CODEN:LUTEDX/(TEIE-5447)/1-39/(2020)

Industrial Electrical Engineering and Automation

Estimation of Load Profiles for Secondary Substations

Generalising AMR Data with a Statistics and Deep Learning Approach

Hans-Christian Nilsson

Division of Industrial Electrical Engineering and Automation
Faculty of Engineering, Lund University

Abstract

The power system today is facing a transformation from fossil and nuclear energy sources to renewable energy sources such as solar- and wind power. At the same time electric vehicles are becoming more common every year. To analyse how these two changes affect the power system it is crucial to understand how loaded the power system is today from traditional loads on a local and regional level. The purpose of this thesis has been to create the best possible representation of today's electricity consumption. This is done by generalising and aggregating hourly collected samples of load from energy meters used at low voltage customers of Kraftringen in Lund. The estimation is done on medium voltage level substations by collecting publicly available features which include information about the expected electricity consumption. The resulting model has combined the strengths of linear regression and artificial feed forward neural networks. The model has a mean absolute percentage error of only 10% when evaluated on unseen data from stations used in training the model and 16% when evaluated on an entirely unseen station. The model has been compared to different implementations of the load curve method (typkurvor) which it outperforms with a mean absolute error 54% smaller than the best load curve implementation. The results are based on data from residential districts only and therefore the accuracy of the methods are limited to residential districts. One of the major advantages of this model is that it should be able to predict the electricity consumption from unseen residential districts with only feature data, no data of the electricity consumption from unseen areas is needed. Besides modelling hourly values of load from substations, an extreme value theory model has been used to model the expected maximum loads that occur for one station. This is done with a combination of block maxima of two week-period blocks and by using the Generalised Extreme Value distribution. The resulting model covers all observed load maxima when including confidence intervals of 95% and can be used to predict the expected maximum load for a given time period. The models shown in this thesis can be used by researchers and utility companies to generate expected load of substations and also to model extreme values of load.

Keywords— Load Estimation, Synthetic Load, Feed Forward Neural Networks, Extreme Value Theory, Block Maxima, Linear Regression, Load Curves, Electricity Consumption Estimation

Acknowledgements

First, I would like to thank my supervisor at LTH Olof Samuelsson for proposing the subject of this thesis and for helping me structure and limit the work. Your knowledge of electricity consumption has helped me in my research for attributes in this thesis and you have helped me find numerous of helpful articles in the subject. I am thankful that you have been able to supervise my work in these times where Covid-19 removed the possibility of having physical meetings.

Thank you Andreas Vikström and Philip Anterup at Kraftringen who have helped me with collecting the data needed for this thesis and who have helped me obtain and explain information about the different stations. I would like to thank you especially for all the time put in helping me since you have not been able to have me as an official thesis worker at Kraftringen as was planned when this work started.

Finally I would like to thank my wife Fanny Nilsson whom I have been able to discuss the contents of this thesis with on a regular basis and also friends and family who have helped me in any way related to this thesis.

Contents

1 Introduction

This introduction gives a background to and a motivation to the work of this thesis. It also states the purpose and research questions of the thesis. The section ends with the limitations that had to be made and an overview consisting of related work.

1.1 Background

The power system today is facing a transformation from fossil and nuclear energy sources to renewable energy sources with solar- and wind power. At the same time electric vehicles are becoming more common every year. Renewables and electrification of vehicles results in more load for the power system unless controlled. Electricity production from solar and wind is highly weather dependent which makes it hard to dimension and run the power grid. Electric vehicles load the power grid when they need to charge which results in them demanding much electricity at the same time. There is a risk that the power system cannot deliver that much power even though the electricity production is sufficient due to the grid being undersized. To analyse how many electric vehicles can charge at the same time or how electric roads would affect the power system both knowledge about charging and knowledge of the variation in solar radiation and wind speed are important. However, it is also crucial to understand how loaded the power system is today from traditional loads on a local and regional level. A research project at Lund University, Faculty of Engineering (LTH) has the ambition to find the best possible information about renewable energy, future electric transports and the load of the power system today. One problem of finding the load of the power system today is that there is no access to available data. Utility companies do traditionally not share data to each other or to researchers for security reasons, which prevents researchers from developing accurate descriptions of the present power system and how much of its current capacity is being used. This results in uncertainty when predicting the impact on the grid from the load of electric roads or electric vehicles. It also results in uncertainty of how regional power grids are loaded unless they are owned by a single company. For these reasons, this master's thesis aims at creating the best possible representation of today's load by generalising hourly collected samples of electricity consumption from load meters used by low voltage customers in Lund to other Swedish cities. By aggregating low voltage customers' hourly collected electricity consumption from the utility company Kraftringen it is possible to find the hourly load from the medium voltage substations. By generalising this data to attributes, a model can be built to estimate the load profile for an arbitrary substation as long as it is supplying a similar load area only by feeding it with attributes, no need of new electricity meter data is needed. Examples of attributes which can be used for predicting load in this sense are number of resident owners per building, hour of day, month, heating means, number of customers etc. The resulting model predictions will consist of hourly load estimates for an arbitrary time period of interest.

1.2 Purpose and research questions

The purpose of this thesis is to estimate the load profile for a general secondary substation with attributes chosen by analysing Kraftringen's hourly collected data. The thesis shall answer the following research questions:

- Is it possible to characterise the load of secondary substations of Kraftringen using a finite set of easily available attributes?

- Is it possible to use electricity meter data from individual customers to build a model that uses these attributes to estimate the load of the stations with a mean absolute error of less than 15%of the average load?
- Is it possible to estimate the load of an unseen station by using the same model and attributes with a mean absolute error of less than 15% of the average load?
- Is it possible to generalise such a model further by adding a stochastic distribution to model not only average values but also extreme values?

1.3 Limitations

The load data available in this thesis consists of stations supplying load to residential districts. This limits the work of this thesis to estimation of stations supplying residential districts. 11 stations have been available in the project and the measurements from these stations span for only 11 months in 2018 from January to November which makes it impossible to estimate load in December. The attributes used in this thesis are neither exhaustive nor optimal and other better attributes may exist. The measured loads are measurements of energy and no attempt is made on estimating the amount of reactive power since it is assumed that it does not affect the stations' load capacity noticeably. The load data available comes from measurements of energy consumed for each customer which are saved each hour. The term electricity consumption is therefore in this thesis used to describe the difference in energy between two hours of sampled load. This quantity is useful for analysing load behaviour for different times of the day but is limited to describing only average power.

1.4 Related work

Using measured energy for estimation of load in order to dimension the power grid's lines and transformers has been done for many years but until recently, only yearly readings of electric energy consumption have been available. Therefore, previous load estimation methods only require the energy consumption of one year to make load profile predictions for the entire year. The electric energy is related to power according to equation 1 where E is the electric energy, t_0 and t_1 defines the time period since E was measured last and $P(t)$ is the power which depends on time.

$$E = \int_{t_0}^{t_1} P(t)\, dt \tag{1}$$

One used method has been the load curve method derived by Svenska Elverksföreningen, [1]. The load curve method have been evaluated in several projects and Peter Eriksson checks the performance of those load curves derived by Svenska Elverksföreningen and concludes that they usually do not estimate load correctly, [2]. In order to model extreme values of load Velander's formula has been and is still being used. The method assumes that the electricity consumption is normally distributed and only requires yearly energy consumption as input. It is evaluated by Oscar Ingvarsson who concludes that the method usually overestimates the load from customers and that it does not work when the load analysed is not normal distributed, [3]. J. Dickert and P. Schegner also concludes that while Velander's formula worked in the past, it has resulted in oversized dimensioning of the power grid and they show that modelling individual maximum loads as normally distributed is an inappropriate approach, [4].

Since hourly collected electric energy consumption often is available today other methods of estimation have been developed for constructing load profiles. A common method consists of clustering customers or stations into a few categories and use profiles for these categories to estimate load. This is done in Gecad's paper which evaluates different clustering algorithms and decide what the optimal number of clusters are for medium voltage customers, [5]. The downside of clustering is that it only gives accurate predictions on groups of the customers that the clusters are based on, it generally performs bad on outliers since clustering algorithms rely on averaging. Another downside is that in order to cluster new customers there must be measured samples of electricity consumption. It is therefore not possible to use a clustering technique to estimate load from an unseen customer or station. An Energiforsk report tries to generalise clustering further by clustering the load based on several attributes derived from the hourly measured load in order to solve the first problem associated by clustering. However, they conclude that combining many attributes using clustering is complicated and requires weighting/normalisation of attributes. They also conclude that using load curves for estimating load is hard since most groups are not homogeneous, [6].

There are numerous methods for short term predictions using hourly measured electricity consumption. For example, in Applied Energy 2018 the Gaussian process is used for short-term prediction (1 - 24 hour long prediction horizon) of load with results in a mean absolute percentage error of 2.4%, [7]. A Long Short-Term Memory (LSTM) Neural Network is proposed by S. Muzaffar and A. Afshari which results in a mean absolute percentage error of 1.5% for 24 hour-ahead predictions, [8]. Y. Wang, Q. Chen, T. Hong and C. Kang have written a review of smart meter data analytics and there is a lack of methods describing load estimation without using previous time steps of load, [9]. That is, the load forecasting methods are focusing on short-term predictions rather than trying to estimate the load of a customer of station for an entire year or similar. They cannot produce arbitrarily long load profiles.

Another area of related work is that of synthetic networks which produces networks, load and generation that match demographic or other data. In IEEE Transactions of Power Systems this is done by first adding substations into a geographic area. The article uses the number of customers and statistical information about how many substations include load, generation and both load and generation to create a number of substations whose load/generation is decided mainly by the number of postal codes in the area, [10]. The downside of this approach is that the loads and geographic locations of stations are not real even though they might be realistic candidates of load and geographic placements. Other methods of generating synthetic load data comes from building models based on real electricity consumption which then can be used to generate load. This is done on a national level of load profiles using an artificial neural network with a mean absolute percentage error between 9.9% to 17.1% depending on which nation that was analysed, [11]. However, the method has not been tested on a regional or local level.

Extreme value theory using the block maxima method combined with the Generalised Extreme Value (GEV) distribution on residuals from short term prediction models have been evaluated in a PSCC paper, [12]. It was found to accurately model extreme events and could be used to evaluate long term risks. No attempt was made of using extreme value theory on the loads directly.

The attributes used in several of the methods described above are the number of customers, temperature, humidity, season, weekday. Beside these some methods include socioeconomic attributes such as average income, age of customers, number of people living in a house and living area.

2 The datasets and attributes

In this section the datasets are described. The attributes used are also defined and explained.

2.1 Description of electricity consumption datasets

The electricity consumption dataset consisted of 11 different stations' customers and their load for each hour during 2018 from January to November. Each station is considered to supply residential customers. Table 1 shows an overview of all stations used in the thesis including the number of sampled measurements used in the analysis per station. After aggregation of customers per station 86636 data points remain for training and validating models, (11 months of hourly data for 11 stations). The energy meters record their measurements with a resolution of ± 0.1 kWh up to 10 kWh and ± 1 kWh for energies bigger than 10 kWh so the precision of the analysis is limited to these resolutions of load.

Table 1: An overview of the stations used in the thesis. Number of customers is defined as the number of energy meters used for billing that the station supplies. The number of hourly values is the sum of all hourly values for 11 months and for all customers supplied by the station. Note that the names of the stations shown are made up in order to keep their identity hidden.

