Capacity demand and climate in Ekerö

Development of tool to predict capacity demand under uncertainty of climate effects

Tong Fan

Degree project in
Electromagnetic Engineering
Second Level, 30.0 HEC
Stockholm, Sweden 2007

XR-EE-ETK 2007:003
Capacity demand and climate in Ekerö
– Development of tool to predict capacity demand under uncertainty of climate effects

Master Thesis
by Tong Fan

Master thesis written at KTH, the Royal Institute of Technology, 2007,
School of electrical Engineering

Supervisors:
Lina Bertling, KTH School of Electrical Engineering
Carl Johan Wallnerström, KTH School of Electrical Engineering
Olle Hansson, Fortum Distribution

Examiner: Lina Bertling, KTH School of Electrical Engineering

XR-EE-ETK 2007:003
Abstract

The load forecasting has become an important role in the operation of power system, and several models by using different techniques have been applied to solve these problems. In the literature, the linear regression models are considered as a traditional approach to predict power consumption, and more recently, the artificial neural network (ANN) models have received more attention for a great number of successful and practical applications. This report introduces both linear regression and ANN models to predict the power consumption for Fortum in Ekerö. The characteristics of power consumption of different kinds of consumers are analyzed, together with the effects of weather parameters to power consumption. Further, based on the gained information, the numerical models of load forecasting are built and tested by the historical data. The predictions of power consumption are focus on three cases separately: total power consumption in one year, daily peak power consumption during winter and hourly power consumption. The processes of development of the models will be described, such as the choice of the variables, the transformations of the variables, the structure of the models and the training cases of ANN model. In addition, two linear regression models will be built according to the number of input variables. They are simple linear regression with one input variable and multiple linear regression with several input variables. Comparison between the linear regression and ANN models will be carried out. In the end, it finds out that the linear regression obtains better results for all the cases in Ekerö. Especially, the simple linear regression outperforms in prediction of total power consumption in one year, and the multiple linear regression is better in prediction of daily peak load during the winter.

Key words: Load Forecasting, System modeling, Simple linear regression, Multiple linear regression, Artificial Neural Network
Acknowledgement

The project is performed at KTH/EE within the Reliability Centered Asset Management (RCAM) group, in co-operation with Fortum Distribution, and Ekerö Energi. During the study, several people have contributed and supported to my work, I would like to express my gratitude as below.

First, I would like to thank my supervisors Lina Bertling (KTH), Carl Johan Wallnerström (KTH) and Olle Hansson (Fortum) for help, support and taking time to answer my questions.

Furthermore, thanks to Anders Sjögren, Stefan Hemström from Ekerö Energi for providing input data. I would also like to thank PhD-student Johan Setrèus for his help.

Finally, I would like to thank my family and my friends for their support during the one and half years.

Stockholm, April 2007
Tong Fan
Contents

Abstract ........................................................................................................................................ iii
Acknowledgement .................................................................................................................. v
Contents ...................................................................................................................................... vii
Chapter 1 Introduction .............................................................................................................. 1
  1.1 Problem description ........................................................................................................... 1
  1.2 Background ....................................................................................................................... 2
  1.3 Overview ........................................................................................................................... 4
Chapter 2 Study of historical data ............................................................................................ 5
  2.1 Movement of power demand ........................................................................................... 5
  2.2 Relationship between different parameters .................................................................... 15
  2.3 Influence of weather parameters on power demand ...................................................... 16
Chapter 3 Model description ..................................................................................................... 21
  3.1 Linear regression model .................................................................................................. 21
  3.2 Artificial neural network ................................................................................................. 23
Chapter 4 Model application and numerical results ................................................................. 27
  4.1 Application of linear regression ...................................................................................... 27
  4.2 Application of artificial neural network ........................................................................... 31
  4.3 Numerical results ............................................................................................................ 32
Chapter 5 Conclusions .............................................................................................................. 43
  5.1 Characteristics of power consumptions in Ekerö ........................................................... 43
  5.2 Models comparison ......................................................................................................... 44
Chapter 6 Future work .............................................................................................................. 46
References ................................................................................................................................. 47
Appendix ..................................................................................................................................... 48
Chapter 1 Introduction

System load forecasting is widely used in different areas for it offers the knowledge of future state of power system. Techniques for the forecasting task have been developed and testified by many researches. From the historical weather data got from SMHI (Swedish Meteorological and hydrological Institute) and power consumption data got from Ekerö Energi, it is interesting to investigate their relationship, and used to predict power demand in the future.

1.1 Problem description

The temperature has become warmer during the beginning of the 21st century. Therefore, the need for heating has been all the time well below a normal year. This fact as measured by SMHI at the airport of Bromma is shown in the Table 1.1:

<table>
<thead>
<tr>
<th>Stockholm - Bromma</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Degree days</td>
</tr>
<tr>
<td>2005</td>
<td>3337</td>
</tr>
<tr>
<td>2004</td>
<td>3435</td>
</tr>
<tr>
<td>2003</td>
<td>3450</td>
</tr>
<tr>
<td>2002</td>
<td>3451</td>
</tr>
<tr>
<td>2001</td>
<td>3425</td>
</tr>
<tr>
<td>2000</td>
<td>2998</td>
</tr>
</tbody>
</table>

Table 1.1: Degree days in recent year compared with normal year

Here, degree days are calculated by SMHI and relate to the energy consumption for space heating. The lower the degree days is, the less electricity energy is needed to keep buildings warm. Ekerö has a relative high portion of heating from electricity, which means it has much higher sensitivity to the climate changes. Accordingly, a tool to predict the need for power with reasonable accuracy when there will be a cold
winter is needed extremely. Compared with other years in 21st century, the temperature in 2003 is quite low. Since it has the largest range of input data, the models are preferred to build based on data in 2003.

Ekerö Energi AB operates the whole power system in Ekerö and the basic information of Ekerö now states as:

Table 1.2: Basic information of Ekerö

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total distribution area</td>
<td>400 km²</td>
</tr>
<tr>
<td>Land area for Ekerö</td>
<td>216 km²</td>
</tr>
<tr>
<td>Number of customers</td>
<td>12,700</td>
</tr>
<tr>
<td>Total conductor length</td>
<td>1270 km</td>
</tr>
<tr>
<td>Voltage levels</td>
<td>20/10/0.4 kV</td>
</tr>
<tr>
<td>Number of substations</td>
<td>8</td>
</tr>
<tr>
<td>Number of transformers (0.4 kV secondary)</td>
<td>486</td>
</tr>
</tbody>
</table>

In this report, the aim of the study is based on the hourly values of temperature, wind, cloudiness and matching values of power demand measured at the connection to regional network, to look for numerical expressions, which can represent the relationship between climate and power consumption. In the end, use the expression to predict the power demand in the future. The main tasks of this project are:

- Analyze the characteristics of power demand in the Ekerö area
- Find out different techniques to model the relationship between power demand and climate
- Compare used models and evaluate the accuracy
- Carry out the result in Excel to solve the forecasting problem

1.2 Background

Power companies and system operators usually face both technical and economical problems when they operate, plan, and control their power systems. This is mainly because they need not only to ensure the security of power utilization for the power consumers, but also to maximize the profits of power trading for their own. In order to achieve these goals, the load forecasting has become one of the most significant issues in decades.

Due to the application of the load forecasting, the period of forecasting may change from hour or day for short-term forecast to week or year for medium and long-term forecast. The short-term forecast is needed for control and scheduling of power system, and also as inputs to load flow study or contingency analysis. The medium-term forecast is used in annual maintenance scheduling and operational planning studies. The long-term forecast is required to determine the capacity of
Generally speaking, it is known that different kinds of aspects such as numbers and sorts of consumers, the price of electricity, people’s daily behaviors and weather conditions will affect the power consumption by means of their own ways. The increasing number of consumers will lead to high power consumptions, at the same time the high price of electricity will decrease the use of power. The power consumption will change in opposite directions in residence and industry during the weekdays and weekends. The weather parameters highly affect the behavior of power consumption for the power utilization of heating and cooling. In this report, it is mainly focus on how the weather conditions affect the change of power consumption and find out their numerical relationships for practical prediction purpose.

Although some authors didn’t consider the weather parameters in their research and solve the forecasting problems by Fourier analysis or time series analysis [2, 3], most researchers tend to find a functional relationship between the weather parameters and the power consumption. The most familiar used weather parameters are temperature, humidity, wind speed, wind direction, cloud cover and sun radiation. Theoretically, the more parameters it considers in the analysis, the more accuracy it obtains. However, in fact, due to the aspects of historical data, the location of investigation, the workload and work time, it is impossible to make a model which can predict the power consumption extremely precisely. Hence, from the point of modeling, the important of this work is to establish empirical relationships between observed variables. Now, the task of prediction can be defined as: Given a description of the system over some period of time and the set of rules governing the change, predict the way the system will behave in the future [4]. This means that the power consumption will be obtained regarding to weather parameters by passing through a certain description, which can be easily expressed as Figure 1.1. Here, x denotes the input variables which are weather parameters, and y denotes the output variables which are power demands in this case.

Different forecasting techniques have been applied and evaluated to solve the problem of prediction [5, 6]. In this report, two techniques are introduced to represent the mentioned description above and use to predict the power demand in Ekerö according to its weather conditions. The first technique is linear regression [7], which apparently knows as linear model. Its benefit is easy to comprehend and has wide facility. The second one is artificial neural network model [8, 9, 10], which is a non-linear model and developed very fast in recent years. Some authors also compared these models in their research work [11, 12]. Both of these approaches will be carried out in the study, and then, after the comparison between these two methods, the better method will be chosen to execute in the Excel.
1.3 Overview

According to the procedure of the study, the thesis is separated to the following steps. The analysis of the characteristics of the power demand in Ekerö will be carried out in the next chapter. Then, chapter 3 states the theories about linear regression and artificial neural network. For linear regression model, it is divided into two parts: simple linear regression and multiple linear regression. Chapter 4 is mainly focus on the application of these models and numerical results are shown after the models are built up. At last, the comparison which is also conclusion will be described in Chapter 5. The suggestions of future work are discussed in Chapter 6.
Chapter 2 Study of historical data

The load changes all the time in power systems. Usually, it is varying year by year according to season changes. People behaviors, for instance, like workday or holiday, weekday or weekend, and day or night, will affect the power consumption since power utilization is a human being activity essentially. The power consumption also varies as a result of sorts of weather conditions. The common characteristics of these phenomena will be studied in the following part.

