MASTER

Improving the electricity distribution loss forecasting accuracy for Enexis B.V.

Heeren, M.E.M.

Award date:
2011

Eindhoven, August 2011

Improving the Electricity Distribution Loss Forecasting Accuracy for Enexis B.V.

By

Heeren, M.E.M.

Michael Heeren BSc
Student Identity Number 0577642

in partial fulfilment of the requirements for the degree of

Master of Science
in Operations Management and Logistics

Supervisors TU/e:
dr. F. Tanrisever (OPAC)
dr. M.J. Reindorp (OPAC)

Supervisor Enexis B.V.
ing. M. Jonkers

TUE. School of Industrial Engineering.
Series Master Theses Operations Management and Logistics

Subject headings: distribution losses forecasting, network losses, electricity demand, electricity distribution, smart metering, electric vehicles, decentralized generation

Preface and acknowledgements

This report is the result of my master thesis project for the master Operations, Management and Logistics at the Eindhoven University of Technology. This master thesis project was conducted at the customer relations department of Enexis B.V. in Weert from February of 2011 until August 2011, and was supervised by the department of Operations Planning and Control (OPAC) of the TU/e.

I enjoyed working at the project for Enexis. It was a challenging project with a couple of very difficult aspects. It definitely has been a great experience to work on a real life business problem, and it contributed to my professional development.

I would like to thank my first supervisor of the Eindhoven University of Technology, Fehmi Tanrisever. I would thank him for his support, guidance and feedback during the final part of my study. I really enjoyed our meetings and discussions about various subjects and especially his useful insight in the use of commodity derivatives. Furthermore, I would like to thank Matthew Reindorp for his feedback which provided me new insight what improved the quality of my theses.

I would also like to thank my colleagues and all the people from Enexis who contributed to my master thesis project. Especially I would like to thank Maikel Jonkers for his support during the thesis, his feedback on my work and his knowledge sharing during the thesis. Furthermore, I would thank Maurice Gadet and Hans van Santen for their critical support during the project.

Last but not least, I would like to thank my family and my girlfriend for their support and encouragement during the project and throughout my period as a student.

Michael Heeren

Eindhoven, Augustus 2011

Abstract

This master thesis describes research for improving the distribution loss forecasting process at Enexis B.V. This research compared two commonly used forecasting methods, namely based on multiple regression and based on the time series approach. The objective was to accurately forecast the distribution losses one year ahead on an hourly basis. It is shown that the simple multiple regression methods is the optimal approach to forecast electricity infeed. The electricity infeed is forecasted using the regression approach, taking into account calendar, meteorological, and historical infeed data as independent variables. Electricity infeed is forecasted for every hour of a day and for every grid of Enexis separately. This approach outperforms the approach of using one model. The importance of the independent variables are described in the thesis. Unfortunately, due to the limited amount of data, no significant effect has been found between economic and demographic data and electricity infeed.

It is concluded that electricity infeed can accurately be forecasted using a simple multiple regression approach for every hour of a day. However, the relationship between electricity infeed and electricity distribution losses is still vague. It is therefore recommended to the company to gain more insight in this complicated part of distribution loss forecasting.

Management Summary

Introduction

In the distribution process of electricity, Enexis is held accountable for electricity losses occurred in their network. Causes for these losses are for example fraud, resistance in the cables, measurement errors, et cetera. The loss percentage, about 4% - 5% of the electricity fed into the network, depends on the properties of the network and the amount of electricity distributed.

To cover the electricity lost, Enexis has to purchase electricity on the market. Large volume volatilities in electricity distributed, together with large price volatilities in electricity prices, result in high financial risks and large costs for the distribution network operator. To be able to make a realistic financial forecast and minimize the risk exposure, Enexis is interested in a reliable distribution losses forecast.

The objective of this thesis is to improve the distribution losses forecast, identify important variables that influence the amount of electricity lost, and to expand the forecasting horizon to be able to buy electricity further in advance.

Problem definition

The problem formulation is defined as: *The current forecasting method is unreliable and cannot be used for forecasts with a longer horizon.*

A first analysis of the problem identified five problem areas;

(i) Lack of knowledge on the amount of distribution losses;
(ii) Unreliable electricity demand forecast;
(iii) New trends are not taken into account;
(iv) No guideline/standard for the forecasting process;
(v) Performance of forecast is not evaluated.

This research is mainly focused on the first three problem areas. For the latter two problem areas, recommendations have been made for further improvements of the forecasting process.

The main research question is defined as: *What are the major causes of the unreliable forecast, and how can the forecast be improved into a reliable forecast that can be used for long-term forecast?*

For the research, infeed data as well as data of several independent variables from January 2007 until December 2009 were collected and analyzed to identify relationships between these variables. The relationships have been identified for all distribution networks separately, since some independent variable significantly influence the distribution losses in one network and does not influence the losses in another network. Several functional and non-functional requirements have been defined to ensure a usable and appropriate solution design for Enexis B.V.

Solution Design

The solution was designed in several steps. First, two electricity infeed forecasting methods have been selected to accurately forecast the electricity infeed per hour for the upcoming year. Second, the relationship between electricity infeed and technical distribution losses has been described. Furthermore, the causes of administrative losses has been investigated. Finally, a number of future trends in electricity demand has been examined.

The electricity infeed forecasting methods selected are based on the multiple regression approach. One method is purely a standard multiple regression, while the other method is a multiple regression model with a time series component based on the Box-Jenkins ARIMA approach. The former model is rather basic, while the latter model is much more complicated to implement and adapt. The two methods were compared to determine whether is it worthwhile to include the time series component in the model.

For every hour of a day, a different model has been generated to avoid complicated modelling of intra-day patterns, assuming that some variables have different effects on different hours of a day. Furthermore, separate models has been generated for each distribution grids of Enexis, since several variables only influence the electricity infeed in a particular region (e.g. carnival). For the model, actual infeed data from January 2007- December 2009 is used to forecast the infeed of 2010 per hour.

The independent variables examined to have a significant relationship with electricity infeed can be categorized in calendar, meteorological, demographic and economic data, and are listed in table A below.

Results

During the thesis, it is found that most of the calendar variables used in the model have a significant effect on electricity infeed. Holidays, ‘bridge days’ as well as days near holidays have a significant effect on the electricity infeed. School holidays, “Bouwvak” and carnival does also influence the electricity infeed.

Two unexpected issues were found in the models, namely that economic growth and population does not have a positive relationship with the electricity infeed. These variable were excluded from the models, since it is assumed that this effect is caused by errors in the data or errors in the variables implemented.

Variable	Hypothesis	Result
Calendar data		
Holiday on Sunday	It is expected that a holiday on Sunday does have a minor effect on the electricity demand	Confirmed
Holiday on	It is expected that a holiday on a weekday does have a major effect on the	Confirmed

weekday	electricity demand	
Holiday on Saturday	It is expected that a holiday on a Saturday does have a moderate effect on the electricity demand	Confirmed
Working day before holiday	Less electricity is used the evening before a holiday	Confirmed
Working day after holiday	Less electricity is used the morning after a holiday	Confirmed
Bridge day before holiday	A more significant effect is expected than in hypothesis H5, since it is in between a holiday and a weekend.	Confirmed
Bridge day after holiday	A more significant effect is expected than in hypothesis H6, since it is in between a holiday and a weekend.	Confirmed
Northern School holiday	Less electricity is used in the North of the Netherlands during school holidays	Rejected
Southern School holiday	Less electricity is used in the South of the Netherlands during school holidays	Rejected
Northern ‘Bouwvak’	Less electricity is used in the North of the Netherlands during the holiday period in the construction industry	Confirmed
Southern ‘Bouwvak’	Less electricity is used in the South of the Netherlands during the holiday period in the construction industry	Confirmed
Carnival - Monday	Less electricity is used on this day in the southern part of the Netherlands.	Confirmed
Meteorological data		
HDD	More electricity is used for heating	Confirmed
CDD	More electricity is used for air-conditioning. This effect is more significant than the HDD effect.	Confirmed
Length of the day	This variable represents the time between sunrise and sunset. The longer the duration of the day, the less electricity is used for lightning.	Rejected
Demographic data		
Population	The larger the population, the more electricity will be used.	Rejected
Economic data		
Economic growth Q to Q	As described in literature it is expected that economic growth will have a positive effect on electricity infeed.	Rejected
Other data		
LNMOVA	Historical data has a positive influence on future infeed	Confirmed
HS	There was more electricity infeed in Brabant in 2007 due to the high voltage grid	Confirmed

Table A: Independent variables

The standard multiple regression model outperforms the more sophisticated model with the time series component, however, the in the latter model the maximum errors were found to be lower than the maximum error in the regression model. The results of the electricity infeed forecasts using standard multiple regression models are listed in table B below

Measurement	Brabant	Friesland	Limburg	Maastricht	Noord
MAPE	2.73%	3.07%	3.00%	4.97%	2.84%
MAE	41945.65	1522.06	20693.59	2670.44	34750.32
RMSE	64752.55	2370.57	29464.31	3366.05	51397.88
RMSPE	3.98%	4.44%	4.12%	6.26%	3.98%
MaxAPE	31.06%	31.08%	28.33%	33.50%	29.34%
(On date)	1/4/2010 08:00	1/4/2010 08:00	1/4/2010 08:00	2/15/2010 08:00	1/4/2010 08:00

Table B: Summary of multiple regression models 2010

The average percentage error of the forecast for 2010 is around 3% for all grids (i.e. Maastricht 5%). It can be seen that the forecast accuracy increases when the size of the grid increases (i.e. the larger the grid, the better the infeed can be forecasted). The maximum error is around 30% for all grids and occur in four of the five grids on January 4th. The reason is, that this date is during a school holiday in 2007-2009, but in 2010 it was a regular day. The moving average of historical values is thus not representative for this day, and actually not representative for the whole week of 2010.

Implementation

After the model to forecast the electricity infeed has been established, the relationship between infeed and distribution losses has been examined. Besides the unavailable data, it is concluded that the electricity flow is unknown during the year, and it is thus inaccurate to use the 'cascade-model' to calculate technical distribution losses.

A regression model has been examined with electricity infeed as independent variable and the total distribution losses as the dependent variable. This approach shows that the larger the grid, the more accurate the regression model fits the data. However, due to the fact that the regression model is fit on relatively outdated data from January 2007 – August 2009, it is concluded that this approach is not very useful.

The four trends in the electricity market investigated to have an impact on the distribution losses in the future, were not included in the model at this moment. It is concluded that Smart Grid, decentralized generation and climate changes does not have a major impact on the distribution losses in the near future. The introduction of electric vehicles may significantly influence the electricity demand, and demand profile in the future. Furthermore, this introduction may change the electricity demand in a short time span. Until 2015, is it expected that the number of electric vehicle is too small to have an impact on the electricity infeed.

Conclusions

Concluding, it is found in this master thesis that the relatively simple multiple regression approach is suitable for forecasting electricity demand with an mid-, long-term horizon and a small interval of 1 hour. Important independent variables to take into account in the forecast are calendar data, temperature data and a moving average of historical data. Future trends does not have to be implemented in the model at this moment, however, it has to be investigated what the effect of electric vehicles will be on the electricity infeed, and when this effect will be visible in the data.

Index

Part I: Introduction and Background

1. Problem Context

1.1 Company profile

Enexis B.V. is the successor to Essent Netwerk B.V. The holding company comprises regulated grid manager Enexis B.V. and the commercial activities of Enexis Meetbedrijf B.V. and Enexis Infraproducts B.V. as separate business units. Enexis B.V. is responsible for the construction, maintenance, management and development of the transportation and distribution networks for electricity and gas in the northern, eastern and southern parts of the Netherlands. Enexis operates, with over 130,000 kilometers of electricity cables and 40,000 kilometers of gas pipes, the largest energy network in the Netherlands, distributing energy to 2.6 million customers.

Approximately 74% of the shares in Enexis Holding N.V. are held by six Dutch provinces (Noord-Brabant, Overijssel, Limburg, Groningen, Drenthe and Flevoland) and approximately 26 percent are held by 130 Dutch municipalities in those provinces and in the province of Friesland[1].

The objective for Enexis is to facilitate the market with a reliable and sustainable electricity network, to be custom-oriented and transport electricity for an affordable price, now and in the future. Since DNO's are in a monopolistic situation, the tariffs they are allowed to ask for their services are determined by the Dutch government. For a detailed description of the tariff structure, see Appendix A. Key figures of Enexis can be found in table 1.1 below.

No. of customers	2,631,000
Employees	4,061
Revenue	€ 1,204,200,000
Cost of distribution losses	€ 96,100,000
Profit	€ 193,700,000
Electricity grid length	132,300 km
Headquarter	Rosmalen

Table 1.1: Key figures Enexis 2010[1]

1.2 Electricity supply chain

Since the deregulation of the Dutch electricity market in 1998, the roles of participants in the market has been changed. Before the deregulation, the market was built upon a couple of major players that produced, distributed and supplied electricity for the whole Dutch market. These companies had a monopoly for a particular region (i.e. the consumers didn't have the opportunity to change their electricity supplier). To encourage the free movement of goods in the European Union, the 'Electriciteitswet 1998' (Electricity Act 1998) has been established. The main goal was to create competitiveness on the market to improve quality of the services and reduce prices for the consumers.

Today, six players are responsible for the primary process of delivering electricity to consumers in the Netherlands, as illustrated in figure 1.1.

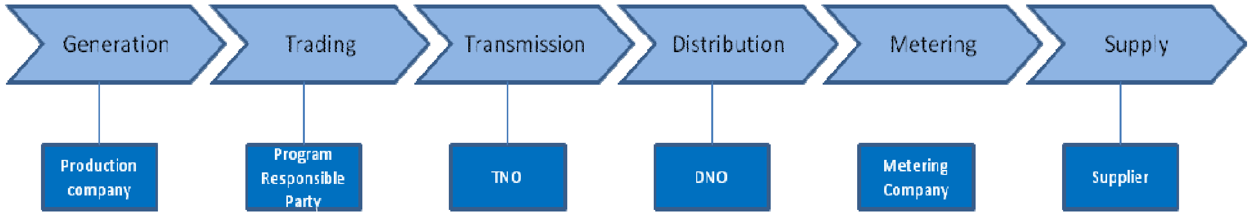


Figure 1.1: Electricity Supply Chain

The **production company** generates electricity out of fossil fuel, uranium, wind etc. The planning of electricity production, transportation and consumption for every 15 minutes of the day is done by the **PV-partij** (program responsible party (PRP)). This planning process is elaborated to ensure a balance on the grid, since electricity cannot be stored and has to be consumed at the time of production. The **Transmission Network Operator** (TNO), Tennet, is responsible for the electricity transmission via the high voltage grid. The **Distribution Network Operators** (DNO) are responsible for the electricity distribution in a region. Today, eight grid operators exist in The Netherlands, divided per region, as illustrated in figure 1.2.

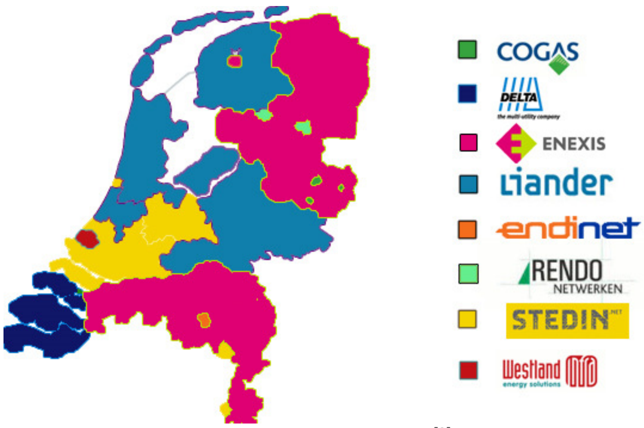


Figure 1.2: Distribution Network Operators [2]

The **metering company** is responsible for the electricity meters. This company collects and sends metering data to the grid operator. The grid operator processes these data into information per consumer, supplier, balance responsible, shipper, etc. The **electricity supplier** is the administrative and commercial contact for the consumer. They deliver the product and charge the electricity used to the consumers. Since the deregulation of the market, a growing amount of electricity suppliers are active in the Netherlands.

1.3 Process steps

Before a description of the project scope will be given, some process steps in the electricity supply chain have to be described. The process steps that will be described are nomination, allocation and reconciliation. A graphical illustration of these process steps can be found in Appendix B.

1.3.1 Nomination

Since electricity cannot be stored, it is important to estimate electricity supply and demand a couple of days in advance for every time span of the days to ensure a balance on the grid. This estimation has to be done to be able to switch on, or switch off electricity generators. This process is called the nomination process and is designed to balance the electricity demand and supply for every moment of the day. This process is carried out by program responsible parties (PRP's). PRP's are the only parties that are allowed to trade on the electricity market.

Any market player is essentially permitted to offer program responsibility services as long as that market player is not a grid company. In fact, the service could be offered by complete outsiders such as insurance companies or financial service providers.

PRP's have to report all electricity agreements (long and medium term contracts) on the electricity market to Tennet, in the form of Electricity-Programs (E-programs). These programs cover 96 program time units (PTU's) (every 15 minutes of a day) and have to be submitted to Tennet one day before the execution at noon.

A couple of days in advance, PRP's start making a forecast of the expected electricity load for the upcoming time units. On the basis of these forecasts they will adjust their long-term positions to match the short-term load forecast.

One day in advance the E-programs have to be delivered to Tennet. Any difference in the E-program and the actual allocated load is called imbalance. To cover this imbalance, the PRP has to buy or sell electricity on the adjustment market.

1.3.2 Allocation

One day after the nomination, the electricity is generated, distributed and consumed. For every PTU the total amount of electricity fed into a distribution network is measured. Furthermore, for every PTU all telemetry connections are measured, which are meters for large consumers. The remaining part (e.g. electricity consumed by households), is estimated by the DNO. This data is not available for each PTU, since these meters are only checked once a year.

The demand for small consumers is estimated using a number of standard profiles (normalised demand profiles) for a couple of categories. Each non-measured consumer is assigned to one of these categories. All profiles are normalised of a year. Together with a consumer's total yearly volume the consumer's demand per PTU can be determined. This demand estimation is called allocation.

The majority of the electricity distributed, about 54%, is consumed by large consumers which are measured every PTU. This data is directly available. A small part is consumed by the government, for example for public lighting. This amount can easily be calculated because it is known when the lights turn on. A third part is assumed to be lost in the grid. These network losses are calculated using a fixed percentage of the electricity distributed. The rest of the electricity distributed is allocated to the small consumers.

Since the profiled consumption and the allocated consumption are known, the DNO is able calculate the ratio between these volumes. This ratio, the measuring correction factor (MCF), is allowed to fall in between 0.8 and 1.2. If the MCF falls outside this range, it is expected that there's an error in the calculation.

After ten working days, all telemetry data is processed and the allocation data is send to Tennet and the PRP's. Tennet calculates the difference between the E-programs and the allocated volumes for every PTU and sends the imbalance and corresponding costs to the PRP's.

1.3.3 Reconciliation

Within 17 months after the PTU, data from households comes available. This takes a couple of months, since the meters are read only once a year. Actual demand is therefore unknown per PTU for every household. However, estimations can be made on a monthly basis using the standard profiles described above.

In the allocation process, the PRP paid for the electricity that was allocated to them. The difference in the amount allocated and the actual amount is known after the reconciliation process. The PRP has to pay or receives the difference in these values. The price they receive is the reconciliation price, which is the weighted market price over the according month. This price is published by Tennet on their website[3].

After 17 months the actual network losses for a particular month are known. The DNO has to pay or receives the difference between the allocated network losses and the actual losses. The reconciliation price is also used for this payment.

2. Project Scope

In the transmission and distribution process of electrical energy, distribution network operators are held accountable for the losses occurred in their network, which are in fact the differences between the amount of electricity that is fed into the grid by electricity generators and the amount actually charged to the consumers. Causes of distribution losses are for example resistance in the cables, measurement errors and fraud.

Enexis' distribution network[1] is divided in six grids, separated by region. These six grids are called:

- Noord; (the provinces Groningen, Drenthe and Overijssel (excl. some small municipalities))
- Friesland; (the city of Leeuwarden)
- Brabant; (the province Noord-Brabant excl. the municipality of Eindhoven)
- Limburg; (the province Limburg excl. the municipalities of Weert and Maastricht)
- Maastricht; (the city of Maastricht)
- Schiphol (the airport). However, "Schiphol" will not be taken into account in this thesis, since this is a private grid and outsourced to Enexis.

[1] See Appendix C for a schematic representation of Enexis' distribution grid

To cover the electricity lost in their grid, Enexis has to purchase electricity on the market. Large volume and price volatilities in electricity distributed, together with the large amount of electricity itself, result in high financial risks for the distribution network operator. To make a realistic financial forecast and minimize the risk exposure, Enexis is interested in a reliable forecast of future distribution losses.

<confidential>

As described in Appendix A, the total income for the DNO is regulated by the government. In order to perform better than other DNO's in the Netherlands, it is important to reduce the costs of distribution losses. For 2010, the total costs for distribution losses were €96,100,000.

The willingness is to improve the distribution losses forecast and to expand the forecast time horizon to be able to buy electricity further in advance. To be able to do that, the current forecasting method needs to be improved, since the method does not take into account several independent variables that effect distribution losses. Furthermore, the influences of future trends on distribution losses are unclear.

The goal of this master thesis is to improve the accuracy of the electricity distribution network losses forecast. Variables that have a significant relationship with distribution losses have to be tested and included in the model. A couple of trends in the electricity supply have to be investigated and taken into account in the forecast if needed.

3. Analysis of the current forecasting method

Currently, the distribution losses are forecasted by the department "Ketenmanagement". This department monitors a couple of processes within the electricity supply chain, analyses the electricity flow, and tries to investigate causes of errors in data. Furthermore, they give an estimation of the network losses in the future based on historical data.

<Confidential>

Figure 3.1: Steps in current forecasting process

4. Problem definition

In this chapter, the specification of the aim of the project will be provided. First, the project formulation will be depicted. Second, the causes and effects of the problem will be defined. Next, a couple of research questions will be provided. Finally, the solution requirements will be provided.

4.1 Problem formulation

The current distribution losses forecast is based on a judgmental forecast (i.e. based on inside knowledge) and a similar day approach. Although the distribution losses forecasting process has been changed in the recent year, it is assumed that the forecasting process can be further improved at some points. In this chapter, some problems related to the distribution network losses forecast are described to be able to create appropriate research questions. The aim of the research is to improve distribution losses forecast to be able to expand the forecast horizon. Therefore, the following problem formulation is defined.

