TAMPERE UNIVERSITY OF TECHNOLOGY

MIKKO JÄRVINEN

DEVELOPING NETWORK LOSS FORECASTING FOR A DISTRIBUTION SYSTEM OPERATOR

Master of Science Thesis

Examiner: Professor Pertti Järventausta

The examiner and the topic approved in the Faculty of Computing and Electrical Engineering council meeting on 8th of May 2013.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Electrical Engineering
JÄRVINEN, MIKKO: Developing Network Loss Forecasting for a Distribution System Operator
Master of Science Thesis, 56 pages, 7 Appendix pages
August 2013
Major: Power Engineering
Examiner: Professor Pertti Järventausta
Keywords: Network losses, meter data management system, automatic meter reading, forecasting, regression

Elenia Oy is a forerunner in Finland in adopting the new automatic meter reading (AMR). By late 2008 most of Elenia's customers were equipped with a new meter that is capable of recording and sending hourly consumption figures. Since then Elenia has been working on ways to utilize this new data. In recent years more attention has been given to network losses. Network losses are one of the largest items of expenditure for distribution system operators (DSO) and as such a good target for cost optimization. In addition the Energy Market Authority is contemplating on possible ways to include network loss costs into the regulation model.

Network losses are formed whenever electric power is transmitted from a place of production to end-users. The losses are formed mainly in the resistances of lines and transformers which are heated up by the loss energy. There are two kinds of losses: no-load losses and load losses. No-load losses are relatively constant and do not depend on the load. Load losses are proportional to the square of the transferred power. Before the large-scale installation of AMR meters the hourly consumption figures were unobtainable and as a consequence also the amount of losses was uncertain.

The main goal of this thesis was to develop a usable Excel-based application for predicting hourly network losses. Loss forecasts can be utilized in procurement and hedging of losses. The application is based on hourly consumption figures acquired from the meter data management system (MDMS) and it formulates the predictive models with the use of multiple linear regression analysis. The application has separate regression models for each month and for the whole year. The main predictor variable is temperature and in addition there are calendar-based indicator variables. Separate models are made for two response variables: network losses and network loss percent.

A forecast was made for January 2013 with the application and the results were compared to the observed values. The results give some promise but also raise questions. In general the loss forecast follows the trend of the hourly losses fairly well but the predicted losses are a bit too high on average with an average error of 2.1 MWh and mean absolute error of 2.8 MWh. The mean absolute percent error is 7.3%. Some of the magnitude of the errors is attributed to data quality issues in the early 2012 data.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Sähkötekniikan koulutusohjelma
JÄRVINEN, MIKKO: Sähkönjakeluverkonhaltijan verkostohäviöiden ennustamisen kehittäminen
Diplomityö, 56 sivua, 7 liitesivua
Elokuu 2013
Pääaine: Sähkövoimatekniikka
Tarkastaja: Professori Pertti Järventausta
Avainsanat: Verkostohäviöt, mittaustiedon hallintajärjestelmä, kaukoluettavat mittarit, ennustaminen, regressio

Elenia Oy tunnetaan Suomessa edelläkävijänä kaukoluettavien AMR-mittareiden käyttöönotossa. Vuoden 2008 loppuun mennessä suurimmalle osalle Elenian asiakkaista oli vaihdettu lukemien tuntikohtaiseen tallentamiseen kykenevä etäluettava mittari. Siitä lähtien Elenia on työskennellyt uuden tarkemman kulutusdatan hyödyntämisen parissa. Viime vuosina verkostohäviöt ovat saaneet enemmän huomiota osakseen. Häviöt ovat yksi verkkoyhtiön suurimmista kulueristä ja siten tärkeä kohde kulujen optimoinnille. Lisäksi Energiamarkkinavirasto pohtii mahdollisia keinoja häviökustannusten sisällyttämiseen valvontamalliin tulevaisuudessa.

Verkostohäviöitä syntyy aina kun sähköä siirretään tuotantopaikasta loppukuluttajille. Häviöt syntyvät pääasiassa johtojen ja muuntajien resistansseissa. Häviöt jaetaan kahteen kategoriaan: tyhjäkäyntihäviöihin ja kuormitushäviöihin. Tyhjäkäyntihäviöt ovat lähes vakioita eivätkä riipu verkon kuormituksesta. Kuormitushäviöt sen sijaan ovat verrannollisia siirretyn tehon neliöön. Ennen AMR-mittareiden laajamittaista asentamista kulutuksien tuntiarvoja ei ollut saatavilla ja siten myös häviöiden määrää ei pystytty selvittämään tarkasti.

Tämän diplomityön päätavoite oli kehittää Excel-pohjainen sovellus häviöiden tuntikohtaiseen ennustamiseen. Häviöennusteita voidaan hyödyntää verkostohäviöiden hankinnassa ja suojauksessa. Sovelluksen lähtödatana oli vuoden 2012 tuntikohtainen kulutusdata, joka saatiin mittaustiedon hallintajärjestelmästä. Tämän datan avulla sovellus muodostaa ennustusmallit käyttäen monen muuttujan lineaarista regressiota. Regressiomalleja muodostetaan jokaiselle kuukaudelle omat ja lisäksi on koko vuoden kattava malli. Tärkeimpänä selittävänä muuttujana käytetään lämpötilaa. Lämpötilan lisäksi käytetään kalenteriin pohjautuvia indikaattorimuuttujia. Mallit luotiin kahdelle selittävälle muuttujalle: häviöiden määrä sekä häviöprosentti.

Sovellusta arvioitiin tekemällä ennuste vuoden 2013 tammikuulle ja vertailemalla ennustetta havaittuihin arvoihin. Tulokset ovat lupausta herättäviä, mutta myös kysymyksiä nousi esiin. Pääsääntöisesti ennustetut häviöt seurasivat toteutuneiden häviöiden trendiä kohtuullisen hyvin, mutta ennustetut häviöt olivat hieman liian suuret. Ennustevirheen keskiarvo oli 2.1 MWh ja ennustevirheiden itseisarvojen keskiarvo oli 2.8 MWh. Prosentuaalisen ennustevirheen keskiarvo oli 7.3%. Osan ennustevirheestä oletetaan syntyvän 2012 alkuvuoden pohjadatassa olevien puutteiden vuoksi.

PREFACE

This thesis was done at Elenia Oy. The thesis was examined by Professor Pertti Järventausta from Tampere University of Technology. The supervisor from Elenia was M. Sc. Matti Halkilahti. I would like to thank them both for great advice and support during the process.

I would also like to thank M.Sc. Ville Sihvola and M.Sc. Matti Halkilahti for the opportunity to work on such an interesting and current topic. Also I would like to thank the co-workers at Elenia for providing a friendly and inspiring workplace.

Last but not least I would like to thank my parents for their support over the years.