2.1 Movement of power demand

The rules of changes of power demand differ from kinds of customers in Ekerö. Hereby, three types of terms will be studied: the total power demand, the greenhouse, and the industry. The movement of power demand of these types will be shown from long time period to short time period.

2.1.1 Movement of power demand for total area

The hourly power demand data and weather parameters are collected from 2003 to 2005 in Ekerö. The main customers here are people who live in detached houses. Therefore, the characteristics of the total area will presumably represent this kind of consumers.

First of all, from the power demand data, the power demand for the whole area in the three years can be drawn as Figure 2.1:
The highest value of the power demand in every year occurs in winter and the use of power decreases when summer comes. It happens because the high proportion of electricity is used for heating. When the weather comes cold, the buildings need more electricity power to keep warm. During the summer, the power consumption decreases for heating facilities shutting down. This differs from some other part of the world, where the summer will be extremely hot that the use of air-conditioning is necessary, which brings highest power consumption in one year.

The power system requires supporting the peak load to ensure the security and reliability manners. Thereby, it is interesting to go deep into the highest load in winter, which indicates January, February, March, November and December in this report. The peak loads and the total power demand of the following month in every year are stated as:

![Power demand from 2003 to 2005](image)

**Figure 2.1: Hourly power demand movement from 2003 to 2005**
Table 2.1: Peak and total power demand in five months during winter for every year

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>March</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peak power demand in a hour (KWh/h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>77050</td>
<td>74670</td>
<td>58050</td>
<td>55170</td>
<td>67570</td>
</tr>
<tr>
<td>2004</td>
<td>70520</td>
<td>68760</td>
<td>63450</td>
<td>63210</td>
<td>63280</td>
</tr>
<tr>
<td>2005</td>
<td>60860</td>
<td>66100</td>
<td>72330</td>
<td>56400</td>
<td>64630</td>
</tr>
<tr>
<td>Total power demand in a month (KWh)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>40186340</td>
<td>33881850</td>
<td>30388100</td>
<td>29047530</td>
<td>34247740</td>
</tr>
<tr>
<td>2004</td>
<td>39237340</td>
<td>33839560</td>
<td>31705700</td>
<td>31862990</td>
<td>35044950</td>
</tr>
<tr>
<td>2005</td>
<td>35012130</td>
<td>33621390</td>
<td>35157830</td>
<td>28714280</td>
<td>36553230</td>
</tr>
</tbody>
</table>

The highest load in the three years reached 77050KWh/h in 2003 corresponding to the lowest temperature -24°C occurred. In 2003 and 2004, the peak loads were occurred in January, but this situation changed to March in 2005. This is caused by the special cold weather happened in March in 2005 when the lowest temperature reached -22.9°C. The peak loads descended from 77050KWh/h in 2003 to 70520KWh/h in 2004, and to 72330KWh/h in 2005, which means low down 8.48% and 6.13% separately. Meanwhile, the monthly total power consumption of 40186340KWh in 2003 was the highest value in these years and happened in January. The power demand reached their highest value of 39237340KWh and 36553230KWh in the following years. By calculating the total power demand of five months in winter and twelve months in the whole year, it gets Table 2.2:

Table 2.2: Total power demand in two periods for every year (KWh)

<table>
<thead>
<tr>
<th></th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>169531890</td>
<td>171690540</td>
<td>170399350</td>
</tr>
<tr>
<td>Whole year</td>
<td>296991890</td>
<td>294633910</td>
<td>293641490</td>
</tr>
</tbody>
</table>

The total power demand of 2004 is higher in winter, but lower in the whole year. For the total power demand in one year, the highest value was obtained in 2003.

Secondly, the medium-term and short-term characteristics of power demand can be gained from weekly power demand shape as Figure 2.2. From now on, because the power demand shapes in three years are similar to each other, the data from 2003 will be extracted to do further analysis. Here, it picks four weeks from four seasons to represent the properties of power consumption in medium and short terms. The days from February 20th to 26th, April 21st to 27th, July 21st to 27th and October 20th to 26th in 2003 are chosen. From Figure 2.2, it is obviously that there are two extremely high values in one day in three seasons expect summer. One of the peak values often occurs during 6:00 to 11:00 in the morning and another occurs during 16:00 to 22:00 in the afternoon and evening. On the other hand, the lower values appear at late night and noon. This phenomenon is highly dependent on human activities, for instance, go to work, have lunch, stay home and go to sleep. One example is the power demand...
reaches its lowest value when people are in deep sleep.

In the summer, the phenomenon of two high peak values appearance is weakened. The long daytime prolongs people behaviors which result in irregular movements of power consumption. More peak values appears in a day, and the power demands increase and decrease sharply in the night.

At last, the comparison of power demands between weekdays and weekends are considered. Usually, the characteristic of power consumption is much different between people at work and not. For investigating this opinion, the power demands are classified in two groups: weekends and weekdays. As these two groups differ from numbers of days, the power demands have to be transformed in advance. The average values of power demand for both weekdays and weekends are calculated for every week for 2003 and 2004. Then the waves of load behaviors in weekdays and weekends can be shown in Figure 2.3:
In Figure 2.3 this figure, there are only slight differences between these two types of power demand. This phenomenon is also shown in Figure 2.2 which represents four weeks power demand in four seasons. They are arranged from Monday to Sunday in a week. Obviously, the power demand of weekends is lower than it's in weekdays in February 20th to 26th, but it is higher than it is in weekdays in April 21st to 27th. The power demand appears almost the same value between weekdays and weekends in July and October shown in Figure 2.2. Accordingly, it is not necessary to apart the power demand into weekdays and weekends in the modeling procedure.

2.1.2 Movement of power demand for greenhouse

Greenhouse is also a special consumer in Ekerö, since its power demand changes highly depending on the development of plants. The daily total load of greenhouse from 2003 to 2005 is plotted in Figure 2.4. It can be seen that the form of power demand is like the total power demand shape in Ekerö. The difference is that a hollow of power consumption exists in 2005. This abnormal exhibition is not brought by the change of the weather conditions, and will not be discussed in this report.
During the growth of plants, it is necessary to provide abundant sun rays to support its photosynthesis. The light, as an energy source, helps green plants to synthesize carbohydrates from carbon dioxide and water. Thus, the power demands during different period of time in 2003, together with the data of sun radiation are shown as Figure 2.5:
From Figure 2.5, three special points are interesting. First, the power demand reaches the top value everyday no matter it is in summer or winter. This is because the sun goes down for a while everyday, then the light energy have to be offered from electricity. Second, the power demand will keep its top value when the sun radiation is low during the summer, and after the sun raises in the morning the power demand decreases sharply. However, the sun radiation is weak and the sun stays much shorter in winter. This leads the power consumption goes down after a certain time delay in greenhouse. The third point is that the strength of the sun radiation does not decide how long the minimum power demand needs in greenhouse, the period of minimum load only depends on the time of the sun appearance.

The comparison between power demand and temperature is shown in Figure 2.6. In summer, the same as sun radiation, the power demand decreases when the temperature increases, and vice versa. This behavior is not clear during the winter. The temperature and power demand don’t seem correlated with each other. This phenomenon is mainly because the power demand of greenhouse is decided by the time of sun arisen, and the temperature is much more affected by the sun during the summer. The temperature differences between day and night are evidently in summer, but it becomes smooth in winter.

For greenhouse reaches its peak power demand everyday, it doesn’t need to predict its peak load. Meanwhile, as the load highly depends on the hours that sun shows but the strength of sun radiation, the power consumption is regularly changing.
day by day and season by season. The load forecasting will not carry out for greenhouse.

![Power demand vs. sun radiation]

Figure 2. 6: Power demand of greenhouse together with temperature for a week in 2003

The appearances of high daily average value are noticed in Figure 2. 4 during the summer, this can be explained in Figure 2. 7:
The cloud covers the sky during the day, which causes the sun radiation is decreasing too intense to act on the plants. The owners of greenhouse artificially increase the electricity consumption to ensure the benefits to the plants.

### 2.1.3 Movement of power demand for industry

Industry is another specific kind of customer in Ekerö. The electricity consumption of industry is strongly correlated to human being’s behaviors, for instance, weekdays, weekends and holidays. The following figures prove these statements about the industry.

From Figure 2.8, it is known that the power consumption is very low during July 20\(^{th}\) to August 10\(^{th}\), and the week after Christmas, it is caused by the summer holiday and Christmas holiday. Also, the power consumption is higher during January and February. It is partly affected by the cold weather, which leads the electrical heating components increasing. From Figure 2.9, the power consumption is lower during the weekends obviously, but it hardly changes on weekdays.
Figure 2.8: Power demand of industry in 2003

Figure 2.9: Power demand of industry in March of 2003
The hourly load and the peak load are not significantly affected by the weather conditions from reading the Figure 2.9, so the modeling process will be neglected.