Problem formulation: *The current forecasting method is unreliable and cannot be used for forecasts with a longer horizon.*

4.2 Problems with the Current Forecasting Process

After the problem formulation has been clearly defined, some meetings were scheduled to gain information about the forecasting process, to understand the problem context and to identify the main causes of the problem. These causes can be classified in the following problem areas:

- **Lack of knowledge on the amount of distribution losses:** <Confidential>
- **Unreliable electricity demand forecast:** The electricity demand forecast is unreliable due to poor usage of exogenous variables and an unreliable forecasting method. For the electricity demand forecast, only historical demand data and a GDP forecast are used; the influences of other exogenous variables are unknown and not taken into account in the forecast. The forecasting method used is unreliable and inappropriate for long-term forecasts since historical data are used incorrect and the impact of holidays is used incorrect.
- **New trends are not taken into account:** New trends that may influence electricity demand or the grid usage in the future are not taken into account in the current forecast, since the effects of these trends are unknown. Examples are: the introduction of electric vehicles and decentralized generation.
- **No guideline/standard for the forecasting process:** The forecasting method has been changed a couple times in the recent years, since there's no guideline what approach to use for the distribution losses forecast.
- **Performance of forecast is not evaluated:** The performance of distribution losses forecasts is not evaluated, while the forecasting methods has been changed a couple of times.

The first two problem areas contribute to the inaccuracy of the distribution losses forecasts, while the other three problem areas contribute to the unreliability and the opinion that the method cannot be used for forecasts with a longer horizon.

Although the last two problem areas are relevant, it won't be taken into account in the research. It will be recommended to the company to set up a forecasting guideline and to evaluate the forecasting process each year to find and remove inaccuracies in the forecast. A cause-and-effect diagram is illustrated in figure 4.1 below.

<Confidential>

igure 4.1: Cause-and-Effect Diagram

4.3 Research questions

Figure 4.1 shows the cause-and-effect diagram of the problem statement. The main research question that follows from this diagram is:

What are the major causes of the unreliable forecast, and how can the forecast be improved into a reliable forecast that can be used for long-term forecasts?

The following sub-questions are used to answer the main research question:

- What is the best forecasting method, taking into account the properties of the variables?
- How is the electricity demand influenced by exogenous variables, and how should these variables be implemented in the forecasting model?
- What are the major trends in electricity demand and distribution grid usage and how will they influence future distribution losses?
- What is the relationship between electricity demand and technical distribution losses?
- What is the amount and what are the causes of administrative losses?

4.4 Solution requirements

Functional and non-functional requirements of the solution that have to be taken into account when designing the solution for the business problem are listed below.

4.4.1 Functional requirements

A functional requirement specifies a function that a system must perform. For the solution design, the following functional requirements are important.

Functional requirement	Description
Solution should solve the business problem	The designed solution should generate a reliable forecast on a PTU basis and should be able to be used for long term forecast.
Show an accuracy level	The designed solution should include an indicator that shows the level of accuracy of the forecast.
Provide a confidence interval	The designed solution should be able to generate a confidence interval of the forecast.
Include exogenous variables	The forecasting model should include exogenous variables and it should be possible to add or remove variable in the future.
Show impact of exogenous variables	The model should be built in such a way that the relationship between independent variables and the distribution losses is visual and understandable.

Table 4.1: Functional requirements

4.4.2 Non-functional requirements

Non-functional requirements ensure that the designed solutions are not only based on its theoretical properties, but also on its practical properties. The following non-functional requirements are important for the solution to be designed.

Non-functional requirement	Description
Ease of use	The method should allow employees of Enexis to operate it with little or no training. The interface should be user-friendly.
Ease of implement	The method should be implemented easily in the information system of Enexis.
Data availability	The method should work with the data currently available.
Adaptability	The method should be easily adapted when the problem environment changes.
Applicability	The designed solution should be applicable in the forecasting problem.
Software	The method should be compatible with Windows, and preferably work on Microsoft Office applications, such as Excel and Access.
Understandability	The method should generate understandable output.

Table 4.2: Non-functional requirements

5. Literature Review

During the master thesis preparation, a literature review is executed to gain insight in studies relating to distribution losses forecasting. The main focus was to describe the external factors that have an effect on distribution losses and electricity demand, and to provide insight in accuracy measures and forecasting models available in literature. In this chapter the most important findings of the literature survey are provided.

5.1 Distribution losses

A lot of studies have been done to estimate distribution losses for various distribution network operators. The majority of the studies distinguish two categories in distribution losses; (i) administrative distribution losses, such as fraud, measurement errors and (ii) technical distribution losses, such as resistances in the cable. Administrative losses are caused by actions external to the power system, while technical losses are caused by the use of the grid.

Technical distribution losses can be calculated exactly by hard measurement or by approximating assuming a homogenous loss behavior within the grid (Nadira et al., 2004). The most important factor that influences the amount of technical distribution losses in the grid is the total amount of electricity distributed in the grid. Other factors, such as the physical properties of the grid, are relative stable for a long period ((Flaten, 1988), (Grainger & Kendrew, 1989), (Oliveira et al., 2001), (Yusoff et al., 2009)).

Administrative distribution losses are caused by external factors and very hard to estimate for the distribution network operator. Kadurek (2010) provides insight into illegal use of electricity in the Netherlands. It proposes a methodology for detection of illegal utilization of electricity in future distribution networks with smart metering infrastructures. Suriyamongkol (2002) provides insights in electricity theft, the methods of electricity thefts and the policies used for dealing with electricity theft. Smith (2004) defines electricity thefts and describes the relationship of electricity theft with governance.

It is generally unknown whether there are measurement errors or when and where fraud takes place in the grid. Furthermore, since the actual losses are not known for every time span (metering data comes available only once a year for households), it is very hard for a distribution network operator to calculate administrative losses during the year.

5.2 Electricity demand forecasts

Electricity demand is the major factor that influences electricity distribution losses. It is therefore very important for the DNO to have an accurate demand forecast to be able to estimate the distribution losses.

A lot of literature can be found on electricity demand forecasting methods since it is very important for several parties in the electricity supply chain to have an accurate forecast of the electricity demand.

Two important properties of the electricity demand forecast has to be taken into account when selecting an appropriate forecasting method; the forecasting horizon and the forecasting interval. Based on the horizon, electricity demand forecasts can be divided into three categories:

(i) Short-term forecast (up to 1 week ahead);
(ii) Mid-term forecasts (1 day to several months);
(iii) Long-term forecasts (more than 1 year ahead).

The forecasts for different time horizons are important for different operations within the electricity supply chain. Short-term forecasts are used for day-day operations of the power system, such as scheduling of electricity transactions, scheduling of start-up times of generators and load flow analysis (Kyriakides & Polycarpou, 2007). Mid-term forecasts are used by several parties for maintenance scheduling, minor infrastructure adjustments and to schedule fuel purchases to reduce financial risks. Long-term electricity demand forecast are mainly used for capital planning, to purchase or build new generation facilities and to decide whether or not to increase the transmission capacity.

The main focus in literature is on short-term forecasts, since these forecasts are an important tool in the day-to-day operations of utility systems (Hahn et al., 2009). This dominance is reflected in the number of survey papers on short-term demand forecast, for example Hippert et al. (2001), Tzafestas & Tzafestas (2001), Kyriakides & Polycarpou (2007), Taylor et al. (2006) and Taylor & McSharry (2008). Since the deregulation of the electricity market, more attention is paid to mid-term and long-term forecast, since market players need to negotiate in long term contracts for electricity transactions. Any significant deviation of the quantity forecasted from the actual electricity demand lead to financial penalties for the trader. Feinberg & Genethliou (2005) included some mid- and long-term forecasting methods in their overview.

The forecasting interval is another important property when selecting an appropriate method. This property is more or less related to the forecasting horizon. For short-term demand forecasts, the interval is small, mainly quarter-hourly, half-hourly, hourly or daily data (Huang & Shih (2003), Taylor (2003), Huang et al. (2005) Soares & Medeiros (2006) and Hyndman & Fan (2009)). Typical outputs of short term forecast are the estimated average load per hour, the daily peak load, and the daily or weekly generation (Kyriakides & Polycarpou (2007)). The intervals for mid-term and long-term forecasts are mainly on monthly or yearly data. Typical outputs for medium term forecasts are peak load demand and total weekly, monthly or yearly demand. Annual peak load demand and annual electricity demand are forecasted for years ahead for capital planning and to decide to build new generation facilities (Egelioglu et al. (2001), Mohamed & Bodger (2003) and Bianco et al. (2009)).

5.3 Independent variables

A lot of research is done to find relationship between electricity demand and various demographic, meteorological, economic, and calendar factors. Exogenous variables found in literature which influence electricity demand are summarized in table 5.1 below.

Calendar data	hour of the day, day of the week, holidays, 'bridge days'[2], daylight saving time, school holidays
Meteorological data	temperature, humidity, cloud cover, luminosity, earth's position in the eclipse, sun's altitude, wind speed, solar radiation, climate change
Economic data	GDP, per capita GDP, consumer price index, average salary earnings, production plans of companies, electricity price, industrial expansion
Demographic data	number of households, population growth, local area development

Table 5.1: Factors that may influence electricity demand

The main dependencies between these exogenous variables and the amount of electricity consumed are straightforward. However, to find more sophisticated variables, the researchers always have to take into account local habits and economic properties to find accurate relationships between these variables.

The variables used in demand forecasts depend on the forecasting horizon. For very short term demand forecasts, periods of six hours or less, it is sufficient to use only historical data of the time series (univariate models) (Taylor, 2006). Huang & Shih (2003) used a univariate time series model to forecast short term electricity demand. Taylor (2003) used an exponential smoothing model to forecast up to one day ahead. For these forecasts it is assumed that the future demand is purely dependent on historical demand.

For lead times of more than six hours, calendar data and meteorological data play an important role and need to be included in the model (Heijnen & Bouwmans, 2008). Huang et al (2005) used a time series model with an exogenous variable, temperature, to forecast hourly load forecast one day and one week ahead. Soares & Medeiros (2006) used several day types (day of the week, holiday, working day after holiday, working day before holiday, 'bridge day', etc.) to forecast the electricity demand in the southeast of Brazil up to seven days ahead. They didn't include meteorological variables, because the area covered by the utility company includes a variety of climates. Mirasgedis et al. (2006) included calendar data, temperature and humidity to forecast daily and monthly electricity demand up to one year ahead.

[2] By 'bridge days', the day between a national holiday and a weekend is meant. These days are thus part of a long weekend. (in dutch: "Brugdagen").

For long term demand forecast, the economic data and demographic data have to be taken into account as well. Bianco et al. (2009) used economic and demographic data, such as GDP, GDP per capita, the annual population, et cetera, to forecast the annual electricity consumption in Italy up to 2030. Mohamed & Bodger (2003) used the GDP, electricity price, and population to forecast the annual load in New Zealand up to 2015. Egelioglu et al. (2001) found that the number of tourists and price of electricity correlate with the annual electricity demand for Cyprus. Bruhns et al. (2005) used historical values, temperature, and calendar data to forecast the half hourly demand for one year ahead. Baker & Rylatt (2008) shows the relationship between the number of televisions, computers and other devices in a region and future electricity demand. Finally, Bakker & Boonekamp (2004) used the expected GDP, population, climate change (hot winters, hot summers), number of households/houses, and numerous industry-specific variables to forecast the electricity demand in the Netherlands up to 2050.

5.4 Forecasting methods

Various approaches are used in literature for load forecasting. Most of the approaches use statistical techniques or artificial intelligence algorithms. These approaches are classified in (Hahn et al., 2009);

(i) Classical time series and regression methods;
(ii) Artificial intelligence and computational intelligence methods;
(iii) Hybrid approaches.

Classical time series and regression methods are statistical techniques which require a mathematical model which gives the relation between the dependent variable and the various independent variables. Several methods applied for demand forecasting are linear regression, time series methods, state space models and Kalman-filtering (Hahn et al., 2009).

Artificial intelligence is a relatively new research field. Support vector machines and artificial neural networks fall in this category. The major disadvantage of these approaches is its ability to implement, their usability and their adaptability.

Hybrid approaches are a combination of two or more approaches to take the advantages of different approaches. For example the combination of the end-use approach and an econometric approach would allow integration of physical and behavioral factors in a common framework (Bharadwaj & Meeta, 2001).

Two of the methods, so called end-use methods and econometric methods are broadly used for long term forecasting, while a variety of approaches, such as similar day approaches, various regression models, times series, neural networks, fuzzy logics and expert systems are used for short term forecasting (Feinberg & Genethliou, 2005).

Since the non-functional requirements of the solution are mainly based on the adaptability, ease of implement and usability of the solution, it is decided not to focus on artificial intelligence algorithms. For these reasons, the focus of the literature survey was on the classical time series and regression methods.

The major difference between times series methods and regression methods is that the former is extrapolative, while the latter is not. Extrapolative means that a model is fitted to a set of data and then used outside the range of the data to which it has been fitted. Time series thus assumes that the future is being like the past.

5.4.1 Time series method

Time series forecasting methods can be classified on the type of input variables and how these variables are used in the method. Chatfield (2000) differentiates three forecasting methods;

(i) Judgmental methods; based on 'inside' commercial knowledge;
(ii) Univariate methods; based on present and past value of the series being forecasted;
(iii) Multivariate methods; based on multiple time series variables.

The most famous judgmental method is the Delphi technique, which aims to find a consensus of opinion for a group of experts. However, the focus of the literature survey lies on the univariate and multivariate methods.

The most commonly used models based on time series are Box-Jenkins ARIMA models (Box et al., 1994) and extensions of the Holt-Winter exponential smoothing models. Exponential smoothing takes the weighted average of previous observations and predictions and can be divided into three categories; constant time series, time series including a trend or seasonality, or both trend and seasonality. Box-Jenkins ARIMA models are a combination of autoregressive models and moving average models. Exponential smoothing doesn't give the opportunity to include independent variables in the model, while it is possible in case of box-Jenkins models. Since electricity demand is influenced by variables described in Section 5.3, the focus will lie on these models.

The order of the Box-Jenkins ARIMA model is represented by (*p*,*d*,*q*), where *p* is the order of autoregressive components, *q*, is the order of moving average components and *d*, is the differencing operator.

A general MA(*q*) process can be written as (Box et al. 1994):

$$X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q} \tag{5.1}$$

Where, X_t is the time series;
Z_t is white noise;
θ_i are the parameters of the model;
q is the autoregressive order.

Or succinctly in the form:

$$X_t = \theta(B) Z_t \tag{5.2}$$

Where,

$$\theta(B) = 1 + \theta_1 B + ... + \theta_q B^q \tag{5.3}$$

The autoregressive AR(p) component can be written as:

$$X_t = \phi_t X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + Z_t \tag{5.4}$$

Where, X_t is the time series;
Z_t is white noise;
ϕ_i are the parameters of the model;
p is the autoregressive order.

Or succinctly in the form:

$$\phi(B) X_t = Z_t \tag{5.5}$$

Where,

$$\phi(B) = 1 - \phi_1 B - .... - \phi_p B^p \tag{5.6}$$

A mixed autoregressive moving average model with autoregressive order p and moving average order *q* ARMA(p,q) may be written as:

$$\phi(B) X_t = \theta(B) Z_t \tag{5.7}$$

The differencing operator, d, is used to remove a trend or seasonality in the data, because the model requires the time series to be stationary. A stationary process has the property that the mean, variance and autocorrelation structure is stable over time. If the time series is not stationary, it can be transformed using the differencing component (d=1 for linear trend, d=2 for seasonality and trend). The ARIMA(p,d,q) model may be written as:

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t \tag{5.8}$$

Where, d is the difference operator.

If the time series include a seasonal pattern, which is the case in the electricity demand pattern, then a seasonal ARIMA may be obtained. A SARIMA model with non-seasonal terms of order (*p*,*d*,*q*) and seasonal terms of (*P*,*D*,*Q*) may be written as (Box et al. 1994):

$$\phi(B)\Phi(B)^s(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)Z_t \tag{5.9}$$

Where, ϕ, θ denote polynomials in *B* of order *p*,*q*

Φ, Θ denote polynomials in B^s of order *P*,*Q*

Once the time series is defined and stationary, the order of the components *p*, *q*, *P* and *Q* has to be determined. The primary tool for doing this is running an autocorrelation plot to find relationships between values. Finally the values of the parameters in the model have to be computed, using statistical software, to minimize the sum of squared errors.

5.4.2 Regression method

Regression methods are commonly used in load forecasting and are used to model the relationship between the electricity demand and external factors, such as meteorological, economic and calendar data. Regression methods are very easy to implement, furthermore the relationship between independent variables and the dependent variable is easy to comprehend.

Regression methods can be separated in simple regression, with one independent variable, and multiple regression, with multiple independent variables. Mainly, linear regression is used, however, the influence of temperature on electricity demand is usually modeled non-linear. Multiple regression can be described as:

$$X = a + b_1x_1 + b_2x_2 + ... + b_nx_n + \varepsilon \tag{5.10}$$

Where, *X* is the dependent variable;

a is a constant;

x_i the independent variable;

b_i the coefficients of the model;

ε the error term.

5.4.3 Hybrid approach

Hybrid approaches combine two or more different methods to combine advantages of the separate methods. Time series and multiple regression can be combined in two ways. The first variant includes exogenous variables in the time series. Another variant is to include a time series in the multiple regression which corrects for autocorrelated errors. For simplicity the models are shown based on ARMA models.

The ARMAX (AutoRegressive Moving Average with eXogenous variables) model can be described as:

$$\phi(B)X_t = \sum_{i=1}^{n} \beta_i x_{ti} + \theta(B)Z_t$$

(5.11)

Where, X_i are the independent variables;

B_i are the coefficients of the independent variables.

Or equivalent,

$$X_t = \sum_{i=1}^{n} \frac{\beta_i}{\phi(B)} x_{ti} + \frac{\theta(B)}{\phi(B)} Z_t$$

(5.12)

In this case, the AR coefficients get mixed up with both the error term and the covariates. This makes the model quite complicated. A more intuitive model is to use the time series to correct for autocorrelated errors in the multiple regression model. This model can be written as:

$$X_t = b_1 x_1 + b_2 x_2 + ... + b_n x_n + \varepsilon_t$$

(5.13)

Where the error term forms an ARMA(p,q) model,

$$\varepsilon_t = \frac{\theta(B)}{\phi(B)} Z_t$$

(5.14)

The model can now be written as:

$$X_t = \sum_{i=1}^{n} b_i x_i + \frac{\theta(B)}{\phi(B)} Z_t$$

(5.15)

The difference between 5.12 and 5.15 is that the AR term doesn't influence the covariates in formula 5.15, so the regression coefficients have its usual interpretation. The ease of interpretation in the second one makes it more attractive than the ARMAX model[4].

5.4.4. The M-Competition

Makridakis et al. (1982, 1993) and Makridakis & Hibon (2000) compared a large number of commonly used time series methods in their studies (M1-,M2- and M3-competition). The series were selected on a basis to include various types of time series data and different time intervals. The major conclusions of the first study where; (i) statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones, (ii) the relative ranking of the performance of the various methods varies according to the accuracy measure being used, (iii) the accuracy when various methods are being combined outperforms, on average, the individual methods being combined and does very well in comparison to other methods, and (iv) the accuracy of the various methods depends upon the lengths of the forecasting horizon involved (Makridakis & Hibon, 2000). The results of the M2-Competition were practically the same as those of the M1-competion.

5.5 Method selection

Several properties of forecasting methods are important when selecting a particular method. These criteria are summarized in table 5.2 below.

Forecasting accuracy	The properties of the series being forecasted
Cost	The way the forecast will be used
Expertise of the analyst	Any other relevant contextual features
Availability of computer software	Input data needed

Table Error! Use the Home tab to apply 0 to the text that you want to appear here.**: Criteria for choosing a forecasting method (Chatfield, 2000)**

The accuracy of forecasting methods can numerically be described in a couple of ways. The MAPE and RMSE are common goodness-of-fit measures and will be described below. In some situations it is difficult to choose one 'best' model in advance, in that case it may be sensible to built several models and choose the best model according to the criteria above.

5.5.1 MAPE

MAPE is the acronym of Mean Average Percentage Error and can be used to compare different forecasting methods since it shows a percentage, i.e. it is scale-independent. The drawback of this method is that it only makes sense for non-negative variables having a meaningful zero (Chatfield, 2000).

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|}{n} \tag{5.16}$$

Where, A_t is the actual value;
F_t is the forecasted value;
n is the number of observations.

5.5.2 RMSE

The Root Mean Squared Error (RMSE) is quite similar to the measures above, except that it takes the squared root of the errors. This means that large errors are more important for this measure. The formula is give by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (A_t - F_t)^2}{n}} \tag{5.17}$$

Where, A_t is the actual value;
F_t is the forecasted value;
n is the number of observations.

5.6 Gaps in literature

Although a lot of research has been done in the field of distribution losses forecasts, electricity demand forecasts and to determine the factors that may influence distribution losses, there are some gaps not considered in literature for distribution losses forecasts for a DNO. The gaps addressed in this research will be described in this section.

- **Long term forecast with hourly demand data:** The main gap in literature is the absence of an accurate mid-, long-term forecasting method (1-2years) with a small forecasting interval. All forecasting methods with a small forecasting interval described in literature are short-term forecasting methods, while mid- and long-term forecasting methods focuses on peak-load and total load per year. Since Enexis tend to trade not during the year of execution, one of the objectives of this thesis is to generate a method that is appropriate to forecast electricity demand with an hourly interval and a 1–year forecasting horizon.

- **Important independent variables for mid-term forecasting methods:** It is generally unknown what variables have to be taken into account in mid-term forecasting. In this thesis it will be investigated whether the forecasting horizon is long enough to include trends in population and economic growth and whether it is important to include meteorological variables, since these variables are difficult to forecast one year in advance.

- **Easy to use forecasting method for a DNO:** Most of the available forecasting methods are created and used by experts in forecasting. One objective of this research is to fill this gap in literature and to create a method that can be used easily by the employees of the DNO.

- **Important accuracy measure for electricity purchasing:** The type of accuracy measure that is important for a DNO is not described in literature. In general, papers on electricity demand forecasting focus on the MAPE, since forecasting methods can be compared easily using this measure. However, for some parties large errors may be more important for instance when penalties have to be paid when large errors occur.

Part II: Validation

6. Validation of the business problem

The main objective of this chapter is to validate the business problem defined in chapter 4, using factual data. Before the business problem is validated, insight will be given in the availability of data needed to validate the problem.

6.1 Data availability

To analyze whether the current distribution losses forecasts are reliable, the forecasts and the actual distribution losses have to be compared. To be able to compare the forecasts over the years, it has to be checked whether the forecasting methods are changed in those years. This will be described in the paragraphs below.

6.1.1 Distribution losses forecasts

<confidential>

6.1.2 Actual distribution losses

Actual data of distribution losses is separated in allocated distribution losses and distribution losses from reconciliation. Both datasets are on a PTU basis. The total amount of distribution losses can be calculated by adding the allocated losses and the losses from reconciliation per PTU. It should be remarked that the distribution losses from reconciliation are estimated for every PTU, because of the lack of data per PTU as described in Section 1.3.3.

Allocation data is available from January 2007 until February 2011[3]. Reconciliation data is available from January 2007 until August 2009[2], since it takes 17 months to complete the reconciliation. The data availability is summarized in table 6.1 below.