Mikko Järvinen
27th May 2013

TABLE OF CONTENTS

1 Introduction 1
  1.1 Elenia Oy 1
  1.2 Previous research 2
2 Background for network losses 3
  2.1 Definition of network losses 3
  2.2 Network loss sources 3
    2.2.1 Lines 3
    2.2.2 Transformers 5
    2.2.3 Other loss sources 7
  2.3 Estimating network losses 7
    2.3.1 Loss function 8
  2.4 Reducing network losses 8
3 Electricity market and regulation 10
  3.1 Nordic electricity market 10
    3.1.1 Power exchange Nord Pool Spot 10
    3.1.2 Elspot 10
    3.1.3 Elbas 11
    3.1.4 Financial market 11
    3.1.5 Balance settlement 12
    3.1.6 Risks at the electricity market 12
  3.2 Laws and regulations concerning network losses 13
  3.3 Energy efficiency 13
4 Network losses at Elenia Oy 15
  4.1 Overview of AMI at Elenia 15
  4.2 Management of network losses at Elenia 16
    4.2.1 Overview of the developed forecasting application 18
  4.3 The need for accurate loss forecasting 20
5 New loss forecasting models 22
  5.1 Linear regression 22
    5.1.1 Predictor variables 23
    5.1.2 Regression diagnostics 24
  5.2 Base forecasting models 26
    5.2.1 Model variables 27
    5.2.2 Discarded variables 28
    5.2.3 Year-based models 29
    5.2.4 Month-based models 33
  5.3 Volume-based monthly loss forecasting 37
6 Evaluating the new forecasting models with January 2013 data 38
  6.1 Model coefficient comparisons 38
    6.1.1 Loss model 38
    6.1.2 Loss percent model 40

6.2 Forecasting January 2013 losses 41
6.2.1 Loss forecast 41
6.2.2 Loss percent forecast 43
7 Investigating the usability of the weekly product on the financial market 46
7.1 Hedging of network loss procurement in general 46
7.2 Weekly products during winter 2012-2013 49
8 Conclusions 51
References 53
Appendix A : Regression variables and coefficients A.1
Appendix B : Weekly charts for month-based loss forecast for January 2013 .... B.1
Appendix C : Data for week futures C.1

ABBREVIATIONS

AMI	Advanced Metering Infrastructure
AMR	Automatic Meter Reading
CfD	Contract for Difference
DSO	Distribution System Operator
EIP	Energy Information Platform by eMeter
EnergyIP	Energy Information Platform by eMeter
EU	European Union
GPRS	General Packet Radio Service
MAE	Mean Absolute Error
MAPE	Mean Absolute Percent Error
MDMS	Meter Data Management System
MSE	Error Mean Square
MUDR	Meter Usage Data Repository
NIS	Network Information System
PLC	Power Line Carrier
SSE	Error Sum of Squares
SSTO	Total Sum of Squares
TSO	Transmission System Operator
VBA	Visual Basic for Applications
VIF	Variance Inflation Factor

1 INTRODUCTION

Electricity distribution business is a natural monopoly. Due to this characteristic the government has seen it appropriate to tightly regulate the distribution sector after the liberation of the electric markets in 1995 in Finland. The regulation is carried out by the Energy Market Authority. Currently the regulation model considers network losses as a non-controllable item of expenditure. Network losses are a large expenditure for DSOs and the Energy Market Authority is interested in including network losses in a broader manner to the regulation model in the future. No matter what method is ultimately chosen to supervise the costs associated with losses it means that DSOs need to be able to forecast network losses more accurately on hourly basis to facilitate the procurement of loss energy and hedging of the prices.

In this thesis the main goal is to develop an Excel-based application for forecasting hourly network losses. The application can be used to forecast losses for a chosen time period with the help of weather forecasts or it can do forecasts based on long time average temperatures. The forecasting is done by utilizing multiple linear regression. The base data for the regression is the hourly losses for the year 2012 obtained from the MDMS at Elenia Oy.

The early part of the thesis concentrates on giving the necessary background information on how network losses are formed. Overview of the Nordic electricity market is presented as well. The background part is finished with a discussion about the management of network losses at Elenia. The latter part of the thesis starts with an overview of multiple linear regression that is used to build the forecasting models. The main part of the thesis is spent analyzing the regression models and their validity. Also a forecast is made for January 2013 which is then compared to observed data. Finally there is a brief overview of hedging and the viability of week future products is investigated.

1.1 Elenia Oy

Elenia Oy is an independent distribution system operator servicing over 410000 distribution network customers in approximately 100 municipalities with a network area of nearly 50000 km^2. Elenia's network is comprised of mostly rural areas and the average line length per customer adds up to around 160 meters. At over 60000 kilometers the total line length is enough to go around the world one and a half times. In addition there are over 100 primary substations and over 20000 distribution transformers to manage. Elenia is known as a forerunner in development and adoption of new technologies for distribution networks. By the end of the year 2008 most of Elenia's customers had AMR meters installed.

Elenia Oy was formed at the start of 2013 through the fusion of Elenia Verkko Oy, Elenia Asiakaspalvelu Oy and Asikkalan Voima Oy. Previously Elenia Verkko Oy was briefly known as LNI Verkko Oy during spring 2012 after Vattenfall sold their Finnish distribution division Vattenfall Verkko Oy.

1.2 Previous research

There has been some research done previously on determining network losses. For example master of science theses by Itäpää (1979), Paloposki (1999), Tyynismaa (2003) and Kuisma (2008). Also one licentiate thesis has been made on the subject by Kinnunen (2002). However these concentrate on calculating or estimating network losses based on modeling the network or its components. The problem has been that the consumption figures have been hard to obtain. Traditionally energy meters have been read only once a year. This means that determining the hourly consumption has been impossible for most of the customers. In recent years the new electricity meters that record hourly consumption and are read remotely have been installed in larger numbers. By the end of 2013 over 80% of customers in Finland should have a new meter installed by regulation. This new availability of hourly consumption data gives opportunities for better estimation and forecasting. Mutanen et al. (2011a; 2011b) have researched the use of hourly consumption data in improving the customer load profiles used by DSOs. Matti Koivisto made a thesis in 2010 on using hourly consumption data to predict electrical loads of residential customers through statistical methods (Koivisto 2010). Koivisto's thesis has given some food for thought while doing this thesis as well.

2 BACKGROUND FOR NETWORK LOSSES

In this section we consider what network losses are and how they are formed. In addition we take a look at how the amount of losses can be estimated and if something could be done to reduce them.

2.1 Definition of network losses

Network losses can be defined simply as the difference between input energy and output energy of the network

$$loss\ energy = input\ energy - \sum loadpoint\ energy \tag{1}$$

In equation (1) input energy is defined as all the energy fed to the network and loadpoint energy is all the energy delivered to customers. Loadpoint energy is all the energy delivered to the customers not including the small loading of electricity meters themselves. The difference between these is the loss energy. (Seppälä et al. 2011)

Network losses are usually divided into two categories: no-load losses and load losses. No-load losses do not depend on the load. The losses vary with voltages but remain relatively constant. Load losses depend on the load in the network. (Itäpää 1979). As can be seen in equation (3) the relationship between load losses and the transferred active power is approximately quadratic.

2.2 Network loss sources

2.2.1 Lines

When a current flows through a line the charge-carrying electrons collide with ions that make up the conductor material and in the process give a part of their kinetic energy to the ions causing the material to heat up. This phenomenon is called resistance and it is the primary source of energy losses in the network.

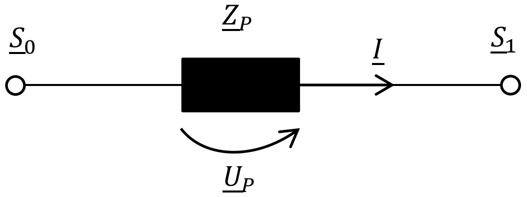

Figure 2.1. *A simple single-phase line.*

Power losses over the line in Figure 2.1 can be calculated

$$\begin{aligned} \underline{S}_{L,P} &= \underline{S}_0 - \underline{S}_1 = \underline{U}_P \underline{I}^* = \underline{Z}_P |\underline{I}|^2 = (R_P + jX_P)(I_p^2 + I_q^2) \\ &= (I_p^2 + I_q^2)R_P + j(I_p^2 + I_q^2)X_P = P_{L,P} + jQ_{L,P} \end{aligned} \tag{2}$$

Where $\underline{S}_0$ is the apparent power at the start of the line
$\underline{S}_1$ is the apparent power at the end of the line
$\underline{S}_{L,P}$ is the power loss over the line
$\underline{Z}_P$ is the impedance of the line
$\underline{U}_P$ is the voltage over the line (phase-voltage)
I_p is the active current in the line
I_q is the reactive current in the line

From this equation we can obtain three-phase active power losses

$$P_L = 3P_{L,P} = 3(I_P^2 + I_q^2)R_P = 3\left(\frac{P^2}{U_P^2} + \frac{Q^2}{U_P^2}\right)R_P = \left(\frac{P}{U}\right)^2 R_P + \left(\frac{Q}{U}\right)^2 R_P \tag{3}$$

In similar fashion we can obtain three-phase reactive power losses

$$Q_L = \left(\frac{P}{U}\right)^2 X_P + \left(\frac{Q}{U}\right)^2 X_P \tag{4}$$

Equation (3) shows that transferring reactive power in the network also causes active power losses. This happens because transferring reactive power increases the total current in the line.