2.2 Relationship between different parameters

For multiple variables measured for Ekerö, a correlation matrix is introduced to examine their relationships. In theory, correlation indicates the sign and strength of linear relationship [13], and the squared correlation describes the proportion of variance in common between two variables [14]. Thus, the relationship between power demand and weather parameters can be evaluated as following:

Table 2.3: The correlations between the power demand and weather parameters

<table>
<thead>
<tr>
<th></th>
<th>Temp</th>
<th>Wind_D</th>
<th>Wind_S</th>
<th>Cloudy</th>
<th>Sun_R</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp</td>
<td>1</td>
<td>-0.0159</td>
<td>0.1326</td>
<td>-0.0787</td>
<td>0.5217</td>
<td>-0.9033</td>
</tr>
<tr>
<td>Wind_D</td>
<td>-0.0159</td>
<td>1</td>
<td>0.2685</td>
<td>-0.0275</td>
<td>0.0853</td>
<td>0.0629</td>
</tr>
<tr>
<td>Wind_S</td>
<td>0.1326</td>
<td>0.2685</td>
<td>1</td>
<td>0.1378</td>
<td>0.2528</td>
<td>0.0247</td>
</tr>
<tr>
<td>Cloudy</td>
<td>-0.0787</td>
<td>-0.0275</td>
<td>0.1378</td>
<td>1</td>
<td>-0.2353</td>
<td>0.1607</td>
</tr>
<tr>
<td>Sun_R</td>
<td>0.5217</td>
<td>0.0853</td>
<td>0.2528</td>
<td>-0.2353</td>
<td>1</td>
<td>-0.3888</td>
</tr>
<tr>
<td>Power</td>
<td>-0.9033</td>
<td>0.0629</td>
<td>0.0247</td>
<td>0.1607</td>
<td>-0.3888</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.3 shows the correlations between all the parameters here, and the data measured for every hour from 2003 to 2005 are used in this proceed. “Temp” is temperature (°C); “Wind_D” is wind direction by degrees, which represent the angle of wind coming direction on rectangular coordinates; “Wind_S” is the wind speed with the unit meters per second; “Cloudy” presents the proportion of the cloud cover, where 8 means the entire sky is covered by cloud and 1 mean only 1/8 of the sky is covered by cloud; “Sun_R” is strength of sun radiation W/m²; “Power” is the power demand in with the unit of KWh/h. From this table, it is known that the correlations between wind conditions and power demand are rather low. The correlation coefficient between load and cloudiness is higher, but it still very low. The squared correlation coefficient of load and cloudiness is 0.0258, which means less than 3 percent of the variance of these two variables is in common. However, on the other hand, the temperature is much relevant to the power demand, which means the influence of temperature on the power demand is more intense than any other parameters. The sun radiation shows its influence on power consumption is smaller compared with the temperature, but higher than any other weather variables. The squared coefficients of temperature and sun radiation with power demand are 0.8160 and 0.1512 separately. Thereby, the temperature will be used as single input parameter in one of the models later for 80 percent of its variance is in common with power demand, although they change in opposite directions. Note that the correlation of any variable with itself is perfect (Correlation =1), and many correlations in Table 2.3 are redundant.
Another technique to convey the information in different parameters is scatter plots. The scatter plot matrix by applying the variables mentioned above is shown as Figure 2.10:

From Figure 2.10, it is shown that the relationship between all the parameters, including power demand, hour, temperature, wind direction, wind speed, cloudiness and sun radiation. The temperature tends to increase slightly at the same time as the wind speed goes up, this behaviors explains why the power demand abnormally increases when the wind speed increase which shows in next segment.

2.3 Influence of weather parameters on power demand

The measured variables are linked by each other. Especially, how the weather parameters affect the power demand is most interesting in the study. First, the power demand is highly connected with temperature and it is shown as Figure 2.11. Notice that the axis of temperature is reversed to decrease order and the data is got from 2003. It is clear that the power demand decreases while the temperature increases. Especially, the power demand curve appears a deep dent when the temperature suddenly lifts up.
Then, wind speed and cloudiness are picked up to analyze their influence on power demand. The results are shown as Figure 2.12:

Here, it is found that the power demand decreases when the wind increases, which is opposite compared with the theory.
From Figure 2.13, we know the power demand will decrease when the cloudiness is high enough (\( \text{Cloudiness} \geq 7 \)) during the winter. This is because the cloudiness will keep warm, which means it firstly affects the temperature directly, and then affect the power demand.
Figure 2.14: Relationship between power demand and wind speed in July and August of 2003

Figure 2.15: Relationship between power demand and cloudiness in January of 2003

From Figure 2.14 and Figure 2.15, the power demand is not affected by these two parameters in summer, or we can say power demand will be affected by other factors.
instead.
Chapter 3 Model description

The linear regression model and ANN model are used to represent the relationship between weather parameters and power demand. Now, general descriptions of these two approaches as well as the theories are stated in this chapter.

3.1 Linear regression model

Regression analysis is the most applied method to model relationships between one or more dependent variables and the independent variables. It can be considered as a conventional way in power demand prediction, and can offer a simple apprehensible way to dig out how the weather parameters affect the power demand. In statistics, regression analysis is the process used to estimate the parameter values of a function, in which the function predicts the value of a response variable in terms of the value of other variables [15]. There are many methods developed to fit functions and these methods typically depend on the type of function being used. For example: linear regression, nonlinear regression, and logistic regression. In this report, linear regression analysis is introduced by two parts.

3.1.1 Theory of simple linear regression

Simple Linear Regression (SLR) attempts to model the relationship between two variables by fitting a “linear” function to the collecting data. In the case, the model can be written as:

\[ y = \alpha x \]

where

- \( y \) is the response variable.
- \( \alpha \) is a vector of coefficients.
- \( x \) is the explanatory variable.
Here, it is necessary to mention that the technique called “linear” above is not the common understanding that the graph of Equation 3.1 is a line. In fact, this relationship can be expressed as below for instance:

\[ y = \alpha x = [\alpha_0 \alpha_1 \alpha_2 \ldots \alpha_n] \cdot [1 x^2 x^3 \ldots x^n]^T \]  

Eq.3.2

where

- \( \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n \) are the coefficients for relevant orders of explanatory variable.
- \( x, x^2, x^3 \ldots x^n \) are the nth order of explanatory variable.

By expressing like that, this case is still one of linear regression, even though the graph is not a straight line. Meanwhile, notice that there is only one explanatory variable with its several orders in the equation. It means it is using one parameter to describe the trend of response variable, so the intensive linear relationship has to be approved before making use of this approach.

After decide to employ the simple linear regression, the most common approach for finding a regression line and the coefficient estimates is mentioned by the method of least squares [16]. The linear regression will works by minimizing the sum of the squares of residuals. Because the deviations are first squared, then summed, there are no cancellations between positive and negative values. The residual for the ith data point \( r_i \) is defined as the difference between the true response value \( Y_i \) and the fitted response value \( y_i \).

\[ r_i = Y_i - y_i \]  

Eq.3.3

Then, the summed square of residuals is given by

\[ S = \sum_{i=1}^{m} r_i^2 = \sum_{i=1}^{m} (Y_i - y_i)^2 \]  

Eq.3.4

where

- \( S \) is the sum of squares error estimate.
- \( m \) is the number of data.

By applying of least squares approach, the coefficients in the equation 3.2 can be obtained, then the linear function can be considered as a tool, which could be used in power demand prediction or other practical applications.

### 3.1.2 Theory of multiple linear regression

As the same as simple linear regression, multiple linear regression (MLR) also attempts to find the relationship between response variable and explanatory variables. From the name, the difference is clear that more than one explanatory variables are used in the multiple linear regression. The multiple linear regression takes the form of

\[ y = \varepsilon + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n \]  

Eq.3.5

where

- \( y \) is still the response variable.
- \( \varepsilon \) is the constant.
$\beta_1, \beta_2, \beta_3, \ldots, \beta_n$ are the coefficients.

$x_1, x_2, x_3, \ldots, x_n$ are the explanatory variables.

In this model, the coefficients indicate the information about how the response variable shifts when the explanatory variables changes. The magnitudes of the coefficients show the strength corresponding to the contribution of certain explanatory variables and the signs of the number point out the direction of the changes. In another word, the coefficients represent the amount of the response variable $y$ changes when the explanatory variables changes 1 unit. Furthermore, the positive coefficients denote the response variable will increase when the explanatory variables increase, vice versa; the negative coefficients means the response variable will decrease when the explanatory variables increase, vice versa.

The least square method will be used to get the best fitting line again.

### 3.2 Artificial neural network

Artificial Neural Network is a powerful tool to find the relationship between multivariate parameters and the targets. It is recently taken attention in several research fields such as function approximation, times series prediction, classification, and pattern recognition etc. It is an effective approach to find the non-linear relationship between several variables. The fact that the trained neural networks can easily solve the difficult problems comparing with conventional approaches in certain applications is obviously recognized. In the following segment, a general introduction of ANN will be stated, which includes the structure of the model, the training process and the test process.

#### 3.2.1 Overview of ANN

Artificial Neural Network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron [17]. It is which can be considered as mathematical models or computational models, inspired from simulating the “intelligent” performance of human brain. From the structure shown in Figure 3.1, an elementary artificial neuron is build as one node with multi-parallel inputs and single output, together with the “weight” which modulates the effect of the associated input signals and the “transfer function” which present the non-linear characteristics of the neuron. The output computation can be described by:

$$O_j = f(s_j) = f\left(\sum_{j=1}^{N} w_j x_j - \theta_j\right)$$

Eq.3.6

where
$x_j$ is the jth input.

$w_{ij}$ is the connection weight from the jth input to the ith output.

$\theta_i$ is the bias value.

$f(\cdot)$ is the transfer function.

$O_i$ is the output of the neuron.

Figure 3. 1: Simple artificial neuron

In this report, the structure of multilayer neural network will be adopted and it is shown in Figure 3. 2.

Figure 3. 2: Three layers neural network

This type of network builds up with one input layer, one hidden layer and one output layer. Each layer employs several neurons and each neuron is connected to the neurons in the adjacent layer with different weights. Signals flow from the input layer, pass through the hidden layer, and arrive at the output. The hidden layer which between the input and output layers is neither receives inputs from, nor send output to, the outer world directly.

In the following, the detailed information about the multilayer artificial neural network will be described.
3.2.2 Multilayer neural network

This class of networks consists of multiple layers of computational units, usually interconnected in a feed-forward way. Each neuron in one layer has directed connections to the neurons of the subsequent layer. In many applications, the units of these networks apply a sigmoid function as an activation function and the learning algorithm is back propagation.

Many transfer functions are used in ANN models. The hard-limit transfer function, linear transfer function and sigmoid function are the most commonly used transfer functions. Furthermore, the sigmoid function is widely used in multilayer, in part because it is differentiable. These functions are used to map a neuron’s net output $s_i$ to its actual output $O_i$, which shows in Figure 3.1.

The log-sigmoid transfer function is often used in these networks. This function generates output between 0 and 1 as the neuron’s net input goes from negative to positive infinity.

$$a = \log\text{sig}(n)$$

Figure 3.3: Log-sigmoid transfer function

Alternatively, multilayer networks can use the tan-sigmoid transfer function. When the inputs have both positive and negative values, it is more convenient than log-sigmoid transfer function.

$$a = \text{tansig}(n)$$

Figure 3.4: Tan-sigmoid transfer function

Occasionally, the linear transfer function is used in back propagation networks as well. It happens when the outputs of last layer of a multilayer network have to take on any values; otherwise, the outputs of the network are limited to a small range.
In order to use the network for practical task, the weights between all connected neurons need to be calculated. This is achieved through a compulsory step called learning. In practice, learning means the following [18]:

Given a specific task to solve, and a class of functions $F$, learning means using a set of observations, in order to find $f^* \in F$ which solves the task in an optimal sense.