Station X	Number of customers	Min (kWh)	Mean (kWh)	Max (kWh)	Number of measurements
1	70	20	51	149	551320
2	114	30	96	244	897864
3	62	30	101	270	488312
4	81	40	124	367	637956
5	39	21	107	358	307164
6	131	21	107	279	1031756
7	131	23	152	358	1031756
8	56	28	100	287	441056
9	131	20	100	219	1031756
10	87	39	104	263	685212
11	131	16	93	225	1031756

The electricity consumption was measured for each hour. In order to use the datasets they had to be cleaned from data errors such as those seen in figure 1. Note that the plot in the figure is corrupted by a single consumption value. The corrupted values were automatically identified as corrupted by setting a threshold that used interpolation on all electricity consumption values above the threshold but also all samples with a consumption underneath the negative threshold value. The corrupted values were replaced using linear interpolation between the neighbouring values. After this cleaning, the datasets could be inspected by visual plots. In figure 2 a single customer's load can be seen for 7 days of data in winter and summer respectively. Note that the hour of the day largely influences the consumption and note that the consumption is higher in the winter graph because of seasonal influence.

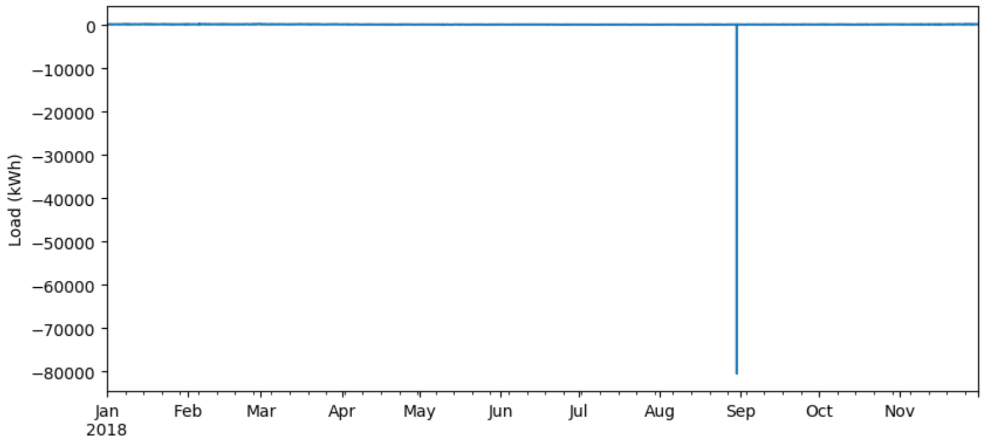

Figure 1: The load of a station with one hourly value corrupting the entire graph. Note that this single outlier also corrupts the average and standard deviation and other statistical metrics on that station.

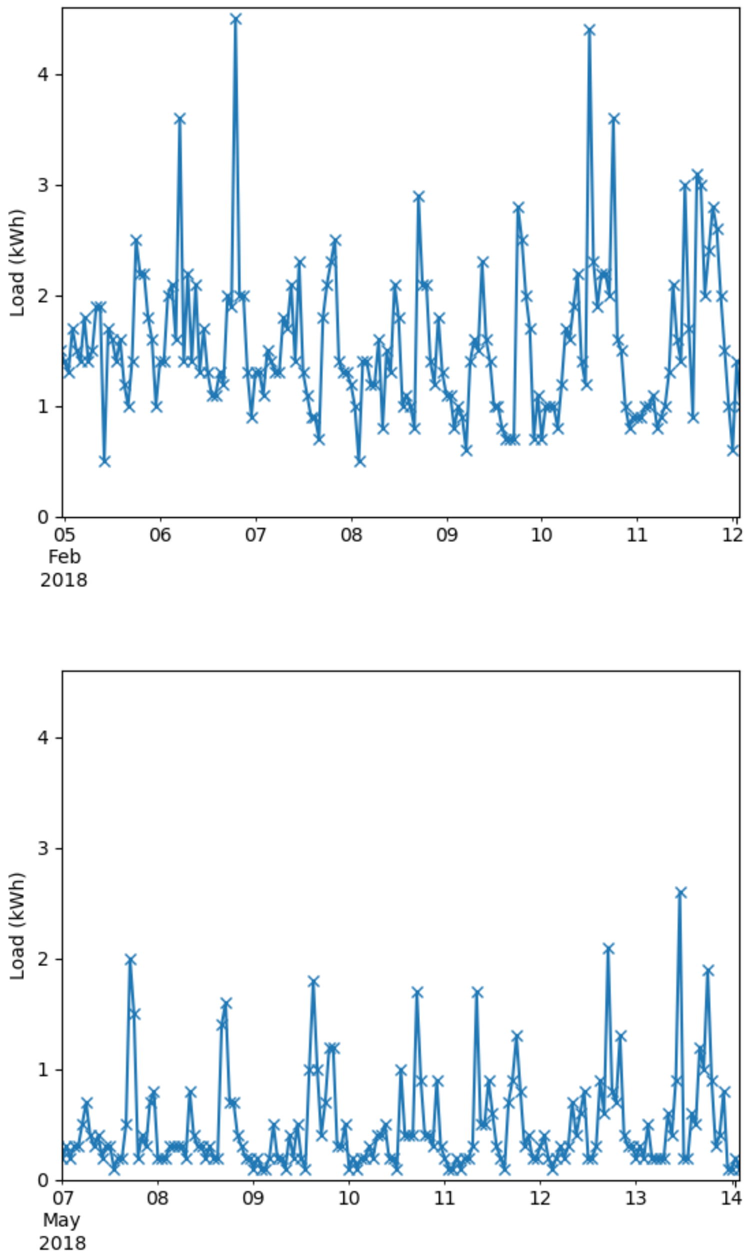

Figure 2: The load of a customer in winter and summer for seven days from Monday 00:00 to Sunday 23:00. The top figure shows the winter days and the bottom figure shows the summer days.

By aggregating the customers by summation, the load profiles of each station could be obtained. This profile is seen for one station in figure 3 and figure 4 in different time scales. Note that the customer seen in figure 2 is supplied by this station. The time influences noted from figure 2 are also seen here. It is also possible to see that there is a clear difference in load during the last two days that correspond to the weekend suggesting that weekdays influence the load consumption.

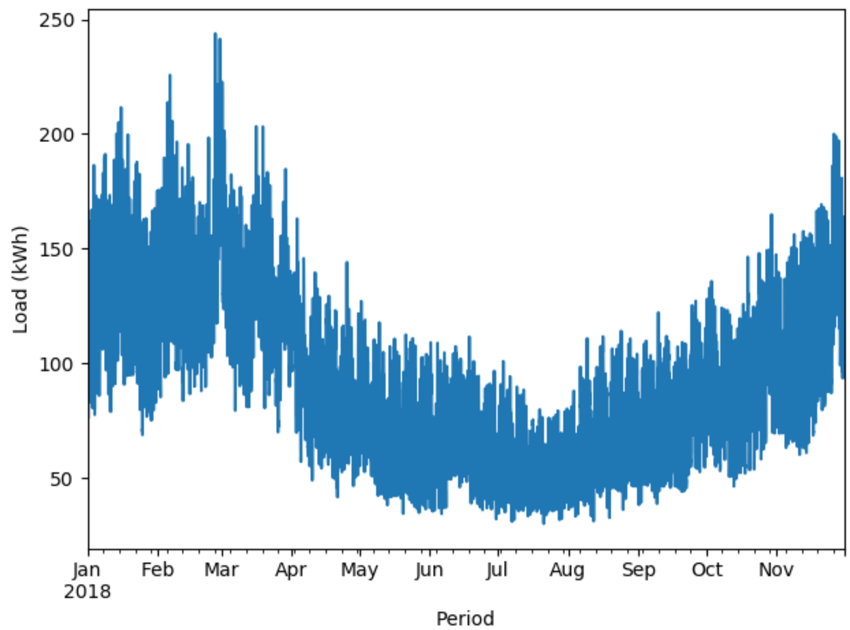

Figure 3: The load of a station for 11 months.

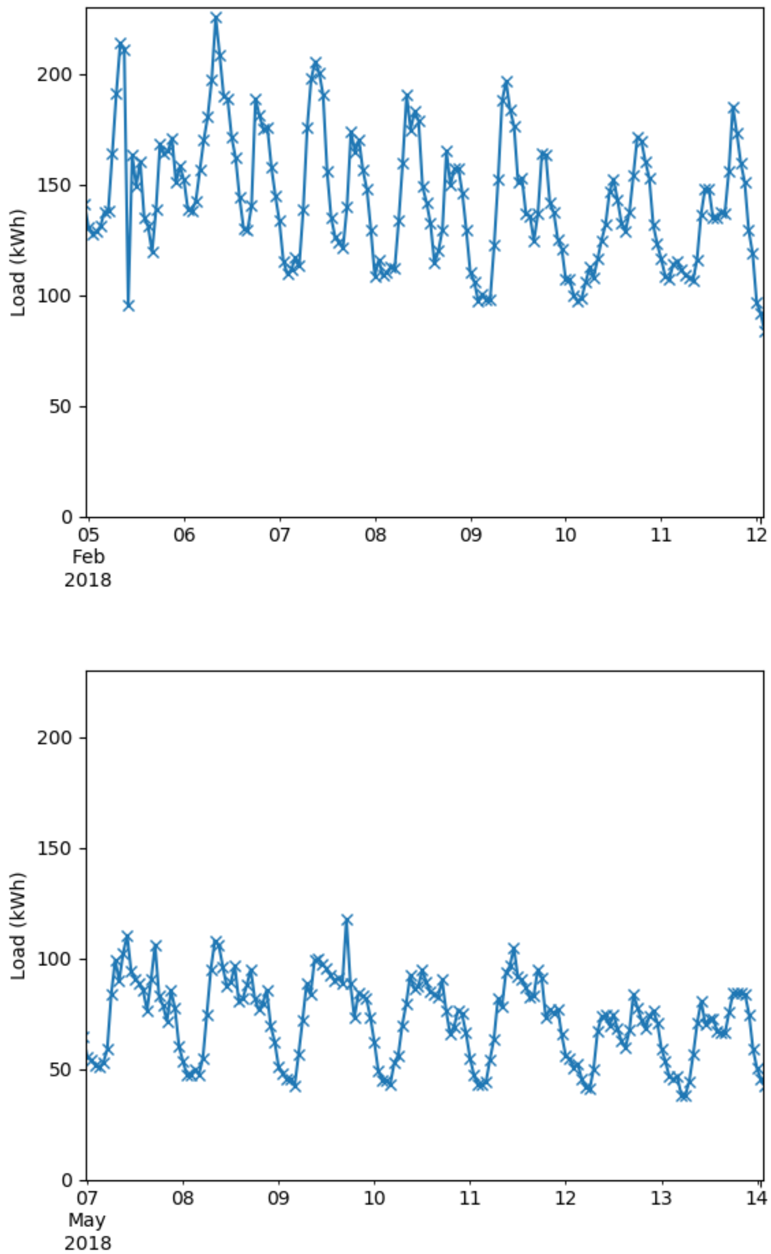

Figure 4: The top figure shows the total load of a secondary substation for one week in February. the bottom figure shows the total load for one week in May. Time period is the same as in figure 2.

Figure 5 shows the load from each customer connected to one station in one day in May. This shows that while most customers are similar there are some that have a hugely different load profile from the rest. To get a better idea of what most customers load looks like, another plot is included underneath where the 8 customers with highest loads have been removed out of the 114 customers. It is likely that the customers with biggest load in figure 5 are not residential owners since they have such high loads. This is because that even though a station is labeled as a residential station it might also feed small local businesses as well as schools, preschools and retirement homes which often lie in residential areas.