Lines have a shunt capacitance and series inductance. Current flowing through the inductance consumes reactive power and voltage on the line produces reactive power in the capacitance. Each line has an operating point where the line consumes all reactive power it produces. It is then said that the line operates at natural power. Natural power of a line depends on the surge impedance and voltage. Table 2.1 has a few examples of natural power for different lines and voltages.

Table 2.1. *Examples of natural power of lines (Elovaara & Haarla 2011a)*

Nominal Voltage (kV)	**Overhead line, 3-phase (MW)**	**Underground cable, 3-phase (MW)**
10	0.26	2.6
20	1.0	10
45	5.4	54
110	32	320

As can be seen from the above table underground cables have ten times the natural power when compared to overhead lines because of their capacitance. Overhead lines are usually operated near natural power. However cables produce large amounts of excess reactive power which needs to be taken into consideration. Natural power of a line can be estimated with equation (5). (Elovaara & Haarla 2011a).

$$P = U^2 \cdot \sqrt{\frac{B}{X}} = U^2 \cdot \sqrt{\frac{C}{L}} \tag{5}$$

High-voltage overhead lines can also experience corona losses where the dielectric strength of the air breaks down and partial discharges start to form on the surface of the conductors. However this is not a concern in distribution networks where the rated voltage is less than 110 kilovolts. (Aro et al. 2003).

Energy losses also happen in the insulators due to leakage current. The conductivity of copper is in the order of 10^{18} to 10^{21} times higher than the conductivity of dielectric materials like porcelain, glass and mineral oil (Elovaara & Haarla 2011b). The resistances of insulators are so high that leakage losses can usually be ignored in distribution networks. As an example of magnitudes in question the leakage losses for a string insulator unit in 110 kV overhead lines are 5 watts in dry air, 50 watts in fog or rain and 100 to 150 watts in drizzle or rime. However if the insulators are very dirty the losses can be significantly higher. (Elovaara & Haarla 2011a).

2.2.2 Transformers

Unlike in the case of lines the no-load losses of transformers are significant in distribution networks. In his thesis Paloposki (1999) found out that in the studied distribution network the no-load losses were over twice as high as the load losses of transformers. This is in part due to the tendency to oversize network components just in case. Also the need to prepare for equipment outages in contingency plans pushes towards higher rated transformers than would be necessary under normal operating conditions.

When a transformer is energized by a voltage a magnetizing current starts to flow through it and two types of no-load losses occur. First type is termed eddy losses and second type is termed hysteresis losses. Eddy losses are caused by currents circulating in the structures of the transformer and these currents are induced by the alternating flux from the magnetizing current. Hysteresis phenomenon is related to the magnetic properties of the ferromagnetic core material. Hysteresis deals with the fact that the magnetic field in the core material can have different values depending on whether the external magnetic field is increasing or decreasing. These no-load losses are also termed iron losses or core losses. Load losses are formed in the resistances of the windings when load current flows through them. Load losses are also sometimes called copper losses. (Nousiainen 2007).

Usually the manufacturer measures the no-load losses and load losses of a transformer. These are given for the rated voltage and rated power of the transformer. For large power transformers the manufacturer might provide measurement data for several operation points in addition to the rated voltage and rated power. The voltage dependency of no-load losses can be estimated with

$$P_0 = \left(\frac{U}{U_N}\right)^{p_{0u}} \cdot P_{0N} \tag{6}$$

Where P_0 is the no-load losses
P_{0N} is the no-load losses at rated power
U is the voltage over the primary winding of the transformer
U_N is the rated voltage of the transformer
p_{0u} is the voltage sensitivity of the transformer's no-load losses

The ranges of values for p_{ou} in equation (6) are demonstrated in Table 2.2. Calculating the no-load losses with these values gives a range from 89% to 120% when compared to no-load losses at rated voltage. As a rule of thumb it can be estimated that a one percent increase in voltage from the rated voltage increases the no-load losses by three percent. (Paloposki 1999).

Table 2.2. *Voltage sensitivity of a transformer's no-load losses (Paloposki 1999)*

Voltage Range U/U_N	**Voltage sensitivity** p_{0u}
0,950 … 0,975	2,35
0,970 … 1,000	2,90
1,000 … 1,025	3,30
1,025 … 1,050	3,80

Load losses can be estimated with equation (7) when the load on the transformer is known. (Nousiainen 2007).

$$P_L = \left(\frac{S}{S_N}\right)^2 \cdot P_{LN} \tag{7}$$

Where P_L is the load losses
P_{LN} is the load losses at rated power
S is the current loading of the transformer
S_N is the rated power of the transformer

Table 2.3 shows a small example of losses in transformers manufactured by ABB. ABB manufactures a wide range of transformers with different power rating, losses and noise levels.

Table 2.3. *Excerpt of losses of liquid filled transformer examples (ABB 2010)*

Rated Voltage (kV)	**Rated Power (kVA)**	**No-load Losses (W)**	**Load Losses (75 °C) (W)**
20	50	125	1350
20	250	650	3250
20	250	425	4200
20	630	1300	6500
20	1600	1700	20000

Using equation (7) we can calculate at which power the load losses match no-load losses for the transformers in the ABB (2010) brochure. From calculations we can see that the load has to be between 29% and 45% of the rated power of the transformer. This is also the point where the transformer is operating at its peak efficiency. However the efficiency stays high for the whole operating window with the exception of very low powers. For example the peak efficiency for the 1600 kVA transformer in Table 2.3 is 99.3% at input power of 466 kVA. At rated power the efficiency is 98.6%. Generally for the example transformers the peak efficiencies range around 98-99%.

2.2.3 Other loss sources

While lines and transformers account for the majority of losses in the network there are several other loss sources. Some of them are true losses and some of them appear as losses while by different reasoning they might not be considered as losses.

Electricity meters also use energy in their operation. Meters use approximately 1-7 watts of power depending on the type of meter at hand. Generally static meters use less energy than inductive meters and single-phase meters use less than three-phase meters. (Kuisma 2008). New AMR meters from Iskraemeco that are used in Elenia’s network use approximately one watt per single-phase and three watts per three-phase meter. Electricity meters in Elenia’s network consume approximately 9.5 GWh per year. (Sievi 2013). Another way meters cause losses is through measuring error. While the energy is not lost in the physical sense it shows up as energy that is input into the network but not delivered to the customer. In his thesis Tyynismaa (2003) estimated that the losses caused by measuring errors in Helsinki Energia’s network were approximately 4 GWh per year.

There is also a lot of other equipment in the distribution network that consume power such as fuses, circuit breakers, switchgear, relays, instrument transformers and other equipment in substations. Generally the energy consumed by these is hard to estimate and their significance to total losses in the network is negligible. There can also be non-metered consumption in the network such as street lighting. In these cases the power consumption and usage hours are known and their total energy consumption can be estimated. Another form of non-metered consumption is electricity theft. In Finland electricity theft is negligible but it can be a major problem in some other countries.

2.3 Estimating network losses

Network losses as defined in equation (1) include all the different loss sources. Input energy includes the energy coming in to the network from other networks and the production inside the network. Input energy is generally readily available from hourly metering at the network’s access points. Loadpoint energy is comprised of three major components. Energy transferred out of the network to other networks, energy delivered to end-users and the remaining part that makes up the network losses. The main difficulty in determining the losses is the estimation of end-user consumption. It is not easily available until AMR meters are installed at every consumption point in the DSO's network.