Year	Objective	Distribution losses forecast	Allocation data	Reconciliation data
2007		-	Available	Available
2008		-	Available	Available
2009		Available	Available	January – August
2010		Available	Available	-
2011		Available	January - February	-

Table 6.1: Data availability

To calculate the accuracy of the current distribution losses forecast, data of the forecast, allocation data as well as reconciliation data are needed. From table 6.1 it can be seen easily that the data needed is only available for the period January 2009 – August 2009.

[3] On March 24, 2011

Although the accuracy of the forecasted total distribution losses could not be analyzed, the forecasts of 2009 and 2010 can be compared with the allocated distribution losses of those years. This won't say anything about the accuracy of today's distribution losses forecast, but some major errors that could still exist in today's forecast, may be visualized.

6.2 Forecast accuracy 2009

<Confidential>

To analyze the accuracy of the forecast of 2009, the forecast will be compared with the actual allocated distribution losses. The data consist of 8760 data points (24h*365d). For each hour the error is calculated by formula 6.1 below. The accuracy measures are calculated as described in Section 5.5. Furthermore, the MaxAPE is calculated to express the maximum error of the forecast.

$$Percentage\ error_t = \frac{A_t - F_t}{A_t} \tag{6.1}$$

Where, A_t are the actual allocated distribution losses
F_t is the forecasted distribution losses to allocate

Figure 6.1 shows the errors of all data points in 2009.

The accuracy measures of 2009 can be found in table 6.2 below.

Measurement	Value
MAE	
MAPE	
RMSE	
RMSPE	
MaxAPE	

Table 6.2: Accuracy measures 2009

6.3 Forecast Accuracy 2010

Measurement	Value
MAE	
MAPE	
RMSE	
RMSPE	
MaxAPE	

Table 6.3: Accuracy measures 2010

Although the forecasting method and the objective have been changed for 2011, it is assumed that the current forecasting method can be improved by eliminating some problem causes which are not eliminated yet. In chapter 7, the potential causes for the unreliable forecasts of 2009 and 2010 will be validated and it will be investigated whether these causes still exist in the current forecast or if they are already eliminated.

7. Validation of the causes of the problem

7.1 Poor time shift

<confidential>

7.2 Inaccurate demand calculation

<confidential>

7.3 Validation of the problem areas

The problem areas described in chapter 5 contribute to the inaccuracy of the forecast and unreliability of the forecasting method. Although the forecasting method has been changed last year, it is expected that some problem areas still exist in the forecasting method.

The problem of the lack of knowledge on the amount of distribution losses cannot be validated since this approach has been changed in the current forecast. There is no data available to check whether this problem is solved. A more sophisticated approach may improve the reliability of the forecasts.

The problem of the new trends that has to be taken into account is not a real problem that has to be solved, since it is more an adjustment on the current forecast. It is a desired system that is not available in the present situation. Further analysis has to be done to investigate the need for these trends in the forecasting method.

The problem area of the changing forecasting method and the absence of evaluations for forecasting methods are relevant, however, these problem will not be examined in this research. It will be recommended to the company to set up a forecasting guideline and to evaluate the forecasting process each year to find and remove inaccuracies in the forecast.

Part III: Solution Design

8. Solution approach

Before a selection is made on the forecasting method and independent variables to be used, the solution approach will be provided. As described in chapter 2, total distribution losses can be divided into technical distribution losses and administrative distribution losses. The technical distribution losses are highly dependent on the total amount of electricity distributed in the grid, while it is basically unknown which factors are related to administrative losses.

The technical distribution losses can be estimated using the total electricity demand data as one of the input variables. Physically, the amount of electricity lost in the grid has a quadratic relationship with the amount of electricity distributed in the grid, however, it is possible that this relationship is not significant in Enexis' grid, for example due to changes in the grid usage during a day (i.e. the relationship is still valid, however, since it is unknown where the electricity is consumed, the loss percentage may vary during a day).

Administrative distribution losses are caused by fraud, measurement errors, et cetera.

A major cause for the inaccurate distribution losses forecast is the inaccurate electricity demand forecast, or inaccurate infeed forecast. The creation of an accurate forecasting method is the main objective of this thesis. It should be investigated during the thesis what variables influence the infeed and how these variables have to be implemented in the model. Furthermore, it has to be investigated in the thesis what methods should be used to forecast the infeed for the upcoming years with a forecasting interval of one hour.

The orange nodes are examples of future trends that may influence the network losses in the future. These trends and their importance will be described at the end of the report.

First, a forecasting method will be created to forecast the infeed for the upcoming years on an interval of one hour. The requirements for the forecasting method have been described in chapter 4. Second, the relationship between infeed and technical distribution losses will be described and it will be investigated whether the current approach can be improved. Furthermore, it will be investigated what causes influence the administrative losses. Finally, a number of future trends will be examined and, when required, implemented in the model.

9. Forecasting method selection

In this section, the forecasting methods to be used will be described. As described in the literature review, a number of different forecasting methods were used in literature to forecast electricity demand. These methods were categorized in classical time series and regression methods and artificial intelligence and computational intelligence methods. A lot of them have been developed for short-term forecasting. For mid- and long-term forecasting only a couple of methods were developed, which are mainly based on econometric methods and the end-use approach.

The end-use approach is not applicable for this problem, since this approach is very sensitive for the quality and amount of end-use data available. A lot of information about customers is needed for this approach and this approach becomes very complicated for the large amount of customers of Enexis.

One important functional requirement of the solution model is to include exogenous variables in the model, to find the relationship between the independent variable and the electricity infeed. This results in an econometric model such as a linear of multiple regression model. However, the infeed pattern shows three seasonal patterns during a year, namely a daily, weekly and yearly season (see appendix F). These seasons are very difficult to correlate with independent variables, since the seasons change over time. The daily season in summer is different from the daily season in winter. No independent variable is found that exactly follow the weekly or yearly pattern. This indicates that it may be appropriate to include a time series component in the forecasting method. However, the inclusion of a time series approach in the model has some major drawbacks.

First, the time series component is much more complicated than the multiple regression model, the times series component is less understandable since the relation between the independent variables and dependent variable is difficult to understand. Second, the time series component requires an extra statistical software package, such as SPSS or SAS, since Microsoft Excel does not support times series. These packages are currently not used by Enexis and it takes some time for the users to get familiar with these software packages. Finally, time series models are rather difficult to generate, since it requires quite a lot of knowledge in and expertise to understand the methodology of time series models, in particular the sophisticated time series models such as ARIMA, SARIMA, etc.

Two econometric models will be compared in this study, namely a standard multiple regression model and a multiple regression model with a time series component. The former model does only take into account exogenous variables. While the latter does also take into account an autoregressive component. These two models will be compared to determine whether it is worthwhile to include the time series component in the model.

Econometric models combine statistical techniques with economic theory. In this report, the electricity demand will be forecasted by finding a correlation between several independent variables, such as weather and economic variables and the electricity demand. The standard multiple regression model will be compared with a econometric model including an autoregressive component. This component is used because it is assumed that future electricity demand shows the same pattern as the historical demand pattern.

Seasons are difficult to forecast with independent variables. It is therefore assumed that the errors of the standard multiple regression model shows an autocorrelated pattern (i.e. the error of observation *h* has a correlation with the error of observation *h-i*, where i is a positive integer). This autocorrelation violates one of the assumptions of the regression model that the errors are uncorrelated. To deal with this autocorrelation, the autoregressive component will be implemented. This forms the first two hypotheses of the project.

Hypothesis 1: The errors of the multiple regression model will show an autocorrelated effect.

Hypothesis 2: The autocorrelated effect can be reduced by incorporating an autoregressive component in the model. The latter model will outperform forecasting accuracy of the former due to the implementation of that component.

For every hour of a day, a different model will be generated. The result is that the daily cycle of the infeed pattern will be removed. This follows the ideas of Soares and Medeiros (2006). By considering a separate model for every hour of a day, complicated modeling of intra-day patterns will be avoided. For example the effect of New Year's day, what could have a positive effect on the electricity infeed during the night and a negative effect during the day.

Hypothesis 3: Exogenous variables will have different effects on the electricity demand at different hours of a day.

As described in the literature review, the seasonality of the electricity demand can be implemented in the time series model. The infeed shows three seasonal patterns as can be seen in appendix F. By separating the models per hour, the daily effect will be removed. However, statistical software does only support one seasonal effect in case of Box-Jenkins time series. The time series component should be a form of the Box-Jenkins family, since these time series allow to include exogenous variable in the model.

Although one season has already been removed, another season has to be removed as well. To deal with this problem another independent variable will be included in the model. To simulate the yearly season, the average hourly infeed of the last three years will be used as an independent variable. This variable is based on the assumption that the future electricity demand follows the same pattern as historical demand, and the future demand is thus correlated with the moving average of historical values.

In total, two models will be compared, a standard multiple regression model, and a multiple regression with a time series component. These models will be generated for every hour of a day and for every distribution grid of Enexis. For readability purposes, only the results of the two largest distribution grids in terms of electricity infeed will be provided in this report. These distribution grids are located in Brabant ("Brabant") and three northern provinces Overijssel, Drenthe, Groningen ("Noord"). By providing the results of two grids, differences in independent variables can be shown, such as carnival and school holidays. A summary of all grids will be provided at the end of the results.

At the time of collecting the data, infeed data was available until February 2011. Data of January 2007 – December 2009 will be used to forecast the infeed of 2010. The independent variables used in the model contains actual values (i.e. actual realized temperature data is used instead of forecasted temperature data). Using actual values, the forecasts can be compared more realistic, since the inaccuracy of forecasted temperature data will not be taken into account.

10. Selection of dependent and independent variables

10.1. Dependent variable

The dependent variable of the model is the total infeed per hour. Infeed is defined as the sum of all electricity that is consumed in the grid, including the electricity that is consumed by the grid itself (technical network losses). This infeed data is available for all five grids of Enexis from January 2007 – February 2011. These amounts will not be summed, since each grid has its own physical properties. To calculate the amount of distribution losses from the infeed data, the amount of infeed per grid has to be known. Furthermore, some variables will do significantly influence electricity demand in one grid and won't influence the infeed of another grid (e.g. carnival).

10.2. Independent variables

In the literature review, the major independent variables used in literature were provided. In this section the independent variables to be used in Enexis model will be described according to their availability, understandability and usability.

Although Enexis stored a lot of data in their database, no input data is available for forecasting the infeed. To develop the forecasting model, a dataset has to be generated including the infeed and the independent variables for every hour of a day from January 2007 – present. The independent variables data has to be forecasted for the upcoming years. For calendar data, this data is fixed, however, input variables for meteorological data have to be forecasted. This can be a very though problem and the accuracy of the final model is influenced by the accuracy of the input variables forecasts. However, the accuracy of the input variables falls outside the scope of this project.

10.2.1. Calendar data

Calendar data is available for a very long period. The variables to include in the dataset are hour of a day, day of the week, all national holidays, 'bridge days', working day near a holiday, and school holidays. The holidays included in the model are listed in appendix G. The national holidays, 'bridge days' and school holidays are separated according to the days of the week, because the effect of these days differs on different days (i.e. a holiday on a Sunday has hardly no effect on the electricity infeed, while a holiday on a weekday does has a significant effect). No distinction will be made between the nature of the holidays. Although Christmas has a different effect on the infeed than Easter, no distinction will be made due to the limited infeed data available. For example, it cannot be determined what the effect is of Queen's day on a Saturday, since this situation did not occur between 2007 and 2010. In the proposed model, the effect of another holiday on a Saturday will be used in the forecast.

Another variable to include is the working day near a holiday. Here a distinction is made between a working day prior to a holiday and a working day after a holiday. It is assumed that these variables affect the electricity infeed the evening before a holiday and the morning after.

A third variable to include in the dataset indicates whether it is a 'bridge day' (i.e. a day in between a holiday and a weekend). It is assumed that a lot of people take an extra day off during 'bridge days' that might influence the electricity infeed.

A fourth calendar variable to include is school holidays. During these days, a lot of people go on vacation and that should have an effect on the electricity infeed. The school holidays included in the model are listen in appendix G. For this variable the school holidays of primary school are used. These holidays are regulated by the Dutch government. School holidays are separated in three regions, the Northern, Middle and Southern part of the Netherlands. The school holiday in the Northern parts of the Netherland will have an effect on the electricity infeed on the 'Friesland' and 'Noord' grid. The other grids of Enexis are situated in the Southern part of the Netherlands, and thus the school holiday of this region is used for these grids.

A fifth calendar variable is called "Bouwvak" in Dutch. This variable indicates the period in which companies in the construction industry are closed. This period is also separated in three regions (i.e. Southern, Middle and Northern part of the Netherlands). A lot of people working in the construction industry go on vacation during this period. It is assumed that this will have an effect on the electricity infeed.

Finally, a variable will be included to indicate carnival. This variable will only have an effect on the electricity infeed in the southern part of the Netherlands. A lot a people take some days off to celebrate carnival. It is not possible to combine this variable with the other holidays, since it is only applicable for the grids in the Southern part of the Netherlands.

10.2.2. Meteorological data

Historical meteorological data is available on the website of a Dutch meteorological institute (KNMI)[5]. This institute measures and offers a lot of daily data, such as cloud cover, humidity, mean daily temperature, wind speed, precipitation amount, etc. The disadvantage of including meteorological data is that it has to be forecasted for the upcoming years. These long-term weather forecasts are inaccurate by definition. Furthermore, for most of the meteorological effects it is unknown what kind of effect it will have on the electricity infeed and whether there is a correlation between several meteorological variables.

All meteorological variables available were tested whether there is a correlation between the particular variable and electricity infeed. Although there were some significant effects between humidity and electricity demand at some hours of a day, it is decided to include only the temperature effect in the model, since the effects were only sporadically significant.

The outdoor temperature data used is the daily mean outdoor temperature. It is plausible that the outdoor temperature in the Southern provinces differs from the outdoor temperature in the Northern provinces, because the data was not available for the different areas, it is decided to use the average temperature of the Netherlands. Furthermore, it might be an improvement to use the average temperature per hour instead of the average temperature per day. This data was also not available.

As described in literature it is assumed that the outdoor temperature has a U-shape effect on the electricity infeed. The relationship is non-linear, increasing for both hot days and cold days. Figure 10.1 shows the relationship between the normalized electricity infeed and outdoor temperature. As can be seen, the electricity infeed has a minimum around a temperature of 17°C. The variable is therefore split into two variables which both have a linear effect on temperature; heating degree days and cooling degree days.

The heating degree days (HDD) and cooling degree days (CDD) are calculated using the following equations:

$$HDD_t = \max(T_{ref} - T_t, 0)$$

$$CDD_t = \max(T_t - T_{ref}, 0)$$

Where T_t is the daily average temperature for day t, and T_{ref} is the reference temperature, which is equal to 17°C.

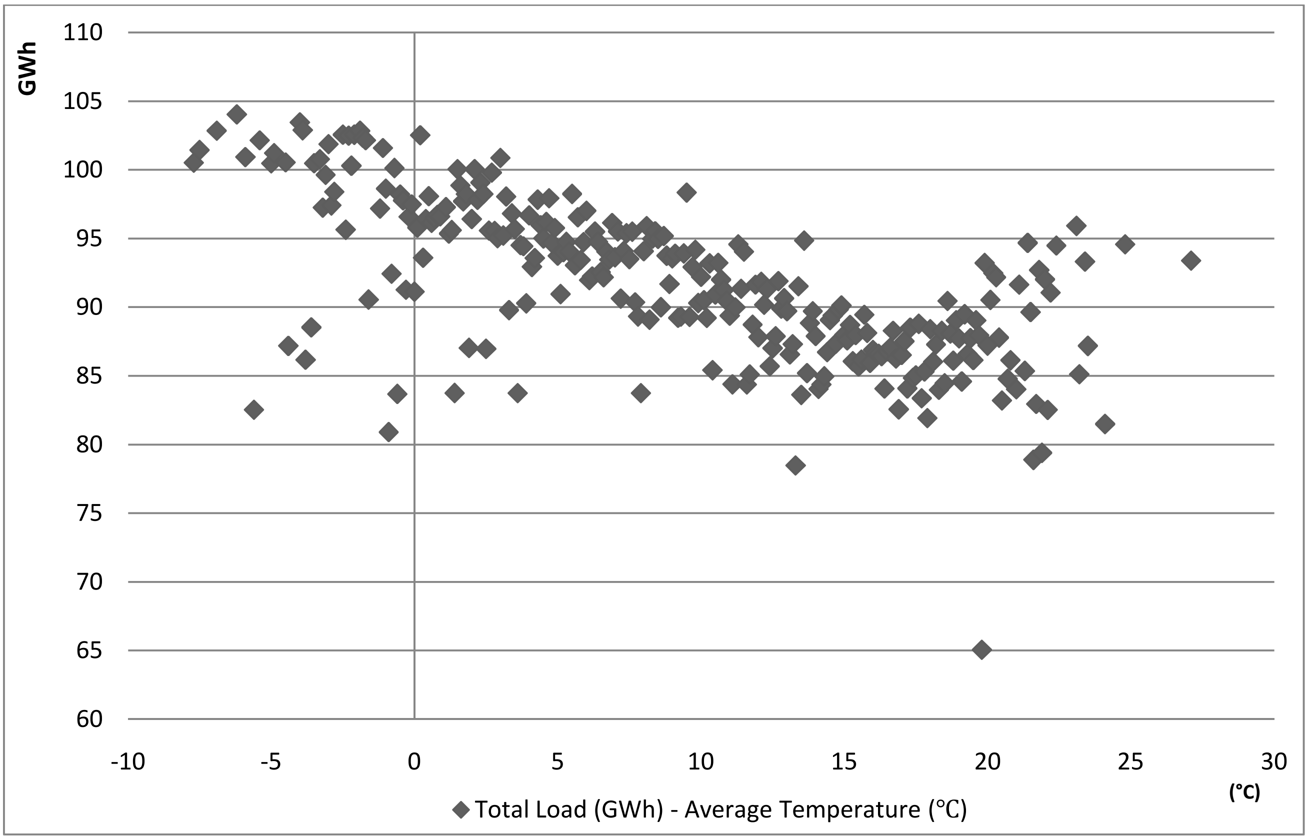


Figure 10.1: Relationship between average daily temperature and total electricity infeed per day

10.2.3. Demographic data

For demographic data, only the national population data is available from Statistics Netherlands (CBS)[6]. This data is available per month for the recent years. There are no prognoses made for the upcoming years. It is assumed that the population has a positive effect on the electricity infeed, however, it is possible that the trend in the national population differs from the trend in the population of the particular provinces in which Enexis operates.

10.2.4. Economic data

For economic data only the economic growth is included in the dataset. This data is available from Statistics Netherlands (CBS)[6] for every quarter of a year. It is assumed that the economic growth has a positive relationship with the total infeed, however, it is plausible that the national economic growth does not give a good representation of the economic growth of the provinces in which Enexis operates. Unfortunately, no other economic growth data was available for these provinces.

10.2.5. Other data

Besides the data described above, two dummy variable were included in the dataset. One of them is a moving average of historical infeed data. This variable is included to simulate the yearly season. The other dummy variable called 'HS' was needed to separate the infeed data in Brabant of 2007 from the other data. Until 2007, the infeed in Brabant has been calculated in a different way, since the high voltage grid was included in the grid until that year.

10.2.6. Total independent variables and hypothesis

The independent variables used in the models are summarized in table 10.1 below. The table includes the hypothesis of the effects of the independent variables on electricity infeed.

	Variable	Scale	Effect	Explanation
Calendar data				
H4	Holiday on Sunday	Nominal	-	It is expected that a holiday on Sunday does have a minor effect on the electricity demand
H5	Holiday on weekday	Nominal	--	It is expected that a holiday on a weekday does have a major effect on the electricity demand
H6	Holiday on Saturday	Nominal	--	It is expected that a holiday on a Saturday does have a moderate effect on the electricity demand
H7	Working day before holiday	Nominal	-	Less electricity is used the evening before a holiday
H8	Working day after holiday	Nominal	-	Less electricity is used the morning after a holiday
H9	Bridge day before holiday	Nominal	--	A more significant effect is expected than in hypothesis H5, since it is in between a holiday and a weekend.
H10	Bridge day after holiday	Nominal	--	A more significant effect is expected than in hypothesis H6, since it is in between a holiday and a weekend.
H11	Northern School	Nominal	-	Less electricity is used in the North of the Netherlands during school holidays

	holiday			
H12	Southern School holiday	Nominal	-	Less electricity is used in the South of the Netherlands during school holidays
H13	Northern 'Bouwvak'	Nominal	-	Less electricity is used in the North of the Netherlands during the holiday period in the construction industry
H14	Southern 'Bouwvak'	Nominal	-	Less electricity is used in the South of the Netherlands during the holiday period in the construction industry
H15	Carnival - Monday	Nominal	-	Less electricity is used on this day in the southern part of the Netherlands.
Meteorological data				
H16	HDD	Metric	+	More electricity is used for heating
H17	CDD	Metric	++	More electricity is used for air-conditioning. This effect is more significant than the HDD effect.
H18	Length of the day	Metric	-	This variable represents the time between sunrise and sunset. The longer the duration of the day, the less electricity is used for lightning.
Demographic data				
H19	Population	Metric	+	The larger the population, the more electricity will be used.
Economic data				
H20	Economic growth Q to Q	Metric	+	As described in literature it is expected that economic growth will have a positive effect on electricity infeed.
Other data				
H21	LNMOVA	Metric	+	Historical data has a positive influence on future infeed
H22	HS	Nominal	+	There was more electricity infeed in Brabant in 2007 due to the high voltage grid

Table 10.1: Independent variables included and hypothesis of their effect.

11. Assumptions of the statistical techniques

Multivariate analysis requires that the assumptions underlying the statistical techniques are met. (Hair et al., 2005). Variables included in the multiple regression model should meet the assumptions of parametric data. Four important statistical assumptions can be distinguished;

- Normality;
- Homoscedasticity;
- Linearity;
- Absence of correlated errors.

Appendix H shows that the assumption of normality of the data is not met. To reduce the impact of heteroscedasticity resulting from the non-normally distributed data, the large amount of data and its high temporal frequency (Pardo, 2002), (Mirasgedis, 2006), the natural logarithm of the dependent variable will be taken.

Although the data is still non-normally distributed, it is improved in relation to the untransformed data. Since the dataset is very large (>200 observations) the impact of non-normally distributed data may be negligible.

The other assumptions summarized above will be verified in the next chapters when using the statistical techniques.

Part IV: Results

After the solution approach, the dependent and independent variables and the hypotheses have been defined, the models to forecast the electricity infeed were generated and tested. In this phase, the results of the two models will be provided. First, the assumptions of multiple regressions will be checked. Second, the results of the multiple regression will be given. Finally, the results of the multiple regression model with the autoregressive component will be provided.

