecasts of a timespan of 24h instead of only half an hour and making use of a rule of thumb when parameter tuning. Hyperparameters that can be tuned in LSTM are: learning rate, lag variable, amount of hidden layers and the amount of hidden nodes.

Method/Scenario	*Avg. MAPE individual forecasts*	*Avg. MAPE Aggregating forecasts*	*Avg. MAPE forecasting the aggregate*
LSTM/2 time steps	**44.39 %**	**8.18%**	**9.14%**
LSTM/6 time steps	**44.31%**	**8.39%**	**8.95%**
LSTM/12 time steps	**44.06%**	**8.64%**	**8.58%**
Empirical mean	136.46%	32.54%	32.54%
MAPE minimisation	**46.00%**	34.91%	27.28%
BPNN-D/1 day	80.02%	11.69%	14.50%
BPNN-D/2 days	75.28%	11.67%	14.48%
BPNN-D/3 days	74.10%	11.66%	14.42%
BPNN-T/2 time steps	49.62%	**8.37%**	9.54%
BPNN-T/6 time steps	49.04%	**8.29%**	9.55%
BPNN-T/12 time steps	49.49%	**8.36%**	9.17%
KNN/2 time steps	74.83%	15.37%	11.23%
KNN/6 time steps	71.19%	14.61%	12.10%
KNN/12 time steps	81.13%	15.23%	15.30%
ELM/2 time steps	122.90%	33.68%	Not tested
ELM/6 time steps	136.49%	35.35%	Not tested
ELM/12 time steps	123.45%	30.05%	Not tested
IS-HF	96.76%	20.43%	32.09%

FIGURE 3.6: Different approaches tried in [4] and their averaged performance of 29,808 individual forecasts of half an hour individual loads.

CNN-LSTM paper
In [3] a novel technique is proposed which makes use of a convolutional neural network from which the outputs are given to a LSTM recurrent network after which a fully connected neural network is used to produce the outputs. The purpose of the CNN is to extract the features that are the main drivers of energy consumption and to

remove the noise that comes initially together with the raw inputs. The CNN is made up out of convolution layers and pooling layers and makes use of the "ReLU" activation function. The main purpose of a convolution layer is to extract features while the pooling layer reduces the number of parameters by making use of the "max pooling principle". Using the "max pooling principle" means taking the max value of each neuron cluster of the previous layer. As discussed in paper [4] LSTM is suitable to alleviate the problem of a vanishing or exploding gradient which characterized a simple RNN. LSTM is able to preserve long-term memory by making use of memory states that is used in the calculation of hidden states. It is therefore suitable to remembering the irregular trend of the electrical load time-serie. Finally, a fully connected time-serie predicts the load forecast.
Paper [3] further showed superiority with respect to only making use of the LSTM layers as can be seen in Table **??**. The Inputs that were used to forecast the household load which is located in France are: three submeters with historical loads, global intensity, voltage, global reactive power, global active power, time, data and month. At last, also an analysis is performed to investigate the influence of the different inputs by calculating the average class activation score over the inputs. The results are shown in Figure **??**. It can be seen that especially "Sub metering 3" has a big influence on the final forecasts. This sub meter corresponds to the the electric water heater and air conditioner of the house. As was shown in Section A.1 the dataset used in this thesis gives only information about the presence of a hot water heater. Discussed limitations in the paper are the definition of the hyper parameters that were set by trail and error instead of using an automated method e.g. a genetic algorithm. A further limitation is the lack of household characteristics e.g. the amount of residents living in the house. It has previously been shown by **C. Beckel et al.** that household occupancy is one of the primarily drivers of electrical consumption in a household.

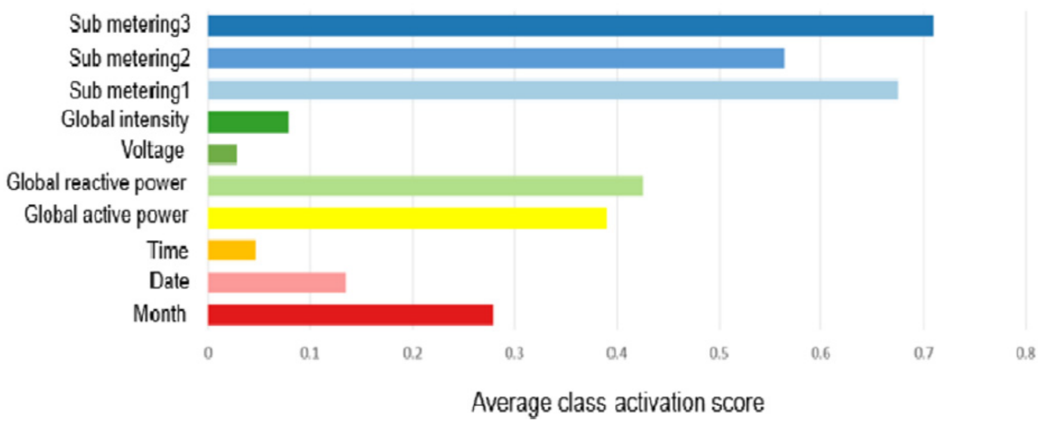


Figure 3.7: The importance of the different inputs as based on the average class activation score. (source [3])

Prediction performance with time resolution change.