Network information systems (NIS) can be used as a help in calculating network losses. NIS has information on the lines and cables in the network and their electrical values. However there are many loss components missing from the systems which need to be taken into consideration when determining total losses. Another problem is obtaining the load information for non-hourly metered consumption points.

2.3.1 Loss function

The loss function estimates hourly energy losses based on the input power to the network. To estimate losses first a loss% is calculated from observed loss energy and input energy data of the network

$$loss\% = 100 \times \frac{annual\ loss\ energy}{annual\ input\ energy} \tag{8}$$

The loss function $P_L(t)$ estimates network losses from the input energy

$$P_L(t) = P_0 + k * P(t)^2 \tag{9}$$

Where P_L is the losses at hour t

P_0 is the no-load losses of transformers in the network

P is the input energy at hour t

From the equation we can see that losses are equal to the no-load losses plus squared input energy multiplied by a coefficient. The coefficient k is defined so that the losses P_L resolve in to the loss% given by equation (8) over a time period T

$$k = \frac{[0{,}01 \times loss\% \times \sum_{t\epsilon T} P(t)] - \sum_{t\epsilon T} P_0}{\sum_{t\epsilon T} P(t)^2} \tag{10}$$

The time period T is usually one or multiple calendar years. (Seppälä et al. 2011)

2.4 Reducing network losses

A simple way to reduce network losses is to increase the conducting cross-section of lines. However, when considering reduction of network losses one has to also consider the total costs. Usually this is done by valuing the future losses with present value method and adding up investment and maintenance costs. The difficulty of estimating the financial aspects of different investments rises from the fact that the life time of electrical equipment in network is generally in the order of tens of years. Combined with the difficulty of choosing appropriate interest rate and cost of electrical energy the comparison can quickly turn into nothing more than a guess.

Other ways to affect network losses are to optimize network configuration, optimizing voltage levels and compensating reactive power near consumption. Network configuration can usually be optimized with the help of network information systems but there is generally no reason to do this more than once a year or every few years. However some companies might employ two different network configurations depending on the season. Optimizing the network configuration is constrained by protection design, usage concerns and so forth. Leeway in changing voltage levels is usually very small or non-existent. Large consumers of reactive power are steered into compensating their own usage by relatively high reactive power tariffs.

In general the Finnish distribution networks are strong already. At around 4% the total network losses are among the lowest in EU. Network losses are already taken into account while choosing the size of the conductors. For medium voltage cables with small cross-section the economical load is only a tenth of the load capacity. (EMV 2010). In his thesis Paloposki (1999) didn’t find viable ways to lower energy losses in Vantaa Energia’s distribution network. One small possibility was to switch off some transformers in the summer during low loading but this would have caused unacceptable reliability risks and potential power quality issues compared to the meager energy savings.

3 ELECTRICITY MARKET AND REGULATION

In this section there is an overview of the electricity market in the Nordic and we also take a look at the financial market for electric power. In addition there is a brief outline of the laws and regulations regarding network losses and an overview on energy efficiency as it pertains to DSOs.

3.1 Nordic electricity market

3.1.1 Power exchange Nord Pool Spot

The power exchange was founded in Norway in 1993 as “Statnett Marked”. Name was changed to Nord Pool in 1996 when Sweden joined. Finland’s turn to join was in 1998. In 2002 the spot market activities were organized as a separate company, Nord Pool Spot. At present Nord Pool Spot is owned by Nordic and Baltic transmission system operators (TSO). Total trade volume in 2011 was 316 TWh. (Nord Pool Spot 2011). At present Nord Pool Spot covers Denmark, Finland, Sweden, Norway, Estonia and Lithuania (Nord Pool Spot 2012a).

Electricity wholesale markets are comprised of several parts. Elspot is a day-ahead market in the Nordic and Baltic region. Elbas is intraday market in the Nordic and Baltic region. Elspot and Elbas are physical electricity markets and they are operated by Nord Pool Spot. The financial market was sold to NASDAQ OMX Commodities in 2008 by Nord Pool. (Nord Pool Spot 2011).

3.1.2 Elspot

Elspot is a day-ahead physical wholesale market for electricity in the Nordic and Baltic region. More than 70% of total energy consumption was acquired through Elspot in the Nordic region in 2011 (Nord Pool Spot 2011). Sellers and buyers must send their offers to the exchange on the previous day before noon (13:00 Finnish time). The smallest unit of trade in the market is 0.1 MWh. At 13:00 Finnish time Nord Pool Spot starts the process of aggregating a price for each hour for the next day based on the received offers. After the calculation has been finished Nord Pool Spot informs the participants how much they bought and sold electricity each hour. This information is also sent to TSOs who need it for balance settlement. (Nord Pool Spot 2012a).

The procedure described above gives the system price which is the price that would be if there were no transmission bottlenecks. The Nordic and Baltic markets are divided in to several different price areas that are connected by various amounts of transmission capacity. For example Sweden is divided into four distinct price areas, Norway into five areas while Finland is one area itself. The transmission capacities between the areas are determined by the TSOs. (Nord Pool Spot 2012b).

After a bottleneck is discovered the prices for areas in question will diverge and form the area prices. The area price is formed by first aggregating the price curves for demand and supply within the areas with the crossing point as the initial price. Then for the area with surplus energy the transfer is reflected as additional demand and for a deficit area the transfer is reflected as additional supply. The area prices are then found from the new crossing points. (Nord Pool Spot 2012a).

3.1.3 Elbas

Elbas is intraday market in the Nordic and Baltic region and it operates around the clock every day of the year. It serves as an aftermarket for Elspot and the products for the following day are published 15:00 Finnish time. Elbas enables one to trade up until one hour before delivery. (Nord Pool Spot 2012). The trade volume in Elbas was 2.7 TWh in 2011 and 2.2 TWh in 2010 (Nord Pool Spot 2011).

3.1.4 Financial market

The financial market for electricity is now operated by NASDAQ OMX Commodities. Only commodity that changes hands on the financial market is money. On the financial market the participants can hedge their selling or buying prices in to the future. The physical electricity market Elspot only operates day-ahead but on the financial market there are products up to six years into the future which allows for appropriate longer term risk management. The reference price used for Nordic market is the Elspot system price. (NASDAQ 2012).

There are several different financial products available in the market. Futures and forwards with base load and peak load products, options and contracts for difference (CfD). Base load contracts are delivered every hour of the week for the duration of the contract while peak load contracts are delivered from 8 to 20 from Monday to Friday. Table 3.1 sums up the available products on the financial market.

Table 3.1. *Available financial products.*

Duration	Base load	Peak load
Day	Future	
Week	Future	Future
Month	Forward, Option, CfD	Forward
Quarter	Forward, Option, CfD	Forward
Year	Forward, Option, CfD	Forward

Futures are available in base load day and week products and peak load week products. The day products are listed for the next week on the last trading day. Thus there are from three to nine day futures available for trade at any one time. There are six base load and five peak load week products available on a rolling cycle. After a trade the future contract is subject to daily mark-to-market settlement until the end of the delivery period. Mark-to-market covers the changes in the future contracts value. During the delivery period there is also a spot reference settlement which covers the difference between the value of the future contract and the spot reference price.

Forward contracts are available in base load and peak load month, quarter and year products. For base load there are six month, from eight to eleven quarter products and five year products. Also for peak load the available products are for the next two months, three quarters and one year. Forwards are similar to futures except that the settlement doesn't start until the delivery period.

Since the area price can differ from the system price which is used as the reference price for the financial products there are also products available that allow the hedging of this price risk. A Contract for Difference (CfD) is a forward product for the difference in area price and system price. The value can be negative or positive depending on whether the market expects the area in question to be a surplus or a deficit area. CfDs are available for the next four months, quarters and years.