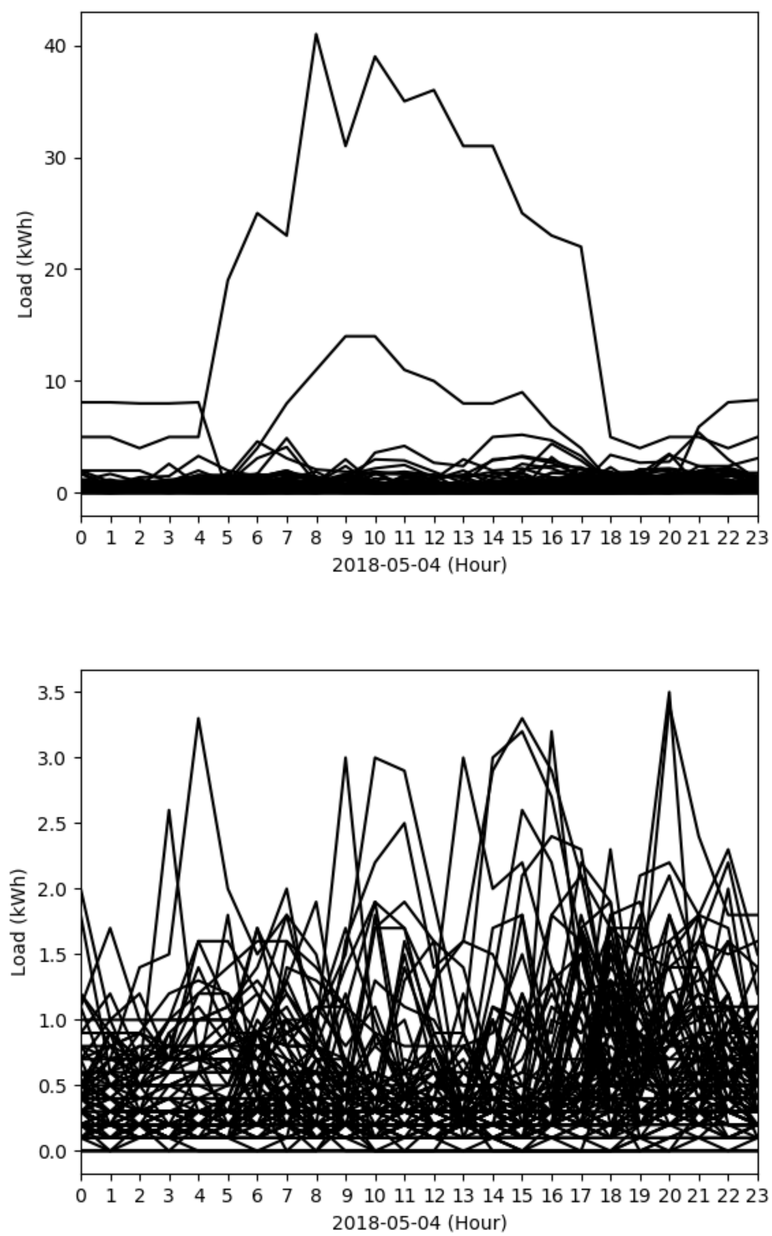

Figure 5: The dispersion of the different customers connected to the same station in one day in May. The upper plot shows the dispersion of all customers and the second plot shows the dispersion of all customers except the 8 customers out of 114 with highest load.

2.2 Description of generalising attributes

Different attributes collected and evaluated during the thesis are described in this subsection. An attribute is defined as a quantity that is able to predict, or at least partly explain the electricity consumption of a substation without the need of knowing the electricity consumption in advance. For example, since a large number of customers connected to a station demands more electricity than a small number, the number of connected customers to a station is useful as an attribute to predict load. The models that are trained in this thesis will find the optimal relationship between used attributes and electricity consumption but no matter how good a model is, it will always rely on the attributes being used. That is, if electricity consumption is explained with an attribute that is not related to the electricity consumption, the model will always fail no matter how ”good” it is. Since the models in this thesis completely relies on attributes in order to predict electricity consumption the concept of time has to be set as attributes. Although this could be done as using one attribute (hour of the year) another choice is to use three attributes: hour of the day, month and weekday. This choice of attributes is useful since the prediction of each hour of load is independent of all other predictions. So in order to capture the periodicity of time it is noted that load is different depending on the hour of day, load is also different depending on the season or month and also different depending on whether it’s an ordinary weekday or weekend. This is observable in figure 3 and 4.

The generalising attributes were chosen based on what was possible to retrieve and based on what was thought to influence the prediction accuracy of the load profiles. This preliminary assumptions on which attributes would be usable for estimating load were based on partly the writer’s prior knowledge of the subject but also of various previous work in the area.

In order to test which attributes were usable, they were included in estimation models described in the theory section of this thesis. The condition for using an attribute in the final model was that it had to improve the result of the estimation models so that smaller residuals were found.

All attributes collected and evaluated during the thesis are summarised in the following list.

- Number of customers
- Hour of the day
- Month
- Weekday
- Temperature (degrees of Celsius)
- Relative Humidity
- Wind speed (m/s)
- Solar radiation (W/m^2)
- Average income (SEK)
- Average price for residential property (SEK/m^2)
- Percentage of highly educated inhabitants

- Average living space (m^2)
- Average main fuse size of customer (A)
- Average year of the buildings in the area being built

Number of customers Defined according to the number of load meters used for billing that are connected to the station. This information was shared from Kraftringen.

Hour of the day Used to capture daily periodic dependence of load. Ranges from the 0th hour to the 23rd hour.

Month Used to capture seasonal influences of load such as higher load in February than May. Ranges from January to November since there is no data collected from December.

Weekday Used to capture weekly periodic influences of load such as differences between load patterns on ordinary weekdays and weekends. Ranges from Monday to Sunday.

Temperature Defined as the outdoor temperature. For the measured time period, this temperature was only available in Malmö which is a city close to Lund. Therefore it is assumed to be the same in Lund. Measured in Celsius and is used to capture the dependence of heating and load. This data is publicly available from SMHI.

Relative Humidity Used to capture load dependence of Weather. Relative humidity has no unit and ranges from 0 to 1. This data is publicly available from SMHI.

Wind speed Used to understand the load dependence of wind. It is assumed that windy days require more heating than windless days. Measured in m/s and cannot be negative. This data is publicly available from SMHI.

Solar radiation Used to understand the relationship of load and sunny days. It is assumed that sunny days require less lighting, especially in winter. Measured in W/m^2. This data is publicly available from SMHI.

Average income Used to include demographic data to load estimation. It is assumed that a higher income results in higher electric load. Measured in SEK. This data is publicly available at hitta.se.

Average price for residential property Used to find whether more expensive houses consume more electricity or not. Measured in SEK/m^2. This data is publicly available at hitta.se.

Percentage of highly educated inhabitants Used to see whether education influences the electricity consumption. Measured as the ratio of people in the area that have a post-secondary education. This data is publicly available at hitta.se.

Average living space Used to understand the dependence of load and the size of the living area. It is assumed that bigger living area results in higher load. Measured in m^2. The living space of most houses is available at hitta.se. In order to find the average living space a random sample of living spaces were used to calculate the average using 15 houses for each substation.

Average main fuse size of customer Used to capture how big the general load consumer is affects the total load. It is believed that a larger average fuse size results in larger loads. Measured in A. Data was shared from Kraftringen.

Average year of the buildings in the area being built Used to understand how buildings built in different decades have different load consumption. It is assumed that different time periods have resulted in separate kinds of isolation for example which would affect heating. This data is publicly available at hitta.se.

In the next section the theory behind the models used to evaluate the named attributes is explained.

3 Theory

The following section describes the theory behind the mathematical models used in the thesis to estimate the load profiles from the attributes described in the previous section. In order to do this the dataset of the previous section are used to train the following models. Note that to these models, the concept of time as season, weekday and hour of the day has to be defined as attributes in order to accurately capture time dependence. The subsections describe how the different methods work and what their advantages and disadvantages are for estimation in general. This section also explains the theory behind the extreme value model used in the thesis. Note that all attribute based models results in equation 2 where different models correspond to different f where $\boldsymbol{x}$ is a vector of attributes defined as one sample and $\hat{y}$ is the estimated consumption of that sample. By feeding the models with many samples of attributes, estimated load profiles can be constructed for an arbitrary period of time.

$$\hat{y} = f(\boldsymbol{x}) \tag{2}$$

3.1 Linear regression

Linear regression (LR) is the simplest regression method and it models the dependent variable (load) as a linear function of the explanatory variables (attributes) explained in the previous section. The method uses the model in equation 3 where n is the number of explanatory variables used in the model, x_i is an explanatory variable, a_i is a coefficient to be estimated from data. b is a bias term and ϵ_j is the residual of the model. Note that in linear regression ϵ_i is assumed to be normally distributed and that each residual is independent of the other residuals. The coefficients are estimated by minimising the mean square error of predictions and the real values from data, [13].

$$y_j = \sum_{i=1}^{n} a_i x_i + b + \epsilon_j \tag{3}$$

Although linear regression is a simple method it has several advantages. It is possible to get statistical information like confidence intervals for the estimated coefficients which explains whether the coefficients and hence the dependence of different explanatory variables are statistically significant or not. In a linear regression model, it is also easy to see how the model will behave to unseen data. It is especially useful if the modeller has an intuition about the dependence of the dependent variable on the explanatory variable since the modeller controls the shape of the estimated function f of equation 2. Another major advantage is that linear regression works well on small datasets where more complex methods like neural networks fail.

There are a few disadvantages of linear regression. It assumes that the residuals are normally distributed. It is known to not work as good in models where big datasets are used as more complex methods. If the modeller does not have a clear intuition of the dependence between the dependent variable and explanatory variables from physics or similar sources it is hard to decide which shape to use for the estimated function f in equation 2, other methods decide the shape automatically. Linear regression cannot model several nonlinear relationships in a good way. There is for example no way to model saturation effects.

3.2 Feed forward Neural Networks

Artificial neural networks are a set of algorithms loosely inspired by the biological neural networks in a nerve system. Feed Forward Neural Networks (FFNN) were the first type of artificial neural network to be developed. A FFNN can be used to estimate a function f from equation 2 without any prior knowledge of the function shape and can be used both for regression tasks like in this thesis but also for classification tasks. A FFNN consists of many simple functions that are combined in order to form layers of different sizes. In equation 4 the estimated function f is divided into layers where $f^{(1)}$ is the first layer of the network that takes the attributes as input and $f^{(3)}$ is the last layer of the network that outputs the approximated output, [14].

$$f(\boldsymbol{x}) = f^{(3)}(f^{(2)}(f^{(1)}(\boldsymbol{x}))) \tag{4}$$

Figure 6 shows a small FFNN with three layers. From the figure it is easy to see that many layers results in deeper networks. Besides deciding how deep a FFNN network should be, figure 6 also shows that each layer can have a different width. The input layer has one node for each feature being used while the number of output nodes depends on how many outputs there are. This restricts the width of these layers to the number of inputs and the number of outputs. The second layer has no fundamental restriction in width and never sees the input or the output and is therefore called a hidden layer. It is up to the modeller to choose the number and width of hidden layers. The value of node Z is shown in equation 5 where n is the number of nodes in the previous layer, W_i is the weight between node Z and the input x_i, b is a bias term used to remove bias and a is an activation function. A common activation function used is the ReLU function seen in equation 6 that outputs the maximum of the input or 0. This activation allows the layer to learn nonlinear functions which is why it is common, [14].