Method	Resolution	MSE	RMSE	MAE	MAPE
Linear Regression	Minutely	0.4046	0.6361	0.4176	74.52
	Hourly	0.4247	0.6517	0.5022	83.74
	Daily	0.2526	0.5026	0.3915	52.69
	Weekly	0.1480	0.3847	0.3199	41.33
LSTM	Minutely	0.7480	0.8649	0.6278	51.45
	Hourly	0.5145	0.7173	0.5260	44.37
	Daily	0.2406	0.4905	0.4125	38.72
	Weekly	0.1049	0.3239	0.2438	35.78
CNN-LSTM	Minutely	**0.3738**	**0.6114**	**0.3493**	**34.84**
	Hourly	**0.3549**	**0.5957**	**0.3317**	**32.83**
	Daily	**0.1037**	**0.3221**	**0.2569**	**31.83**
	Weekly	**0.0952**	**0.3085**	**0.2382**	**31.84**

FIGURE 3.8: Comparison between LSTM and CNN-LSTM. (source: [3])

CNN-GRU paper
[7]
See oneNote for the summary of the paper and say that it is showed that CNN-GRU performs even better than CNN-LSTM

3.3 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

Forecasting the daily electricity consumption

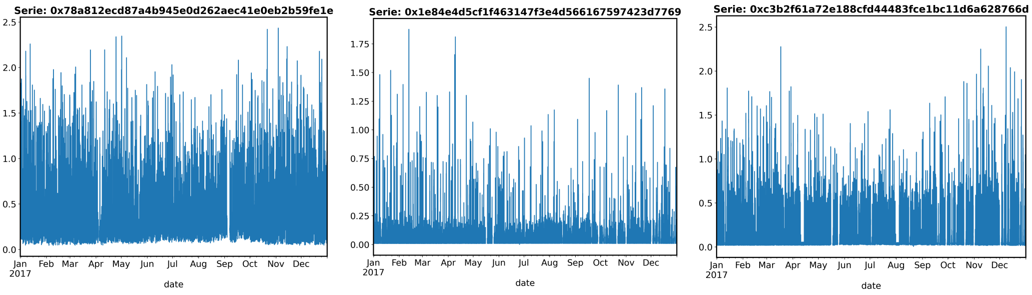
In this chapter the different forecasting techniques to perform a 24 hour prediction for an individual household are discussed. One daily prediction has a data granularity of an half hour, which means that 48 data points have to be estimated for each prediction.
update −> discussing data First pre-processing is done in Section 4.1 and the time series used are discussed. Afterwards the baseline models are looked into in Section 4.2. These models are characterised by a low calculation load during training and they therefore serve as an easy to obtain result where a more complex model can be compared with. Next, more complex models based on a neural network philosophy are discussed in Section 4.3. "Long Short-Term Memory" and "Gated Recurrent Unit" are most suitable to process time series and therefore serve as the core model which is analysed with different design choices. Finally, a parameter search is conducted. Here, an analysis is made of the sensitivity of the choice of different parameters is made.

4.1 Pre-processing

The data that is available is summarized by Table 2.1.

4.1.1 Data

In order to reduce the calculation load to do the parameter search in Section **??**, three series are selected from the *consumption.csv* that is listed in Table **??**. The series are chosen based on the least missing values of the historic electrical consumption serie and the absence of a big shift. Figure **??** shows the three figures and Table **??** summarizes their characteristics.



(A) Serie 1 (B) Serie 2 (C) Serie 3

Figure 4.1: The consumption of 2017 for the three selected series.