12.Assumptions of multiple regression methods

The assumptions to be checked before the multiple regression can be used are (Hair et al., 2005);

- Normality of dependent variable;
- No multi-collinearity;
- Homoscedasticity of the errors;
- Linearity of the relationship between the dependent and independent variables;
- Errors are normally distributed;
- Independence of errors.

In chapter 11 it has been shown that the dependent variable, electricity infeed, was not normally distributed. To reduce the impact of heteroscedasticity resulting from non-normally distributed data, the large amount of data and its high temporal frequency (Pardo, 2002), (Mirasgedis, 2006), it has been decided to transform the dependent variable using the natural logarithm. After the dependent variable has been transformed, it was found that is was still not normally distributed, however, since the amount of observations was very large (>200), this assumption could be violated (Hair et al, 2005).

A verification of the other assumptions can be found in appendix I. This verification is performed on only two regression models (Brabant, Hour3 & Hour18). It is assumed that the other models show comparable results.

Appendix I shows that all assumptions are met, accept for the assumption of independence of errors. The Durbon-Watson test in SPSS shows that the errors are not independent. As described in hypothesis 1, it is assumed that the errors of the model show an autocorrelated effect. This hypothesis is confirmed as can be seen in the autocorrelation plots in figure 12.1 below. The autocorrelation plots for Hour3 and Hour18 show that the error is highly correlated with all prior errors. For now, the assumption of uncorrelated errors will be violated. However, another models will be created with an autoregressive component included to deal with the problem of autocorrelation.

The consequence of violating this assumption is that the confidence intervals, if generated, will be estimated too small. Furthermore, the use of the F-test and t-tests is not strictly valid and the estimate of the coefficients may be unstable (Makridakis et al, 1998). The autocorrelation problem can therefore lead to misleading results.

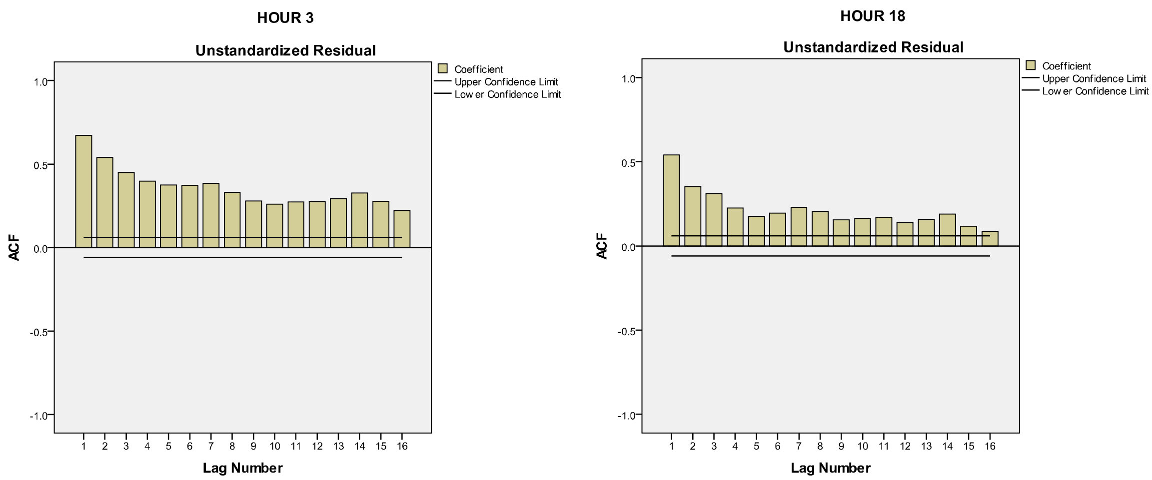


Figure 12.1: Autocorrelation plot Hour3 and Hour18

13. Results of the regression model

In this chapter, the results of the regression model will be shown. For readability purposes, only two of the five models will be shown, namely 'Brabant' and 'Noord'. These models are interesting, since they face the two largest grids in terms of electricity infeed, and the grids are geographically different, resulting in some different independent variables needed in the model.

The coefficients obtained by using the multiple regression for each hour of a day are presented in the tables in appendix J for both grids. These tables are the summary of all the models used (24 per grid). The independent variables were chosen on the significance level $\alpha=0.05$.

An important modification in the model is that two important independent variables, which were described in chapter 10 were not taken into account in the model, namely 'economic growth' and the Dutch 'population'.

For economic growth, no relationship has been found for most of the models. In other cases, the relationship has been found to be negative, indicating that the electricity infeed decreases if the economy growths. This negative relationship is unexpected and must have to do with inaccurate data. The relationship between electricity demand and GDP for the last 25 years in the Netherlands is shown in appendix K. This relationship has been positive for the last decades, and it is thus unexpected that this relationship is negative for the electricity infeed for Enexis. A possible reason for this phenomenon is that the economic growth for the particular provinces where Enexis operates is different from the total economic growth in the Netherlands. Unfortunately no information could be found about the economic growth per province. Another cause of this phenomenon can lie in the fact that the ratio of large companies and small consumers could be different in these provinces. Economic growth can influence the electricity demand for business consumers, but may have a minor effect on the electricity demand for households.

The independent variable ‘population’ had a significant effect on electricity infeed, however the correlation has been found to be negative as well. Including this variable in the model leads to very inaccurate forecast for 2010. It therefore has been decided to not include this variable. The reason for the negative effect could be the same as for the economic growth, namely that the population for the provinces had a different trend than the total Dutch population.

A not included variable may also play a significant role and may influence the effect of population and economic growth on electricity infeed, for example a trend in a more efficient use of electricity by companies and consumers.

13.1. Evaluation of hypotheses

At this point, most of the hypothesis described in chapter9 and 10 can be tested. **Hypothesis 1** is already **confirmed** in chapter 12. It has to be checked whether the model with an autoregressive component outperforms the model without the autoregressive component. This will be evaluated in chapter 14.

From the tables in appendix J h**ypothesis 3** can be **confirmed**. The tables show that some independent variables have a significant effect at a particular time of a day, and no effect on other times of a day. For example the variable ‘Bridge day before holiday’ and ‘Working day before holiday’. These variables only have an effect on the electricity infeed in the evening. From the confirmation of this hypothesis it can be concluded that modeling each hour of the day separately is an important improvement for the model.

Hypothesis 4 can be **confirmed**, however, the occurrence of a national holiday on a Sunday does only have an effect on the electricity infeed during daytime. The effect of a holiday on a weekday is more pronounced as expected, so **hypothesis 5** can also be **confirmed**. A holiday that occur on a Saturday does also have a negative effect on the electricity infeed (**hypothesis 6 confirmed**).

Hypotheses 7 and 8 can be **confirmed** as well. There seems to be a positive effect on the electricity infeed during the afternoon the day after a holiday. This effect is quite strange and cannot be explained.

Hypotheses 9 and 10 can be **confirmed** as well. The occurrence of a bridge day before a holiday does only have an effect on the electricity infeed in the evening, while the electricity infeed during a bridge day after a holiday is affected during the whole day.

Hypothesis 11 and 12 are **rejected**. As expected, the occurrence of school holidays have an effect on the electricity infeed, however this effect seems positive. More electricity is fed into the grid during school holidays. This effect is strange and cannot be explained.

During the holidays in the construction industry, there is less electricity fed into the grid, as can be seen in the tables in appendix J. Therefore **hypotheses 13 and 14** are **confirmed**.

Carnival does have a negative effect on the electricity infeed. **Hypothesis 15** is **confirmed.**

As expected, the derivatives of the outdoor temperature, CDD and HDD, both have a positive effect on the electricity infeed. The effect of cooling degree days is as expected more pronounced than the effect of heating degree days. **Hypotheses 16 and 17** are **confirmed.**

Hypothesis 18 is **rejected**. It was expected that the duration of the day would have a negative effect on the electricity infeed, however the tables show a positive effect.

The effects of population and economic growth, as described in hypotheses 19 and 20, were not significant. These variables were left outside the model, as already mentioned above. **Hypotheses 19 and 20** are **rejected**.

The two other variables, a moving average of historical data and the effect of the high voltage network in Brabant were significant and had the same effect on the electricity infeed as expected. **Hypotheses 21 and 22** are **confirmed.**

13.2. Validation of model

The regression models were generated using the data from January 2007 until December 2009, also called training data. To validate the robustness of the model, the electricity infeed in the five grids are forecasted using the coefficients generated in the regression models. The 24 models per grid are used to generate the forecasts for each hour of the day in 2010. All 120 model results are used to validate the forecast of 2010. The results of the forecast are described in table 13.1.

Measurement	Brabant	Friesland	Limburg	Maastricht	Noord
MAPE	2.73%	3.07%	3.00%	4.97%	2.84%
MAE	41945.65	1522.06	20693.59	2670.44	34750.32
RMSE	64752.55	2370.57	29464.31	3366.05	51397.88
RMSPE	3.98%	4.44%	4.12%	6.26%	3.98%
MaxAPE	31.06%	31.08%	28.33%	33.50%	29.34%
(On date)	1/4/2010 08:00	1/4/2010 08:00	1/4/2010 08:00	2/15/2010 08:00	1/4/2010 08:00

Table 13-1: Summary of multiple regression models 2010

The average percentage error of the forecast for 2010 is for all grids around 3% (i.e. Maastricht 5%). It can be seen that the forecast accuracy increases when the size of the grid increases (i.e. the larger the grid, the better the infeed can be forecasted). The maximum error is around 30% for all grids and occur in four of the five grids on January 4^{th}. The reason is that this date is during a school holiday in 2007-2009, but in 2010 it was a regular day. The moving average of historical values is thus not representative for this day, and actually not representative for the whole week of 2010 as can be seen in figure 13.1 and 13.2. The maximum error in Maastricht is on February 15^{th}, during carnival.

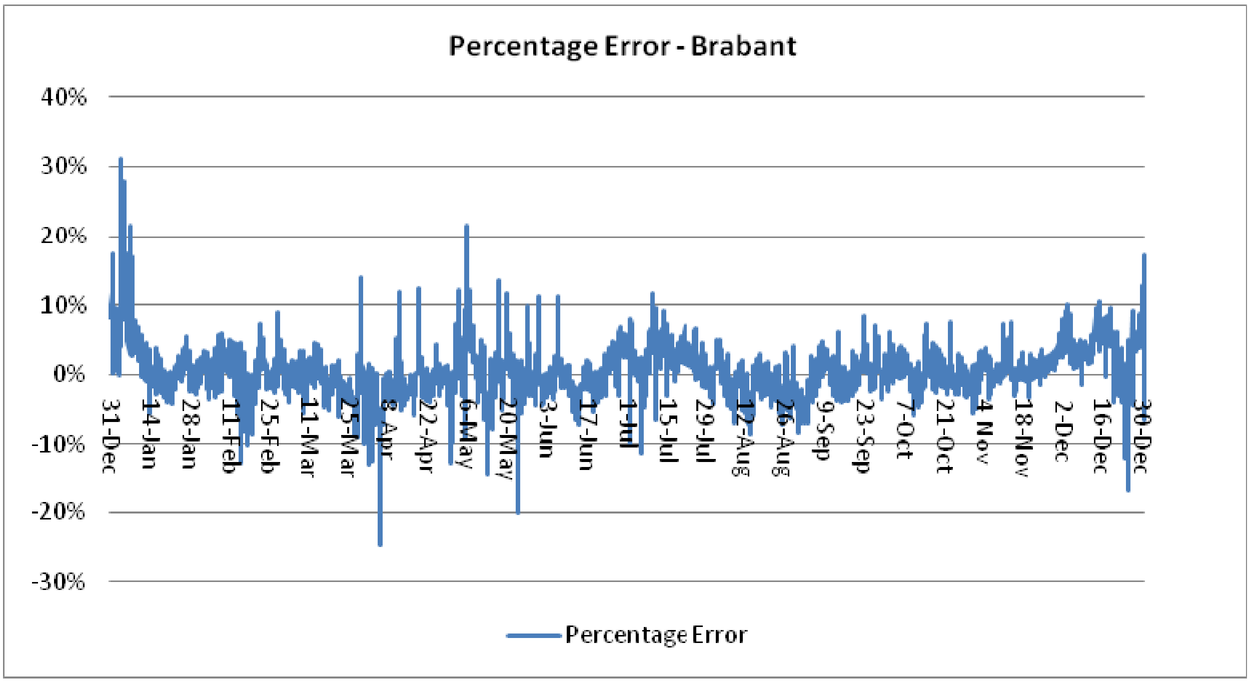


Figure 13.1: Percentage Error Multiple Regression model Brabant 2010

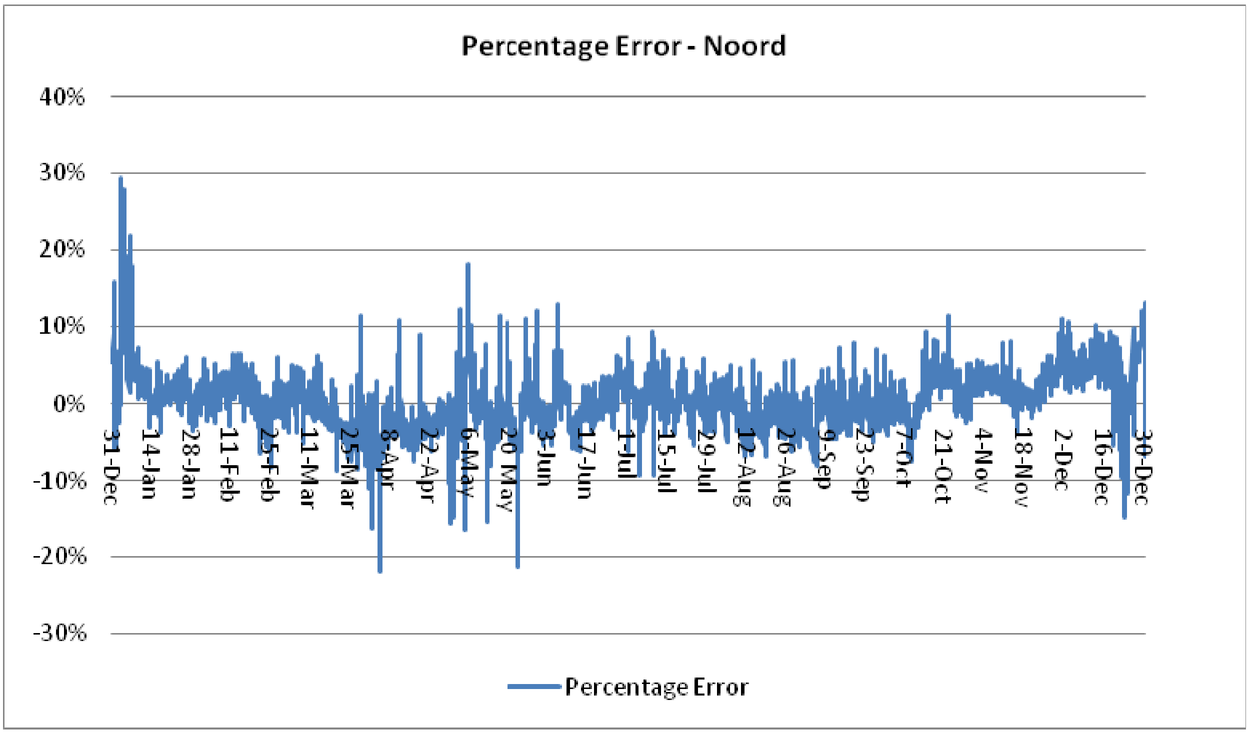


Figure 13.2: Percentage Error Multiple regression model Noord 2010

It seems that large errors also occur during national holidays. The reason for these errors could be the limited amount of data of these holidays and the assumption that the effect of different holidays are equal during the year. In the future, it should be investigated whether it is possible, and an improvement, to included a separate variable for each holiday.

Another major error can be found on May 5th, independence day. This day is a holiday once in five years in the Netherlands, however, the infeed is not as much effected as during other holidays, since most of the companies are not closed that day. Since there was no historical data of independence day five years ago, it could not be implemented in the model.

14.Results of time series model

To deal with the autocorrelation of the errors in the regression model, a regression model including an autoregressive component will be generated.

Normally, the order of autoregressive components can be determined by analyzing the autocorrelation function and the partial autocorrelation function, however, in this case the autocorrelation functions of the 120 models show different orders. To improve the usability of the models, it is decided to include the same autoregressive and moving average components for all models. After some trial and error, it seems that the ARIMA (1,0,1)(0,1,1) model is appropriate for all models. The advantage of this model is that it is relatively simple and relatively easy to understand.

The ARIMA (1,0,1)(0,1,1) model is a model that includes two moving average components, a non-seasonal and a seasonal (season = 7 days) component, a differencing factor to meet the assumption of stationarity, and a non-seasonal autoregressive component.

For Brabant and Noord, the results of the time series model can be found in table 14.1 and figures 14.1 and 14.2 below. It seems that this model does not outperform the regression model, since the average percentage error is larger for this model. The advance of the time series model is that the maximum errors are smaller, as can be seen in table 14.1.

Since Brabant and Noord are the two largest and most important grids, and no improvements has been found in the forecasting accuracy of the time series models, it is decided not to run the 24 models for Friesland, Limburg and Maastricht.

Measurement	Brabant	Friesland	Limburg	Maastricht	Noord
MAPE	3.22%	N/A	N/A	N/A	3.72%
MAE	49587.27	N/A	N/A	N/A	46900.20
RMSE	66089.00	N/A	N/A	N/A	60782.45
RMSPE	4.16%	N/A	N/A	N/A	4.66%
MaxAPE	16.86%	N/A	N/A	N/A	20.12%
On date	1/4/2010 07:00	N/A	N/A	N/A	5/5/2010 14:00

Table 14-1: Summary of time series models 2010

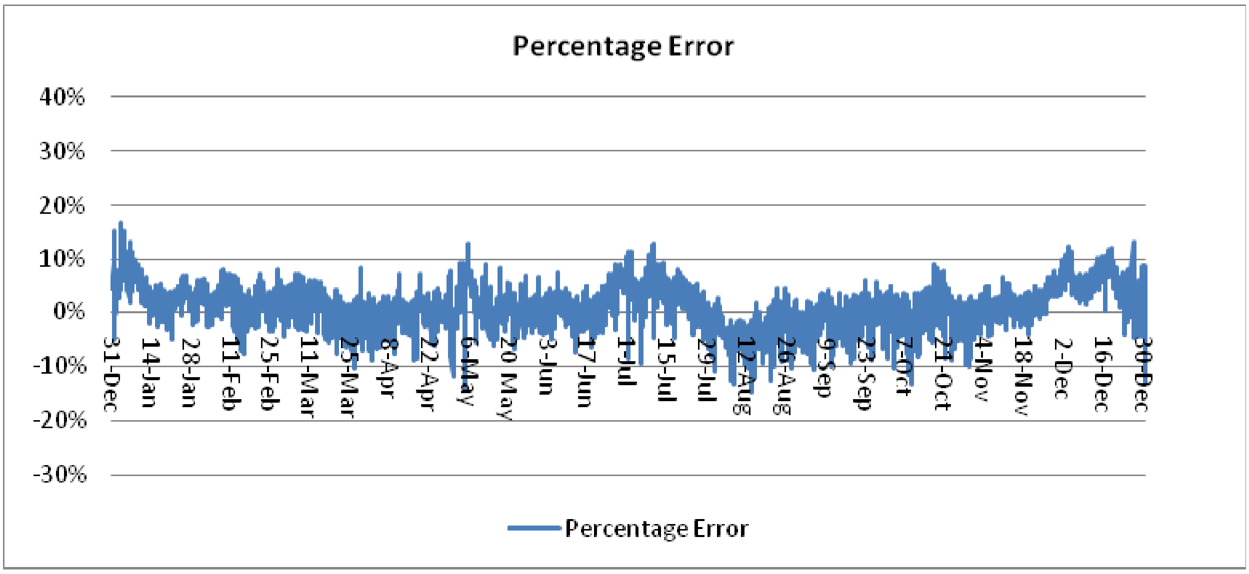


Figure 14.1: Percentage Error Time Series Model Brabant 2010

The model including the autoregressive component does meet the assumption of uncorrelated errors. This assumption has been violated in the multiple regression model, but as can be seen in the autocorrelation functions of the residuals of Brabant Hour3 and Hour18 in figure 14.3, the residuals are almost uncorrelated.

The coefficients of the variables used in the model can be found in Appendix L. The tables in this appendix show equal results as the tables of appendix J.

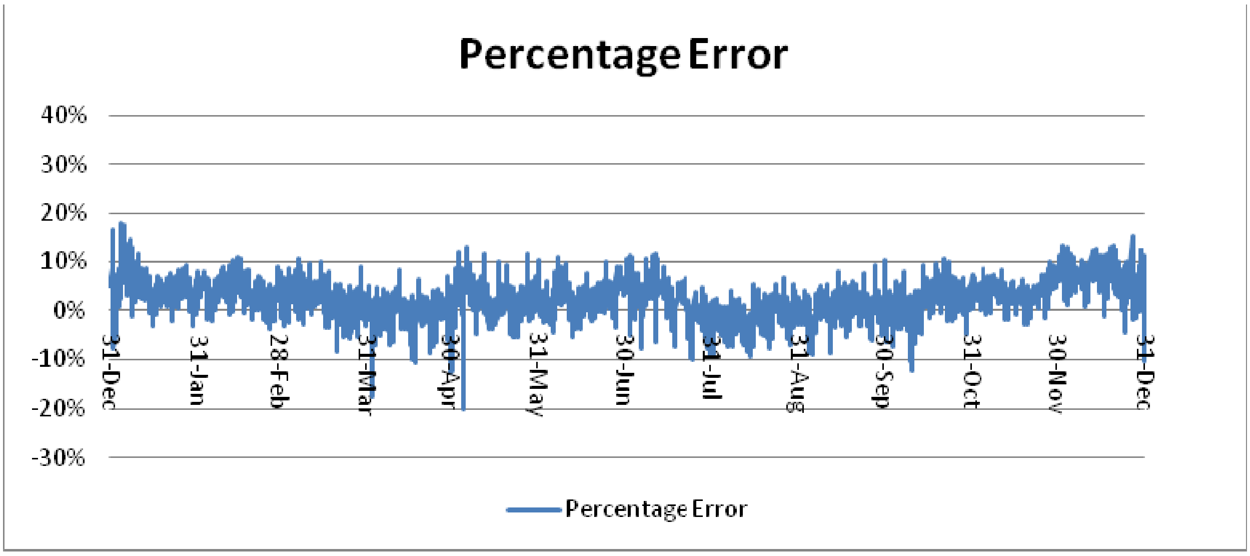


Figure 14.2: Percentage Error Time Series Model Noord 2010

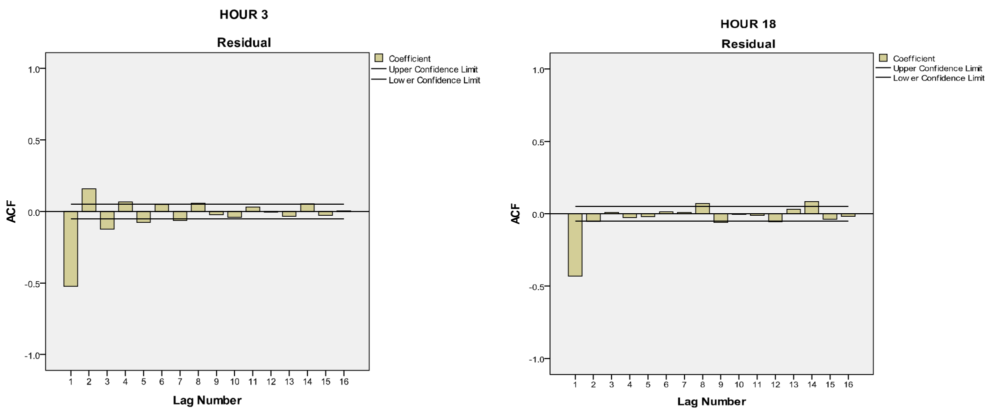


Figure 14.3: Autocorrelation function Hour3 and Hour18

Part V: Implementation

Before one of the methods can be implemented by the company, the relationship between the dependent variable and distribution losses has to be determined. Furthermore, it will be investigated whether some new trends in the electricity market will influence the distribution losses in the future and whether these trends have to be taken into account in the forecast. Finally, one of the forecasting methods elaborated above has to be selected based on their characteristics, taken into account the current environment, the purchasing process and the functional and non-functional requirements of the solution.