A variety of learning techniques are used in multi-layer networks, the most popular being back propagation. It is a supervised learning technique used for training artificial neural networks. In this kind of learning technique, the training data consist of both input variables and desired outputs. This differs from the unsupervised learning technique which the prior outputs are not supplied. During the training process, the output values calculated by network are introduced to compare with the desired output. Then, the mean square errors between network outputs and desired outputs are fed back through the network from the output layers to the input layers. Using this information, several different algorithms can use to adjust the weights of each connection in order to reduce the value of the error. After repeating this process for a sufficiently large number of training cycles the network will usually converge to some state where the error of the calculations is minimized. In this case one says that the network has learned a certain target function, which can be used for prediction and validation.
Chapter 4 Model application and numerical results

Based on the theory of the linear and non-linear models, the power demand and weather parameters in 2003 are selected and transferred as inputs and outputs. Then the relationship between power demand and climate will be identified. The modeling task will be carried out in the software MATLAB. Eventually, the accomplished model will be used to predict the power demand in 2004 and 2005; afterwards, the accuracy evaluation will be estimated by comparing the predicted value with the real power consumption.

4.1. Application of linear regression

The power demand will be expressed as several functional equations by using linear regression. The selection and transformation of input variables will be described in the following segment.

4.1.1 Application of simple linear regression

According to the high correlation coefficient between power demand and temperature, it is reasonable to find a linear functional expression which can calculate the value of power demand correspond to temperature as the dependent variable. The model of power demand using this approach is expressed in the form as:

\[ P_i(T) = \alpha_0 + \alpha_1 T + \ldots + \alpha_{n-1} T^{n-1} + \alpha_n T^n \]

Eq.4.1

where

i is the ith hour in one year.
P_i(T) is the power demand in ith hour.
T is the temperature in the ith hour, and correlates with the hourly power demand.
\[ \alpha_0, \alpha_1, \ldots, \alpha_{n-1}, \alpha_n \] are simple linear regression coefficients.
Based on the model, three types of power demand need to be investigated: hourly power demand for one year, yearly total power demand and daily peak power demand in winter. Since the yearly power demand is the sum of hourly power demand for one year, these two types of solution will solved together.

The model is based on the yearly data of power demand and temperature from 2003 and then tested by the data in 2004 and 2005. At the beginning, the distribution of the real power demand versus its relevant temperature in 2003 need to plot. Then the polynomial fit curves can be obtained by different orders in MATLAB. There are maximal 10 degrees polynomial fit can be implement in MATLAB. After applied from the linear fit to 10th degree polynomial fit, the 5th degree polynomial fit is chosen. The motive of choosing 5th degree fit curve mainly because of two reasons. One is that higher polynomial fit is closer to the real curve than the lower polynomial fit; the other reason is that the fit curves have sharply increasing or decreasing trends at the end of the edges, which usually obtain quite high inaccuracy in the prediction. Thereby, as a trade-off, the 5th order polynomial fit is adopted. From the 5th order polynomial fit curve, the function of temperature to power demand can be written as:

\[ P_i(T) = \alpha_0 + \alpha_1 T + \alpha_2 T^2 + \alpha_3 T^3 + \alpha_4 T^4 + \alpha_5 T^5 \]  

Eq.4.2

The coefficients are computed from MATLAB as:

\( \alpha_0=45736; \alpha_1=-1745.1; \alpha_2=-29.123; \alpha_3=1.8472; \alpha_4=0.051824; \alpha_5=-0.0016493. \)

The real scatter plot of power demand versus its relevant temperature together with the 5th degree polynomial fit curve is shown in Figure 4.1:

![Scatter plot of power demand versus temperature in 2003](image)

**Figure 4.1:** Scatter plot of power demand versus its relevant temperature together with the 5th degree polynomial fit curve
After built the simple linear model, the hourly power demands can be calculated by its relevant temperature, and the yearly total demand can be gained from:

\[ P_{\text{total}} = \sum_{i=1}^{n} P_i(T) \]  

Eq.4.3

where

- \( P_{\text{total}} \) is the yearly total power demand.
- \( n \) is the number of hours in a year.

As mentioned before, the peak power consumption definitely happened during the winter. Accordingly, the specific day’s data, which are from January 2003 to March 2003, November 2003 to December 2003, are selected to set up a model to predict daily peak power demand in winter next two years. In this case, the input variable is the average temperature in one day and the output variable is the peak power demand in that day. By performing the same method mentioned previously, the polynomial fit curve is shown in Figure 4.2.

![Figure 4.2: Scatter plot of daily peak power demand versus daily average temperature and its 5th degree polynomial fit curve](image)

The relationship between temperature and peak load can be expressed as:

\[ P_{i,\text{peak}}(T_a) = \alpha_0 + \alpha_1 T_a + \alpha_2 T_a^2 + \alpha_3 T_a^3 + \alpha_4 T_a^4 + \alpha_5 T_a^5 \]  

Eq.4.4

where

- \( P_{i,\text{peak}} = \max(P_1, P_2, \ldots, P_{24}) \), which is the maximum hourly power demand in a day.
\[ T_a = \text{mean}(T_1, T_2, \ldots, T_{24}), \] which is the mean value of hourly temperature in a day.

The values of these coefficients are:
\[ \alpha_0 = 53778; \alpha_1 = -1426.2; \alpha_2 = -5.3002; \alpha_3 = -1.1001; \alpha_4 = -0.13765; \alpha_5 = -0.002792. \]

After the model is built up, the same concerned data, which are from January 2004 to March 2005, November 2004 to December 2004, January 2005 to March 2005 and November 2005 to December 2005 are used to test the efficiency and accuracy of the model.

### 4.1.2 Application of multiple linear regression

Besides the temperature, the power demand is affected by the wind speed, wind direction, cloudy cover and sun radiation as well. The multiple linear regression modeling the relationship between all these weather conditions and power demand will be investigated.

The hourly climate data together with the power demand in 2003 are used to establish the multiple linear model. In this case, the number of input variables are 5 and output variables is 1. Then the quantitative relationship between climate and power demand can be written by:

\[ P_i = \varepsilon + \beta_1 T + \beta_2 V_{w,d} + \beta_3 V_{w,s} + \beta_4 A_{\text{cloud}} + \beta_5 R_{s,r} \]

Eq. 4.5

On the left side of this equation, \( P_i \) represents the hourly power demand. Its corresponding climates data and their respective coefficients are states on the right side. Here, \( T \) is the temperature, \( V_{w,d} \) is the wind direction, \( V_{w,s} \) is the wind speed, \( A_{\text{cloud}} \) is the cloudiness and \( R_{s,r} \) is the sun radiation. The coefficients are obtained after data fitting in MATLAB:

\[ \varepsilon = 38680; \beta_1 = -1518; \beta_2 = 1.005; \beta_3 = 617.3; \beta_4 = 581.9; \beta_5 = 10.13. \]

For predicting the daily peak power demand in winter, the forecasting period is the same as it is in simple linear regression model. Thereby, the weather parameters and the power demand parameters from January to March and from November to December are extracted from all the data. In advance, the average values of weather parameters, such as temperature, wind speed, cloud cover and sun radiation, in one day are calculated to apply on this model. The wind direction is neglected for its average value is meaningless in this case. Then the model appears like:

\[ P_{i, \text{peak}} = \varepsilon + \beta_1 T_a + \beta_2 V_{w,s,a} + \beta_3 A_{\text{cloud},a} + \beta_4 R_{s,r,a} \]

Eq. 4.6

In this equation, \( P_{i, \text{peak}} \) on the left side is the maximum hourly power demand in a day. On the right side, the parameters are \( T_a = \text{mean} (T_1, T_2, \ldots, T_{24}) \), \( V_{w,s,a} = \text{mean} (V_{w,s1}, V_{w,s2}, \ldots, V_{w,s24}) \), \( A_{\text{cloud},a} = \text{mean} (A_{\text{cloud}1}, A_{\text{cloud}2}, \ldots, A_{\text{cloud}24}) \), \( R_{s,r,a} = \text{mean} (R_{s,r1}, R_{s,r2}, \ldots, R_{s,r24}) \). They are the mean value of hourly temperature, wind speed, cloud cover and sun radiation in a day respectively. After define the input and output variables, the result of execution of the model is:
\[ \varepsilon = 55680; \beta_1 = -1310; \beta_2 = 385.4; \beta_3 = -368.0; \beta_4 = -38.65. \]

From the coefficients of multiple linear regression models, it also proved that the temperature has the strongest influence to the power demand. By using these equations above, the power consumption can be predicted for all the cases, i.e. hourly power demand, totally power demand and daily peak power demand.

### 4.2 Application of artificial neural network

After introduced the application of linear regression models, the utilization of the neural network are introduced in the prediction procedure. The transformation of the variables and the setting of the networks in MATLAB will be mentioned.

For the reason of comparison with former models, an easy structure of ANN model is built to complete the forecasting task. In the prediction of hourly power demand, there are 5 inputs regarding to 5 weather input variables like in multiple linear regression and 1 output as power demand. As a fact that there is no known technique to determine the exact number of neurons in hidden layer to optimize the solution, the number of neurons of hidden layer is fixed to 3, which is equally to the average of inputs and outputs as a rule of thumb. Before training the network, all data are normalized from real value to ANN input values between 0 and 1. It is necessary because the sigmoid transfer function (Figure 3. 3) is employed between different connected layers in the back propagation algorithm. The normalization is done by using the following equation:

\[ I_{\text{nor}} = \frac{I_r - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \quad \text{Eq.4.7} \]

where

- \( I \) is one of the class of variables, can be weather parameters of input or power demand of output.
- \( I_{\text{nor}} \) is the normalized value.
- \( I_r \) is the real value.
- \( I_{\text{min}} \) and \( I_{\text{max}} \) are the minimum and maximum values in the certain class.

Then, the stop condition needs to be set; they are named as “epochs” and “goal” in MATLAB separately. “Epochs” is the iteration times in the back propagation algorithm and “goal” is the mean error of the fit curve to the real curve. Theoretically, huge numbers of epochs and minimum goal get the best fit relationship. However, that will bring the training time too long to wait, and it is impossible to obtain a perfect fit. In this project, the epochs and the goal are set to 1000 and 0.001 in MATLAB.