Characteristic	**Serie 1**	**Serie 2**	**Serie 3**
Mean daily consumption [kWh]	14.55	3.17	6.58
Standard deviation daily consumption [kWh]	3.21	0.99	2.57
Median daily consumption [kWh]	14.09	2.96	5.88
Maximum daily consumption [kWh]	30.08	6.60	17.15
Minimum daily consumption [kWh]	7.51	1.83	1.50
Total missing days (consumption)	4	25	26
Days in validation set (November)	30	29	29
Days in Test set (December)	31 23	31	23
Days in training set (Rest)	300	288	287
Mean average temperature [$°C$]	10.37	10.61	10.22
Standard deviation average temperature [$°C$]	5.09	5.20	5.01
Median average temperature [$°C$]	10.46	10.61	10.35
Maximum average temperature [$°C$]	23.99	24.51	22.95
Minimum average temperature [$°C$]	-1.07	-1.28	-1.42
Missing average temperature days	0	0	0
Amount of holidays in 2017	8	8	8

Table 4.1: summarizing characteristics about the selected series.

The three time series are divided in a training, validation and test set. As validation and test set are respectively the months November and December chosen. Because in this chapter the time series are not aggregated to a single signal as was the case in Section 2.3 the min-max normalization can be used. The missing values in the consumption are substituted by making use of following baseline models in order: "previous week" model, "previous day" model and the mean model. When a baseline can't make a forecast because the data is not available, a next baseline model is used. Similar, the missing values of the temperature are substituted by first looking at the temperature yesterday, then tomorrow and as last the mean temperature.

The simulations that are done in the chapter are performed on a virtual machine through the Microsoft Azure service. Table 4.2 shows the different features of the hired machine.

Name	**Logical cores**	**RAM (GB)**	**Storage (GB)**
F4s v2	4	8	32

Table 4.2: Specifications of the virtual machine.

4.2 Baseline models

4.2.1 Models

As earlier discussed, the baseline models are characterised by a low calculation load during training and therefore serve as a baseline to compare more complex models with. The different baseline models tried are listed as follows:

- Model 1: "closest day forecast"
- Model 2: "1 day ago forecast"
- Model 3: "7 days ago forecast"
- Model 4: "Mean forecast"
- Model 5: "MAPE forecast"

For all the models listed here, the training set for the forecast of the next day are all the days before this day of the year 2017. These models can therefore be categorized as "lazy learning models" because they only do work when they are asked a query. In contrast, the models discussed in Section 4.3 generalize the training data without knowing the actual query. They belong to the "eager learning methods" class. The 24 hour predictions made by the 5 models are done all at once. This is in contrast to the models described in Section 4.3, where the prediction is made sample per sample and where the predictions done for an earlier hour of the day are taken into account.

Model 1: "closest day forecast"
This model looks for the most similar day in the training set based on following metrics to make a prediction:

- Holiday
- Day of the week

All the days in the training are categorized according to these metrics. Then it is looked in which category the desired day belongs i.e. which day of the week and if it is a holiday. Inside the selected category, an assessment of the difference in average temperature is made for all days with respect to the desired day. It is assumed that the average temperature of the desired day is already available, which is a very plausible assumption. Finally, the day with the closest euclidean distance in temperature is selected and the electrical consumption signal is copied to serve as the prediction of the desired day. It should be noted that there are only a maximum of 8 holidays as can be seen in Table 4.1. Therefore, when a holiday should be predicted all Sundays are also included in the training set because a Sunday behaves most similar to a holiday as can be seen in Figure 2.7.

Model 2: "1 day ago forecast"
This model simply looks at the consumption of the day before the desired day. The philosophy of the model is that the most recent consumption data serves as a a good predictor.

Model 3: "7 days ago forecast"
This model looks at the most recent household consumption of the previous corresponding day of the week. It is expected that people have a reasonably fixed routine during the week and therefore it is likely that this routine will also be found back in the electrical consumption.

Model 4: "Mean forecast"
In the mean forecast the different days are again categorized as was done in Model 1, but instead of selecting a single day out of the group of days, a mean day is calculated and used as prediction of the desired day. No extra Sundays are included to forecast a holiday.

Model 5: "MAPE forecast"
This model solves for each half hour of the desired day a small non-linear optimization problem displayed by (Eq. 4.1). This model served as baseline model in paper [4].

$$objective = \sum_{i=1}^{K} \zeta_{pi} \left| \frac{(\hat{y} - p_i)}{pi} \right| \tag{4.1}$$

Again a group of days of size M, corresponding the desired day is selected based on the metrics of weekday and if the desired day is a holiday. Also Sundays are added to the group of holidays for this model. Next, the consumption at time t is extracted out of this group of days, which gives a list of length M with historic consumption values. From these values an empirical probability mass function ζ_{pi} is derived by making use of a histogram using "Freedman-Diaconis rule" to decide the bin size. Figure B.1 shows an example of such an histogram. The amount of discretized values p_i is equal to the K bins and taken as the midpoint of two bin edges. From the count in the histogram the probability mass function for each discretized value is found. $\hat{y}$ is found by minimizing equation 4.1.