Options come in two varieties. Seller of a put option agrees to buy the underlying contract of the option and seller of a call option agrees to sell it. While the seller has the obligation to sell or buy the buyer of the option has the right to buy or sell it. This means that the buyer doesn't have to do it if the prices have developed unfavorably. For this the buyer pays the seller a risk premium. The underlying contracts for options are quarter and year forward products.

3.1.5 Balance settlement

In a sense Elspot is also only a financial market. The buyer gets the electricity even if the producer cannot generate the power due to a sudden fault and the buyer has to pay to the producer. The producer then has to acquire the electricity he had sold to settle his balance. For this reason each market actor needs an open supplier who sells or buys the electricity required. At the highest level the balancing supplier in Finland is Fingrid who is also responsible for electrical balance in the grid. (Partanen et al. 2012).

3.1.6 Risks at the electricity market

There are many risks involved in the electricity business. After the deregulation of the Nordic electricity market the risks have gone up. Some risks are due to the nature of the commodity and some due to the structure of the market. Partanen et al. (2012) list some of the major risks as follows:

- Price risks arise from the volatility of the market. The major factors behind price risks are the weather dependency of production and demand and the fact that storage of electric power is not viable.
- Demand risks are caused by customer’s ability to change suppliers.
- Volume risks are formed when procurement and sales differ.
- Political risks arise from the whims of the politicians. Political terms are short when compared to the timescales involved in electricity business. Varying and uncertain politics create unknown risks for long term investments. One example is the emissions trading system in the EU.
- Operational risks involve miscalculations in the planning of procurement and sales.

In addition to these risks there can be currency risks, credit risks and strategic risks. Open position is the part of procurement that is not secured by bilateral agreements or hedged with financial products. Karjalainen (2006) lists few additional risks. Area price risk means that the price in the area differs from system price due to insufficient transmission capacity. Profile risks are formed because the financial products have a constant volume but the actual consumption varies with time.

3.2 Laws and regulations concerning network losses

Distribution network operation is a natural monopoly as the building of several physical networks in the same area is not feasible. For this reason the Electricity Market Authority regulates the transmission and distribution business. The goal of the regulation is to keep the prices reasonable for consumers while facilitating the further development of electricity networks.

Article 15 b of Electricity Market Act says that network operators must acquire loss energy for their network through open, non-discriminating and market-based procedures (Sähkömarkkinalaki 1995). The current regulation model for years 2012-2015 (EMV 2011) does not include network losses in any special way. Based on a consultation work by Pöyry Management Consulting Oy (EMV 2010) network losses are included in uncontrollable operating costs. However Electricity Market Authority does monitor that DSOs procure energy losses in accordance with the law.

New legislative proposal concerning electricity and natural gas markets states in the justifications portion that an obligation for a bidding competition on providing the loss energy should be set (Government 2013a).

3.3 Energy efficiency

The European Union (EU) has set a goal to decrease the amount of primary energy used in EU by 20 percent by the year 2020 as a part of the so called “20-20-20” target. In practical terms this means that the amount of primary energy used in 2020 should be no more than 17.1 PWh or the final energy used should be less than 12.5 PWh. In 2007 the forecasted primary energy consumption for 2020 was 21.4 PWh so to meet the goal a reduction of 4.3 PWh in primary energy consumption needs to be achieved. (EU 2012).

In Finland the government set a goal of reducing final energy use by 37 TWh or about 11 percent compared to what it would be according to forecasts if the efficiency measures would not be implemented by 2020. In addition the use of electricity needs to be made more effective by 5 TWh or in other words by about 5 percent. (Government 2010). The goal for final energy use in 2020 is 310 TWh. With the current measures in place the projected final energy use would be 325 TWh which means that further measures need to be taken. (Government 2013b)

A cornerstone of meeting the requirements imposed by EU is the energy efficiency agreements. The goal of the voluntary agreements is to reduce the usage of energy that is not included in the emissions trading system by 9 percent by year 2016. The reference level is the average consumption during the years 2001-2005. By signing the agreement the company or community agrees to set goals to improve energy efficiency, implement the measures to achieve these goals and finally to report on the implemented measures and planned improvements. The duration of the contracts is from 2008 to 2016. (EEA 2013)

The electricity distribution sector's goal in the agreement is to reduce losses by 150 GWh during the time period. The only realistic way for a DSO to reduce its losses is to replace a network component with a more efficient one. A big challenge is the verification of the achieved loss reductions for reporting. In addition to the problem of knowing the exact losses before and after the change a big problem is the possibly huge amount of components changed. Current information systems do not have adequate support for the needs of the energy savings reporting. (Seppälä & Trygg 2011)

Another problem in achieving the target of 5 percent reduction in consumption is that network losses make up a vast majority of a DSO's energy consumption. For example Elenia's network losses are approximately 250 GWh per year at a loss percent of less than four. In comparison all the substations in the network use approximately 3 GWh in total per year. Small reductions in the energy consumption at substations or at the DSO's other premises will not be enough to meet the goal by a long shot. As discussed in chapters 2.4 and 4.2 a large reduction in network losses is not economically feasible as the conductor cross-sections in the network are already fairly robustly sized.

4 NETWORK LOSSES AT ELENIA OY

This section gives on overview on how Advanced Metering Infrastructure (AMI) has been progressing at Elenia. Also there is a brief outline on how network losses are managed at Elenia Oy and some discussion on why forecasting network losses is important. In addition there is a description on the forecasting application developed as part of this thesis.

4.1 Overview of AMI at Elenia

AMI can be thought of consisting of six different functionalities: Data Acquisition, Data Transfer, Data Cleansing, Data Processing, Information Storage and Information Delivery. New smart meters take care of the data acquisition. Data transfer is handled by many techniques such as PLC (Power Line Carrier), GPRS (General Packet Radio Service), radio links and so forth. Meter Data Management System (MDMS) is a part of AMI and it is involved in rest of the functionalities. Its main job is to validate the incoming data, store it, analyze it and share it. (Mäkelä 2011).

The MDMS used at Elenia is Energy Information Platform by eMeter (EnergyIP or EIP for short). EnergyIP is comprised of several parts. There are two databases called Meter Usage Data Repository (MUDR) and AMI Database. MUDR stores the large amounts of data coming from the meters. In Elenia’s network the AMR meters generate approximately 10 million hourly consumption figures each day. AMI Database holds the asset information such as accounts, meters, service delivery points, premises and so forth. The application part of EnergyIP is modular. There is no single big application but instead there are many different applications that have different purposes. The applications communicate with each other and the databases through EnergyIP Message Bus. EnergyIP is mainly used through a web browser.

For Elenia the AMI project started in the early 2000s. After a few pilot projects the main AMR project designated Santra started in 2005 and ended in 2008. The goal of Santra was to change the electricity meters of all the residential customers to new AMI meters. After the Santra project the MDMS project was started in 2009. In 2012 the MDMS project had progressed to the point where Elenia started to send hourly measurement data to suppliers. With AMI the most important issue to take into consideration is data quality. Without good data quality all the analyses done with the data will be flawed and the largest benefits of advanced meters will be lost. Also the figures sent to the suppliers will contain balance errors. To facilitate good data quality EnergyIP has applications that process all the incoming data according to set rules before it is stored into the database. (Halkilahti 2013)

When considering network losses the biggest issue in data quality is missing data. All the consumption data that is missing from the database shows up as additional losses. Data missing from a customer usually means that there has been either a data input error which means that the MDMS cannot determine the correct target for the incoming data or that the meter has become faulty. Another possibility is that the reception of the meter is so poor that the meter cannot be contacted.