After all, with the ready network and data, the ANN model is trained to find the relationship between weather and power demand and the information are memorized in the weights between connected neurons and the bias in the neurons. The prediction of power demand can be achieved by using the trained network.
To predict the daily peak power consumption, the number of inputs data is replaced by 4 for neglecting the wind direction. The input data is the same as the data used in multiple linear regression model for prediction of daily peak power demand, and the normalized work will be carried out. The structure of the model and the initial setting are the same as the ANN model mentioned above, but the input layer changes to 4 neurons instead. By applying the training step, the trained network can be used for load forecasting and evaluation.

For the trained ANN model, the output of the network will be settled to a fixed value when the input variable is chosen. The weights between the layers and the bias in the neurons are memorized and fixed to express the relationship between inputs and outputs. However, these weights and bias will change if the network is trained again, and will compute different outputs while the same input imports to the network. This is because the initialization of training procedure is setting random values to all the weights and bias, then search an adaptive group of weights and bias to achieve the goal by using back propagation algorithm. Therefore, the values inside the neuron network will shift by different initialized weights and bias. Despite the changing neural networks, the output of the network changes slightly from another point of view. After large numbers of iterations, the goal is achieved and the network can represent the relationship between inputs and output reasonably. For the relationship between power demand and weather parameters don’t change, the output from the different trained neural networks becomes similar. In order to evaluate the model objectively, the ANN model will be trained 10 times for each case, and calculate the mean prediction error to represent its accuracy.

4.3 Numerical results

The former linear regression models and artificial neural network models are executed and testified firstly. Then, the modifications are carried on to get better results. The confidence level is discussed as well.

4.3.1 Model discussion

After finish the modeling work, the prediction of power demand in 2004 and 2005 performs in all the mentioned models. To compare the accuracy of these models, the mean absolute relative error (MAE) is defined as:

\[
Error = \frac{|actual \ power \ demand - predicted \ power \ demand|}{actual \ power \ demand} \times 100\%
\]

In Eq.4.8, the actual power demand and its prediction can be hourly, daily peak or yearly total power demand by means of the problems. Table 4.1 shows the MAE
obtained from these models:

<table>
<thead>
<tr>
<th></th>
<th>SLR</th>
<th>MLR</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2004</td>
<td>2005</td>
<td>2004</td>
</tr>
<tr>
<td>Hourly Load</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>13.70</td>
<td>13.54</td>
<td>15.20</td>
</tr>
<tr>
<td>2005</td>
<td>21.27</td>
<td>11.84</td>
<td>21.27</td>
</tr>
<tr>
<td>Total Load</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>0.84</td>
<td>0.25</td>
<td>0.59</td>
</tr>
<tr>
<td>2005</td>
<td>18.08</td>
<td>4.74</td>
<td>18.08</td>
</tr>
<tr>
<td>Daily peak Load</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>4.06</td>
<td>3.87</td>
<td>3.45</td>
</tr>
<tr>
<td>2005</td>
<td>11.21</td>
<td>13.27</td>
<td>11.21</td>
</tr>
</tbody>
</table>

The error of 11.84% of hourly power demand prediction is lowest in 2005 by using ANN model, but the error of ANN model used to predict hourly load in 2004 are too high to unacceptable. Hence, the SLR model appears better results than the others in prediction of hourly power consumption. Its average error is 13.62% for the two years which is lower than 14.955% of MLR model and 16.555% of ANN model. Meanwhile, SLR model have very high accuracy to predict yearly total power demand. The error of load forecasting is only 0.84% and 0.25%, which is very close to the real value. The predictions of total power by SLR model are 2971057798KWh in 2004 and 292912703KWh in 2005. Compared to the real power consumption in Table 2.2, it is clear that the prediction value in 2004 is a little higher and the prediction value in 2005 is lower. For the relative high accuracy of prediction by SLR model, it is chosen to be used for hourly load forecasting and yearly total load forecasting. On the other hand, 3.45% and 3.20% errors are produced by MLR model in the prediction of daily peak load, which is better than 4.06% and 3.87% errors obtained from SLR model. The MLR will be chosen to predict daily peak power demand as a result.
Figure 4.3: Hourly power demand prediction and real power demand in 2004

Figure 4.4: Hourly power demand prediction and real power demand in 2005
The shapes of hourly power demand prediction compared with the real value are shown in Figure 4.3 and Figure 4.4. From these figures, the predictions of power demand curves have almost the same fundamental waves as the real curves. However, the errors are still too high to use in practice. Thereby, a detailed SLR model is introduced in next segment and testified by the historical data.

For prediction of peak load, the MLR model has the lowest errors and the results are very encouraging. The predictions are shown as Figure 4.5 and Figure 4.6. They are quite close to the real curves.

Figure 4.5: Daily peak power demand prediction and the real power demand in winter of 2004
4.3.2 Improvement

The best hourly load forecasting is obtained from the simple linear regression model. However, it still has more than 10% mean absolute error in this case and the accuracy is not acceptable. In order to decrease the errors, a more detailed model is invoked. As it is known that human behaviors affect the power consumption intensively, this contribution can be treated by dividing the simple linear regression models into 24 parts. Every part indicates 1 hour in a day, and people’s behaviors will be indirectly reflected by the time flow. Then, the model can be expressed as:

$$P_i(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \alpha_4 t^4 + \alpha_5 t^5 + \alpha_6 t^6$$

Eq.4.9

Here, the only difference between this equation and Eq.4.1 is the “t” which represents 24 hours in a day. Therefore, there are 24 equations used to predict hourly power consumption. By inputting the historical temperature and power demand data of 2003 into MATLAB, the models are built up. In this case, the 6th degree polynomial fit curve is decided as the trade-off between the closer approximations and no sharply edges.

$$P_i(T) = \alpha_0^i + \alpha_1^i T + \alpha_2^i T^2 + \alpha_3^i T^3 + \alpha_4^i T^4 + \alpha_5^i T^5 + \alpha_6^i T^6$$, (t = 1,2,…24) Eq.4.10

The coefficients are arranged in the following:
Table 4.2: Coefficients of the detailed simple linear regression model

<table>
<thead>
<tr>
<th>hour</th>
<th>a0</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>a5</th>
<th>a6</th>
</tr>
</thead>
<tbody>
<tr>
<td>hour1</td>
<td>39727</td>
<td>-1665.8</td>
<td>-43.525</td>
<td>0.84828</td>
<td>0.19547</td>
<td>-2.85E-05</td>
<td>-0.00021146</td>
</tr>
<tr>
<td>hour2</td>
<td>39315</td>
<td>-1652</td>
<td>-42.828</td>
<td>0.77877</td>
<td>0.20106</td>
<td>0.00017201</td>
<td>-0.00022178</td>
</tr>
<tr>
<td>hour3</td>
<td>39337</td>
<td>-1626.1</td>
<td>-41.52</td>
<td>0.79268</td>
<td>0.20018</td>
<td>0.00018192</td>
<td>-0.00022063</td>
</tr>
<tr>
<td>hour4</td>
<td>39537</td>
<td>-1631.5</td>
<td>-45.017</td>
<td>0.82298</td>
<td>0.21749</td>
<td>0.00011759</td>
<td>-0.00024505</td>
</tr>
<tr>
<td>hour5</td>
<td>41003</td>
<td>-1638.2</td>
<td>-41.069</td>
<td>0.80007</td>
<td>0.17316</td>
<td>0.00013099</td>
<td>-0.00018355</td>
</tr>
<tr>
<td>hour6</td>
<td>43369</td>
<td>-1617.4</td>
<td>-39.152</td>
<td>0.74547</td>
<td>0.14517</td>
<td>0.00013999</td>
<td>-0.00015162</td>
</tr>
<tr>
<td>hour7</td>
<td>46205</td>
<td>-1662.7</td>
<td>-36.366</td>
<td>0.70269</td>
<td>0.11046</td>
<td>0.00031071</td>
<td>-0.0001972</td>
</tr>
<tr>
<td>hour8</td>
<td>48935</td>
<td>-1744.6</td>
<td>-36.052</td>
<td>0.46908</td>
<td>0.10905</td>
<td>0.0011655</td>
<td>-0.00012039</td>
</tr>
<tr>
<td>hour9</td>
<td>49172</td>
<td>-1733.7</td>
<td>-35.21</td>
<td>0.32943</td>
<td>0.10799</td>
<td>0.0014756</td>
<td>-0.00012176</td>
</tr>
<tr>
<td>hour10</td>
<td>49952</td>
<td>-1582</td>
<td>-30.331</td>
<td>-0.00145</td>
<td>0.099082</td>
<td>0.0018557</td>
<td>-0.00011725</td>
</tr>
<tr>
<td>hour11</td>
<td>49973</td>
<td>-1613.7</td>
<td>-29.731</td>
<td>-0.63362</td>
<td>0.1134</td>
<td>0.0029782</td>
<td>-0.00014964</td>
</tr>
<tr>
<td>hour12</td>
<td>49904</td>
<td>-1571.8</td>
<td>-28.153</td>
<td>-1.1013</td>
<td>0.11663</td>
<td>0.0038709</td>
<td>-0.00016831</td>
</tr>
<tr>
<td>hour13</td>
<td>49610</td>
<td>-1544.1</td>
<td>-22.397</td>
<td>-1.4367</td>
<td>0.10494</td>
<td>0.0045025</td>
<td>-0.00017025</td>
</tr>
<tr>
<td>hour14</td>
<td>48985</td>
<td>-1490.7</td>
<td>-29.702</td>
<td>-1.4059</td>
<td>0.12474</td>
<td>0.0041464</td>
<td>-0.00017512</td>
</tr>
<tr>
<td>hour15</td>
<td>48902</td>
<td>-1438.8</td>
<td>-31.191</td>
<td>-2.0847</td>
<td>0.14184</td>
<td>0.0057057</td>
<td>-0.00022339</td>
</tr>
<tr>
<td>hour16</td>
<td>50092</td>
<td>-1442.1</td>
<td>-38.453</td>
<td>-2.028</td>
<td>0.16453</td>
<td>0.0050376</td>
<td>-0.00022223</td>
</tr>
<tr>
<td>hour17</td>
<td>51370</td>
<td>-1467.4</td>
<td>-52.683</td>
<td>-1.9278</td>
<td>0.21527</td>
<td>0.0046985</td>
<td>-0.0002598</td>
</tr>
<tr>
<td>hour18</td>
<td>52314</td>
<td>-1535.1</td>
<td>-63.246</td>
<td>-1.3587</td>
<td>0.23886</td>
<td>0.0037045</td>
<td>-0.00026573</td>
</tr>
<tr>
<td>hour19</td>
<td>50543</td>
<td>-1503.4</td>
<td>-56.958</td>
<td>-0.86302</td>
<td>0.21221</td>
<td>0.0026074</td>
<td>-0.0002714</td>
</tr>
<tr>
<td>hour20</td>
<td>48780</td>
<td>-1550</td>
<td>-50.597</td>
<td>-0.23988</td>
<td>0.19227</td>
<td>0.0017816</td>
<td>-0.00021381</td>
</tr>
<tr>
<td>hour21</td>
<td>46456</td>
<td>-1535.3</td>
<td>-48.404</td>
<td>-0.15643</td>
<td>0.20645</td>
<td>0.0017735</td>
<td>-0.00024805</td>
</tr>
<tr>
<td>hour22</td>
<td>43744</td>
<td>-1563.7</td>
<td>-51.739</td>
<td>-0.23107</td>
<td>0.25421</td>
<td>0.0023793</td>
<td>-0.00032746</td>
</tr>
<tr>
<td>hour23</td>
<td>42077</td>
<td>-1663.9</td>
<td>-55.459</td>
<td>0.17432</td>
<td>0.27752</td>
<td>0.0016725</td>
<td>-0.00034253</td>
</tr>
<tr>
<td>hour24</td>
<td>40728</td>
<td>-1710.6</td>
<td>-55.159</td>
<td>0.47338</td>
<td>0.26337</td>
<td>0.00096362</td>
<td>-0.00030196</td>
</tr>
</tbody>
</table>