The metrics used to evaluate the predictions performance of the base line are $RMSE$ (Eq. 3.3), $NRMSE$ (Eq. 3.4), MAE (Eq. 3.5), MSE (Eq. 4.2) and $MAPE$ (Eq.

4.3).

$$MSE = \frac{\sum_{t=1}^{N} (\hat{y}_t - y_t)^2}{N} \tag{4.2}$$

$$MAPE = \frac{\sum_{t=1}^{N} |\hat{y}_t - y_t| / y_t}{N} \tag{4.3}$$

4.2.2 Results of baseline models

The results of the different forecasts of the month December are summarized by Table **??**. In order to make a fair comparison only the days of December where all models could produce a forecast are included in the error metrics.

Error metric	**Closest day**	**1 day**	**7 day**	**mean**	**MAPE**
Mean absolute error	0.2049	0.1954	0.1896	0.1542	0.1920
Mean squared error	0.1148	0.1090	0.1011	0.0701	0.1079
Normalized root mean squared error	0.1591	0.1550	0.1493	0.1243	0.1542
Root mean square error	0.3389	0.3302	0.3180	0.2648	0.3285
Mean absolute percentage error					

TABLE 4.3: Evaluation results for Serie 1 tested on 31 days of December.

Error metric	**Closest day**	**1 day**	**7 day**	**mean**	**MAPE**
Mean absolute error	0.0559	0.075	0.0693	0.0473	0.0507
Mean squared error	0.0123	0.0264	0.0188	0.0085	0.0125
Normalized root mean squared error	0.0823	0.1205	0.1017	0.0681	0.0828
Root mean square error	0.1111	0.1625	0.1373	0.0919	0.1117
Mean absolute percentage error					

TABLE 4.4: Evaluation results for Serie 2 tested on 12 days of December.

Error metric	**Closest day**	**1 day**	**7 day**	**mean**	**MAPE**
Mean absolute error	0.1267	0.1370	0.1323	0.1038	0.1130
Mean squared error	0.0824	0.0846	0.0895	0.0521	0.0743
Normalized root mean squared error	0.1453	0.1472	0.1514	0.1155	0.1380
Root mean square error	0.2871	0.2909	0.2991	0.2282	0.2726
Mean absolute percentage error					

TABLE 4.5: Evaluation results for Serie 3 tested on 12 days of December.

It can be seen that model 1,2 and 3, which copy the consumption of another day, show more peaks in their prediction than model 4 and 5 that make use of the mean and Mape techniques. Also, a practical downside of model 1,2 and 3 is that it is possible that they give no output e.g. no temperature or consumption yesterday is available.

- show all four plots –> MAPE was not possible because of the division by zero - show pictures –> three peak - show picture –> two mean

Base model Can't plot the Mape figure because sometimes the real value is equal to zero. 99 time-series have sometimes a zero consumption during half an hour. The different base models can behave quite different on the different time-series. Sometimes the closest day works better than previous week or day and sometimes works much worse. Doesn't make a lot of sense to average the base models over all the possible time-series.

Experiment of forecasting three time-series

Looking at the bar plots

Mean forecast behaves best for the three time series and for all the different metrics. In this serie the prev week and day behave worse than the other techniques. Similar day and MAPE work here clearly better.

Three other techniques which is best, depends a lot about which time serie is forecasted. When look at mse –> the amount that mean performs better gets bigger. Indication that mean is better in predicting also when occuring peak, Because this is what the mse penalizing more severly than the other methods. Verify this with the individual plots in the tekst. This is possibly due to the square. The MAE gives a more all similar behaviour than the other metrics.

Looking at the day plots The previous day and week and similar forecast have typically much peaks but shifted. The MAPE and the mean forecasts follow more the trends.

Important that when do predictions –> only look at the values in December where there is a day forecast for all the different techniques. This is in order to make a fair prediction. serie 0x78a–> all 31 days serie 0x1e–> 12 days serie 0xc3b –> 12 days Disadvantage of previous day, previous week and similar week is that there can be not a forecast available! This is mostly the case for prev day and prev week. This issue is not occuring for the mean and MAPE options.

Experiment to look at all the time series All the 261 time-series are set to be equally important. In every time-serie a comparison with the performance of the worst technique is made. The worst forecasting technique is set to have a value one and the rest of the techniques get a value how much better they perform than the worst signal. Which technique performs worst, depends on every time-serie. If the mean performance is better for one technique than the other this means that the to expect performance for a random time-serie for one technique is better than the other technique. This is because this techniques behaved more often or more proportional better than the worst technique in comparison of the other technique over all the time series.

for one technique the performance in comparison with the worst technique Was on average better than the other technique.