Since 2010 on average 200-300 faulty AMR meters have been replaced each month. The number varies a lot depending on the amount and severity of thunderstorms in Elenia’s network area. Naturally when a meter becomes faulty it isn’t possible to obtain the consumption figures from it. Generally it can be seen from the systems fairly quickly when data isn’t received from a meter. Some difficulties are caused by so called main-switch targets such as summer cottages where the power is turned off when the residents are away. If the meter is installed after the main-switch it is hard to distinguish whether the meter cannot be contacted because it has become faulty or because it simply doesn’t have power. Luckily the main-switch targets do not use a lot of power usually so the error is not big when it comes to loss calculations. Some companies install shunt wires that keep the meter energized even when the main-switch is turned off but Elenia does not do this because it is seen as risky especially when dealing with old switchboards. New instructions that were given in late 2009 call for switchboard manufacturers to have a place for the meter before the main-switch. (Sievi 2013)

4.2 Management of network losses at Elenia

Network losses in Elenia’s network were 245 GWh in 2011. Energy losses compared to input energy were 3.85%. Compared to other Finnish DSOs this loss percent is the median value with values ranging from under 1% to over 10%. (EMV 2012). Combining the fairly low overall loss percent with the fact that Elenia’s network is mostly in rural areas we can judge that overall the network is already fairly robust. Company’s inner estimates have also come to the conclusion that increasing the cable sizes or transformer sizes to reduce losses is not economically feasible. (Halkilahti 2013)

In early 2012 Elenia moved to utilizing the MDMS data in determining the network losses. Figure 4.1 displays an overview of network loss management at Elenia on a general level. On the left side the current process is illustrated and on the right side a possible use of the forecasting application developed in this thesis is displayed. The process starts at the AMR meters that measure consumption. The meters are read by a service provider who sends the figures to Elenia’s MDM system EnergyIP. After checks the data is stored in to the MUDR database. Based on the data EnergyIP calculates the network loss report. In addition to MDMS calculations a loss formula similar to equation 9 is used to estimate network losses which are then compared to the network loss report to

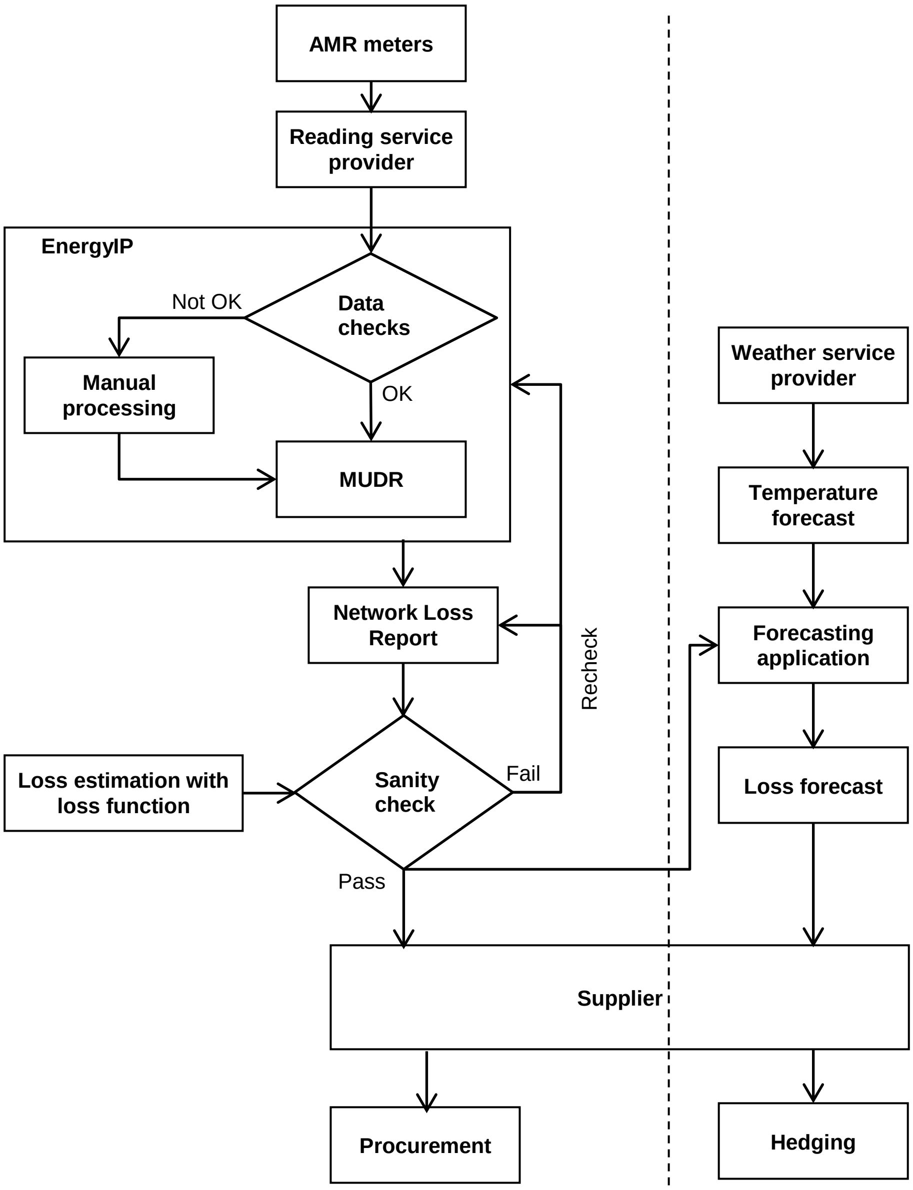

Figure 4.1. *Diagram of the network loss management at Elenia.*

provide a sanity check in case of serious data quality issues which were mentioned in chapter 4.1. If the report seems ok it will be sent to the supplier. (Halkilahti 2013)

Loss energy is acquired through a supplier who is chosen by a bidding competition to minimize the costs and to comply with the law as stated in chapter 3.2. In addition to the physical energy procurement the supplier is required to provide a portfolio management service for hedging the electricity prices. Elenia's hedging policy is to fully hedge the forecasted volume in advance over a lengthy time period to spread the price risk. The main goal of hedging is to have stable and predictable network loss costs. Trying to minimize the costs is important as well but not at the expense of predictability. Elenia does not speculate with the financial products. (Halkilahti 2013)

The right side of Figure 4.1 shows a possible use of the forecasting application developed as a part of this thesis. Network loss data is inserted in to the application as base data from which the forecasting models are formulated. With the help of temperature forecasts the models can be used to forecast network losses. The supplier could then possibly make additional hedging based on the forecast.

4.2.1 Overview of the developed forecasting application

The forecasting application was developed as an Excel application with Visual Basic for Applications (VBA). The application contains roughly 3500 lines of code. Chapter 5 explains how the forecasting models implemented in the application have been formulated. The application is divided into four sections: temperature estimation, basic forecasting, monthly forecasting and model updating and analyzing. The main interface of the application reflects these operations. Figure 4.2 displays the Model tab of the application.

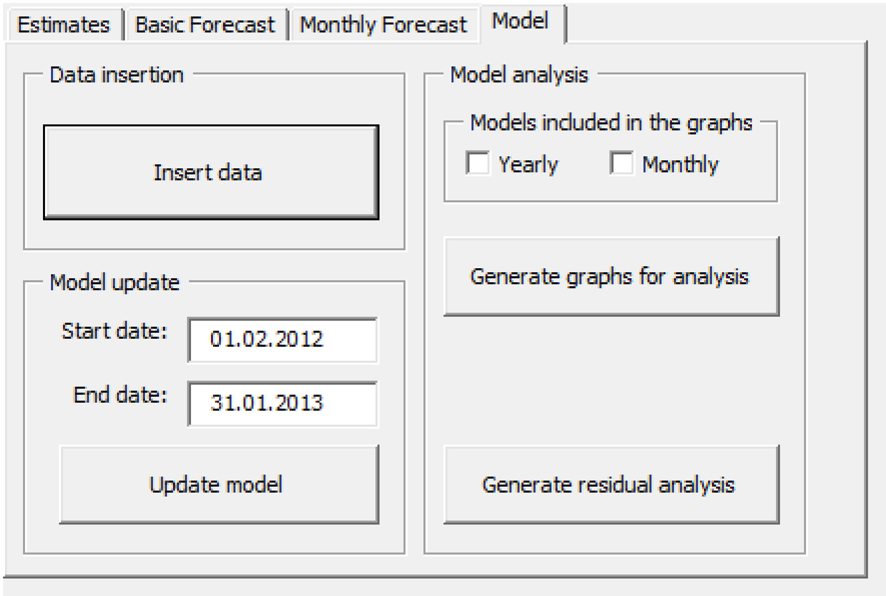

Figure 4.2. *Model section of the application.*

From this tab the user can choose to insert data, update all the models to use a different date range of the data or graph charts and residuals for analyzing the regression models. Insert data button takes the user to a sheet where the hourly energy measurement and temperature data is entered. The models can be updated to use any range of the data with the restriction that the length is a minimum of 360 days and the range is continuous. Graphs can be generated to analyze how well the formulated models fit the data. Also distributions of residuals can be generated to analyze the normality like in Figure 5.4.