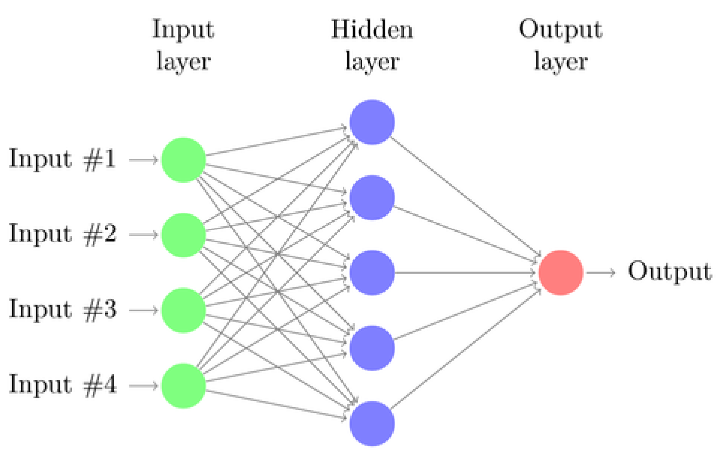

Figure 6: A FFNN with 4 inputs forming the input layer, 5 neurons forming the hidden layer and a single neuron forming the output layer. Note that each neuron is connected to all neurons in the previous layer. Source: texample.net

$$Z = a\left(\sum_{i=1}^{n} W_i x_i + b\right) \tag{5}$$

$$a = g(x) = \max(0, x) \tag{6}$$

To summarise, each node is connected to the previous layer through weights $\boldsymbol{W}$. It is these weights that are trained in order to estimate an optimal function for regression or classification. In order to quantify what a good estimator is an objective function is needed. This function is minimised by a variant of stochastic gradient descent optimization algorithm which requires evaluation of the objective function as well as gradients of the objective function with respect to the weights W which can be found through backward propagation. There are numerous of different objective functions and one common for regression is the mean absolute error function seen in equation 7 where n is the number of samples used for evaluating the neural network, $\hat{y}_i$ is the output from sample i of the estimator while y_i is the true value of sample i, [14].

$$\text{MAE} = \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n} \tag{7}$$

When training and evaluating a neural network, it is important to divide the available data into three sets called training set, validation set and test set. The training set is used to train the network by determining weights. The validation set is used to tune hyper-parameters such as the number of layers and the width of each layer while the test set is used to evaluate the fully trained network. It is important to evaluate on the test set which contains entirely unseen data in order to check whether the network has succeeded in learning to generalise on the training data so it has not only learnt to map training input to training output which is a much easier task.

The advantages of using a FFNN as an estimator include the following: they are superior to other known approaches for many different problems involving huge amounts of data. There is no need to have any prior knowledge of the function to be approximated. The FFNN models nonlinear functions as easily as it models linear functions.

A disadvantage using a FFNN is that they tend to overfit to the data being trained on instead of generalising the information from the training data when little data is available. In cases where the solution is remarkably simple neural networks often fail to learn the simple solution. Another disadvantage is that it is hard to find statistical information from a neural network, these kinds of information must be calculated separately. It is hard to know how a neural network reacts on data that is not similar to the training data since the approximated function usually becomes very big and complex.

3.3 Load curves

A method used by Swedish utility companies to analyse electricity consumption of its customers is called the load curve method. It was introduced by the Swedish electricity association (Svenska Elverksföreningen) in 1991, [1]. The method divides individual customers into different groups and derives standard load curves for each of these groups. Each load curve consists of a daily curve divided into 24 hourly values. For each hour, the curve shows the percentage of average annual energy that hour consumes in average for a year. In order to derive the curves the following assumptions were made: the loads of a customer for each hour are normally distributed and independent of each other. However, each load is dependent on time and the outdoor temperature, [15]. In order to choose a load curve the following input data is needed:

- Customer category
- Season (Winter, Spring/Fall, Summer)
- Weekday (Workday or weekend)
- Temperature of given day
- Average annual power of the customer

Note that the method depends on the load curves derived by the Swedish electricity association in order to be applicable in the entire country of Sweden. These curves are only given for up to three different temperatures per season and therefore a new curve must be created from the given curves when a new temperature is given. This is done through linear interpolation/extrapolation between the curves whose temperatures are closest to the given temperature. In order to clarify how type curves are used an algorithm is given below which shows how to make predictions with load curves for a specific customer on a specific day.

1. Normalise the measured yearly energy consumption for the customer by transforming it into what it would have been at the reference point (Arlanda, Stockholm). This is done with equation 8 where E_{an} is the normalised annual energy, E_{meas} is the measured annual energy, φ is a number between 0 and 1 representing the amount of energy that is temperature dependent at the reference point, G_{meas} is the average annual heating degree day number at the measured point and G_{np} is the average annual degree day number at the reference point. Degree day, or heating degree day is used to quantify the big amount of electricity heating in Sweden. A large number of degree days for a certain year and geographic location means that a large amount of heating has been needed that year at that location which in average results in higher load.

$$E_{an} = \frac{E_{meas}}{1 + \varphi_{np}(\frac{G_{meas}}{G_{np}} - 1)} \tag{8}$$

2. Find the average load P_{av} with equation 9 where h is the amount of hours of given year and E_{an} is the normalised annual energy.

$$P_{av} = E_{an}/h \tag{9}$$

3. Decide which load curve is appropriate depending on category, season, workday or weekend. Transform each value on the curve to the temperature of the given day with equation 10 where T is the current temperature, T_1, T_2 are the two closest given temperatures, $P_{type}(h, T_2)$, $P_{type}(h, T_1)$ are the loads from the already known load curves while $P_{new}(h, T)$ is the interpolated or extrapolated load from the new load curve corresponding to T.

$$P_{new}(h, T) = \frac{P_{type}(h, T_2) - P_{type}(h, T_1)}{T_2 - T_1} \cdot (T - T_1) \tag{10}$$

4. Multiply the average load P_{av} with all hours of the type curve $P_{new}(h, T)$ to get predicted load of all hours of the day.

3.4 Extreme value theory

The previously described methods have all been developed for estimating load at each hour from attributes like time and temperature. Such models are good for estimating normal electricity consumption but fail to estimate extreme cases such as the minimum load and the maximum load which are highly important quantities for dimensioning of the power grid. In order to answer questions such as ”What is the maximum or minimum expected load in 3 years?” in a mathematical acceptable way extreme value theory is needed.

Equation 11 shows the maximum of all measured loads for a limited time period where X_1 is the first measured load for a substation and X_n is the last measured. If the real probability distribution behind the loads were known it would be possible to calculate the probability of M_n being bigger than the value of z with equation 12 where $F(z)$ is the distribution function of the loads and n is the number of observations of loads. However, this function is not known and estimating it by standard statistic methods is not reliable since a small error in the estimated $F(z)$ becomes big for $F(z)^n$, [16].

$$M_n = \max\{X_1, ..., X_n\} \tag{11}$$

$$P(M_n < z) = F(z)^n \tag{12}$$

Another approach is to look at approximations of $F(z)^n$ which can be estimated based on only extreme data. Finding these approximate distributions always results in the Generalised Extreme value (GEV) distribution which is a family of three distributions that can be written as one. The members are the Gumbel, Weibull and Fréchet distributions. The GEV distribution is the extreme analogue of the normal distribution for sample means. Equation 13 shows the GEV-distribution where ϵ is the shape parameter, u is the location parameter and σ is the scale parameter. When $\epsilon = 0$ the GEV-distribution results in a limit which is the Gumbel distribution. These parameters are fitted to extreme data by maximum likelihood. The weakness of this approach is that it assumes that the number of observed extremes n is sufficient which is hard to guarantee, but the advantage is that if n is large enough the GEV-family will always be a good approximation for the extreme values, [16].

$$G(z) = \exp\left\{-\left[1 + \epsilon\left(\frac{z-u}{\sigma}\right)\right]^{\frac{-1}{\epsilon}}\right\} \tag{13}$$

In order to use the GEV-distribution for extreme value analysis of hourly collected load all available data must be put into n blocks. For each block, the block maximum is found and used to fit the parameters of the GEV-distribution. This needs to be done since the GEV-distribution takes an extreme value as input and outputs the probability for that input being the maximum value for the next time period, the time period is defined by the size of the sequential blocks. There is no general guidance about how much data should be used in each block. Huge blocks that use large periods of data fulfil the assumptions of using the GEV-distribution better and therefore results in less model bias. However, much data is lost when using large blocks resulting in large variance of the model which can be reduced by using smaller blocks of data. Small blocks do not fulfil the assumptions of using the GEV-distribution as much as bigger blocks since a block maximum might not be a true extreme value and therefore there is a risk for model bias using small blocks, [16]. The choice of amount of blocks and their size is a compromise and the modeller has to try different choices to find the optimal choice.

4 Implementation

In this section the implementation of all methods used in the thesis are described. Note that in order to choose the best implementations, some results had to be used. Although a few implementation choices are made from preliminary results, the results shown in the Results and discussion section only show the results of the final implementation of different models. In this section it is also explained how the models whose theory is described in the theory section are used in practice. All models have been implemented using Python 3.

4.1 LR and FFNN

A combination of the linear regression method and a feed forward neural network method was developed. This model is called ”LR and FFNN” and is constructed in the following way:

1. Use linear regression according to equation 14 where $\hat{y}_{lin}$ is a first approximation of the station load based on the average living area A_m and number of customers connected to the station N. There are mainly two reasons for why linear regression is used with these attributes. First, it is expected that a bigger average area and a bigger number of connected customers scale the load in a linear way, second these attributes differ only between stations and hence there are only 11 possible unique values to learn from since there are only 11 stations in the dataset. This is far too few stations to make use of Neural Networks or other complicated regression models in a robust way and therefore linear regression is the best model to use for these attributes.

$$\hat{y}_{lin} = a \cdot A_m + b \cdot N + c \tag{14}$$

2. Use a feed forward neural network to model the contribution of time in the form of month, weekday, hour of the day and also temperature since these attributes are expected to affect the load regardless of which station is being modelled in a similar way. These attributes have a lot of different possible values and there are numerous observations of the load from different stations to learn from which justifies the use of a neural network on these attributes. Another reason to use a neural network is that it is not certain that these attributes models load linearly. Beside these attributes the predicted load from the linear regression model is passed on as input to the neural network.

4.1.1 Preliminary analysis

Before finding the optimal implementation of LR and FFNN, the predictive power of each attribute described previously had to be tested. This was done by first including the time attributes in a simple neural network model. Then weather attributes were added one by one and kept only if they reduced the mean absolute error of the model. This resulted in keeping temperature as the only weather attribute since the others did not improve the model accuracy, see table 2 for details.

The other attributes take different values depending on which station is being used and therefore they can only have 11 unique values since there are only 11 stations in the dataset. For example, number of customers can only differ between stations since the number of customers is constant during the entire analysis period for a particular station. Because of this these attributes were tested with a simple linear regression model that used only number of customers connected to each station first. Then attributes that depend on which station is chosen were added one by one and kept only if they reduced the mean absolute error of the linear regression model. This resulted in keeping only number of customers and average living space as attributes, see table 2 for details.

Table 2: The resulting mean absolute error of LR and FFNN by using different setups of attributes for estimation. Final conf. is the model used with the final attributes.

Attribute configuration	mean absolute error (kWh)
Final conf.	14.8
Final conf. and solar radiation	15.2
Final conf. and wind speed	25.3
Final conf. and relative humidity	23.1
Final conf. and average income	32.5
Final conf. and average price for residential property	17.4
Final conf. and percentage of highly educated inhabitants	18.9
Final conf. and average main fuse size of customer	14.4
Final conf. except including temperature	22.8
Final conf. expect including average living space	32.8

In table 2 it is seen that the final configuration of attributes results in the smallest mean absolute error except using the same configuration and average main fuse size of customer. Therefore this model was chosen for three reasons. Average main fuse size of customer did not improve the mean square error by even 1kWh and the attribute is not easily obtainable.

Note that the attributes used in the other subsections were the ones used in the final models, all others have been removed since they do not improve the estimations.

4.1.2 Details of LR and FFNN

The neural network was implemented using the Keras library, [17] with TensorFlow, [18] as back-end. It consisted of an input layer of 5 attributes which are repeated in the following list.