From these detailed models, the mean absolute relative errors are computed for prediction power demand of every hour in a day. Table 4.3 shows the errors calculated from these models in the following. Obviously, most of the errors are below 10%, and compared to the previous model, the detailed models improve the accuracy of prediction to make the error around 8%. For example, it is only 7.16% of the mean relative error to predict the power demand during 9:00 to 10:00 in 2005.
Table 4.3: Error(%) of prediction for hourly simple linear regression models

<table>
<thead>
<tr>
<th></th>
<th>Error_04</th>
<th>Error_05</th>
<th>Error_04</th>
<th>Error_05</th>
<th>Error_04</th>
<th>Error_05</th>
</tr>
</thead>
<tbody>
<tr>
<td>hour1</td>
<td>11.27</td>
<td>11.09</td>
<td>8.44</td>
<td>8.58</td>
<td>9.03</td>
<td>8.76</td>
</tr>
<tr>
<td>hour2</td>
<td>9.56</td>
<td>9.76</td>
<td>7.32</td>
<td>7.16</td>
<td>8.54</td>
<td>8.55</td>
</tr>
<tr>
<td>hour3</td>
<td>8.8</td>
<td>8.86</td>
<td>7.41</td>
<td>7.29</td>
<td>8.37</td>
<td>7.95</td>
</tr>
<tr>
<td>hour4</td>
<td>9.53</td>
<td>9.42</td>
<td>7.45</td>
<td>7.58</td>
<td>8.49</td>
<td>7.93</td>
</tr>
<tr>
<td>hour5</td>
<td>7.65</td>
<td>7.53</td>
<td>7.93</td>
<td>7.72</td>
<td>8.77</td>
<td>8.12</td>
</tr>
<tr>
<td>hour6</td>
<td>8.97</td>
<td>8.08</td>
<td>8.56</td>
<td>7.86</td>
<td>8.41</td>
<td>8.94</td>
</tr>
<tr>
<td>hour7</td>
<td>11.81</td>
<td>10.72</td>
<td>8.9</td>
<td>8.24</td>
<td>9.06</td>
<td>10.45</td>
</tr>
<tr>
<td>hour8</td>
<td>10.84</td>
<td>10.34</td>
<td>9.05</td>
<td>8.4</td>
<td>10.15</td>
<td>10.96</td>
</tr>
</tbody>
</table>

After the models are settled, where the hourly power demand prediction is done by the detailed SLR models and the daily peak power demand prediction is done by MLR models. It is interesting to investigate the reliability of load forecasting. For examining the reliability of these models, the distributions of the absolute errors are analyzed. Here, the absolute errors are the difference between predicted values and the real power demand. Figure 4.7 shows the distribution of the errors for four cases. One of them is extracted from MLR models and the others are arbitrarily picked up from the detailed SLR models.

Figure 4.7: Distributions of errors for different models

From these distributions and after tested in MATLAB with the function “jptest”, it is reasonable to suppose that they are fit or close to the normal distribution. In
statistics, about 68% of values drawn from a standard normal distribution are within 1 standard deviation away from the mean; about 95% of the values are within two standard deviations and about 99.7% lie within 3 standard deviations [19]. Hence, the standard deviations are calculated from the errors of power demand prediction for 2003 in certain cases, which can be expressed as:

\[
\text{Standard error} = \left( \frac{1}{n-1} \sum_{i=1}^{n} (\text{error}_i - \text{mean error})^2 \right)^{\frac{1}{2}}
\]

Eq.4.11.

In this equation, error\(_i\) is the difference between real power demand and the predicted power demand. Mean error is the mean value of these errors. n is the numbers of the errors in certain case and i is ith input. Consequently, the standard deviations for all the cases are shown as Table 4.4:

Table 4.4: Standard deviation (KWh/h) of all the models

<table>
<thead>
<tr>
<th>Detailed SLR models to predict hourly power demand</th>
<th>hour1</th>
<th>hour9</th>
<th>hour17</th>
<th>hour12</th>
<th>hour18</th>
<th>hour19</th>
<th>hour20</th>
<th>hour21</th>
<th>hour22</th>
<th>hour23</th>
<th>hour24</th>
<th>hour25</th>
</tr>
</thead>
<tbody>
<tr>
<td>hour1</td>
<td>4065</td>
<td>3439</td>
<td>3548</td>
<td>3585</td>
<td>3718</td>
<td>3585</td>
<td>3251</td>
<td>3247</td>
<td>3642</td>
<td>4033</td>
<td>4074</td>
<td>4227</td>
</tr>
<tr>
<td>hour2</td>
<td>3831</td>
<td>2973</td>
<td>3718</td>
<td>3585</td>
<td>3585</td>
<td>3251</td>
<td>3247</td>
<td>3642</td>
<td>4033</td>
<td>4033</td>
<td>4074</td>
<td>4227</td>
</tr>
<tr>
<td>hour3</td>
<td>3624</td>
<td>2965</td>
<td>3718</td>
<td>3585</td>
<td>3585</td>
<td>3251</td>
<td>3247</td>
<td>3642</td>
<td>4033</td>
<td>4033</td>
<td>4074</td>
<td>4227</td>
</tr>
<tr>
<td>hour4</td>
<td>2698</td>
<td>3026</td>
<td>3251</td>
<td>3247</td>
<td>3642</td>
<td>4033</td>
<td>4074</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
</tr>
<tr>
<td>hour5</td>
<td>3089</td>
<td>3071</td>
<td>3251</td>
<td>3247</td>
<td>3642</td>
<td>4033</td>
<td>4074</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
</tr>
<tr>
<td>hour6</td>
<td>3161</td>
<td>3215</td>
<td>3642</td>
<td>3642</td>
<td>4033</td>
<td>4074</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
</tr>
<tr>
<td>hour7</td>
<td>3942</td>
<td>3326</td>
<td>4033</td>
<td>4033</td>
<td>4033</td>
<td>4074</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
</tr>
<tr>
<td>hour8</td>
<td>4074</td>
<td>3450</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
<td>4227</td>
</tr>
</tbody>
</table>

MLR model to predict daily peak power demand: 2217

From the calculated standard deviation, the confidence band of the prediction can be easily obtained. The Figure 4.8 and Figure 4.9 show the prediction of daily peak load with its confidence band and give examples of this application. The confidence bands are drawn from the predicted values within 3 standard deviations, which suppose to bring 99.7% of the predicted values within the confidence area. Although the errors are not exactly distributed as the normal distribution, the results are still pleasant that almost all the real values keep in the confidence bands. This is also clearly shown in Figure 4.10 and Figure 4.11. The data are normalized by dividing the real power demand.
Figure 4.8: Daily peak load forecasting with its confidence band in 2004

Figure 4.9: Daily peak load forecasting with its confidence band in 2005
Figure 4.10: Normalized value of daily peak load forecasting with its confidence band in 2005

Figure 4.11: Normalized value of daily peak load forecasting with its confidence band in 2005
From Figure 4. 10 and Figure 4. 11, one point of real power consumption is not located in the confidence bands in 2004 and two points are not located in the confidence bands in 2005, but the unexpected confidence bands are very close to the real value. Although the unexpected values are existed, the accuracy with the confidence bands is still high, where the confidence bands covers 99.01% of the real values in these two years.

For the detailed SLR models, the confidence bands can be obtained as the same as the way of MLR model.

After all the models are completed, it is easy to input the equations into Excel and use the predicted weather conditions to execute load forecasting procedures.
Chapter 5 Conclusions

In this report, the characteristics of power consumption in Ekerö are studied from the historical data obtained from Fortum Distribution. Three types of customers are analyzed during different period. The prediction of power demand of total area is then investigated. Two ideas, linear regression analysis and artificial neural network, are chosen to forecast the load, which are classified to three subjects: hourly power demand, yearly total power demand and daily peak power demand. The comparison between these methods is discussed.

5.1 Characteristics of power consumptions in Ekerö

The study of characteristics of power consumption in Ekerö is divided into three categories: whole area of Ekerö, greenhouse and industry.