4.3 Neural network models

- predicting each time one sample - discuss the inputs to the models -assumption is made that you are waiting untill midnight to do the prediction of the next day... Typical variables used in a forecasting model are: past electricity consumption loads, weather information, calendar information and error-correction terms [1].

4.3.1 LSTM

- model - as metric for the model during training –> chosen for mse.

4.3.2 GRU

4.3.3 Parameter search

Multithreading is used to use the full potential of the CPU listed in Table 4.2 and simultaneously use all the available logical cores to run different threads.

4.3.4 Temperature model

A MLP model for developing a regresion model to forecast a temperature signal.

4.3.5 Evaluation of different designs

4.4 Conclusion

Chapter 5

Pooling Strategy

5.1 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

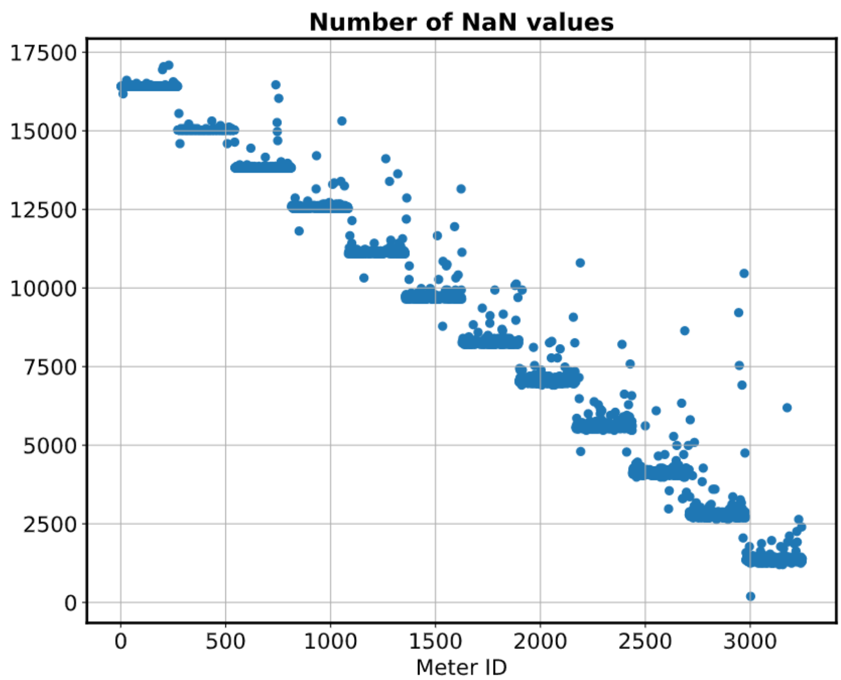


FIGURE A.1: The amount of NaN values in all the 3248 smart meters.

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

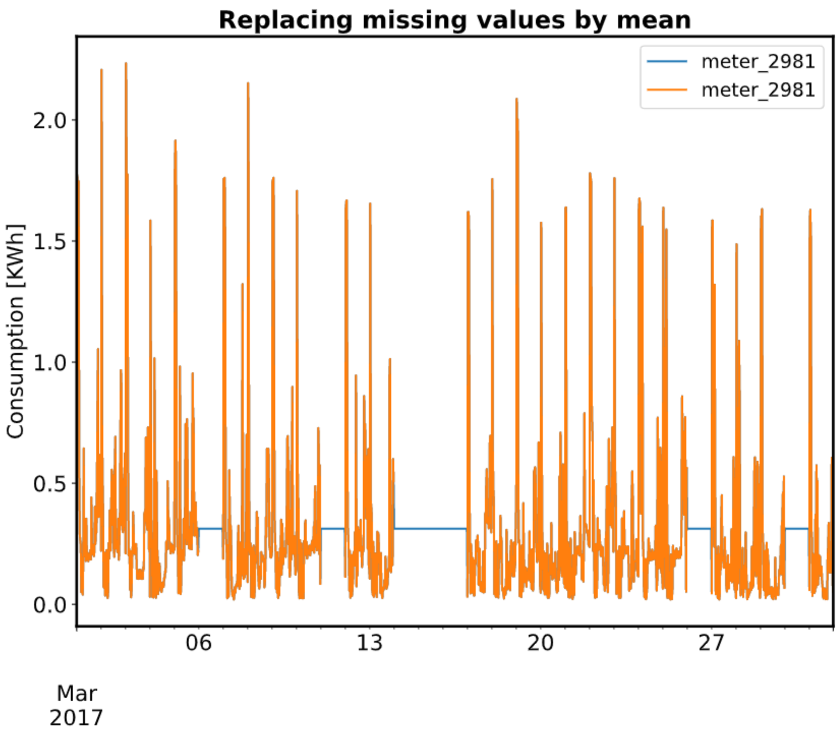


FIGURE A.2: Resulting month of March after substitution of the missing values by the mean value of the measurements.