15.Distribution losses

As described in the solution approach, the next step is to find the relationship between electricity infeed and distribution losses per hour. This seems a rather difficult step, since the actual distribution losses per hour are unknown for the company (i.e. actual demand per hour is unknown since households are only metered once a year, as described in Section 6.1.2).

In the paragraphs below, information about distribution losses will be provided. First, the characteristics of technical distribution losses will be investigated. Second, the characteristics of administrative losses will be described.

15.1. Technical distribution losses

The amount of electricity lost in the grid due to resistances in the cables and transformers can be traced by hard measurement or by approximation, assuming a homogenous loss behaviour within the grid.

<confidential>

15.2. Administrative distribution losses

<confidential>

15.3. Regression model

As described in the two sections above, it is impossible at this moment to find an accurate relationship between technical and administrative losses and electricity infeed. However, it is possible to relate the total distribution losses with electricity infeed. The scatterplots of total distribution losses and electricity infeed per grid can be found in figure Appendix M.

Grid	Function	R^2
Brabant		0.84
Limburg		0.78
Maastricht		0.41
Noord		0.70
Friesland		0.83

Table 15-1: Loss percentage per grid

Taken into account that these loss percentages are generated from infeed and loss data from January 2007 until august 2009, and thus somewhat outdated, and the relative poor fit of the linear trend line, it can be concluded that this approach is not very useful for forecasting.

16.Trends in the electricity market

Since the amount of distribution losses are highly influenced by the total electricity demand and the use of the grid, it is important to investigate how these factors will change in the future. In this chapter an overview will be given about possible trends in these factors.

16.1. Electric Vehicles

The demand from plug-in hybrid electric vehicles (PHEV) is an important issue for future distribution losses. It is assumed that the introduction of these vehicles will influence the total electricity demand, the load shape of the day as well as the electricity flow during the day. In the most optimal situation, the vehicles can even be used as electricity storage to balance between the electricity supply and demand.

The predictions of the influence of the PHEV introduction on electricity demand differ a lot between several researchers. In the ProgNET-model (Bakker & Boonekamp, 2004), the demand predictions of the Dutch TNO made in 2004 it is not even mentioned. In a report made by the Dutch government in 2009[7] a prediction is made for four different time-scales as illustrated in table 16.1.

Year	Number of PHEV
2009-2011	10-100
2012-2015	15,000 - 20,000
2015-2020	200,000
>2025	1,000,000

Table 16.1: Estimation of number of PHEV in the Netherlands [7]

As can be seen in table 16.1 above, the introduction of electric vehicle will take a couple of years. In the first couple of years, this introduction will not be visible in the electricity demand, because of the small amount of electric vehicles. It is therefore not needed to implement this trend in the forecasting method.

In the future it has to be investigated how the introduction of electric vehicle will affect the electricity infeed.

16.2. Smart Grid

Another major trend in the electricity market that may influence the distribution losses in the future is the transition to a Smart Grid. Smart Grid, in general terms, is the concept of connecting information technology with the distribution network to make the network more intelligent. An intelligent distribution network aims to increase the reliability, security and efficiency through the use of information and communication technology in combination with storage, demand side management and controllable equipment.

Electricity storage, for example in PHEV’s in the future, together with demand side management and controllable equipment can lead to a more stable electricity flow in the future. The more stable electricity flow during a day will directly lead to a lower percentage of electricity losses, since the peak-load will be smaller.

To gain more stability in the electricity demand profile in the future, it is important to have: (i) enough electricity storage capacity, (ii) controllable equipment, and (iii) the ability to centrally manage the electricity demand. The first requirement can be fulfilled by the electricity storage capacity of the PHEV’s in the future. The second requirement can also be fulfilled by the PHEV’s. In this situation the vehicles can be recharged when the electricity demand is low. At peak load the electricity stored in the vehicles can be used instead of generating electricity. The third requirement can be fulfilled by the implementation of smart meters in the distribution grid. These meters can keep track of the demand during the day and shut down the controllable equipment during peak load.

Besides the decrease in distribution losses due to the stabilization of the electricity demand, the introduction of smart metering will lead to a reduction of distribution losses due to a better detection of illegal use of electricity (Kadurek, 2010).

At this moment it is not recommended to implement the effect of smart metering in the forecasting method. The major effect of smart metering will be visible when the whole grid is ‘smart’, what will take some decennia. The introduction of smart meters will not dramatically influence the amount of electricity losses, however, data will be available earlier what will enhance the distribution losses calculation.

16.3. Decentralized generation

A third trend is the decentralization of electricity production in the Netherlands, mainly due to the growing amount of combined heat and power (CHP) generators. Figure 16.1 shows the amount of electricity produced in the Netherlands the past eleven years central and decentralized.

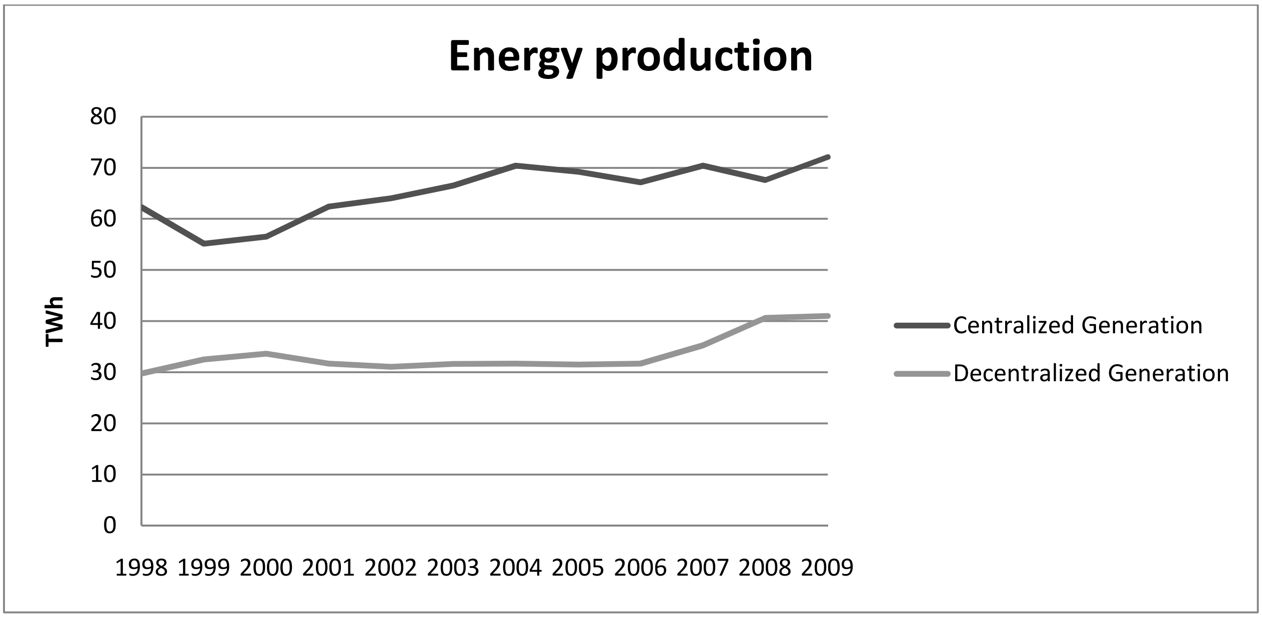


Figure 16.1: Electricity generation in the Netherlands

The effect of decentralized generation on distribution losses depends on the intensity of CHP generators in a region. If the CHP generators are spread equally in a region, the transportation distance will be small compared to the situation in which the electricity is produced at central plants in the Netherlands. This results in smaller amounts of distribution losses. However, if the CHP plants are concentrated in one region, for example in the "Westland" (a region south of The Hague with a large amount of greenhouses), the electricity has to be transformed to a high voltage and transported to another region. This transformation is very inefficient and leads to a higher percentage of distribution losses in the grid. It can be seen in figure 16.1 that the amount of electricity generated decentralized is relatively stable in the recent years. This amount may increase dramatically in the future, what may influence the distribution losses.

Before this trend should be taken into account in the model, it has to be investigated what the effect of decentralized generation is on the distribution losses in Enexis' grid in the future. And how the generators should be situated in the region to optimally distribute the generated electricity.

16.4. Climate changes

A fourth important trend that may influence future electricity demand is the climate change in the Netherlands and the rest of the world. Pardo et al. (2002) and Mirasgedis et al. (2006) show in their paper that the relationship between the outdoor temperature and the electricity demand follow a U-shaped and non-linear relation. However, in recent years this positive relationship between the outdoor temperature above 18°C and electricity demand is becoming more pronounced.

The electricity demand pattern shows a clearly seasonal pattern in all countries of the word. Three different seasonal patterns may be observed: (i) a yearly pattern with a peak in winter, (ii) a yearly pattern with a peak in summer, or (iii) a yearly pattern with a peak in both summer and winter (Hekkenberg, 2009). Figure 16.2 shows the yearly pattern of Norway, The Netherlands, Spain and Cyprus.

The Norwegian pattern shows an obvious peak during winter months. The Dutch pattern shows a peak during winter months too, although less obvious. The Cyprus' pattern has a clear peak during summer months, while the Spanish pattern has both a peak in summer and winter months.

Hekkenberg et al. (2009) show in their paper that the changing outdoor temperature in the Netherlands results in a changing demand pattern during the year. Historically, a higher outdoor temperature results in a decreasing electricity demand, however, in the last couple of years a new trends has been seen, what means that a higher temperature now leads to an increased electricity demand in the summer months. This may result in a yearly pattern in the Netherlands that is similar to the Spanish pattern with both a peak in summer and winter months.

Today, this change in the electricity demand pattern is not visible in Enexis' data, and it will take a lot of time before the electricity pattern will be changed. It is therefore not needed to implement this trend in the forecasting method.

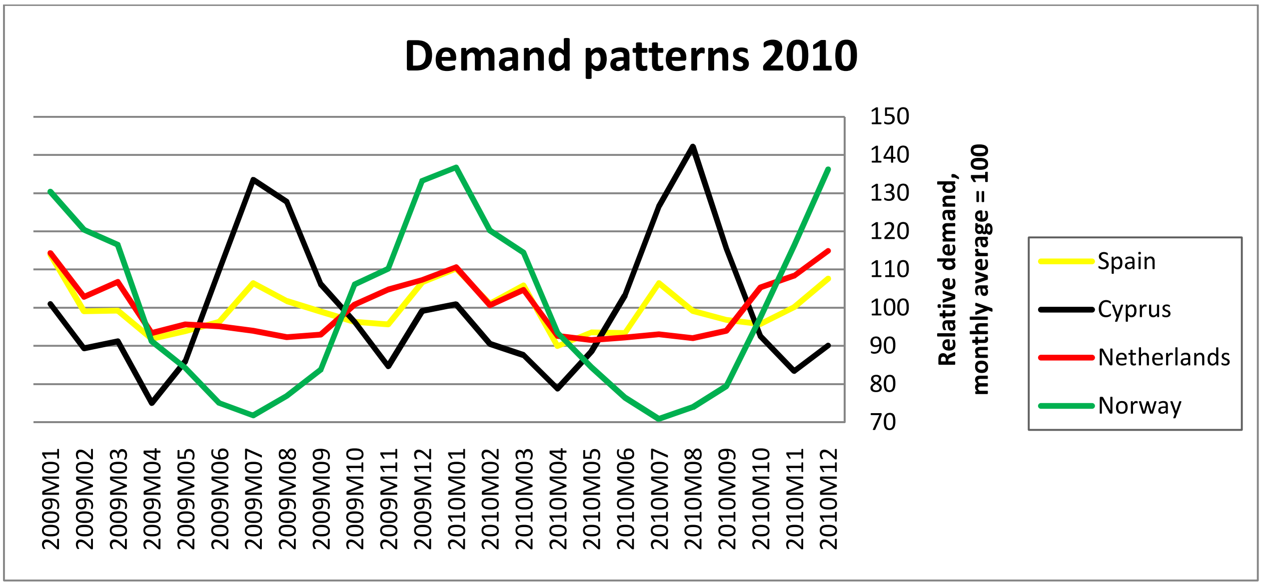


Figure 16.2: Demand Pattern of several EU-Countries, Source: Eurostat [8]

17.Purchasing process

Before a selection will be made between the two forecasting methods, the distribution losses purchasing process will be described to be able to take the point of view of this stakeholder into account.

<Confidential>

17.1. Important criteria for the purchasing process

<Confidential>

When the multiple regression model is used for the forecast, it is recommended to investigate whether the method can be improved to eliminate the major errors in the forecast. These errors will be visible for the supplier and harm the contract negotiation. The time series method is more stable in the forecasting accuracy and may be better for negotiation, but performs worse overall.

18.Selection of the forecasting method

In chapter 4, a number of functional and non-functional requirements have been described for the forecasting method. In this chapter the two forecasting methods will be evaluated on these requirements. Furthermore, the distribution losses purchasing process will be described to be able to make a good selection between the forecasting methods. Finally, one of the methods will be selected based on the requirements.

18.1. Criteria

Looking at the functional requirements, it can be concluded that both of the forecasting methods meet all functional requirements. The regression model performs somewhat better in the forecasting accuracy, as can be seen in table 18.1 below. Furthermore, the multiple regression model is easier to understand and the exogenous variable are easier to understand. Regarding the other functional requirements, both methods perform equal.

Functional requirement	Multiple regression model	Time series model
Solution should solve the business problem	++	+
Show an accuracy level	+	+
Provide a confidence interval	+	+
Include exogenous variables	+	+
Show impact of exogenous variables	++	+

Non-functional requirement		
Ease of use	+	-
Ease of implement	+	-
Data availability	+	+
Adaptability	+	+
Applicability	+	+
Software	-	-
Understandability	+	-
- : Method doesn't meet requirement, + : Method does meet requirement, ++ Method does meet requirement and performs better than the other method		

Table 18.1: Requirements of the solutions

Regarding the non-functional requirements it can be seen that the multiple regression model performs much better than the time series model. The main reason for this is that it is much easier to understand and generate the multiple regression model.

One of the requirements is not met by both of the models, namely the ability to generate the model in Microsoft Office software. For both models statistical software, such as SPSS, is needed for the generation of the model, since the number of independent variables is too large for the regression function in Microsoft Excel. It is therefore recommended to the company to use statistical software in the future to be able to generate accurate infeed forecast.

18.2. Method selection

Based on the purchasing process and the functional and non-functional requirements, it can be concluded that the multiple regression model outperforms the characteristics of the time series model. The regression model is much easier to understand and implement in the company. It is easier to investigate what causes forecasting errors in the multiple regression model and the model is easier to adapt in the future.

Part VI: Conclusions

19.Research findings

At the start of this thesis, Enexis was unsatisfied about the distribution losses forecasting process, the forecasting methods used, and the factors included in the forecasting method. During this thesis, insights were provided in the optimal forecasting method to be used by the company, factors to be included in the method and trends that may influence the losses in the future. In this chapter, the new insight found in this research will be described.

19.1. Forecasting method

In this research it is shown that the optimal forecasting method to be used by the company to forecast the distribution losses for the upcoming years on an hourly basis, is based on the multiple regression method.

To improve the forecasting accuracy and avoid difficult daily patterns in the forecast, it is decided to forecast the electricity demand for each hour of a day in a separate model. Furthermore, since the factors that influence electricity demand are related to particular regions (e.g. carnival), it is decided to forecast the electricity infeed for each electricity grid.

During the thesis it has been found that the multiple regression method outperforms the time series method, based on the forecasting accuracy. This finding is similar to the findings in the M-competitions, where the main conclusion was that statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones (Makridakis et al. 1982).

With MAPE values of 4.97% in Maastricht and lower in other regions it is shown that the multiple regression is suitable for forecast with a horizon of 1 year on an hourly basis.

Although a one-to-one comparison with the current forecasting method cannot be made, it can be concluded that the major errors in the current forecast are caused by a poor time shift. These errors can easily be eliminated by including the independent variables described in this thesis in the regression model.

19.2. Independent variables

As described in the section above, it is decided to separate the models according to the electricity grid and the hour of the day. In this research it has been found that this decision improves the forecast accuracy for electricity infeed. It is shown that the independent variables carnival, school holidays and 'bouwvak' effects on the electricity infeed in different regions.

It is furthermore shown that holidays have an effect in the electricity infeed on the day before the holiday and the day after the holiday. However, this effect is not significantly present during the whole day. This shows that it is an improvement to forecast every hour of a day separately.

Besides the independent variables described above, it is shown in the research that the outdoor temperature significantly influences the electricity infeed for all grids. It is shown that this relationship has a U-shape effect on the electricity infeed. Furthermore, it is shown that this variable can be included in the model using two variables with a linear effect, namely cooling degree days (CDD) and heating degree days (HDD). As described in chapter 9, actual data is used to generate the forecast for a realistic comparison of the models. In the future, the company should use the average daily temperature in the forecast provided by the KNMI[9].

Finally, it is shown that the total electricity infeed is significantly lower during holidays in comparison with normal days. A holiday on a Sunday does only have an effect on the electricity demand during the day, while the effect of a holiday on a weekday or Saturday is significant for all hours of the day.

The effect of economic growth and the growth in population could not be found. This may be caused by the fact that no information could be found on economic growth and population statistics of the regions in which Enexis operates. It is recommended to further investigate what the effect is of future economic growth on electricity infeed.

19.3. Future trends

The four trends investigated in this research, smart grid, electric vehicles, climate change and decentralized generation will not have a significant effect on the distribution losses in the near future. However, since the introduction of electric vehicles can dramatically change the electricity demand pattern in a short time span, it should be investigated in the near future what the effect of electric vehicles will be during the introduction.

The other trends, for example climate change, will take a long time before the effect can be seen in the electricity infeed. It is assumed that these effects cannot be distinguished from other effects, such as economic growth, population and the efficiency of electricity usage.

20. Recommendations to the company

During the thesis, a couple of problems were found what may cause the current problem of the unreliable distribution losses forecast. The following recommendations are made which may improve the forecasting process in the future.

<Confidential>

21.Contribution to literature

In chapter 5, a couple of gaps in literature were depicted which were not considered in literature before. In this research some additions were made to the literature in mid-term forecasting and distribution losses. The additions made will be described in this section.

The first addition to the literature is the use of the medium term forecasting model with small time intervals. As described in the literature review, no research has been done in midterm electricity demand forecasting on an hourly basis. This research showed that it is possible to generate a forecasting model with an horizon of 1 year. It furthermore showed that generating a model for each hour of a day improves the forecasting accuracy. The overall MAPE of the five electricity grids was below 5% for a one year forecast. The larger the electricity grid, the lower the average percentage error.

The second addition to the literature is the identification of significant independent variables for mid-term electricity demand forecasting in the Netherlands. The research showed that meteorological data, historical data and calendar data should be used for mid-term forecast. No significant effect has been found between economic or demographic variables on electricity infeed.

A third addition to the literature is the selection of an appropriate forecasting methods for a distribution network operator. In literature, most of the forecasting methods are very sophisticated (e.g. neural networks, artificial intelligence models) and used by experts in forecasting. These sophisticated methods are in most of the cases not appropriate in practical situations. In this research, an excellent forecasting method is selected based on several criteria, such as accuracy, ease of use, adaptability, et cetera.

22.Research opportunities

Although more insight is given in distribution losses forecast for Enexis, a lot of research could still be done to further improve the forecast and purchasing process of the company. Suggestions for further research in distribution losses forecasting are described in this chapter. The suggestions are categorized according to their importance in my opinion.

- In the data used for this research, no significant correlation was found between economic or demographic data and electricity infeed, whereas this relationship is found in many other researches on electricity demand. It should be investigated what causes this error and how these variables should be included in the forecasting method.
- In the forecasting method used in this research, historical infeed data of the past 3 years is used. This moving average does not take into account whether the historical data was a special day, such as a holiday. The results is, that it is possible that the moving average is not representative for a particular forecasting day. It should be investigated how this moving average could be improved to increase the forecasting accuracy.

- In the forecasting method generated in this research it is decided to include only one variable for all holidays, assuming that the effect of a holiday is similar during the year. This decision was made because of the limited amount of historical data. It should be investigated whether the holidays have a similar effect during the year and whether including separate variables for each holiday would improve the forecasting accuracy. For example, the effect of New Year's Day will be different than the effect of Christmas Day, since the electricity demand on New Year's Day will be higher during the night. The significance of the relationship should be investigated when more data is available.
- The historical data used in this research to train the forecasting methods is 3 years of data (Jan 2007 – December 2009). When more data is available, especially data of special days (e.g. holiday on Saturday) it should be investigated what the optimal amount of data is to forecast the demand.
- In addition to the amount of historical data to be used in the forecast, it should be investigated whether the forecasting accuracy increases when using more recent data. If the accuracy improves, it should be investigated whether the forecasts should be made more frequent, for example every month.
- Finally, it should be investigated whether the model is useful for forecasts with a longer horizon (e.g. 2-4 years). For these forecasts, is should be noticed that demographic and economic variables becomes more important, and it is expected that these variables should be implemented in the model to gain accurate distribution losses forecasts.