- Output of linear regression model
- Hour of the day
- Month
- Weekday
- Temperature

The neural network used two hidden layers where the first hidden layer had a width of 90 nodes. The second hidden layer had a width of 18 nodes and both hidden layers were activated by the ReLU activation function and included a bias term. These choices were made by trying different amount of layers first, resulting in 2 as the optimal amount before severe overfitting was observed. Then the width of the first hidden layer was varied to find an acceptable width based on mean absolute error and overfitting. Note that finding an optimal width is a well known problem of using neural networks since there is no way to guarantee that an optimal width has been found, [19]. After this the width of the second hidden layer was chosen in the same way. The output layer consisted of one node with a bias term but no activation function applied. This node had as output the final estimation of load. The model was trained using 70 epochs. An epoch is defined as the number of times the optimization algorithm goes through all available training and validation samples. All hyper parameters used in the neural network such as number of layers and the width of layers have been chosen to minimise the mean absolute error function. This is also the objective function used by the neural network. The optimiser used for optimising the network is RMSprop, which is a widely used optimization algorithm. As mentioned before, the attributes finally used in the model were chosen based on how much they reduced the mean absolute error.

In order to avoid overfitting, the dataset was divided into training set, validation set and test set. This was done in two different ways. First all samples from the different stations were shuffled. The shuffle was done so that the time series from the 11 different series were shuffled between each other, not only reordered within each time series. This means that the samples belonging to different stations were mixed together. Then the training set was chosen as 60% of the dataset samples and validation and test set was chosen as 20% samples each. This was done to evaluate how good the network models existing stations.

After this the dataset was divided so that the test set consisted of all available data of one station and so that the training set and validation set consisted of 80% and 20% respectively of the remaining stations. This was done to evaluate how good the neural network could predict load of an entirely unknown station. All sets in both implementations were normalised by equation 15 where A_{nij} is the normalised ith attribute of sample j, A_{ij} is the original value of the ith attribute of sample j, μ_i is the average of attribute i and σ_i is the standard deviation of attribute i. The normalisation is done to simplify the learning process of the neural network.

$$A_{nij} = \frac{A_{ij} - \mu_i}{\sigma_i} \tag{15}$$

4.2 Load curve methods

The implementation of load curves is divided into two parts. The method is first implemented with load curves from Svenska Elverksföreningen and will be referred to as LC1. The curves used are from the following four categories:

- Resident owner without direct heating
- Resident owner with direct heating with a building older than 1980.
- Resident owner with direct heating with a building newer than 1980.
- Apartment owner without direct heating

Note that resident owner in the list above is referred to as house owner. These curves were chosen since they were the ones available that contained categories that are reasonable to find in a station supplying a residential district.

Since the loads in this thesis are on station level instead of customer level it was concluded that using only a single curve would not give good results since every station is connected to several types of customers. This was validated by using only one curve when predicting loads according to the algorithm described in the theory section. To solve this issue the predictions were made for all four categories and then a new prediction was made as a weighted sum of the four individual curves according to equation 16 where A, B, C, D are constants and P_{Ci} are the predictions from curve i.

$$P_{optimal} = AP_{C1} + BP_{C2} + CP_{C3} + DP_{C4} \text{ where } A + B + C + D = 1 \text{ and } A, B, C, D > 0 \tag{16}$$

In order to find optimal values of A, B, C and D one million different combinations were tested and the combination that resulted in the smallest average absolute error were kept. This method was repeated a few times in order to check that the results were consistent.

Since hourly values of the stations are known in this thesis the load curve method was also implemented by constructing new load curves especially fit for the station data. This implementation is referred to as LC2. A new load curve was made for each station and this was done in the following steps:

1. Divide the hourly load values depending on which season they are in and depending on if the hourly value is in a workday or weekend. This results in six different tables of loads when using the seasons winter, spring/fall and summer.
2. For each table divide the table further into two separate tables where one table consists of hourly values corresponding to days where the average temperature is lower than the median average temperature. The other table consists of hourly values corresponding to the days with higher average temperature than the median average temperature of a day.
3. For each temperature divided table, calculate the average temperature. This is the temperature that will be used for the calculated curve.
4. In order to calculate the load curve, calculate the average load of every hour in the temperature divided table. This will result in 24 values for each temperature value which in turn gives two curves to use for interpolation and extrapolation described in the theory section of load curves.

The constructed load curves are used for predictions in the same way as the given ones from Svenska Elverksföreningen.

4.3 Extreme value model

The extreme value model can be implemented on customer level, station level or as a sum of all stations-level. It was evaluated on a single station with 114 customers in this thesis. The blocks with sequential data was set to include two weeks of data each. The maximums of these blocks were used with maximum likelihood to fit the GEV-distribution with a Python library called sci-kit extremes, [20]. Since the most extreme loads in the dataset always occur in winter, only winter-data between October and April is used. This is an attempt to make the data stationary to improve the results of the analysis. The length of the blocks was decided from analysis of QQ-plots and PP-plots as well as power density plots/histograms of the resulting GEV-distribution for different block lengths. The block length that resulted in best fit for the data was chosen.

5 Results and discussion

In this section the results from the different implementations are shown and discussed. The results are divided into two subsections where the first subsection shows results from the regression models and the second subsection shows results from the extreme value analysis model.

5.1 Regression models

In table 3 and table 4 the resulting residuals and absolute value of the resulting residuals can be seen respectively from the load curve methods and LR and FFNN trained to estimate unseen data from seen stations and LR and FFNN trained to estimate unseen data from an unseen station. Note that the residual r_i is defined according to equation 17 where $\hat{y}_i$ is the model predicted value of sample i and y_i is the correct value of sample i which is common in literature. A positive residual means that the model has underestimated the real load and a negative residual means that the model has overestimated the load.

$$r_i = y_i - \hat{y}_i \tag{17}$$

Table 3: The residual distributions from the different implementations described in the implementation section. The count row shows the amount of residuals used for calculating the other metrics.

Information statistic	LC1	LC2	LR and FFNN	LR and FFNN predicting unseen station
count	88176	88176	17328	7876
mean (kWh)	-6.9	19.4	1.4	0.6
std (kWh)	28.9	37.0	13.9	18.3
min (kWh)	-203.0	-209.0	-139.1	-50.5
25% (kWh)	-21.2	-1.8	-6.2	-13.4
50% (kWh)	-4.2	16.0	0.3	-2.3
75% (kWh)	11.3	39.4	8.0	12.9
max (kWh)	169.6	219.7	87.5	97.5

Table 4: The absolute value residual distributions from the different implementations described in the implementation section. The count row shows the amount of residuals used for calculating the other metrics.

Information statistic	LC1	LC2	LR and FFNN	LR and FFNN predicting unseen station
count	88176	88176	17328	7876
mean (kWh)	21.7	30.9	9.9	14.8
std (kWh)	20.2	28.1	9.7	10.7
min (kWh)	0.0	0.0	0.0	0.0
25% (kWh)	7.6	10.2	3.1	6.8
50% (kWh)	15.9	22.3	7.0	13.2
75% (kWh)	29.5	44.1	13.8	20.3
max (kWh)	203.0	219.7	139.1	97.5

In figure 7, 8, 9 and 10 the ordered values of the residuals of the load estimation can be seen for the different methods. The different plots do not share the same figure since not all use the same amount of samples when plotting the curves.

Figure 7: The residuals of LC1 when estimating the load of all stations.

Figure 8: The residuals of LC2 when estimating the load of all stations.

Figure 9: The residuals of LR and FFNN when estimating the load of all stations part of the test set. The data estimated comes from stations that have been used for training LR and FFNN. However, the hours which are estimated have not been seen by LR and FFNN in training

Figure 10: The residuals of LR and FFNN when estimating the load of an unseen station.

In figure 11, 12, 13 and 14 the ordered absolute values of the residuals of the load estimation can be seen for the different methods. The different plots do not share the same figure since not all use the same amount of samples when plotting the curves.

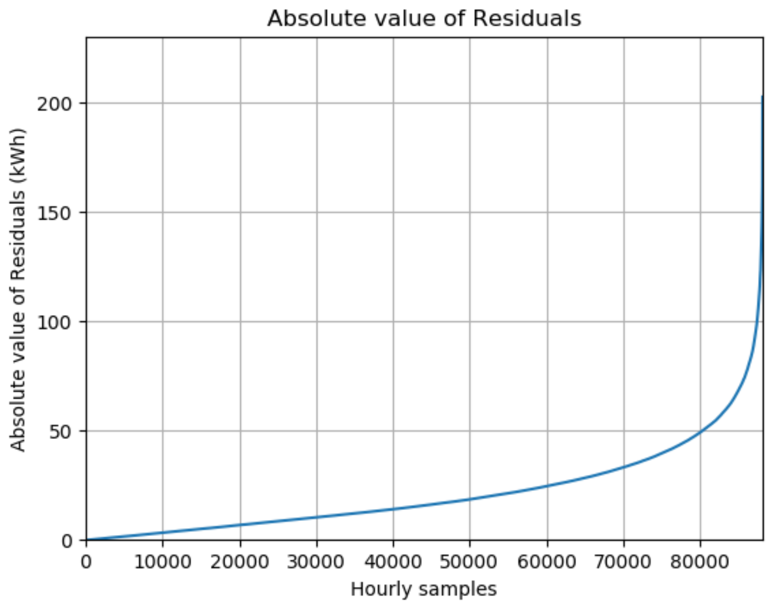

Figure 11: The absolute value residuals of LC1 when estimating the load of all stations.

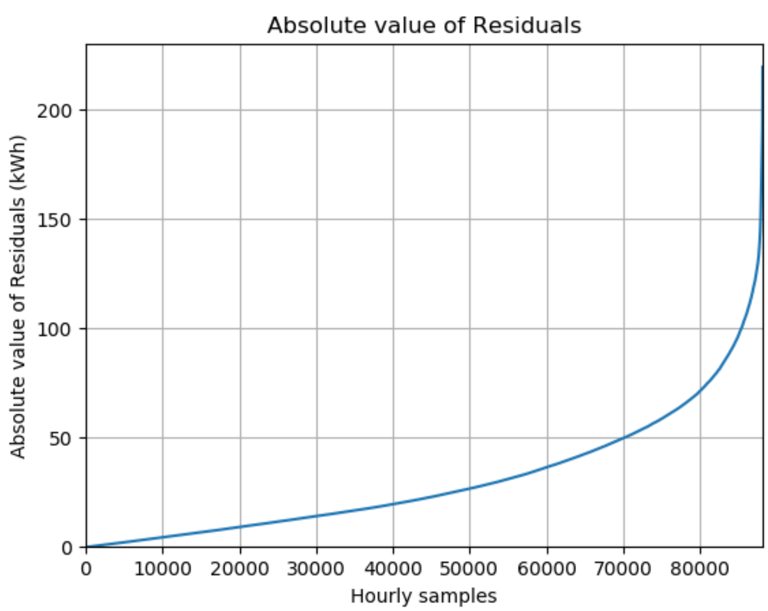

Figure 12: The absolute value residuals of LC2 when estimating the load of all stations.