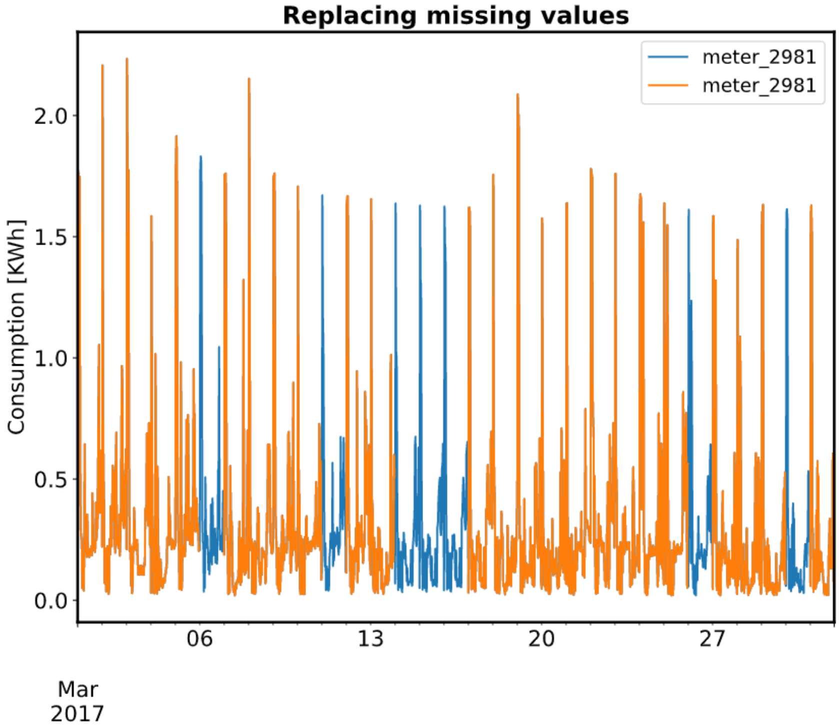


Figure A.3: Resulting month of March after substitution of the missing values by the mean value of the same moment on the next and previous day.

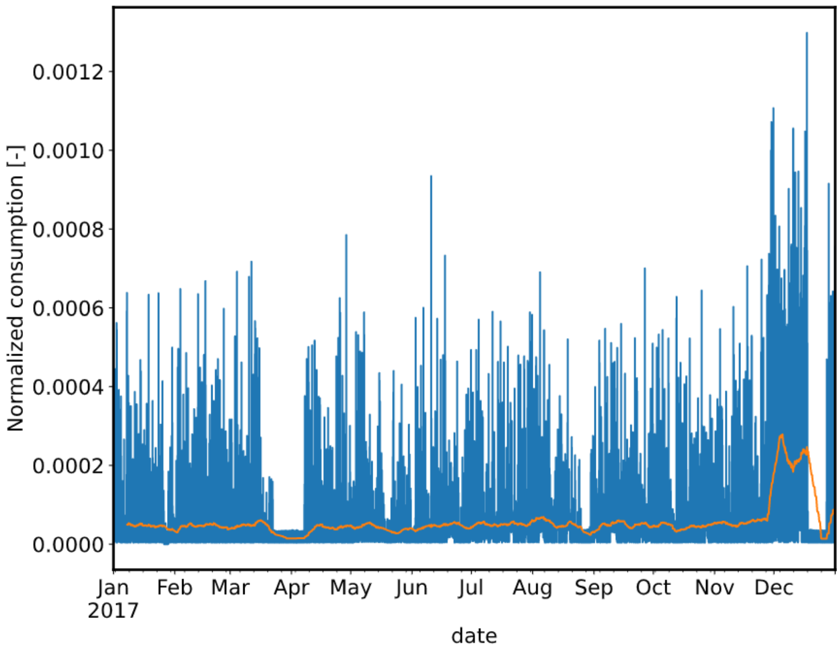


Figure A.4: The time-serie with the original maximum difference between the minimum and maximum weekly rolling averages.