5.1.1 Whole area of Ekerö

In Ekerö, the power consumption has obviously seasonal properties. It increases when the temperature goes down in winter and decreases when the temperature goes up in summer. The peak load usually happens during January to March, because the coldest weather appears during this period. The change of power demand also highly depends on human being behaviors in a day. It reaches it bottom value in the night and increase in the morning. Usually, there are two peak values in a day, but in summer, this phenomenon takes off, because the day is very long to affect the normal work and rest time of people. The power consumptions don’t changes between weekdays and weekends in Ekerö. Thereby, it is not necessary to separate these two cases in the modeling procedure.
5.1.2 Greenhouse

The greenhouse reaches its highest capacity every day except out of work. The number of hours that highest capacity reaches is strongly affected by the number of hours that sun shows up. The more time sun is in the sky, the less power it needs in greenhouse. Hence, the power demand is high in winter and low in summer.

5.1.3 Industry

In this case, the power demand correlates with people’s work activities. The power demand will go down in holidays, such as Christmas and summer holiday. In weekends, the power demand is obviously lower than it is in weekdays. However, its small share of power consumption make this phenomenon is weak in the whole area.

5.2 Models comparison

The results from the models are stated in chapter 4. All of the models are using the predicted weather parameters as inputs, which leads uncertain factors existed by using this model. However, due to the periodic properties of weather conditions as well as the power demand, the prediction tasks of them become much more successful. Based on the models used above, the results of the work are encouraging. The models are fit for Ekerö area and need to refresh data and coefficients if they apply to other area.

5.2.1 Simple linear regression model

This model expresses the relation between power demand and temperature as single input. It is easy to understand and utilize. After tested by data in the following two years, the high accuracy of this model are obtained when it is used to predict yearly total power demand. In order to improve the accuracy of this model, it employs 24 detailed models to represent 24 hours in a day to predict hourly power consumption. From the numerical results, the accuracy enhances evidently.

5.2.2 Multiple linear regression model

Multiple linear regression model plays an important role in daily peak power demand prediction in winter. It uses all the weather parameters in the model and finds the best fitting expression based on the historical data in 2003. In the task of hourly power demand prediction and yearly total power demand prediction, it shows just a little deficiency.
5.2.3 Artificial neuron network model

ANN model proves a great improvement in order to predict power demand recently. However, it does not exhibit its advantages in this report. The reason can be the simple structure of network and numbers of input variables. Large structure of ANN model with detailed classified input variables is successfully used to predict short-term power demand [20], and it can be adopted in future work.
Chapter 6 Future work

The benefits of power demand prediction are presented everywhere of system planning, therefore, the concerned research on prediction are developed very well. Based on this report, the following works are recommended to enhance the accuracy of prediction:

1. Data inputs. It is only used one year historical data to build up the model and it is just focus on the relationship between power demand and weather parameters. In the following, two additional preparations are necessary: Increase the number of historical data and consider other parameters besides weather. The electricity price, number of customers and historical power demand affects the power consumptions. They can be added to multiple linear regression model and artificial neural network model. Further, the relationship between any one of these mentioned parameters and forecasting power demand is interesting to investigate.

2. Complex models. Simple models are applied in this report, especially, the artificial neural network. More detailed model can be achieved by adding the number neurons in every layer, considering historical power demand and using the data which is more close to prediction days.
References

[6]. Mikael Amelin, Valerijs Knazkins, Lennart Söder, ”Uppskattning av maximal elförbrukning i Fortums näatområde Stockholm”, A-ETS/EES-0503, 2005
[8]. Yuuichi Mizukami, Toshiro Nishimori, “Maximum electric power demand prediction by neural network”, IEEE, 1993
Appendix

1. MATLAB codes

Here, the MATLAB codes are shown below. The arrangement of the input data are introduced and the codes have the procedure of building the SLR, MLR and ANN models. It can be used to predict the power demand in 2004 directly, and need to change the number of input data to predict other cases.

The detailed simple linear regression models have the same code as the previous SLR model. The differences are the coefficients and the input data.

data() contains 7 columns and arranges as: hour in a day, temperature, wind direction, wind speed, cloudiness, sun radiation and power demand. It has 26304 rows which represent hourly data from 2003 to 2005.

meanpara() contains 5 columns and arranges as: daily mean temperature, daily mean wind speed, daily mean cloudiness, daily mean sun radiation, daily mean power demand. It has 454 rows which represent 454 days in the winter from 2003 to 2005. The winter means January, February, March, November and December.

peakpara() is arranged as the same as meanpara(), the difference is that all the daily mean values change to daily peak values.

% Simple linear regression model

% Hourly power demand prediction

% Polynomial coefficients
p1 = -0.0016493;
p2 = 0.051824;
p3 = 1.8472;
p4 = -29.123;
p5 = -1745.1;
p6 = 45736;

% Input data
x04 = data(8761:17544,2);

% Hourly power demand prediction
y04 = p1*x04.^5 + p2*x04.^4 + p3*x04.^3 + p4*x04.^2 + p5*x04.^1 + p6;
% Total power demand prediction
pre04 = sum(y04);
real04 = sum(data(8761:17544,7));

% Error analysis
err04 = mean(abs(y04-data(8761:17544,7))./data(8761:17544,7));
S_D=std(abs(y04-data(8761:17544,7))./data(8761:17544,7));
Err04 = mean(abs(y04-data(8761:17544,7)));
E04 = (pre04-real04)/real04;

% Figure plot
figure;
plot(1:8784,y04,'r',1:8784,data(8761:17544,7),'b:');
title('Hourly power demand in 2004');
xlabel('Hour (h)');
ylabel('Power demand (KWh/h)');
legend('Prediction of power demand','Real power demand');

% Daily peak power demand prediction

% Polynomial coefficients
p1 = -0.002792;
p2 = -0.13765;
p3 = -1.1001;
p4 = -5.3002;
p5 = -1426.2;
p6 = 53778;

x04 = meanpara(152:303,1);
y04 = p1*x04.^5 + p2*x04.^4 + p3*x04.^3 + p4*x04.^2 + p5*x04.^1 + p6;
err04 = mean(abs(y04-peakpara(152:303,5))./peakpara(152:303,5));

% Multiple linear regression model
% Hourly power demand prediction
% Model Constructing
% Input data
x1 = [ones(size(data(1:8760,2))) data(1:8760,2) data(1:8760,3) data(1:8760,4)
data(1:8760,5) data(1:8760,6)];
% Output data
y = data(1:8760,7);

% Coefficients
a1 = x1\y;

% Model Testing
x1_pre = [ones(size(data(17545:26304,2))) data(17545:26304,2) data(17545:26304,3) data(17545:26304,4) data(17545:26304,5) data(17545:26304,6)];
y_pre = data(17545:26304,7);

% Prediction of power demand
Y1_pre = x1_pre*a1;

% Error analysis
Err1_pre = abs(Y1_pre - y_pre);
err1_pre = Err1_pre./y_pre;
Mean_E=mean(err1_pre);

% Daily peak power demand prediction

% Model Constructing
x2 = [ones(size(meanpara(1:151,1))) meanpara(1:151,1) meanpara(1:151,2) meanpara(1:151,3) meanpara(1:151,4)];
y = peakpara(1:151,5);
a2 = x2\y;

% Error distribution analysis of model
Y2 = x2*a2;
Err2 = abs(Y2 - y);
Err2_r = Y2 - y;
StdErr2 = std(Err2_r);
err2 = Err2./y;
figure;
hist(Err2_r,-8000:500:6000);

% Model Testing
x4_pre = [ones(size(meanpara(152:303,1))) meanpara(152:303,1) meanpara(152:303,2) meanpara(152:303,3) meanpara(152:303,4)];
y4_pre = peakpara(152:303,5);
Y4_pre = x4_pre*a2;
Err4_pre = abs(Y4_pre - y4_pre);
Mean4_Err = mean(Err4_pre);
err4_pre = Err4_pre./y4_pre;
mean4_err=mean(err4_pre);

% Figure plot
figure;
plot(1:152,Y4_pre,'g',1:152,y4_pre,'b');
title('Daily peak power demand in 2004');
xlabel('Day');
ylabel('Power demand (kW)');

% Confidence band
hold on
plot(1:152,Y4_pre+StdErr2*3,'r:',1:152,Y4_pre-StdErr2*3,'r:');
legend('Prediction of power demand','Real power demand','Confidence band');

% Artificial neural network model
% Hourly power demand prediction

% Network input data
P=[data(1:8760,2:6)];
% Network target data
T=[data(1:8760,7)];

% Network test input data
P_test=[data(8761:17544,2:6)];
% Network test target data
T_test=[data(8761:17544,7)];

% Normalize the whole data
for j=1:5
    P_norm(:,j)=(P(:,j)-min(P(:,j)))/(max(P(:,j))-min(P(:,j)));
end
T_norm=(T-min(T))/(max(T)-min(T));
for j=1:5
    P_test_norm(:,j)=(P_test(:,j)-min(P_test(:,j)))/(max(P_test(:,j))-min(P_test(:,j)));
end
P_input=P_norm.';
T_input=T_norm.';
P_test_in=P_test_norm.';

[a1 b1]=size(P_input);
\( [a_2 \ b_2] = \text{size}(T_{\text{input}}); \)

% Network input range
threshold=repmat([0,1],a1,1);

% Create the neural network
a1=ceil((a1+a2)/2);
net=newff(threshold,[a1,a2],{'tansig' 'tansig'},'trainlm');
net.trainParam.epochs = 1000;
net.trainParam.show = 100;
net.trainParam.goal = 0.001;
LP.lr=0.1;
net = train(net,P_input,T_input);

% Power demand prediction
P_out_norm=sim(net,P_test_in);
P_out=P_out_norm*(max(T_test)-min(T_test))+min(T_test);

% Error analysis
Err=abs(P_out.'-T_test)./T_test;
Err_average=mean(Err)

% Daily peak power demand prediction
Q=[meanpara(1:151,1:4)];
U=[peakpara(1:151,5)];
for j=1:4
    Q_norm(:,j)=(Q(:,j)-min(Q(:,j)))/(max(Q(:,j))-min(Q(:,j)));
end
U_norm=(U-min(U))/(max(U)-min(U));

Q_input=Q_norm.';
U_input=U_norm.';

[a1 b1]=size(Q_input);
a2 b2]=size(U_input);
threshold=repmat([0,1],a1,1);

a1=ceil((a1+a2)/2);
net=newff(threshold,[a1,a2],{'tansig' 'tansig'},'trainlm');
net.trainParam.epochs = 1000;
net.trainParam.show = 100;
net.trainParam.goal = 0.001;
LP.lr=0.1;
net = train(net,Q_input,U_input);

Q_test=[meanpara(304:454,1:4)];
U_test=[peakpara(304:454,5)];
for j=1:4
    Q_test_norm(:,j)=(Q_test(:,j)-min(Q_test(:,j)))/(max(Q_test(:,j))-min(Q_test(:,j)));
end
Q_test_in=Q_test_norm.';
Q_out_norm=sim(net,Q_test_in);
Q_out=Q_out_norm*(max(U_test)-min(U_test))+min(U_test);

Err_U=abs(Q_out.'-U_test)./U_test;
Err_average_U=mean(Err_U);

2. Fit curve

When decide the order of the polynomial fit curve, the trade-off method is applied for the following analysis. Here give an example of the inappropriate fit curves with very low and very high orders fit.
3. **Prediction tool in Excel**

The SLR models and MLR model can be executed to predict power demand in Excel. The processes are state as below.