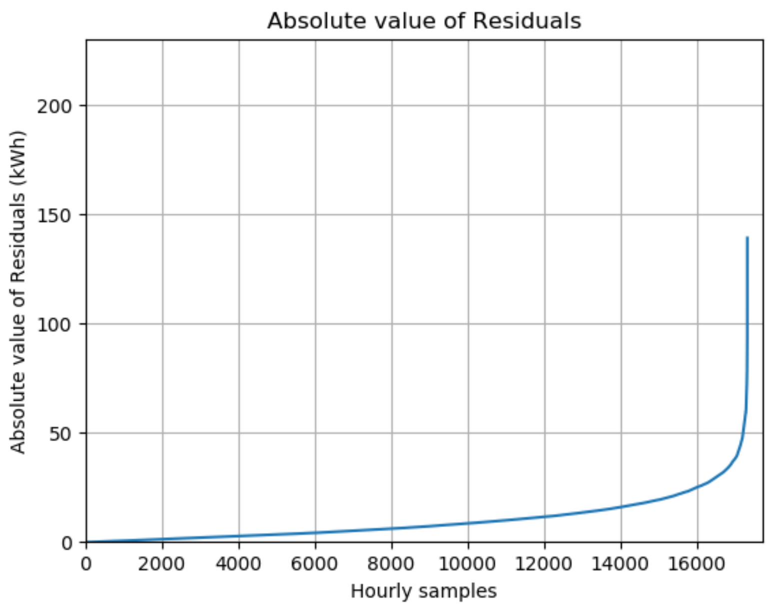

Figure 13: The absolute value of the residuals of LR and FFNN when estimating the load of all stations part of the test set. The data estimated comes from stations that have been used for training LR and FFNN. However, the hours which are estimated have not been seen by LR and FFNN in training

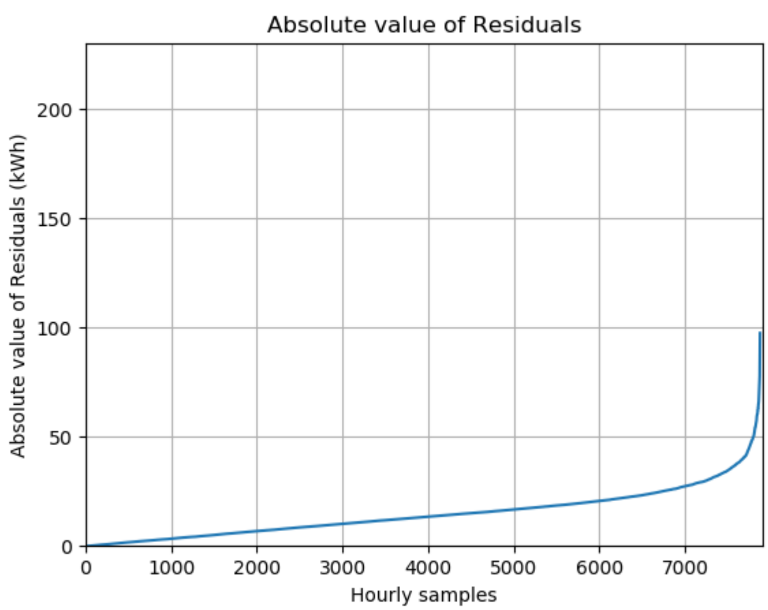

Figure 14: The absolute value of the residuals of LR and FFNN when estimating the load of an unseen station.

In order to gain information of the uncertainty of the model it was noted that the residual distribution was close to normal distributed and independent of each other, except in the tails of the distribution as seen in figure 15. Beside predictions a confidence interval of 95% was therefore added based on the residuals being normally distributed and independent of each other for each season.

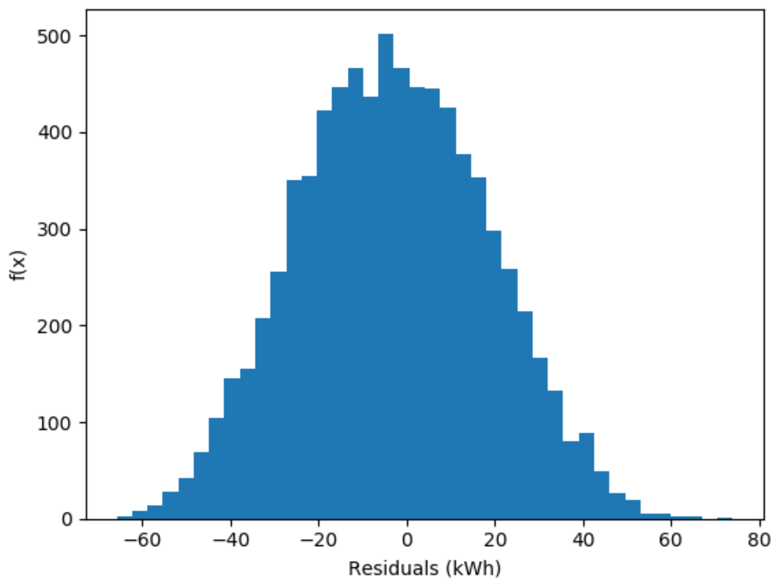

Figure 15: A histogram of the residuals of LP and FFNN. Note that the histogram shape is close to normal but the tails are a bit skewed in contrast to a real normal distribution.

In figure 16 the predictions made by LR and FFNN can be seen for the unseen station including approximate 95% prediction intervals. The predictions are made on the station whose values are plotted in figure 3 and 4. Figure 17, 18, 19 and 20 shows realisations of all four estimation methods together with the real load for the same weeks as figure 4 shows. By comparing the figures it is noted that LR and FFNN is never able to predict the highest loads. One possible reason for this is that none of the attributes collected contains the information needed for explaining loads deviating far from the mean load. It is noted that LR and FFNN misses the common load peaks in the morning and also that it estimates to much load during night. From both figures and table 4 it is noted that despite this LR and FFNN is performing much better than the load curve method since LR and FFNN is able to predict the average load accurately. It is also noted that by including the confidence interval almost all loads are covered by LR and FFNN in figure 16. The median error is 15.3% of the real load of the station and the average error is 16.0% of the real load for using LR and FFNN for an entirely unseen station.

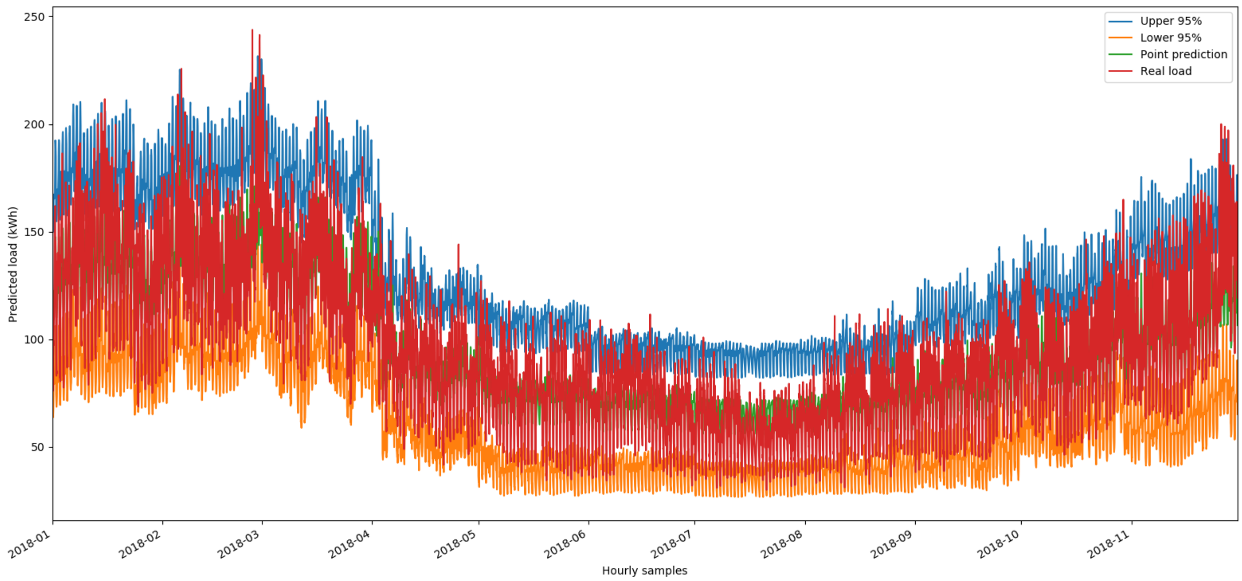

Figure 16: The predictions of LR and FFNN when estimating the load of an unseen station. The upper and lower graphs defines an approximate prediction interval covering the real load in 95% of all cases. The green graph shows the most likely value of the load and is the output of LR and FFNN. The red graph shows the real loads.

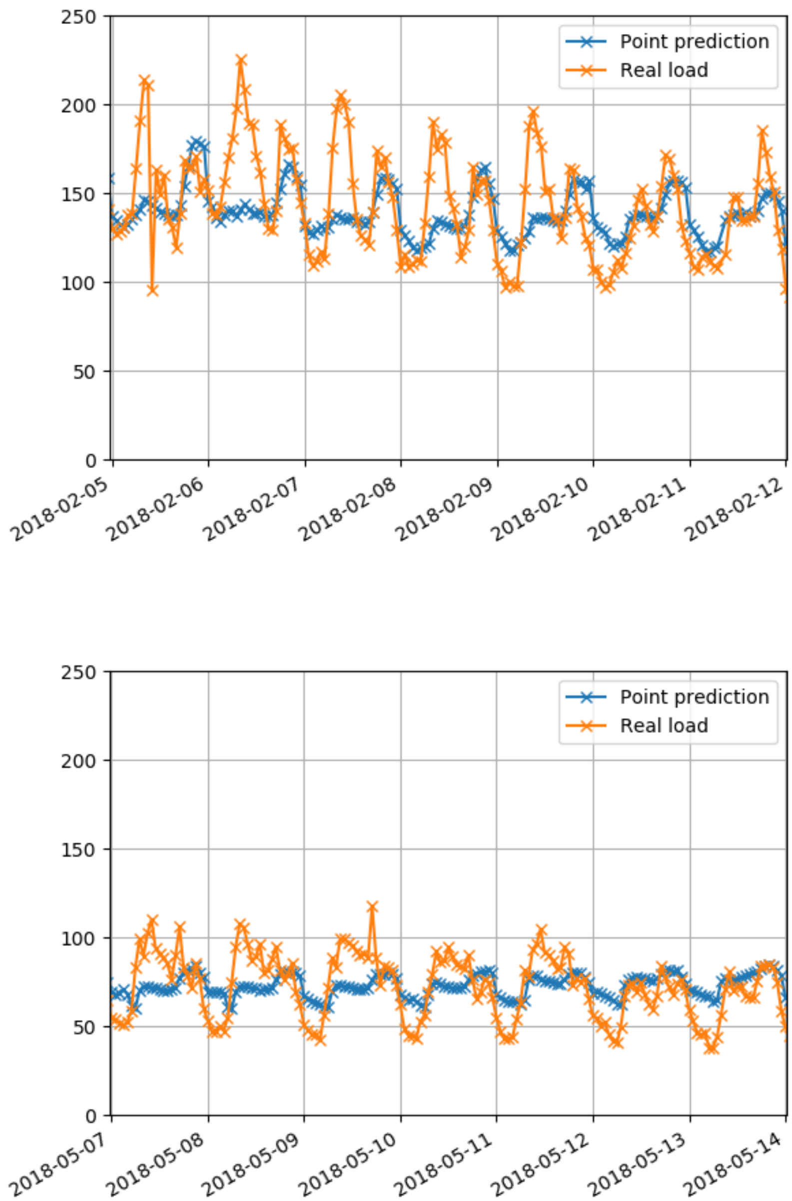

Figure 17: The predictions of LR and FFNN when estimating the load of an unseen station together with the real loads for a week in February and a week in May.