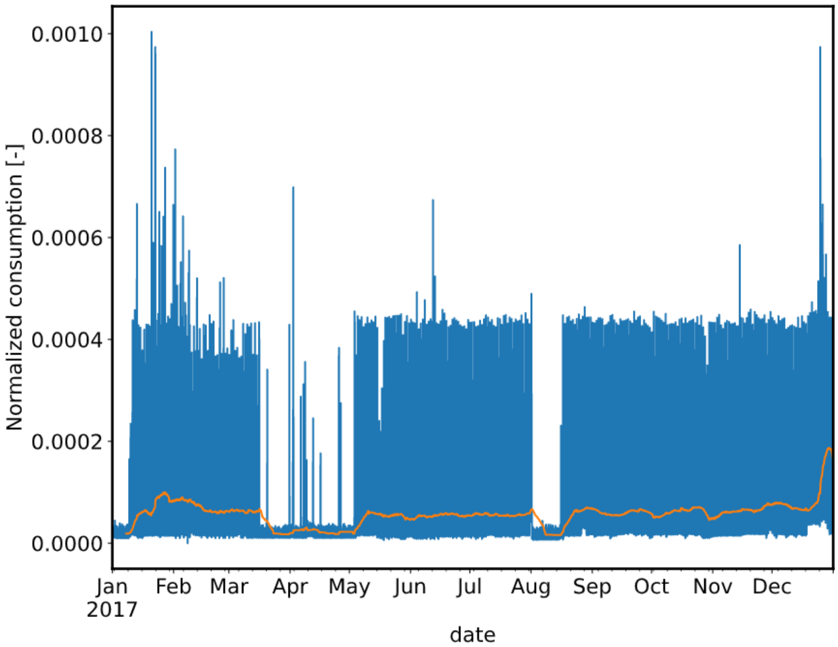


FIGURE A.5: The time-serie with the new maximum difference between the minimum and maximum weekly rolling averages.

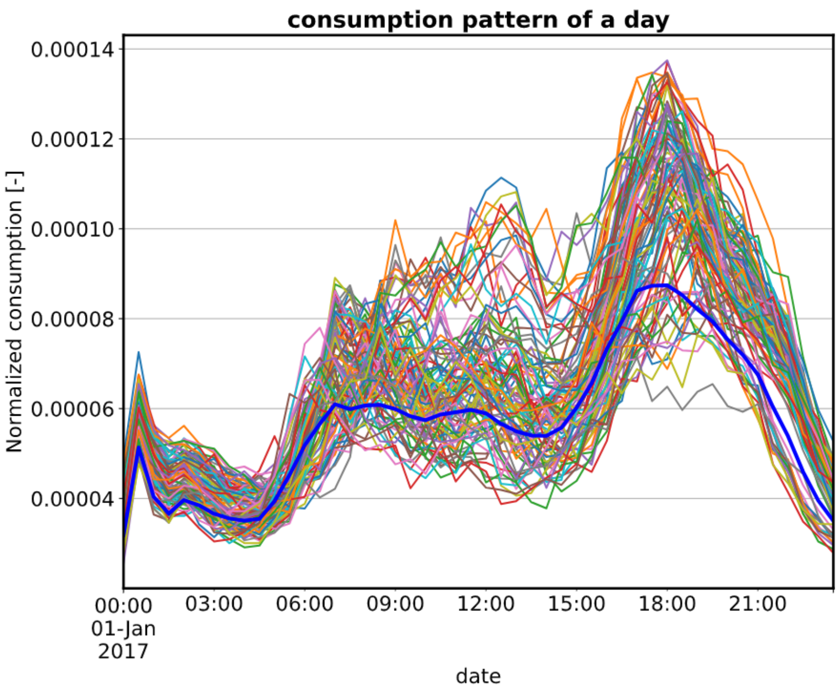


FIGURE A.6: Figure that shows the seasonality of the electrical load during the day.

Appendix B

Forecasting the daily electricity consumption - extra

In this appendix extra information and Figures are added that are not necessary to understand the work discussed in Chapter 4.