Figure 4.3 shows the Estimates tab of the main interface. First field is used to specify a start date. Hours per data point is used to specify how many hours one temperature value will cover. The specified number of days is used to generate the date stamps for the insert page that opens after pressing the button. Temperature offsets lets the user to specify an offset value for the temperature. The application then generates temperature estimates for the given time range from the long time average temperature plus offset value to the insert sheet.

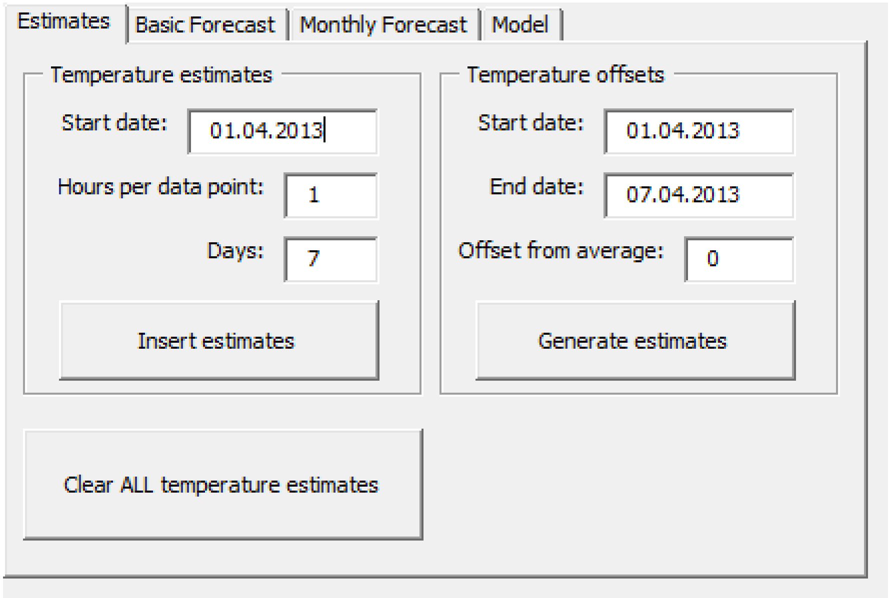

Figure 4.3. *Estimates section of the application.*

Figure 4.4 displays the basic forecast interface. The user simply defines the time period and the application then calculates the forecast. The temperature used for the forecast is the given estimates or if there is no estimate given for an hour then the long time average value is used.

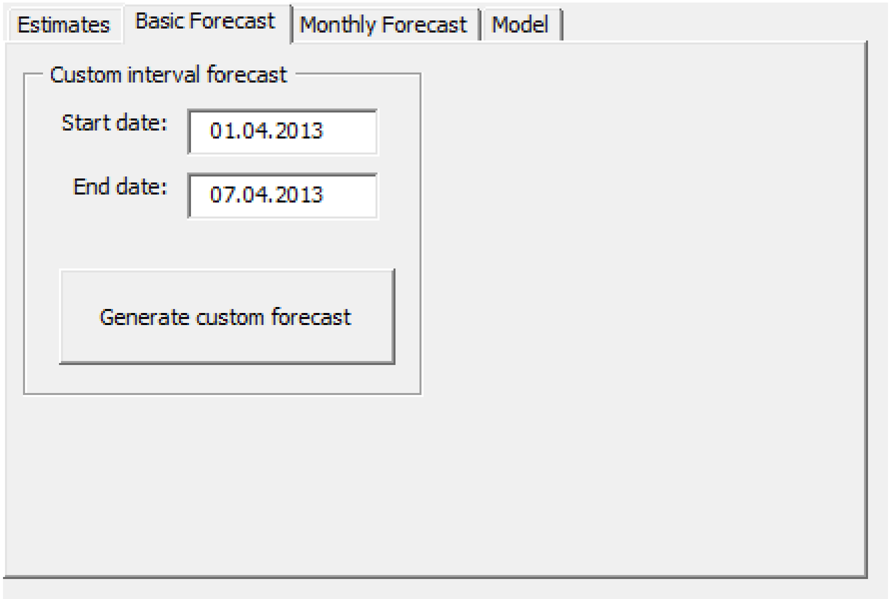

Figure 4.4. *Basic Forecast section of the application.*

Figure 4.5 displays the monthly forecast interface. Target volume means the forecasted total distribution volume for the month. After the user gives the required target values the application then calculates the basic forecast for losses and loss percent for the given month and then scales the losses as described in chapter 5.3.

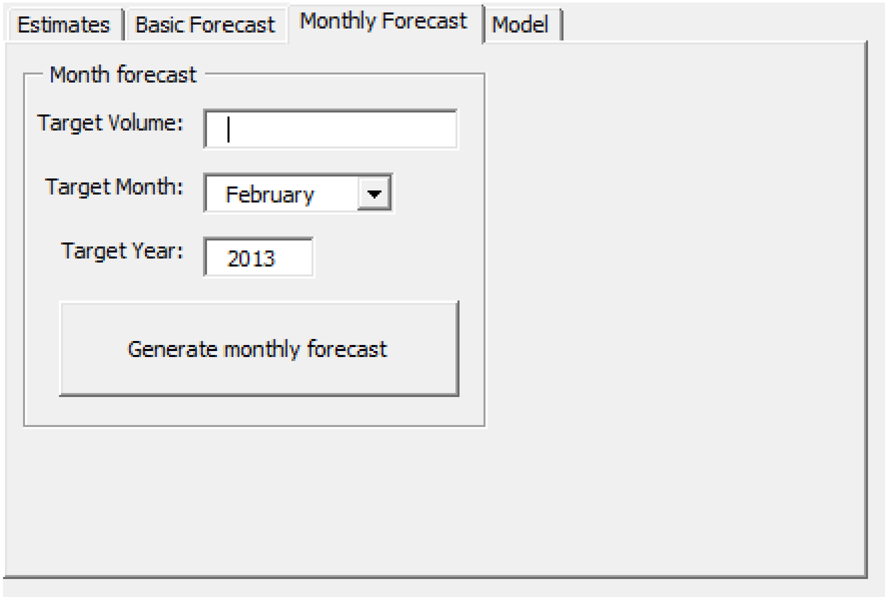

Figure 4.5. *Monthly Forecast section of the application.*

Basic operation of the application is simple. To update the model the user must gather the relevant data from other systems and insert it to the application. Then he chooses the time period for the models, instructs the application to recalculate the model coefficients and then checks that the results look valid. Updating the models is done relatively infrequently. To forecast the user enters the temperature estimates, chooses the forecast time period and finally instructs the application to generate the forecast. (Järvinen 2013)

4.3 The need for accurate loss forecasting

Accurate energy loss forecasting is important so that the hedging levels can be more accurately set. The losses are highest during cold winter days and the electricity price is also at its highest in the power exchange at the same time. The amount of losses in Finland is heavily influenced by the weather due to heating load. During the winter months in 2012 Elenia's network losses were over 30 GWh per month while in June the losses were as low as 13 GWh. During exceptionally cold weather in the Nordic region the spot-price can spike up and DSOs have very little control over the energy loss amounts. During winter 2009-2010 there were three massive price spikes in the 1000-1400 €/MWh range (NordREG 2010). Figure 4.6 shows the Finnish area price during the winter in question. Large open position during such price anomalies can result in significant extra costs.