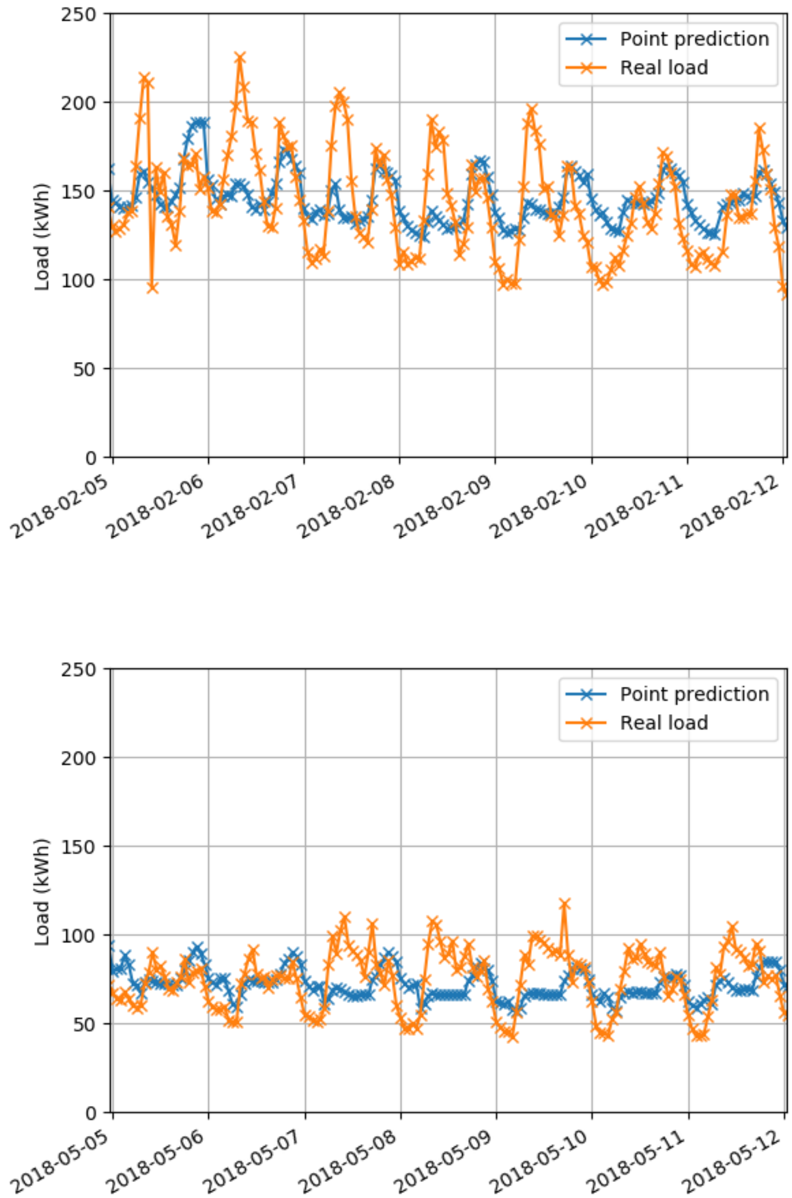

Figure 18: The predictions of LR and FFNN when estimating the load of a seen station but with unseen data together with the real loads for a week in February and a week in May.

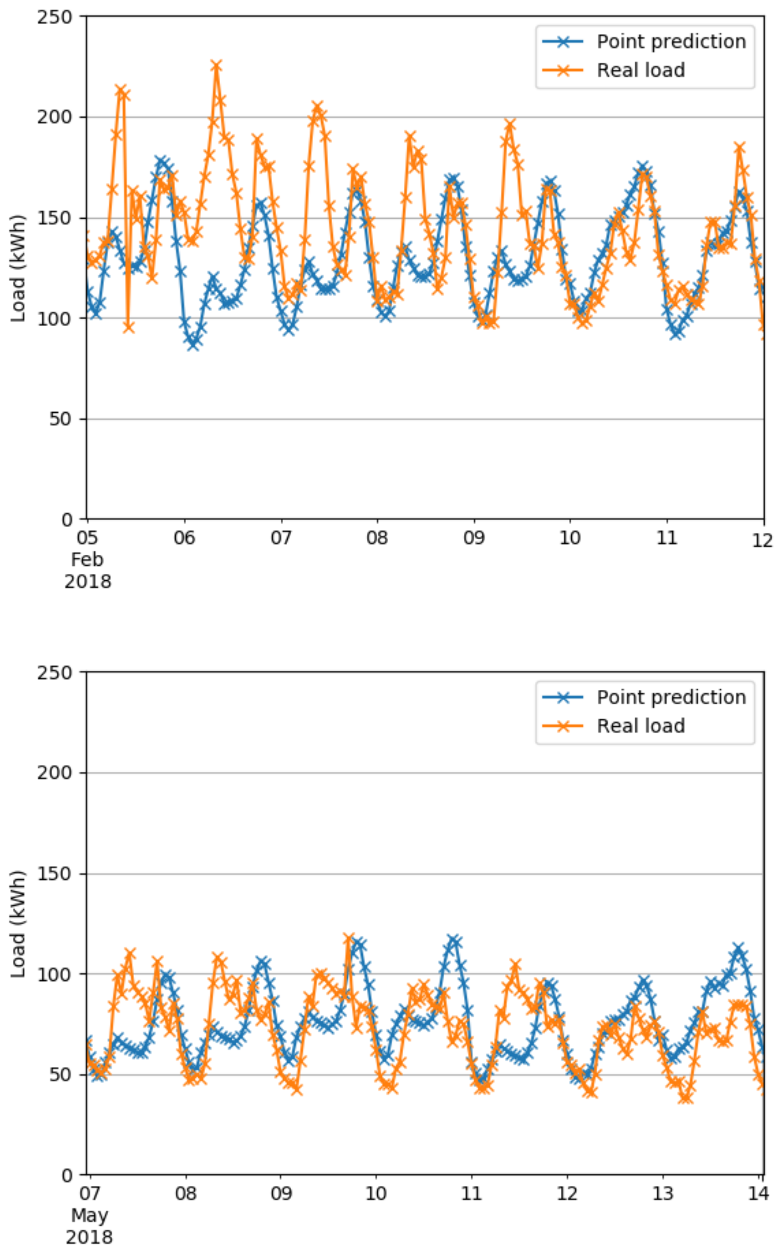

Figure 19: The predictions of LC1 when estimating the load of a seen station but with unseen data together with the real loads for a week in February and a week in May. Note that LC1 is predicting night time load better than LR and FFNN but performs worse at other times of the day.

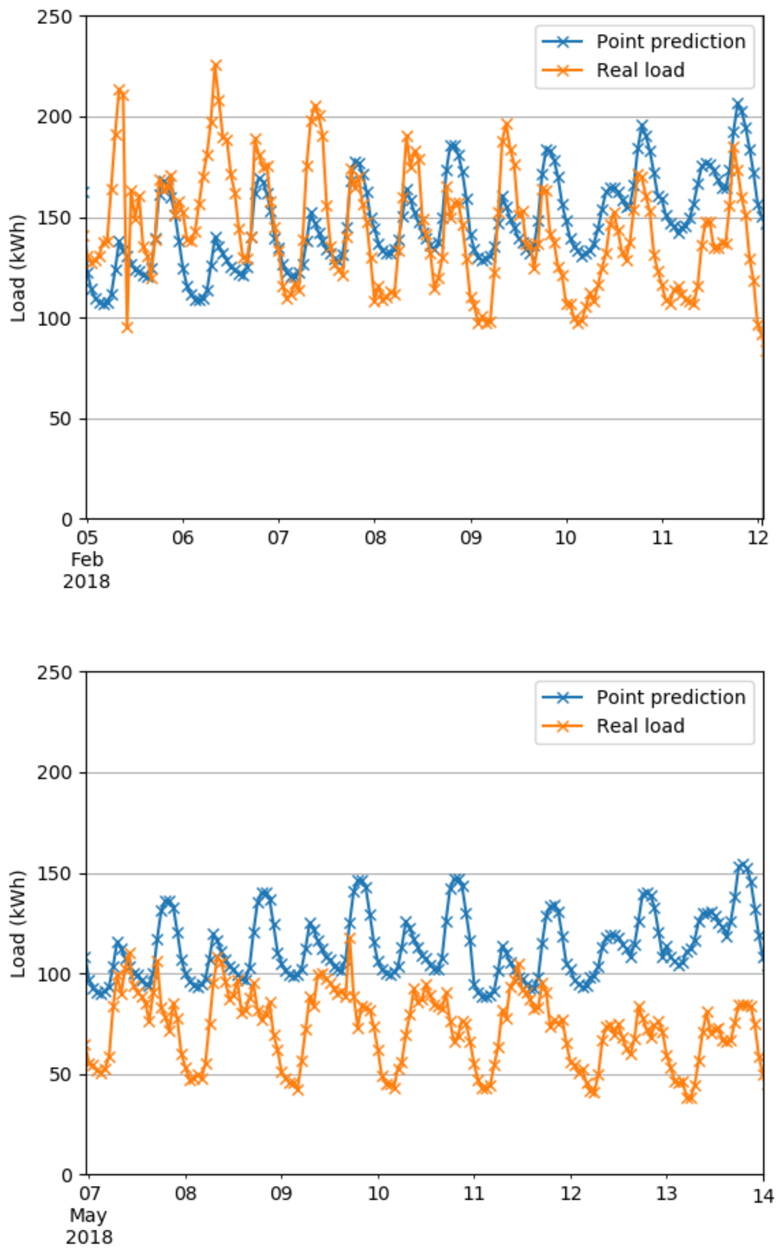

Figure 20: The predictions of LC2 when estimating the load of a seen station but with unseen data together with the real loads for a week in February and a week in May. The method performs considerably worse than other methods and misses the average load during both week and day.

LR and FFNN performs better when estimating load of an entirely unknown station than LC1 and LC2 does when estimating load on the same stations they were trained to estimate load on. The mean absolute error is 29.4% lower when using LR and FFNN for predicting load compared to using LC1 which was the best performing load curve method. This is likely caused by the fact that the load curve method only uses daily temperatures. Another explanation could be that the loads in the load curve method are assumed to be normal distributed which is not the case. When evaluating LR and FFNN on stations it has seen before the comparison shows a more accurate performance difference and in this case the mean absolute error is 54% smaller compared to LC1. Besides comparing the mean absolute error it can be seen that LR and FFNN outperforms the other methods by having smaller residuals for all quantiles of the residual distribution, see table 3 and 4. The maximum residual is for example considerably smaller for LR and FFNN.

The resulting attributes used in this thesis that gave smallest residuals for LR and FFNN was found to be month, weekday, hour of the day, temperature, number of customers and average living area. All other attributes were discarded when evaluating different models. The resulting attributes might have been more if there had been more stations available. Attributes with smaller predictive power than those used might not be significant in a small dataset and could be more significant when evaluating models that use fifty or hundreds of stations. The attributes average income, average price for residential property and percentage of highly educated inhabitants and average year of the buildings in the area being built was collected from Hitta.se and it was not possible to see how large groups were used to compute those attributes. There is a risk that the groups were so big that they covered several stations which could explain why they had no significant predictive power in this study.

One hypothesis was that using LC1 would give an indication of how many customers were apartment owners, resident owners without direct heating, resident owners with direct heating with buildings older than 1980 and resident owners with direct heating with buildings younger than 1980 by examining the constants obtained by LC1 from equation 16. However, different combinations of constants resulted in the same optimal mean absolute error distribution so the method could not be used to estimate the relative amount of customers belonging to each of these categories.

5.2 Extreme value model

The results from the extreme value model for one station are seen in figure 21. The return level plot shows the expected return in the return period defined by the x-axis. In other terms it shows the expected maximum load observed in a time period defined by the x-axis. Since the x-axis has a unit of two weeks this means that $x = 10^1$ shows the expected return in 10 times two weeks' time. However, since the model is based on winter data only the real return period is even longer because there are no extreme values in summer. For example, if $x = 10^2$ then the return period is theoretically 200 weeks but a whole year of extremes is characterised only by the winter data of 17 two-week periods, which means that on a yearly basis 200 weeks is equal to $\frac{200}{17} = 11.8$ years. The dashed lines in the return level plot shows the upper and lower 95% approximate confidence interval of the expected return level and are computed with the Bootstrap method. The power density plot shows the fitted GEV-distribution and the histogram of the data. By drawing random values from the fitted distribution, synthetic maximum loads can be shown as in figure 22. However, the return level plot shows more information since it explains what the highest load is expected to be in any time period, not just a possible sequence of extreme values defined by the block size of the block maxima. However, realisations can be used to create synthetic extreme value data.