B.1 Baseline models

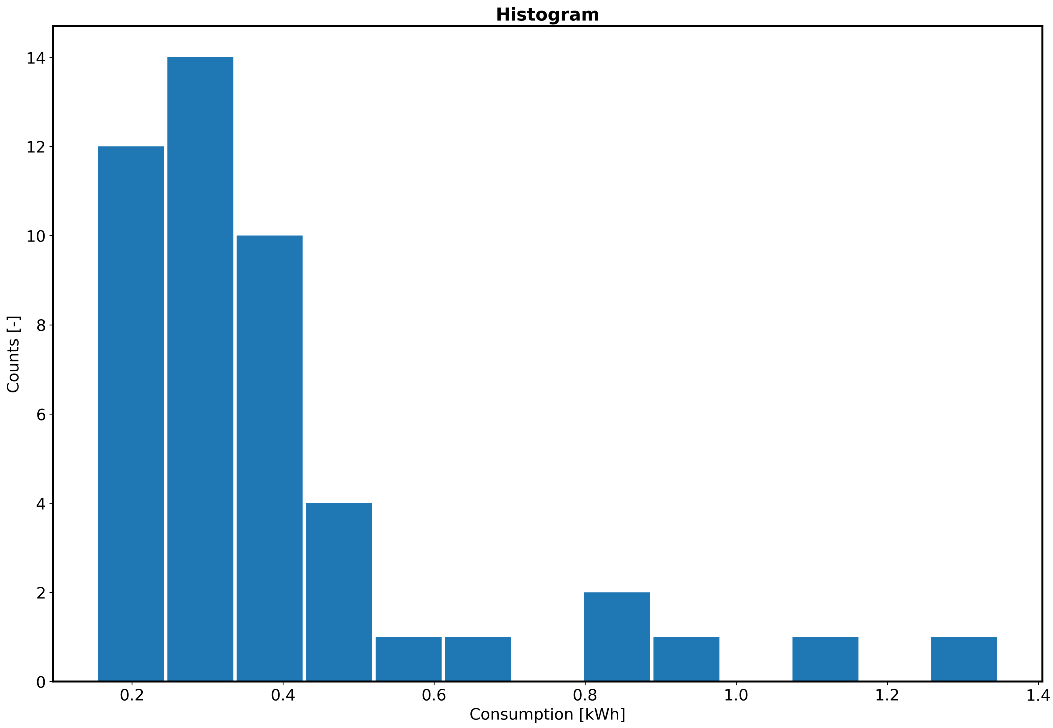


Figure B.1: An example histogram of the consumption in [kWh] versus count [-] used during MAPE forecast.

Appendix C

Old things

C.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 $|z - score| < 3$.

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

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

C.0.2 Normalization of the data

Normalization is necessary because while absolute consumption differs, relative patterns of human behaviour are more similar [5]. 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 C.2.

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

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.

C.1 ARIMA

What is ARIMA. Assumptions of ARIMA...

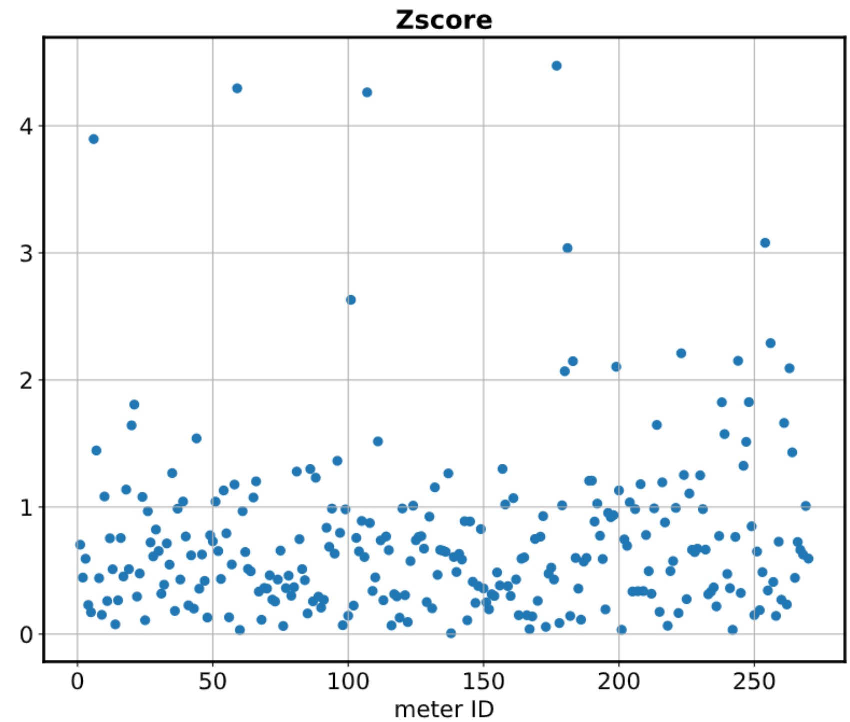


Figure C.1: Z-scores calculated from the yearly consumptions.

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] T. Y. Kim and S. B. Cho. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, 182:72–81, 2019.

[4] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Transactions on Smart Grid*, 10(1):841–851, 2019.

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

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

[7] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik. A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting. *IEEE Access*, 8:143759–143768, 2020.

[8] 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.

[9] J. Teuwen and N. Moriakov. *Convolutional neural networks.* Elsevier Inc., 2019.

[10] M. Zhang, Aston Lipton, Zachary Li and A. Smola. *Dive Into Deep Learning*, volume 17. 2020.