23.References

[1] http://www.enexis.nl/site/Images/halfjaarbericht-Enexis-NV-2010%20UK.pdf

[2] http://www.leerwiki.nl/Wie_is_mijn_netbeheerder,_in_een_oogopslag

[3] http://www.tennet.org/bedrijfsvoering/Systeemgegevens_afhandeling/Reconciliatieprijs/index.aspx

[4] http://robjhyndman.com/researchtips/arimax/

[5] http://eca.knmi.nl

[6] http://statline.cbs.nl/StatWeb/dome/?LA=EN

[7] http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/kamerstukken/2009/07/03/plan-van-aanpak-elektrisch-rijden/plan-van-aanpak-elektrisch-rijden.pdf

[8] http://epp.eurostat.ec.europa.eu/portal/page/portal/energy/data/database

[9] http://www.knmi.nl/kd/daggegevens/download.html

[10] http://www.nma.nl/images/103549_7-10_Enexis22-149551.pdf

[11] http://www.nma.nl/images/-_ENEXIS_-_103547_5_NXIS_NE-TAR(i)-10-09_Tarievenvoorstel22-146512.xls

[12] http://www.apxendex.com/?id=281

[13] Baker, K.J. & Rylatt R.M., “Improving the prediction of UK domestic energy-demand using annual consumption-data”, *Applied Energy 2008*, Vol. 85, p. 475-482.

[14] Bakker, S.J.A. & Boonekamp, P.G.M., “Het ProgNET-model”, *Tennet 2004.*

[15] Bharadwaj A.G. & Meeta K.M., ”Demand Forecasting for Electricity: the Indian experience”, In: Sarkar S.K. & Deb K., (Eds.), *Regulation in Infrastructure Services: progress and the way forward.* TERI, 2001, p. 175-192.

[16] Bianco, V., Manca, O. & Nardini, S., “Electric consumption forecasting in Italy using linear regression models”, *Energy 2009*, Vol. 34, p. 1413-1421.

[17] Box, G.E.P., Jenkins, G.M. & Reinsel, G.C., *Time series Analysis: Forecasting and Control*, third ed., New Jersey, Englewood Cliffs, Prentice Hall, 1994.

[18] Bruhns, A., Deurveilher, G. & Roy, J.S., “A non-linear regression model for mid-term load forecasting and improvements in seasonality”, 2005, 15th Power System Computation Conference.

[19] Chatfield, C., *“Time series forecasting”*, Chapman & Hall/CRC, Florida, 2000

[20] Deng, S.J., & Oren, S.S., “Electricity derivatives and risk management”, *Energy 2006,* vol. 31, p. 940-953.

[21] Egelioglu, F., Mohamad, A.A. & Guven, H., “Economic variables and electricity consumption in Northern Cyprus”, *Energy 2001*, Vol. 26, p. 355-362.

[22] Feinberg E.A. & Genethliou D., “Load Forecasting”, In: Chow J.W., Wu, F.F. & Momoh, J., *Applied Mathematics for restructured electric power systems,* Springer 2005, Chapter 12, p. 269-285.

[23] Flaten, D.L., “Distribution system losses calculated by percent loading”, *Transactions on power systems 1988,* Vol.3, No.3.

[24] Grainger, J.J. & Kendrew, T.J., “Evaluation of technical losses on electric distribution systems”, *Cired 1989.* International Conference on Electricity Distribution.

[25] Hahn, H., Meyer-Nieberg, S. & Pickl, S., “Electric load forecasting methods: Tools for decision making”, *European Journal of Operational Research 2009,* vol. 199, p. 902-907.

[26] Hair, J.F., Black, B., Baban, B., Anderson, R.E. & Tatham, R.L., *Multivariate Data Analysis (6th ed.)*, Prentice Hall, 2005.

[27] Heijnen P. & Bouwmans I., “Supporting the electricity network operator in the allocation process”, *Infrastructure Systems and Services: Building Networks for a Brighter Future 2008,* p. 1 -5.

[28] Hekkenberg, M., Benders, R.M.J., Moll, H.C. & Schoot Uiterkamp, A.J.M., “Indications for a changing electricity pattern: The temperature dependence of electricity demand in the Netherlands”, Energy Policy 2009, Vol. 37, p. 1542-1551.

[29] Hippert, H., Pedreira, C. & Souza, R., “Neural networks for short-term load forecasting: A review and evaluation”, *IEEE Transactions on Power Systems 2001*, vol. 16, p. 44-55

[30] Huang S.J., & Shih, K.R., “Short-term Load Forecasting Via ARMA Model Identification Including Non-Gaussian Process Considerations”, *Transactions on power systems 2003*, Vol. 18, No.2, p. 673 – 679.

[31] Huang C.M., Huang C.J., & Wang, M.L., “A particle swarm optimization to identify the ARMAX model for short-term load forecasting”, *IEEE Transactions on Power Systems 2005,* vol. 20, p. 1126-1133.

[32] Hyndman, R.J. & Fan, S., “Forecasting long-term peak half-hourly electricity demand for South Australia”, 2009, Monash University Australia.

[33] Kadurek P., Blom J., Cobben J.F.G. & Kling W.L., “Theft detection and Smart Metering Practices and Expectations in the Netherlands”, 2010, University of Technology Eindhoven.

[34] Kyriakides, E. & Polycarpou, M., “Short term electric load forecasting: A tutorial.”, In: Chen, K. & Wang, L. (Eds.), *Trends in Neural Computation, Studies in Computational Intelligence vol. 34,* Springer, p. 319-418.

[35] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & Winkler, R., “The accuracy of extrapolative (time series) methods: results of a forecasting competition”, *Journal of Forecasting 1982,* Vol. 1, p. 111-153.

[36] Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K. & Simmons, L.F., “The M2-Competition: A real-time judgmentally based forecasting study”, *International Journal of Forecasting 1993,* Vol. 9, p. 5-22.

[37] Makridakis, S., Wheelwright, S.C. & Hyndman, R.J., *Forecasting Methods and Applications*, New York, Wiley, 1998.

[38] Makridakis, S. & Hibon, M., “The M3-Competition: results, conclusions and implications”, *International Journal of Forecasting 2000,* Vol. 16, p. 451-476.

[39] Mirasgedis, S., Sarafidis, Y., Georgopoulou, E., Lalas, D.P., Moschovits, M., Karagiannis, F. & Papakonstantinou, D., “Models for mid-term electricity demand forecasting incorporating weather influences”, *Energy 2006*, Vol. 31, p. 208-227.

[40] Mohamed, Z. & Bodger, P., “Forecasting electricity consumption in New Zealand using economic and demographic variables”, *Energy 2005*, vol. 30, p. 1833-1843.

[41] Nadira, R., Benchluch, S. & Dortolina C.A., “A novel approach to computing distribution losses”, 2004, Transmission and Distribution Conference and Exposition.

[42] Oliveira, C.C.B., Kagan, N., Meffe, A., Jonathan, S., Caparroz, S. & Cavaretti, J.L., “A new method for the computation of technical losses in electrical power distribution systems”, *Electricity distribution 2001,* vol. 5, p. 1-5.

[43] Pardo, A., Menue, V. & Valor, E., “Temperature and seasonality influences on Spanish electricity load”, *Energy Economics 2002,* Vol. 24, p. 55-70.

[44] Smith, T.B., “Electricity theft: a comparative analysis”, *Energy Policy 2004,* Vol. 32, p. 2067-2076.

[45] Soares L.J., & Medeiros M.C., “Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data”, *International journal of forecasting 2008,* vol. 24, p. 630-644.

[46] Suriyamongkol, D., “Non-technical losses in electrical power systems”, 2002, Ohio University.

[47] Taylor, J.W., “Short-term electricity demand forecasting using double seasonal exponential smoothing”, *Journal of the operational Research Society 2003*, Vol. 54, p. 799-805.

[48] Taylor J.W., Menezes de, L.M. & McSharry, P.E., “A comparison of univariate methods for forecasting electricity demand up to a day ahead”, *International Journal of Forecasting 2006,* Vol. 22, p. 1-16.

[49] Taylor J.W. & McSharry, P.E., “Short-term load forecasting methods: An evaluation based on European Data”, *IEEE Transactions on Power Systems 2008,* vol. 22, p. 2213-2219.

[50] Tzafestas, S. & Tzafestas, E., “Computational intelligence techniques for short-term electric load forecasting”, *Journal of Intelligent & Robotic Systems 2001,* vol. 31, p. 129-147.

[51] Yusoff, M., Busrah, A., Mohamad, M. & Au, M.T., “A simplified approach in estimating technical losses in TNB distribution network based on load profile and feeder characteristics”, 2008, Power and Energy Conferences.

Part VII: Appendices

Appendix A: Tariff structure DNO

Every electricity consumer in the Netherlands is able to choose their electricity supplier, but they cannot choose between grid operators, since the DNO is related to the region of delivery (i.e. there is only one DNO per region, and the DNO's are in a monopolistic situation).

To ensure a reliable and affordable high quality network, the transmission network operator and the regional grid operators are regulated by the NMa Energiekamer. The NMa decide what tariff the network operator is allowed to ask to the electricity consumers. The sum of tariffs allowed to ask is stated in the Electriciteitswet 1998 and described below.

The formula for the total income is given by [10]:

$$TI_t = \left(1 + \frac{cpi \pm x + q}{100}\right) TI_{t-1}$$, where

TI_t = Total income allowed from customers in year t;

cpi = The inflation factor (Consumer Price Index) published by the CBS (Statistics Netherlands);

x = the discount to the benefit of efficient operations;

q= the quality-factor.

Enexis' total allowed income from electricity transportation for 2010 was €796,581,465 the CPI was equal to 1.5, the x-factor is -6.2 and the q-factor 0.02. This results that the total income from electricity transportation for Enexis for 2011 is €858,077,554. [11]

The total income is the sum (price * quantity) of tariffs the DNO ask to consumers (households, small companies, etc.). The costs for a connection depends on the kind of connection. The amount of electricity consumed is not taken into account in the costs. Households pay around 200 euro per year per 25 ampere connection. The costs for companies depends on the connection used.

Appendix B: Graphical illustration of allocation & reconciliation process

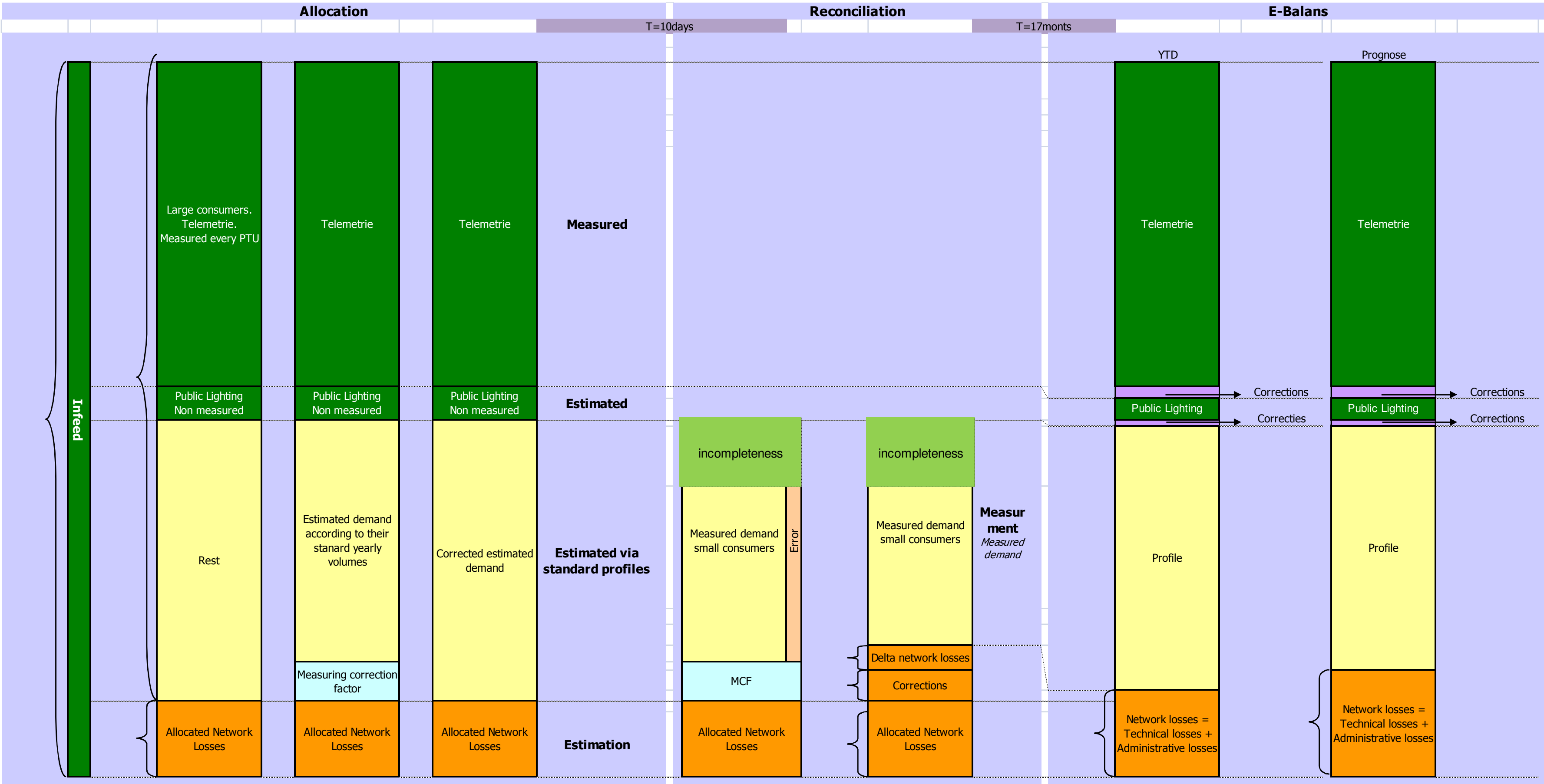


Figure B.1: Allocation and reconciliation process

Appendix C: Schematic representation of Enexis' distribution grid

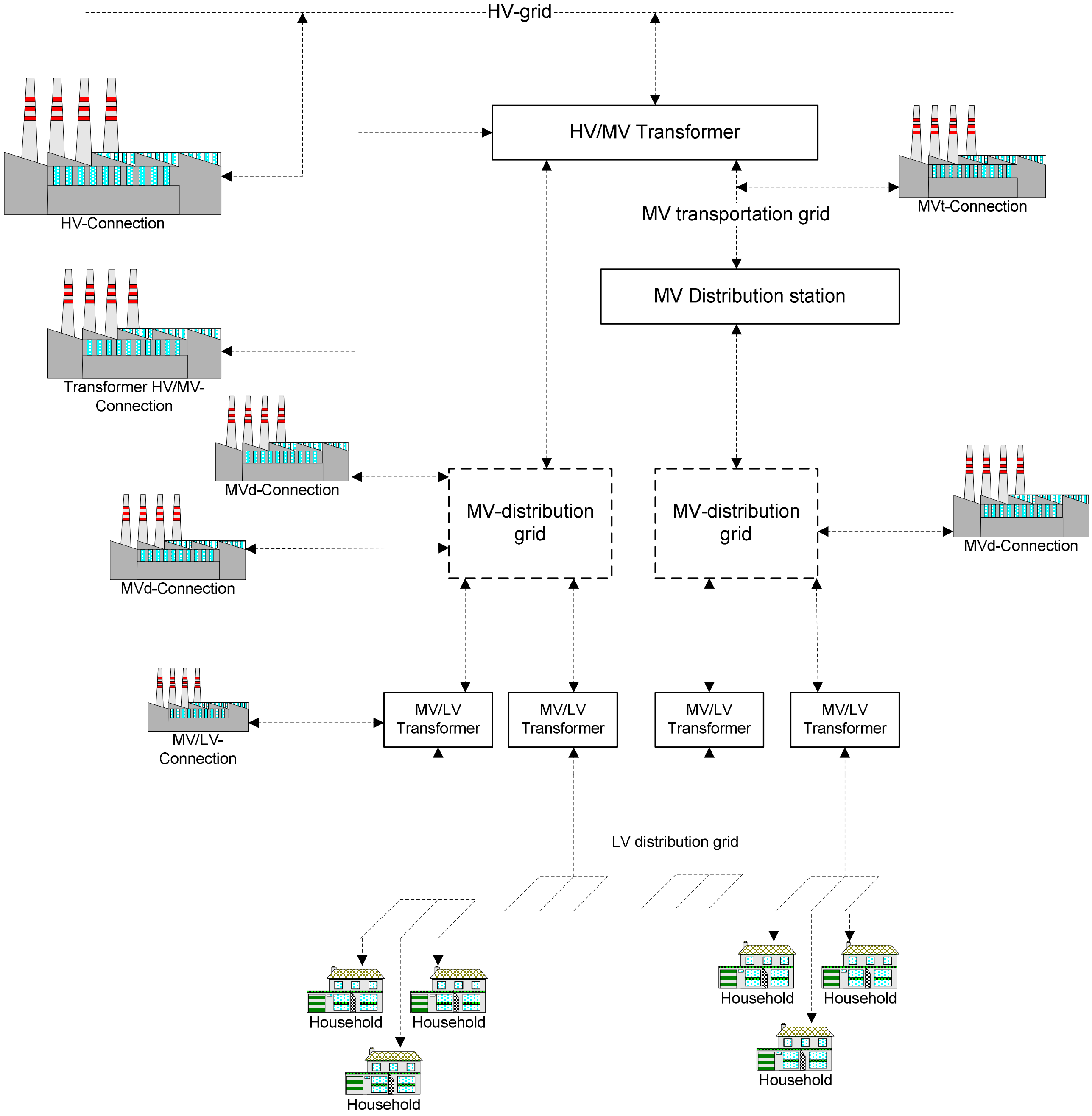


Figure C.1: Schematic representation of Enexis' grid

Appendix D: Numerical values of distribution losses 2011 (Confidential)

Appendix E: Cascade Model

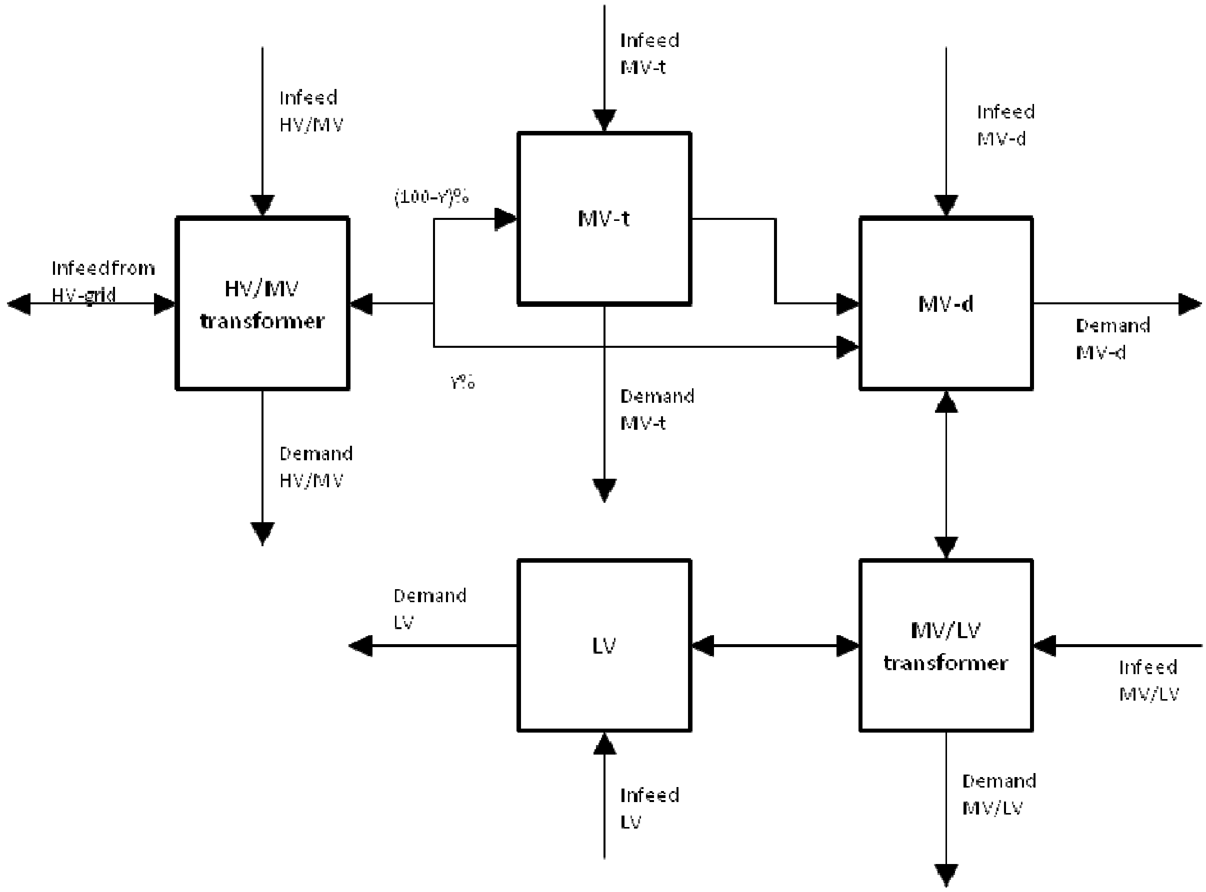


Figure E.1: Cascade Model

Appendix F: Sequence charts of electricity infeed

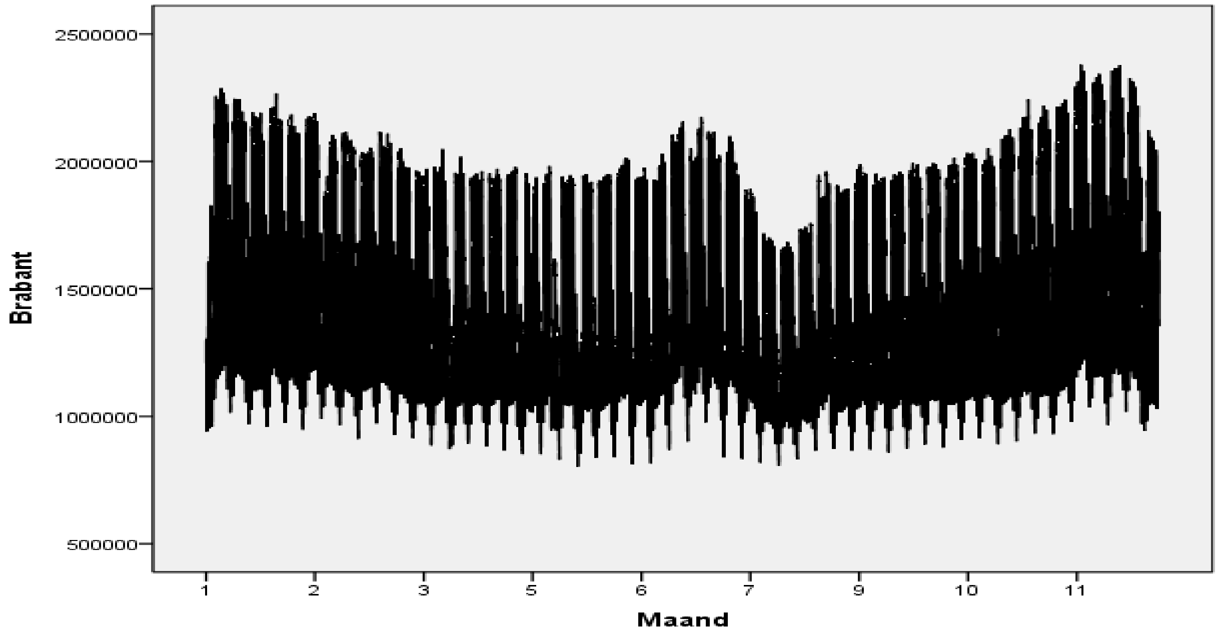


Figure F.1: Infeed pattern 2010 Brabant

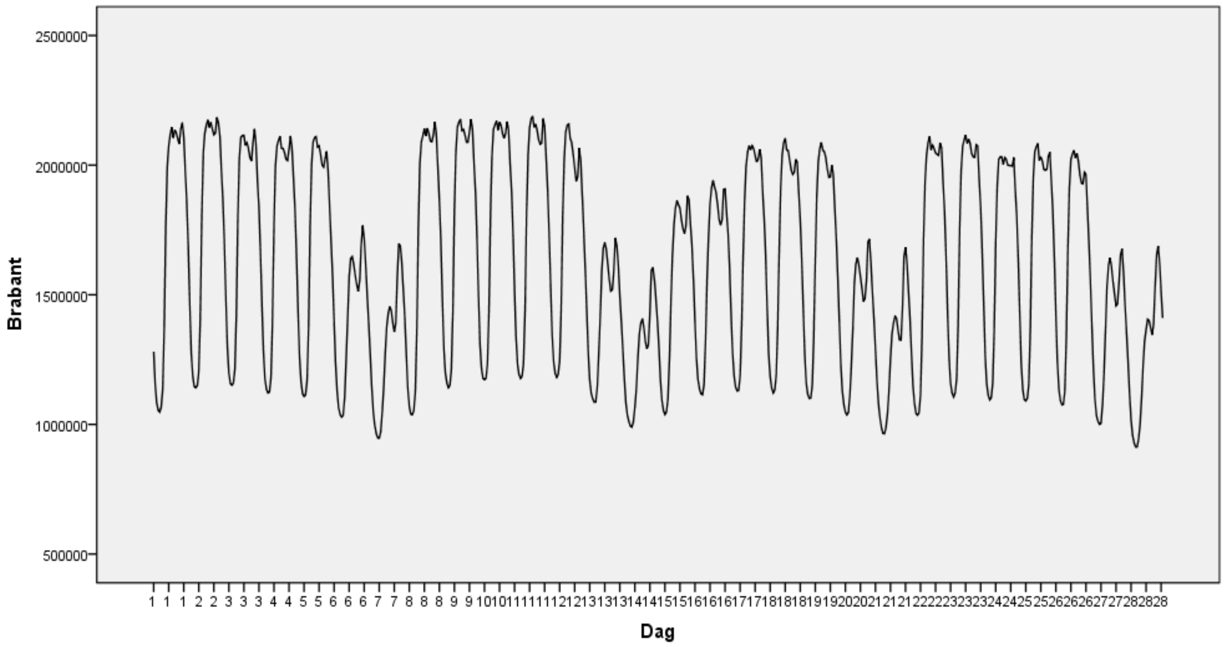


Figure F.2: Infeed pattern February 2010 Brabant

Appendix G: Holidays

Holidays

- New Year’s Day
- Easter Sunday
- Easter Monday
- Queen’s day
- Ascension day
- Whit Sunday
- Whit Monday
- Christmas Day
- Boxing Day