To predict hourly power demand, the input data are temperatures in every hour, and then use equation 4.10, and the coefficients in table 4.2 for every hour, the power demand can be calculated. The confidence bands can be expressed as upper limit and lower limit. In this model, the predicted power demand plus and minus 3 times standard deviation which is shown in table 4.4, can represent 99% confidence bands.
### Input data

- **DATE**
  - 2005-01-01
  - 2005-01-02
  - 2005-01-03
  - 2005-01-04
  - 2005-01-05
  - 2005-01-06
  - 2005-01-07
  - 2005-01-08
  - 2005-01-09
  - 2005-01-10
  - 2005-01-11
  - 2005-01-12
  - 2005-01-13
  - 2005-01-14
  - 2005-01-15
  - 2005-01-16
  - 2005-01-17

### Prediction of power demand

- Power
- 99% Upper limit
- 99% Lower limit

### Confidence bands

- Power prediction bands

### Simple linear regression model

- Predicted value
- Confidence intervals
<table>
<thead>
<tr>
<th></th>
<th>DATUM</th>
<th>UTC</th>
<th>TEMP</th>
<th>Power</th>
<th>99% Upper limit</th>
<th>99% Lower limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2005-01-01</td>
<td>0</td>
<td>-6,1</td>
<td>48336,3</td>
<td>66531,26217</td>
<td>36141,26217</td>
</tr>
<tr>
<td>4</td>
<td>2005-01-01</td>
<td>1</td>
<td>-5,6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2005-01-01</td>
<td>2</td>
<td>-5,0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2005-01-01</td>
<td>3</td>
<td>-4,5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2005-01-01</td>
<td>4</td>
<td>-4,3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2005-01-01</td>
<td>5</td>
<td>-4,2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2005-01-01</td>
<td>6</td>
<td>-4,0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2005-01-01</td>
<td>7</td>
<td>-3,7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2005-01-01</td>
<td>8</td>
<td>-3,5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2005-01-01</td>
<td>9</td>
<td>-3,2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2005-01-01</td>
<td>10</td>
<td>-2,4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2005-01-01</td>
<td>11</td>
<td>-1,5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2005-01-01</td>
<td>12</td>
<td>-0,8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2005-01-01</td>
<td>13</td>
<td>-0,2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>2005-01-01</td>
<td>14</td>
<td>0,5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2005-01-01</td>
<td>15</td>
<td>1,1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2005-01-01</td>
<td>16</td>
<td>1,3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2005-01-01</td>
<td>17</td>
<td>1,8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Input data:

<table>
<thead>
<tr>
<th>DATUM</th>
<th>TEMP</th>
<th>VINDAarhus/or</th>
<th>NORDMAD</th>
<th>GLOBALSTRAINING Wind</th>
<th>Power</th>
<th>99% Upper limit</th>
<th>99% Lower limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-01-01</td>
<td>1,37</td>
<td>2,375</td>
<td>5,375</td>
<td>5,525</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-02</td>
<td>1,9</td>
<td>6,003</td>
<td>7,125</td>
<td>9,766*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-03</td>
<td>1,75</td>
<td>6,412</td>
<td>4,547</td>
<td>6,843</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-04</td>
<td>3,15</td>
<td>7,5</td>
<td>4,493</td>
<td>13,519</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-05</td>
<td>2,465</td>
<td>4,169</td>
<td>2,493</td>
<td>16,853</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-06</td>
<td>2,973</td>
<td>3,6333</td>
<td>8,904</td>
<td>8,904</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-07</td>
<td>7,55</td>
<td>5,3333</td>
<td>7,5</td>
<td>3,300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-08</td>
<td>1,708</td>
<td>4,5033</td>
<td>7,687</td>
<td>2,950</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-09</td>
<td>1,042</td>
<td>8,375</td>
<td>4,547</td>
<td>14,107</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-10</td>
<td>5,46</td>
<td>4,25</td>
<td>7,375</td>
<td>3,7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-11</td>
<td>6,65</td>
<td>4,526</td>
<td>5,047</td>
<td>15,267</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-12</td>
<td>5,6542</td>
<td>4,7919</td>
<td>6,25</td>
<td>14,986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-13</td>
<td>4,922</td>
<td>5,26</td>
<td>4,547</td>
<td>9,290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-14</td>
<td>2,293</td>
<td>3,0033</td>
<td>3,5</td>
<td>24,079</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-15</td>
<td>2,103</td>
<td>2,8033</td>
<td>3,959</td>
<td>20,597</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-16</td>
<td>2,187</td>
<td>3,025</td>
<td>7,625</td>
<td>9,135</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-17</td>
<td>3,133</td>
<td>35</td>
<td>7,169</td>
<td>16,525</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-18</td>
<td>3,042</td>
<td>5,7625</td>
<td>4,2033</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-19</td>
<td>3,008</td>
<td>0,5417</td>
<td>7,0013</td>
<td>14,6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-20</td>
<td>1,607</td>
<td>3,9167</td>
<td>7,125</td>
<td>11,975</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-21</td>
<td>2,292</td>
<td>4,975</td>
<td>7,5</td>
<td>4,508</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Multiple linear regression model

<table>
<thead>
<tr>
<th>DATUM</th>
<th>TEMP</th>
<th>VINDSTÄNGT (m/s)</th>
<th>MOLMÄNGDO</th>
<th>GLOBALSTRÄLING (W/m²)</th>
<th>Power</th>
<th>99% Upper limit</th>
<th>99% Lower limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-01-01</td>
<td>-1.36</td>
<td>2.35</td>
<td>5.35</td>
<td>5.825</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-02</td>
<td>-3.15</td>
<td>6.0333</td>
<td>7.125</td>
<td>7.9865</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-04</td>
<td>0.3125</td>
<td>2.6</td>
<td>4.893</td>
<td>13.5125</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-05</td>
<td>2.4933</td>
<td>4.9486</td>
<td>2.9433</td>
<td>18.9065</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-06</td>
<td>2.9792</td>
<td>3</td>
<td>6.8333</td>
<td>8.3342</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-07</td>
<td>5.618</td>
<td>4.9486</td>
<td>7.5687</td>
<td>10.528</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-08</td>
<td>0.1708</td>
<td>5.2909</td>
<td>7.8632</td>
<td>29.166</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-10</td>
<td>6.4117</td>
<td>4.26</td>
<td>7.376</td>
<td>8.3932</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-11</td>
<td>0.185</td>
<td>4.56</td>
<td>6.8417</td>
<td>19.289</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-12</td>
<td>0.06</td>
<td>4.7917</td>
<td>0.35</td>
<td>14.917</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-13</td>
<td>2.836</td>
<td>2.0333</td>
<td>3.9933</td>
<td>20.967</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-15</td>
<td>0.524</td>
<td>6.4917</td>
<td>1.35</td>
<td>4.9911</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-16</td>
<td>0.2098</td>
<td>0.5441</td>
<td>7.6933</td>
<td>1.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-17</td>
<td>0.2098</td>
<td>4.3167</td>
<td>7.125</td>
<td>11.795</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-18</td>
<td>0.2292</td>
<td>4.3167</td>
<td>7.125</td>
<td>11.795</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Standard deviation

<table>
<thead>
<tr>
<th>DATUM</th>
<th>TEMP</th>
<th>VINDSTÄNGT (m/s)</th>
<th>MOLMÄNGDO</th>
<th>GLOBALSTRÄLING (W/m²)</th>
<th>Power</th>
<th>99% Upper limit</th>
<th>99% Lower limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-01-01</td>
<td>-1.36</td>
<td>2.35</td>
<td>5.35</td>
<td>5.825</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-02</td>
<td>-3.15</td>
<td>6.0333</td>
<td>7.125</td>
<td>7.9865</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-04</td>
<td>0.3125</td>
<td>2.6</td>
<td>4.893</td>
<td>13.5125</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-05</td>
<td>2.4933</td>
<td>4.9486</td>
<td>2.9433</td>
<td>18.9065</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-06</td>
<td>2.9792</td>
<td>3</td>
<td>6.8333</td>
<td>8.3342</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-07</td>
<td>5.618</td>
<td>4.9486</td>
<td>7.5687</td>
<td>10.528</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-08</td>
<td>0.1708</td>
<td>5.2909</td>
<td>7.8632</td>
<td>29.166</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-10</td>
<td>6.4117</td>
<td>4.26</td>
<td>7.376</td>
<td>8.3932</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-11</td>
<td>0.185</td>
<td>4.56</td>
<td>6.8417</td>
<td>19.289</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-12</td>
<td>0.06</td>
<td>4.7917</td>
<td>0.35</td>
<td>14.917</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-13</td>
<td>2.836</td>
<td>2.0333</td>
<td>3.9933</td>
<td>20.967</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-15</td>
<td>0.524</td>
<td>6.4917</td>
<td>1.35</td>
<td>4.9911</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-16</td>
<td>0.2098</td>
<td>0.5441</td>
<td>7.6933</td>
<td>1.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-17</td>
<td>0.2098</td>
<td>4.3167</td>
<td>7.125</td>
<td>11.795</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-01-18</td>
<td>0.2292</td>
<td>4.3167</td>
<td>7.125</td>
<td>11.795</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Confidence bands

Confidence bands are indicated by surrounding the predicted values with a range. For example, the predicted value for the first row is indicated by the range from 3.106 to 5.825.

Peak load prediction

The peak load prediction is the highest value predicted in the dataset, which is indicated by the highest value in the Power column.