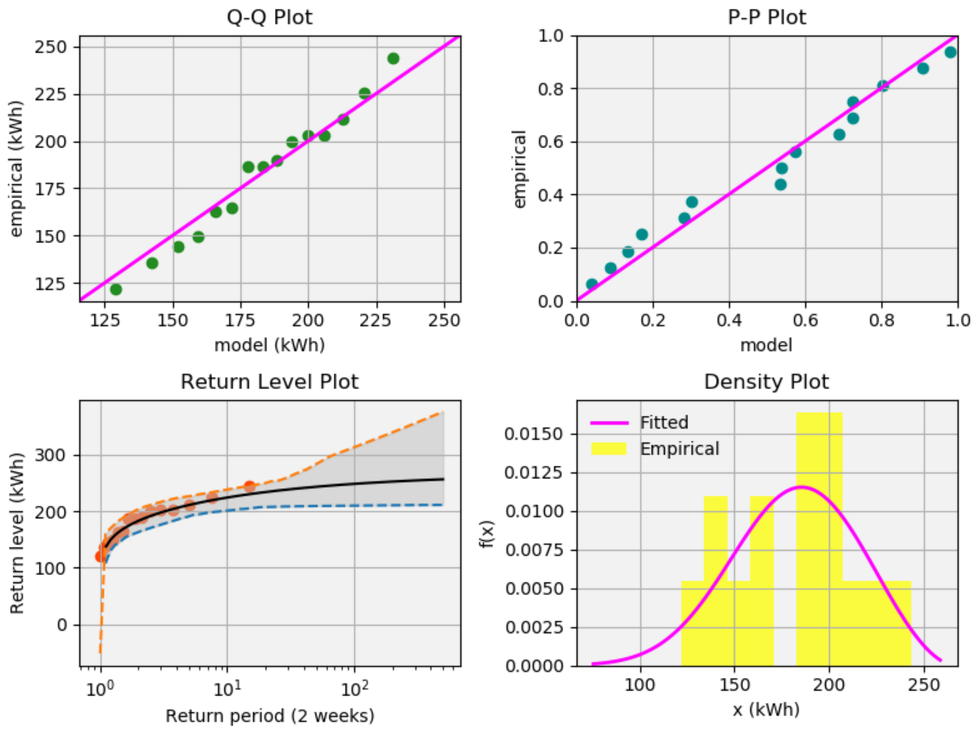

Figure 21: The result of the extreme value model in terms of a return plot, QQ-plot, PP-plot and power density plot.

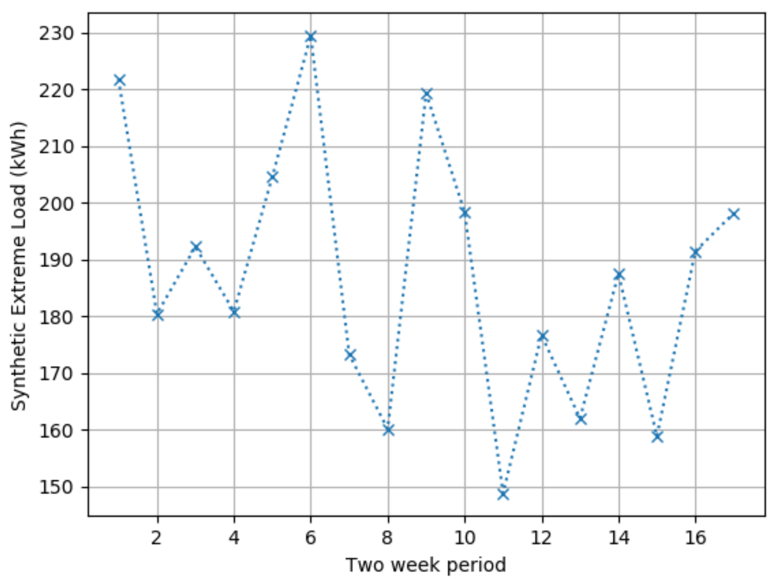

Figure 22: A realisation of the fitted GEV-distribution corresponding to synthetic maximum data of one winter-period consisting of 17 two-week periods for one station. It corresponds to likely extreme loads of the winter period of figure 16 when looking at the red graph.

Although the author of this thesis is convinced that using extreme value theory is the most appropriate approach for modelling extreme loads when hourly load observations are available, there are several sources of errors that influence the results obtained in this thesis. There were only data available for a time period of 11 months which is remarkably little data when modelling extremes and no observations are seen from the month of December. It is so little data that it is not certain that any truly extreme loads have been seen during the observed period which makes estimation of maximum return level for periods like 10 or 20 years unreliable. The small amount of data is also shown in the results especially in the power density plot which shows that the fitted distribution does not match the empirical loads perfectly but also in the confidence intervals that grow considerably for longer return periods. The reader is reminded that using too little data risks violating the assumption that the number of sequential blocks used in the analysis are enough for the GEV-distribution to be a perfect approximation. Although these problems should be taken seriously the resulting model fits the data good enough to be a valid model for estimating likely extremes in a short span of time like three to five years, especially when taking the confidence interval into consideration.

6 Conclusions

This thesis has presented a model which combines the strengths of linear regression and feed forward neural networks to estimate load which outperforms the previously used method of load curves. It has also presented an extreme value theory model using block maxima and the GEV-distribution which does not need any of the assumptions necessary for using an older model like Velander's formula. The objective of this thesis was to answer the following:

- Is it possible to characterise the load of secondary substations of Kraftringen using a finite set of easily available attributes?
- Is it possible to use electricity meter data from individual customers to build a model that uses these attributes to estimate the load of the stations with a mean absolute error of less than 15%of the average load?
- Is it possible to estimate the load of an unseen station by using the same model and attributes with a mean absolute error of less than 15% of the average load?
- Is it possible to generalise such a model further by adding a stochastic distribution to model not only average values but also extreme values?

The results have shown that it is possible to estimate the electricity consumption of the medium voltage power stations by using temperature, number of customers connected to the station, average living area of the customers connected to the station, month, weekday and hour of the day with a mean absolute error of only 10%. The best model could only achieve a mean absolute error of 16% on an entirely unseen station but since the results are based on generalising data from only 11 stations it is believed that the same model can show results underneath 15% by including more stations in the process of training the model.

An extreme value model based on block maxima and the GEV-distribution has been used to successfully model all extreme values observed from a station by capturing them in a return level plot with 95% confidence intervals using blocks of two weeks based only on winter data. The models shown in this thesis can be used by researchers and utility companies to generate expected load of substations and also to model extreme values of load.

7 Future work

In this section ideas to future work are presented.

Estimate load on other utility companies

The dataset used in this thesis consisted of only 11 stations spanning for 11 months of the year of 2018. In a future study LR and FFNN should be trained and evaluated on a bigger dataset to see its full potential of estimating load of an unseen station, preferably on a station owned by a different utility company than Kraftringen. An accurate model could be used by small utility companies that do not collect hourly values of load to obtain accurate hourly load profiles. However, the model could also be used by a utility company that does collect hourly values of load to construct reliable load profiles for new settlements when planning for new stations.

It is important to include training of December data to get a reliable model that can estimate the load of all months of the year accurately.

Improve LR and FFNN at night and during morning peak

As noted in the discussion LR and FFNN misses the morning peak and outputs to high load during night. It would be interesting to further study why this happens and try to improve the estimations during these periods, especially since a morning peak in a residential district is common. With more data available it would be possible to validate whether adding a night bias and a morning bias respectively would improve the results in a robust manner.

Include other types of stations

It would be interesting to include stations with different kinds of customers such as industries, farms and shopping malls and fit a similar model to LR and FFNN to include these stations in the same model. This would extend the existing model of this thesis to accurately model the entire power system in a selected region of interest.

Optimize attributes

Finding other attributes than those used in this thesis would probably improve the predictive accuracy of the model. It was not possible to get reliable demography information about the customers like age although other studies have concluded that these can improve predictions of load, [9]. Another feature that was not possible to use in this thesis was the amount of customers for each station that had electricity heating. If information about heating was collected the predictive power could increase further. Another proposal would be to build a model that automatically classifies customers as customers with or without electricity heating to get access of this attribute.

Comprehensive extreme value analysis

During this thesis it has not been possible to get load information for more than 11 months. When possible, the extreme value model should be evaluated on all measured hourly data available to get a robust extreme value model. If trends are discovered when using years of data it is possible to extend an extreme value model consisting of the GEV-distribution and block maxima to include trends and non-stationary scale parameters.

8 References

[1] Svenska Elverksföreningen. Belastningsberäkning med typkurvor. Svenska Elverksföreningen report, Stockholm, 1991.

[2] P. Eriksson. Automatiserad kundkategorisering och anpassade typkurvor. Bachelor thesis, Karlstads Universitet, 2014.

[3] O. Ingvarsson. Hur dimensionerar vi framtidens elnät? Master thesis, Lund University, 2018.

[4] J. Dickert and P. Schegner. Residential load models for network planning purposes. Institute of Electrical Power Systems and High Voltage Engineering, 2010.

[5] S. Ramos, J. M. Duarte, F. J. Duarte, Z. Vale, and P. Faria. A data mining framework for electric load profiling. GECAD – Knowledge Engineering and Decision-Support Research Center, 2013.

[6] J. Helbrink, M. Lindén, M. Nilsson, D. Pogosjan, and J. Ridenour. Kategorisering av elkunder utifrån förbrukningsprofil. ENERGIFORSK, February 2017.

[7] M. Shepero, D. van der Meer, J. Munkhammar, and J. Widén. Residential probabilistic load forecasting: A method using gaussian process designed for electric load data. *Applied Energy*, 218(15):159–172, 2018.

[8] S. Muzaffar and A. Afshari. Short-term load forecasts using lstm networks. Masdar Institute, Khalifa University, Abu Dhabi, 54224, UAE, August 2018.

[9] Y. Wang, Q. Chen, T. Hong, and C. Kang. Review of smart meter data analytics: Applications, methodologies, and challenges. Institute of Electrical and Electronics Engineer (IEEE), 2019.

[10] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye. Grid structural characteristics as validation criteria for synthetic networks. *IEEE Transactions on Power Systems*, 218(15):159–172, 2016.

[11] G. G. Pillai, G. A. Putrus, and N. M. Pearsall. Generation of synthetic benchmark electrical load profiles using publicly available load and weather data. *Electrical Power and Energy Systems*, 61:1–10, 2014.

[12] C. Hor, S. Watson, D. Infield, and S. Majithia. Assessing load forecast uncertainty using extreme value theory. 16th PSCC, 2008.

[13] J. Lübeck. *Matematisk Statistik AK för F, I, ; C, D, E mfl Föreläsningsanteckningar i Statistikteori*. Lunds universitet, Matematikcentrum, Matematisk statistik, April 2020.

[14] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[15] G. Brännlund. Evaluation of two peak load forecasting methods used at fortum. Master thesis, KTH Electrical Engineering, 2011.

[16] S. Coles. *An Introduction to Statistical Modelling of Extreme Values*. Springer, 2001. `https://www.gbv.de/dms/goettingen/332343839.pdf`.

[17] F. Chollet et al. Keras. `https://keras.io`, 2015.

[18] M. Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[19] H. Sildir, E. Aydin, and T. Kavzoglu. Design of feedforward neural networks in the classification of hyperspectral imagery using superstructural optimization. *Remote Sensing*, 12:7, 2020.

[20] K. Correoso. Scikit extremes, 2015. Software available from github.com/kikocorreoso/scikit-extremes.