Working days near holidays

- Day after New Year’s Day (if Monday - Thursday)
- Day after Easter Monday
- Day before Queen’s day (if Tuesday - Friday)
- Day after Queen’s day (if Monday - Thursday)
- Day before Ascension day
- Day after Whit Monday
- Christmas Eve (if Tuesday – Friday)
- Day after Boxing day (if Monday – Thursday)
- New Year’s Eve (if Tuesday – Friday)

‘bridge days’

- Day after New Year’s Day (if Friday)
- Day before Queen’s day (if Monday)
- Day after Queen’s day (if Friday)
- Day after Ascension day
- Christmas Eve (if Monday)
- Day after Boxing day (if Friday)
- New Year’s Eve (if Monday)

Carnival

- Carnival Monday
- Carnival Tuesday

School holidays

- School holidays during carnival

- School holidays in May
- School holidays in summer
- School holidays in autumn
- School holidays during Christmas

Appendix H: Normality of dependent variable

The most fundamental assumption in multivariate analysis is normality. Normality of the dependent variable can be tested by analyzing the normal probability plot, the histogram or by calculating the skewness and kurtosis values.

In this appendix the electricity infeed for Brabant will be tested whether these dependent variables meet this assumption. It is assumed that the infeed data of other grids are similar to this dependent variable. For readability purposes only two models (hours) were picked to show non-normality of the dependent variable.

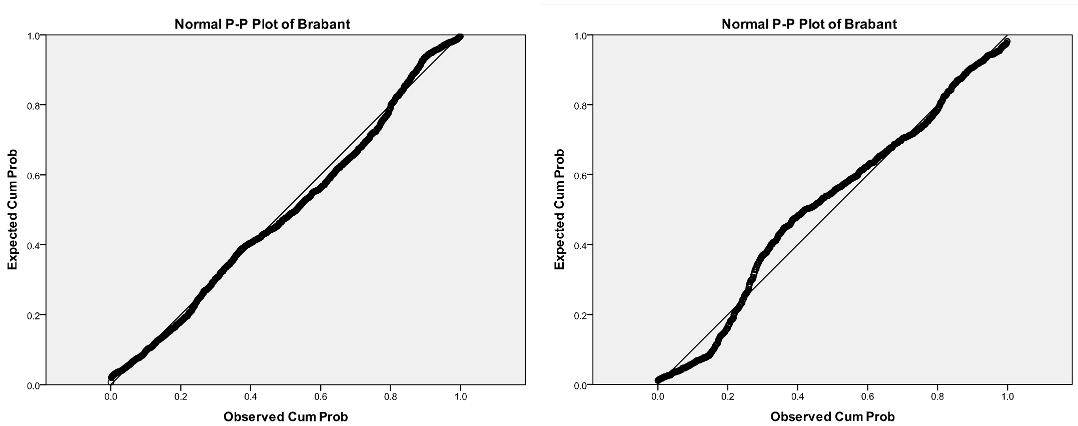


Figure H.1: Normal probability plots Infeed Brabant Hour3 and Hour18

Figure H.1 above show the normal probability plots of the infeed of Brabant for the Hour3 and Hour18. It can be seen that the values of Hour3 are more normally distributed than Hour18. This can be explained easily, since the electrify demand is more stable during night than during day times. Electricity demand during day time is highly dependent on the time of the year and whether it is a weekend or working day, as can be seen in the histograms in figure H.2 below.

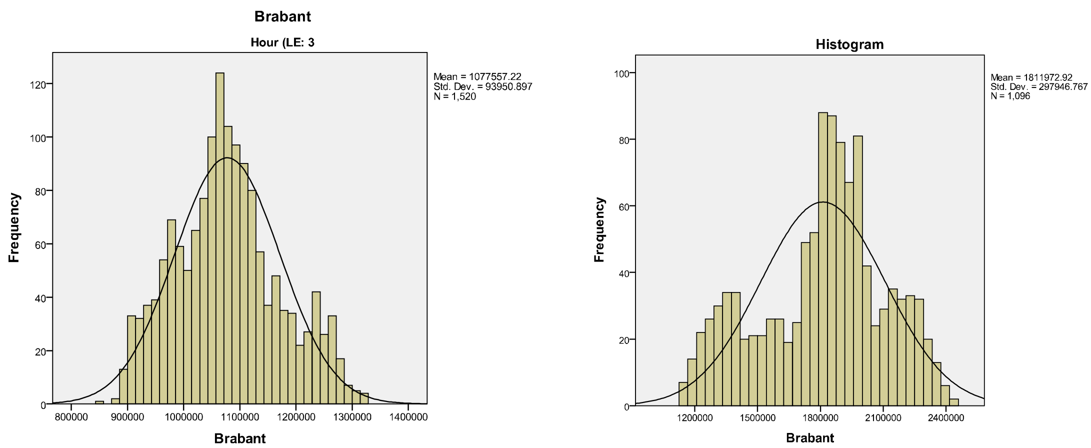


Figure H.2: Histogram infeed Brabant Hour3 and Hour18

In sum, none of the models were normally distributed. For small sample sizes, this non-normality would be a serious problem. However, since the dataset is very large (>200 observations), the impact of non-normality is negligible (Hair et al., 2005).

To reduce the impact of heteroscedasticity resulting from the non-normally distributed data, the large amount of data and its high temporal frequency (Pardo, 2002) (Makridakis, 2006), the natural logarithm of the dependent variable will be taken.

The normal probability plots of the new variables are shown in figure H.3 below.

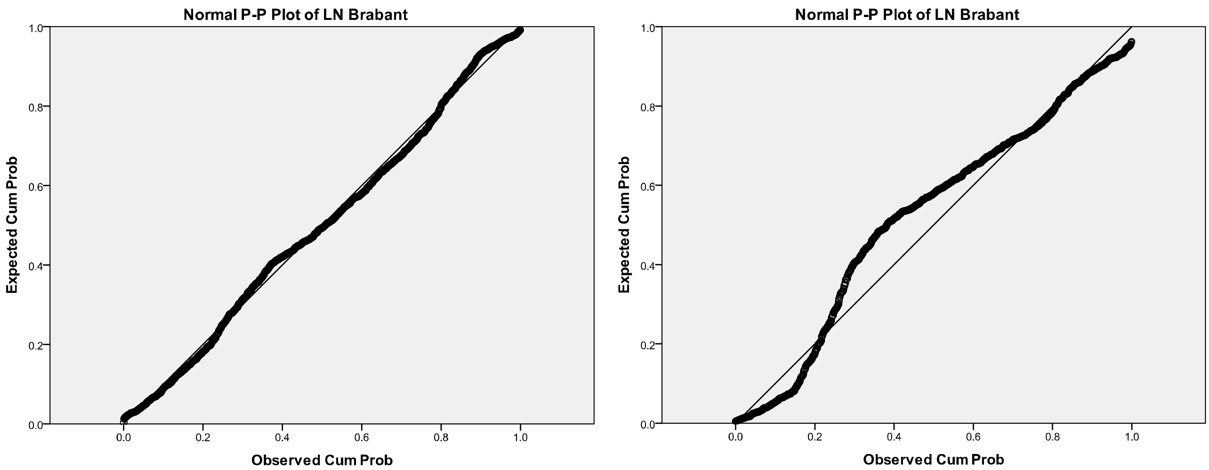


Figure H.3: Normal Probability Plots of Hour3 and Hour18 with natural logarithm

Appendix I: Assumptions of Multiple Regression

Normality of dependent variable

In chapter 11 it has been shown that the dependent variable, electricity infeed, was not normally distributed. To reduce the impact of heteroscedasticity resulting from non-normally distributed independent variables, it has been decided to transform the dependent variable using the natural logarithm. After the dependent variable has been transformed, it was still not normally distributed, however, since the amount of observations was very large (>200), this assumption could be violated (Hair et al, 2005).

No multi-collinearity

Multi-collinearity means that there exists a strong relationship between two or more independent variables. This is unfavorable, since it is difficult to get good results. Multi-collinearity can be tested by analyzing the VIF-factors while running the regression. These factors should be lower than 10. In the tables below, the results are shown for the two regression models (Brabant, hour 3), (Brabant, hour 18). It can be seen that the maximum VIF-factor is equal to 1.510, what means that this assumption is met, and that there's no significant correlation between two or more independent variables.

The constant factor in the second model is not significant on a $\alpha = 0.05$ level. This factor could not be eliminated by SPSS. Since the significance level is not dramatic, the variable is not deleted.

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	.879	.201		4.382	.000		
	HS2007	.139	.002	.707	78.858	.000	.974	1.027
	HolidaySunday	-.023	.011	-.019	-2.097	.036	.980	1.020
	HolidayMonFri	-.054	.006	-.078	-8.537	.000	.930	1.076
	Workingdayafterholiday	-.068	.009	-.071	-7.931	.000	.989	1.011
	Bridgedayafterholiday	-.099	.012	-.073	-8.205	.000	.987	1.014
	SchoolholidaySouth	.009	.003	.035	3.313	.001	.703	1.422
	Carnival	-.037	.010	-.032	-3.583	.000	.971	1.030
	bouwvaksouth	-.039	.005	-.086	-7.872	.000	.662	1.510
	CDD	.009	.001	.092	9.055	.000	.759	1.317
	HDD	.001	.000	.090	8.982	.000	.784	1.276
	LNMOVABrabant	.933	.014	.593	64.544	.000	.928	1.077

a. Dependent Variable: LNBrabant Hour 3

Table I.1 Descriptive Statistics Hour 3

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.186	.100		1.849	.065		
	HS2007	.076	.002	.204	34.052	.000	.974	1.027
	Workingdaybeforeholiday	-.029	.012	-.014	-2.363	.018	.959	1.043
	HolidaySunday	-.047	.014	-.020	-3.364	.001	.972	1.028
	HolidayMonFri	-.220	.008	-.170	-27.891	.000	.941	1.062
	HolidaySaturday	-.077	.034	-.013	-2.277	.023	.998	1.002
	Workingdayafterholiday	.028	.011	.015	2.547	.011	.989	1.011
	Bridgedaybeforeholiday	-.068	.024	-.017	-2.803	.005	.979	1.022
	Bridgedayafterholiday	-.083	.015	-.032	-5.412	.000	.986	1.014
	SchoolholidaySouth	.014	.003	.030	4.079	.000	.663	1.509
	Carnival	-.078	.013	-.036	-5.946	.000	.969	1.032
	bouwvaksouth	-.067	.006	-.077	-10.512	.000	.655	1.526
	CDD	.009	.001	.049	7.303	.000	.761	1.315
	HDD	.001	.000	.042	5.895	.000	.683	1.465
	LNMOVABrabant	.985	.007	.926	140.389	.000	.802	1.247

a. Dependent Variable: LNBrabant Hour 18

Table I.2 Descriptive Statistics Hour 18

Homoscedasticity of the errors

To check for the assumption of homoscedasticity of the errors, the scatterplot of the studentized and standardized residuals and standardized predicted variable for both models (Hour3, hour 18) is analyzed. As shown in the figures below, the scatterplot give no concern for heteroscedasticity and shows that this assumption is met.

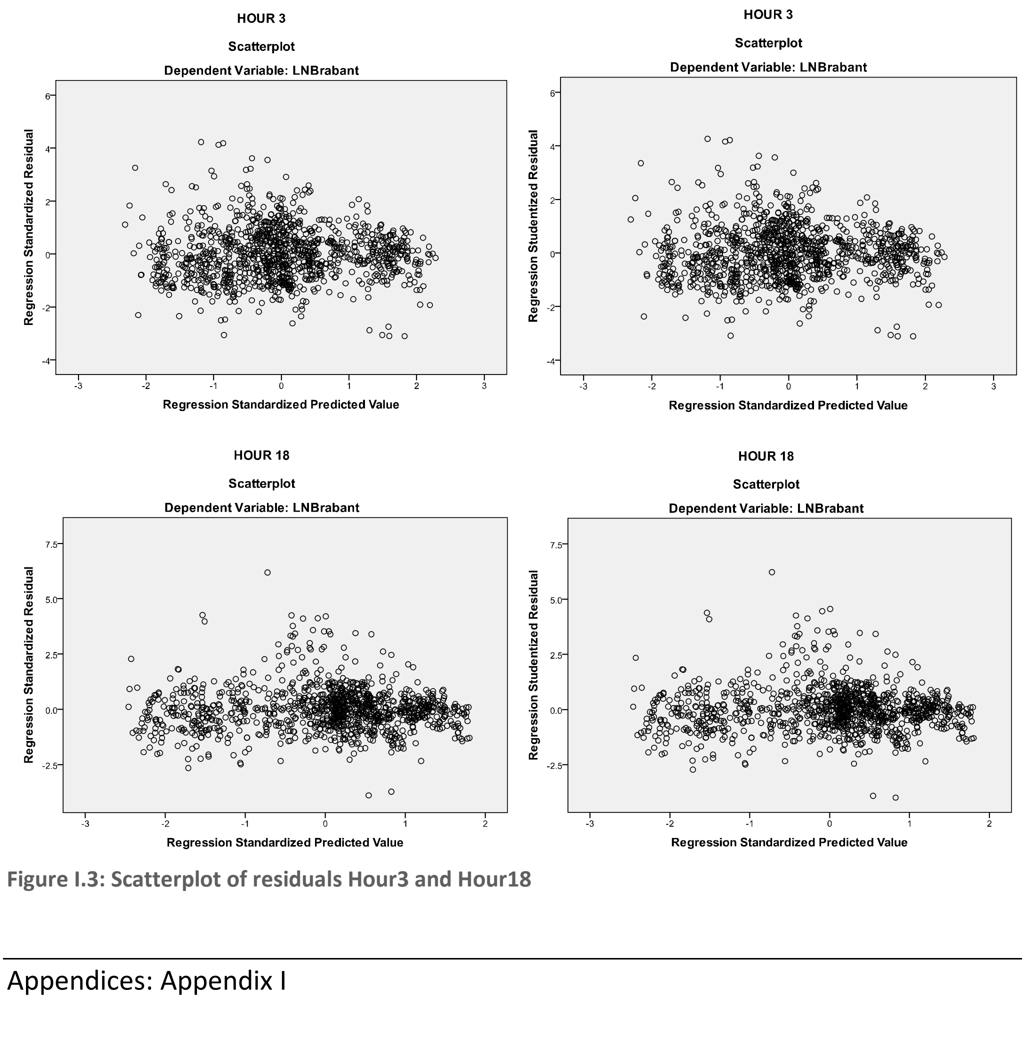
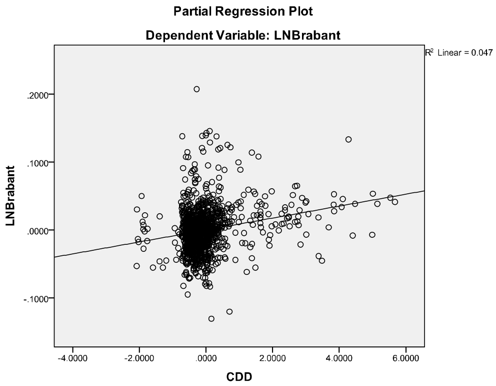
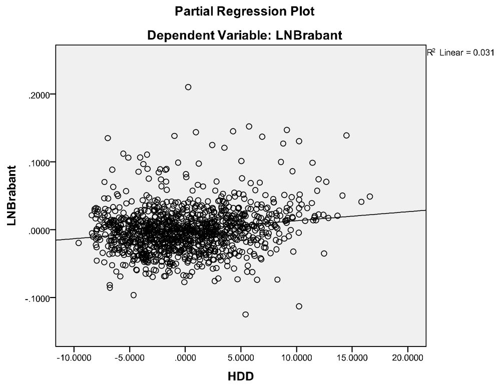


Figure I.3: Scatterplot of residuals Hour3 and Hour18

Linearity of the relationship between the dependent and independent variables

In the figures below, the partial regression plots for the metric independent variables are shown to investigate whether the relationship between the dependent (Brabant, hour 18) and independent variables is linear. It can be seen from these figures that the relationships are linear and that there is no cause for concern for this assumption. The linearity of the variable CDD is somewhat difficult to see, since the majority of the data is concentrated at 0, caused by the fact that most of the days in the Netherlands, the temperature is below 17°C.





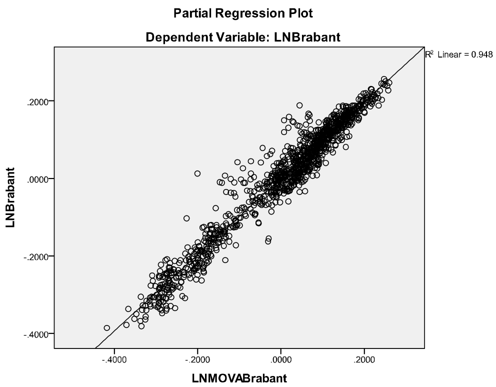


Figure I.4: Partial scatterplot of independent variables

Errors are normally distributed

To check for this assumption, the normal probability plots of the residuals have to be analyzed. These plots are shown in the figures below. These figures show that the residuals are normally distributed. This assumption is met.

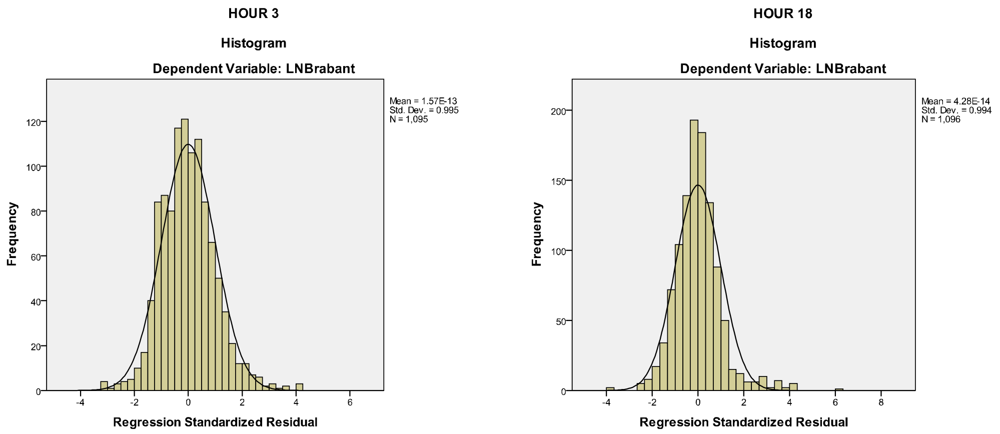


Figure I.5: Histogram of residuals Hour3 and Hour18

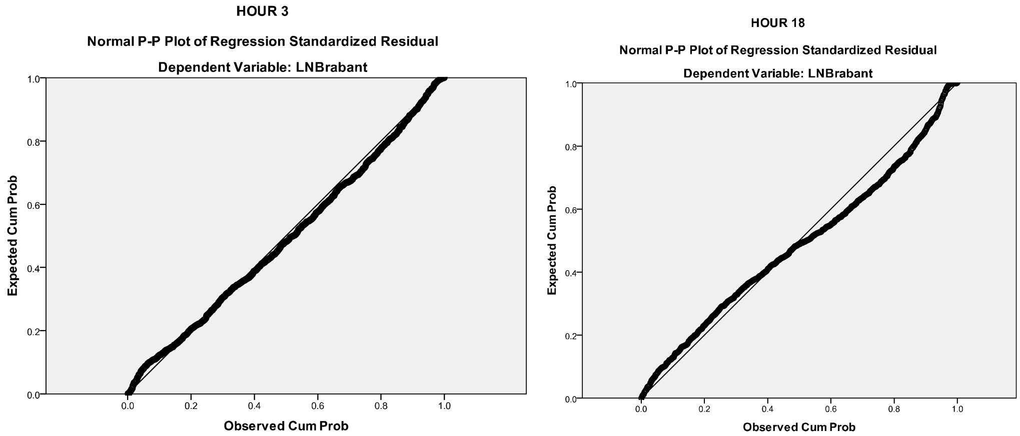


Figure I.6: Normal probability plots of residuals

Independence of errors

This assumption means that it has to be checked whether two errors are correlated. Having information about the value of a residual should not give any information about the value of another residual. This assumption can be checked with the Durbin-Watson test, which checks for serial correlations between errors. The test can vary between 0 and 4. A value of 2 means that the residuals are uncorrelated. As a rule of thumbs, values below 1 and greater than 3 are a cause for concern.
This assumption is usually violated when observations are nested within groups or in time series. In our data, the observations are nested within time series. It is thus expected that the residuals are correlated, as described in hypothesis 1.

The values for the Durbin-Watson test of the two models are shown in the table below. It can be seen that the Durbin-Watson value for hour3 is equal to 0.647 and for hour18 is equal to 0.917, what means that the errors are correlated.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.957[a]	.915	.914	.0267760	.647

a. Predictors: (Constant), LNMOVABrabant, Workingdayafterholiday, Bridgedayafterholiday, Carnival, HS2007, bouwvaksouth, HolidaySunday, HolidayMonFri, HDD, CDD, SchoolholidaySouth

b. Dependent Variable: LNBrabant Hour 3

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.981[a]	.962	.962	.0339630	.917

a. Predictors: (Constant), LNMOVABrabant, Bridgedayafterholiday, HolidaySaturday, Bridgedaybeforeholiday, Workingdaybeforeholiday, HolidayMonFri, Workingdayafterholiday, Carnival, bouwvaksouth, HS2007, HolidaySunday, CDD, HDD, SchoolholidaySouth

b. Dependent Variable: LNBrabant Hour 18

The correlated errors are caused by an effect which is not included in the model. For time series, this can be caused by any season or pattern in the data. It is plausible that this is the case in the data studied. The consequence of violating this assumption is that the confidence intervals, if generated, as estimated too small.

Appendix J: Coefficients and p-values of multiple regression models

BRABANT	HOUR0	HOUR1	HOUR2	HOUR3	HOUR4	HOUR5	HOUR6	HOUR7	HOUR8	HOUR9	HOUR10	HOUR11	HOUR12	HOUR13	HOUR14	HOUR15	HOUR16	HOUR17	HOUR18	HOUR19	HOUR20	HOUR21	HOUR22	HOUR23
Constant	.45	.70	.87	.88	.81	.72	.49	.24	.09	.08	.10	.15	.16	.17	.19	.17	.16	.14	.19	.36	.52	.54	.32	.35
HS 2007	.11	.13	.13	.14	.14	.14	.14	.12	.10	.09	.09	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.09	.10
Working day before holiday		.02				.02	.02												-.03	-.03	-.05	-.05	-.06	-.06
Holiday (Sunday)	-.05			-.02	-.02	-.02	-.02	-.03				-.04	-.04	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05
Holiday (Mon-Fri)	-.08	-.04	-.04	-.05	-.07	-.08	-.11	-.19	-.29	-.33	-.31	-.28	-.27	-.26	-.27	-.28	-.29	-.27	-.22	-.17	-.16	-.14	-.14	-.11
Holiday (Saturday)										-.09	-.09	-.08	-.08	-.08	-.08	-.09	-.09	-.09	-.08		-.06			
Working day after holiday	-.04	-.08	-.08	-.07	-.06	-.06	-.05	-.04							.03	.03	.03	.03	.03	.03	.03	.03	.02	.02
Bridge day before holiday																			-.07	-.08	-.10	-.12	-.12	-.10
Bridge day after holiday	-.06	-.10	-.10	-.10	-.10	-.10	-.11	-.13	-.17	-.18	-.15	-.14	-.14	-.13	-.13	-.13	-.13	-.11	-.08	-.06	-.06	-.06	-.06	-.05
School holiday South	.01	.01	.01	.01	.01	.01	.01	.02	.01	.02	.02	.02	.02	.02	.02	.02	.02	.02	.01	.01	.01	.01	.01	.01
Carnival	-.04	-.03	-.03	-.04	-.04	-.04	-.05	-.07	-.09	-.10	-.09	-.08	-.07	-.07	-.07	-.08	-.09	-.09	-.08	-.07	-.07	-.06	-.06	-.05
bouwvak south	-.05	-.04	-.04	-.04	-.04	-.04	-.05	-.06	-.08	-.09	-.08	-.08	-.08	-.08	-.08	-.08	-.08	-.08	-.07	-.06	-.06	-.06	-.06	-.05
CDD	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
HDD	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Length of the day					.03	.03	.03	.04	.05	.05	.05	.04	.04	.03									.02	.02
LNMOVA Brabant	.97	.95	.93	.93	.94	.94	.96	.98	.99	.99	.99	.99	.99	.98	.98	.99	.99	.99	.98	.97	.96	.96	.97	.97

Table J.1: Coefficients for Brabant

BRABANT	Hour0	Hour1	Hour2	Hour3	Hour4	Hour5	Hour6	Hour7	Hour8	Hour9	Hour10	Hour11	Hour12	Hour13	Hour14	Hour15	Hour16	Hour17	Hour18	Hour19	Hour20	Hour21	Hour22	Hour23
Constant	.01	.00	.00	.00	.00	.00	.00	.02	.03	.04	.06	.11	.10	.08	.05	.07	.09	.08	.06	.00	.00	.00	.01	.01
HS 2007	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Working day before holiday		.02				.04	.04												.02	.00	.00	.00	.00	.00
Holiday (Sunday)	.00			.04	.05	.04	.05	.03				.02	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday (Mon-Fri)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday (Saturday)										.04	.02	.03	.03	.02	.02	.02	.01	.02	.02		.04			
Working day after holiday	.00	.00	.00	.00	.00	.00	.00	.00							.02	.02	.01	.01	.01	.01	.00	.00	.01	.01
Bridge day before holiday																			.01	.00	.00	.00	.00	.00
Bridge day after holiday	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
School holiday South	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Carnival	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
bouwvak south	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
CDD	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
HDD	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Length of the day					.01	.00	.00	.00	.00	.00	.00	.00	.01	.01									.03	.01
LNMOVA Brabant	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Table J.2: p-values for Brabant

NOORD	HOUR0	HOUR1	HOUR2	HOUR3	HOUR4	HOUR5	HOUR6	HOUR7	HOUR8	HOUR9	HOUR10	HOUR11	HOUR12	HOUR13	HOUR14	HOUR15	HOUR16	HOUR17	HOUR18	HOUR19	HOUR20	HOUR21	HOUR22	HOUR23
Constant	.35	.49	.54	.53	.52	.39	.29	.19	.15	.16	.18	.24	.26	.30	.29	.32	.30	.29	.35	.55	.67	.66	.35	.46
Working day before holiday																		-.03	-.04	-.04	-.05	-.05	-.05	-.05
Holiday (Sunday)	-.03											-.04	-.04	-.05	-.05	-.05	-.05	-.05	-.04	-.04	-.05	-.05	-.05	-.04
Holiday (Mon-Fri)	-.07	-.03	-.03	-.05	-.06	-.08	-.11	-.18	-.26	-.30	-.28	-.26	-.26	-.25	-.26	-.27	-.27	-.25	-.20	-.15	-.14	-.13	-.11	-.08
Holiday (Saturday)										-.09	-.11	-.10	-.11	-.10	-.10	-.11	-.11	-.10	-.09	-.07	-.08	-.06		
Working day after holiday	-.03	-.08	-.08	-.07	-.06	-.06	-.05	-.03							.02	.03	.03	.03	.03	.02	.02	.02		
Bridge day before holiday																			-.08	-.09	-.11	-.11	-.11	-.09
Bridge day after holiday	-.05	-.10	-.10	-.10	-.10	-.10	-.11	-.13	-.15	-.15	-.13	-.12	-.13	-.12	-.12	-.12	-.12	-.10	-.08	-.06	-.05	-.05	-.05	-.04
School holiday north	.01			.01	.01	.01	.01	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.01	.01	.01	.01	.01
Bouwvak north	-.02	-.02	-.02	-.03	-.03	-.03	-.03	-.03	-.04	-.04	-.04	-.04	-.04	-.04	-.04	-.04	-.04	-.04	-.03	-.03	-.03	-.03	-.03	-.03
CDD	.01	.00	.00	.00	.00	.00			.00	.00	.00	.00	.00	.01	.01	.01	.01	.01	.01	.01	.01	.00	.00	.00
HDD	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Length of the day						.04	.04	.04	.04	.04	.04	.04	.04	.04	.03								.03	.03
LNMOVA Noord	.97	.96	.96	.96	.96	.97	.98	.98	.99	.99	.99	.98	.98	.98	.98	.98	.98	.98	.97	.96	.95	.95	.97	.96

Table J.3: Coefficients for Noord

NOORD	Hour0	Hour1	Hour2	Hour3	Hour4	Hour5	Hour6	Hour7	Hour8	Hour9	Hour10	Hour11	Hour12	Hour13	Hour14	Hour15	Hour16	Hour17	Hour18	Hour19	Hour20	Hour21	Hour22	Hour23
Constant	.18	.04	.06	.06	.05	.05	.04	.02	.03	.04	.03	.01	.01	.05	.04	.04	.02	.03	.02	.02	.00	.00	.13	.11
Working day before holiday																		.02	.00	.00	.00	.00	.00	.00
Holiday (Sunday)	.01											.02	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday (Mon-Fri)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday (Saturday)								.03	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.04	
Working day after holiday	.00	.00	.00	.00	.00	.00	.00	.00											.05					.04
Bridge day before holiday																			.02	.00	.00	.00	.00	.00
Bridge day after holiday	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
School holiday South											.04	.01	.01	.01	.01	.01	.02	.04	.03	.02			.01	.03
Carnival	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
bouwvak south	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
CDD	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
HDD	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Length of the day	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.01	.01	.01	.00	.01	.00	.00	.00	.00
LNMOVA Limburg	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Table J.4: p-values for Noord

Appendix K: Relationship between economic growth and electricity demand

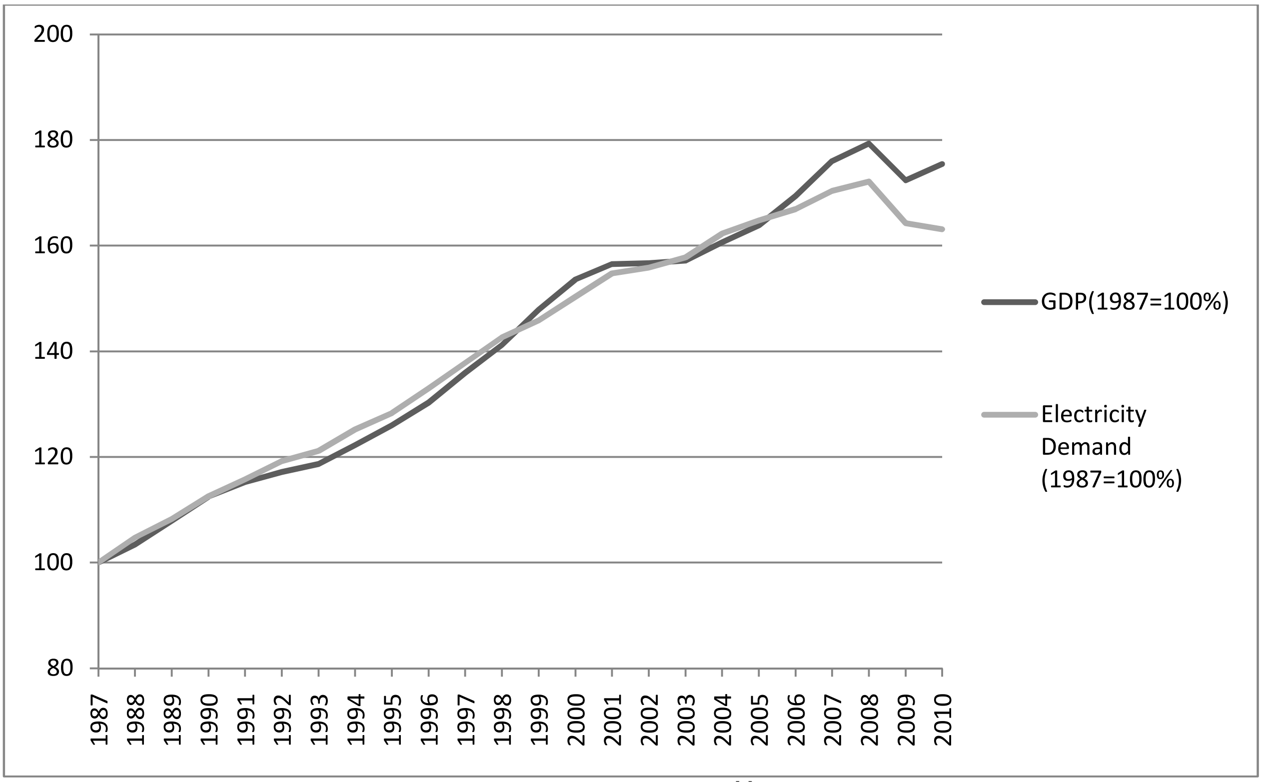


Table K.1: Relationship between economic growth and electricity demand[4].

Appendix L: Coefficients of time series models

BRABANT	HOUR0	HOUR1	HOUR2	HOUR3	HOUR4	HOUR5	HOUR6	HOUR7	HOUR8	HOUR9	HOUR10	HOUR11	HOUR12	HOUR13	HOUR14	HOUR15	HOUR16	HOUR17	HOUR18	HOUR19	HOUR20	HOUR21	HOUR22	HOUR23
AR(1)	.30	.23	.16	.13	.14	.87	.82	.78	.81	.82	.82	.79	.78	.81	.78	.76	.75	.83	.88	.89	.88	.89	.88	.86
MA(1)	-.49	-.55	-.48	-.49	-.52	.19	.18	.19	.23	.25	.24	.21	.23	.30	.28	.26	.29	.38	.38	.39	.29	.36	.34	.21
MA(7)	.99	.92	.84	.89	.89	.89	.89	.85	.84	.84	.82	.79	.78	.79	.79	.80	.80	.81	.78	.83	.87	.89	.87	.83
HS2007	.08	.10	.10	.02	.12	.06	.07	.07	.06	.05	.04	.04	.04	.04	.04	.04	.05	.03				.03	.03	.03
Working Day before holiday	-.02						-.01	-.02	-.03	-.04	-.03	-.03	-.04	-.04	-.05	-.06	-.07	-.07	-.07	-.07	-.08	-.09	-.10	-.09
Holiday Sunday	-.03	-.01				-.02	-.02					-.02	-.03	-.04	-.04	-.04	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05
Holiday Mon-Fri	-.10	-.05	-.05	-.01	-.08	-.10	-.15	-.27	-.40	-.46	-.42	-.38	-.37	-.35	-.38	-.39	-.40	-.37	-.30	-.24	-.22	-.20	-.19	-.16
Holiday Saturday	-.04			.00	-.05	-.07	-.08	-.12	-.17	-.22	-.22	-.21	-.20	-.19	-.19	-.20	-.21	-.19	-.15	-.11	-.10	-.08	-.06	-.06
Working Day After Holiday	-.04	-.09	-.08	-.01	-.08	-.07	-.07	-.05	-.04	-.02														
Bridge Day Before Holiday		-.04	-.07	-.01	-.07	-.05	-.06	-.09	-.14	-.14	-.11	-.10	-.11	-.11	-.14	-.14	-.15	-.15	-.16	-.15	-.17	-.19	-.19	-.16
Bridge Day After Holiday	-.05	-.09	-.09	-.01	-.09	-.10	-.11	-.15	-.21	-.22	-.18	-.17	-.16	-.16	-.17	-.16	-.15	-.13	-.10	-.07	-.07	-.06	-.06	-.04
School Holiday South								-.01	-.04	-.03	-.02	-.01	-.01		-.01	-.01	-.01	-.01						
Carnival					-.02	-.02	-.03	-.04	-.06	-.07	-.06	-.05	-.05	-.05	-.06	-.07	-.07	-.08	-.07	-.05	-.05	-.05	-.04	-.03
Bouwvak South		-.02	-.03	.00	-.03	-.02	-.03	-.05	-.06	-.06	-.05	-.05	-.05	-.05	-.06	-.06	-.06	-.06	-.04	-.04	-.03	-.03	-.04	-.03
CDD	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01
HDD	.00	.00	.00	.00	.00		.00	.00						.00	.00	.00	.00	.00	.00					
Length of the Day	-.06						-.07	-.17	-.27	-.24	-.21	-.19	-.17	-.18	-.18	-.18	-.19	-.26	-.37	-.36	-.37	-.33	-.26	-.14
LNMOVA Brabant	.65	.74	.80	.09	.67	.67	.57	.42	.34	.29	.28	.31	.34	.39	.37	.37	.38	.42	.47	.53	.54	.55	.50	.47

Table L.1: Descriptive Statistics Time Series Model 2010 Brabant

Brabant	Hour0	Hour1	Hour2	Hour3	Hour4	Hour5	Hour6	Hour7	Hour8	Hour9	Hour10	Hour11	Hour12	Hour13	Hour14	Hour15	Hour16	Hour17	Hour18	Hour19	Hour20	Hour21	Hour22	Hour23
AR(1)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
MA(1)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
MA(7)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
HS2007	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00	.04				.03	.02	.01
Working Day before holiday	.00						.04	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday Sunday	.00	.03				.01	.01					.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday Mon-Fri	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday Saturday	.02			.04	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Working Day After Holiday	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00														
Bridge Day Before Holiday		.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Brigde Day After Holiday	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
School Holiday South								.00	.00	.00	.00	.00	.01		.03	.01	.00	.01						
Carnival					.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Bouwvak South		.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
CDD	.00	.00	.00	.00	.00	.00	.00	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
HDD	.00	.00	.00	.00	.00		.03	.01						.02	.01	.00	.00	.00	.00					
Length of the Day	.00						.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
LNMOVA Brabant	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Table L.2: p-values Time Series Model 2010 Brabant

Noord	HOUR0	HOUR1	HOUR2	HOUR3	HOUR4	HOUR5	HOUR6	HOUR7	HOUR8	HOUR9	HOUR10	HOUR11	HOUR12	HOUR13	HOUR14	HOUR15	HOUR16	HOUR17	HOUR18	HOUR19	HOUR20	HOUR21	HOUR22	HOUR23
AR(1)	.926	.872	.844	.876	.851	.841	.836	.845	.861	.884	.869	.839	.816	.796	.785	.778	.798	.867	.890	.889	.906	.924	.931	.889
MA(1)	.474	.186	.204	.180	.110	.090	.165	.240	.304	.356	.325	.308	.327	.327	.356	.370	.418	.502	.463	.418	.462	.500	.471	.296
MA(7)	.927	.906	.887	.924	.904	.891	.873	.849	.877	.875	.847	.815	.799	.803	.806	.816	.820	.825	.816	.872	.912	.942	.948	.890
Working Day before holiday	-.016						-.015	-.028	-.040	-.039	-.027	-.026		-.034	-.048	-.057	-.065	-.067	-.070	-.064	-.057	-.078	-.078	-.067
Holiday Sunday	-.024	-.019	-.017	-.022	-.018	-.016	-.017	-.014				-.029	-.038	-.052	-.053	-.051	-.046	-.045	-.043	-.044	-.053	-.056	-.051	-.045
Holiday Mon-Fri	-.088	-.035	-.038	-.055	-.068	-.083	-.131	-.237	-.352	-.400	-.374	-.347	-.339	-.323	-.342	-.352	-.360	-.331	-.263	-.196	-.184	-.160	-.135	-.105
Holiday Saturday	-.068	-.044	-.038	-.038	-.042	-.053	-.068	-.092	-.134	-.170	-.192	-.171	-.173	-.170	-.168	-.170	-.174	-.162	-.135	-.104	-.122	-.070	-.054	-.051
Working Day After Holiday	-.041	-.107	-.087	-.086	-.078	-.072	-.065	-.049	-.031	-.020			.001											
Bridge Day Before Holiday		-.023	-.055	-.044	-.036	-.025	-.038	-.085	-.130	-.128	-.095	-.090	-.096	-.097	-.122	-.130	-.135	-.126	-.136	-.123	-.139	-.146	-.139	-.105
Bridge Day After Holiday	-.041	-.097	-.092	-.091	-.088	-.086	-.103	-.142	-.188	-.193	-.158	-.148	-.144	-.138	-.147	-.145	-.142	-.118	-.082	-.056	-.052	-.039	-.039	-.024
School Holiday North		.008	.007	.006	.005			-.010	-.029	-.020	-.013													
Bouwvak North	-.012	-.012	-.013	-.013	-.014	-.013	-.021	-.034	-.045	-.043	-.036	-.043	-.045	-.041	-.048	-.052	-.053	-.044	-.027	-.019	-.026	-.015	-.015	-.013
CDD	.006									.003	.003	.004	.004	.005	.005	.005	.006	.006	.006	.005	.005	.005	.004	.005
HDD					.001	.001	.001	.001													.001	.001	.001	.001
Length of the Day	-.103						-.084	-.196	-.289	-.286	-.264	-.239	-.222	-.205	-.200	-.198	-.209	-.280	-.363	-.361	-.348	-.287	-.217	-.134
LNMOVA Noord	.684	.752	.943	.798	.766	.722	.613	.451	.380	.340	.329	.373	.403	.441	.446	.460	.463	.484	.549	.609	.544	.626	.629	.618

Table L.3: Descriptive Statistics Time Series model 2010 Noord

Noord	Hour0	Hour1	Hour2	Hour3	Hour4	Hour5	Hour6	Hour7	Hour8	Hour9	Hour10	Hour11	Hour12	Hour13	Hour14	Hour15	Hour16	Hour17	Hour18	Hour19	Hour20	Hour21	Hour22	Hour23
AR(1)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
MA(1)	.00	.00	.00	.00	.00	.02	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00
MA(7)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00		.00	.00	.00
Working Day before holiday	.00						.01	.00	.00	.00	.00	.00		.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday Sunday	.00	.00	.01	.00	.00	.00	.01	.03				.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday Mon-Fri	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Holiday Saturday	.00	.00	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Working Day After Holiday	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00			.87											
Bridge Day Before Holiday		.01	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Brigde Day After Holiday	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
School Holiday North		.00	.01	.03	.04			.00	.00	.00	.00													
Bouwvak North	.00	.00	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
CDD	.00									.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
HDD					.01	.00	.00	.03													.00	.00	.00	.00
Length of the Day	.00						.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
LNMOVA Noord	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Table L.2: p-values Time Series Model 2010 Noord

Appendix M: Infeed per hour versus distribution losses per hour (Confidential)

Appendix N: Purchasing process (Confidential)