Finland's own generating capacity is not able to satisfy domestic demand. In February 2011 the peak hourly demand was nearly 15000 MWh while the domestic production was only a bit above 12000 MWh with the difference being covered by import from neighboring countries (Fingrid 2012). This reliance on electricity imports leaves Finland at risk in case of faults or some other unexpected incidents. During winter 2005-2006 there were few such incidents. First the Swedish TSO Svenska Kraftnät abruptly low-

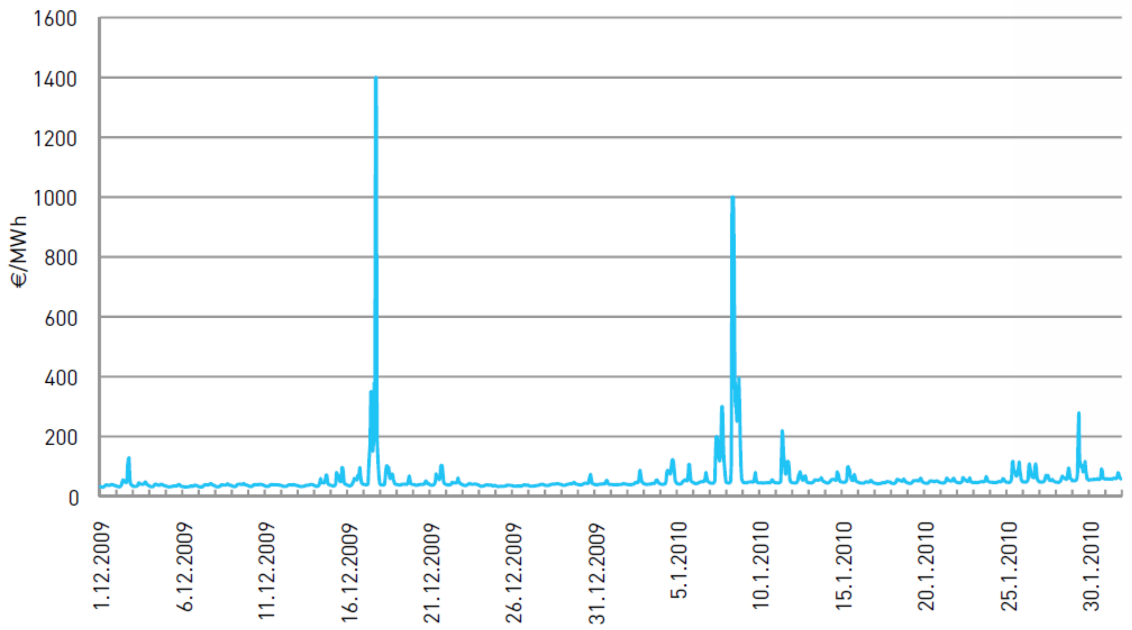

Figure 4.6. *Finnish area price in December 2009 and January 2010 (Fingrid 2010)*

ered its electricity transmission to Finland without prior notice which caused a price spike of 1147 €/MWh during the hour from 16 to 17 on 8th of December 2005. The second event happened during January 19th and 20th when Russia lowered its electricity exports to Finland by a third on a very short notice. The spot price in Finland rose to over 300 €/MWh on 19th and over 200 on 20th. The price of balancing power rose to 1800 €/MWh at its highest. (Energiateollisuus 2006).

Accurate forecasting is also needed for procurement of electricity and not only for hedging purposes. When staging a bidding competition for loss energy procurement the suppliers are very interested in more accurate forecasting of losses. If the losses cannot be forecasted in any reasonable accuracy by the supplier the offers given will have a higher risk margin applied in them which means extra costs for the DSO. The supplier needs to acquire the electricity from Elspot or other sources and poor forecasting exposes the supplier to large open position during consumption peaks.

5 NEW LOSS FORECASTING MODELS

In this section we go over the basic principles of multiple linear regression and analyze the developed forecasting models. Linear regression is a widely used and studied method for statistical inference. It was chosen as the method of choice for this thesis for its relative simplicity, ease of implementation and relative clarity of the results.

5.1 Linear regression

The exposition of multiple linear regression and statistics in this chapter and its subchapters has been adapted from the textbooks by Kutner et al. (2005) and Laininen (2000).

Regression analysis is a statistical method for predicting a response variable based on one or several predictor variables. These variables are also termed as dependent and independent variables respectively. The general linear regression model with *p-1* predictor variables and *n* observed values can be expressed as follows

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \tag{11}$$

Where Y_i is the *i* th observed value of the response variable
β_{p-1} is the *p-1* th regression coefficient
β_0 is the intercept term
$X_{i,p-1}$ is the *i* th value of the *p-1* th predictor variable
ε_i is the value of the *i* th error term

By defining the following matrices

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 X_{11} X_{12} \cdots X_{1,p-1} \\ 1 X_{21} X_{22} \cdots X_{2,p-1} \\ \vdots \; \vdots \; \vdots \; \vdots \\ 1 X_{n1} X_{n2} \cdots X_{n,p-1} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

The general linear regression model (11) can be expressed in matrix notation simply as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{12}$$

The model assumes that the random error terms ε_i have a mean of zero, constant variance σ^2 and that the error terms are uncorrelated.

To find good estimators for the regression coefficients **β** in equation (12) the method of least squares is employed. The least squares method means minimizing the sum of squared deviations between the observed value and the expected value. This means minimizing Q in the following equation

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_{p-1} X_{i,p-1})^2 \tag{13}$$

It can be shown that the least squares estimators, denoted here as **b**, that minimize Q can be obtained by the following equation

$$\boldsymbol{b} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y} \tag{14}$$

According to Gauss-Markov theorem the least squares estimators **b** are unbiased and have minimum variance among all unbiased linear estimators. Furthermore it is usually assumed that the error terms ε_i have normal distribution. Under this assumption the estimators **b** are also the maximum likelihood estimators and also they are consistent and sufficient.

The fitted values can be obtained by

$$\widehat{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{b} \tag{15}$$

and residual terms by

$$\boldsymbol{e} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b} \tag{16}$$

The residuals have an interesting property: the sum of the residuals equals zero. This means that the mean of the residuals is zero and also that the sum of the fitted values is equal to the sum of the observed values when calculated over the base data.

5.1.1 Predictor variables

There are few basic types of variables that can be employed in regression analysis. Quantitative variables are interval scaled numerical variables that can have different values freely. Qualitative variables can represent different things such as gender or day of the week. In regression qualitative variables are usually represented by indicator variables (also called dummy variables). For example in the case of gender the indicator variable can be defined to get the value 1 when the gender is male and 0 if female.

Indicator variables can also be used to represent a qualitative variable with several classes. For example in the case of a weekday variable one would need to use six indicator variables. Generally speaking there has to be one less indicator variable than there are classes in the qualitative variable. This is because if there is an indicator variable for each class then the columns in the predictor value matrix **X** are linearly dependent which leads to the matrix $\boldsymbol{X}^T\boldsymbol{X}$ having columns that are linearly dependent. This means that in equation 14 the inverse cannot be calculated and no unique estimators of the regression coefficients can be found. The class without an indicator variable can be interpreted to be the base case on which the other classes are compared to.

Alternative to using indicator variables when describing qualitative variables is to use a single variable with allocated codes. For example in the case of weekdays one could assign the codes as in Table 5.1

Table 5.1. *Example of code allocation for a qualitative variable.*

Class	**Code**
Monday	1
Tuesday	2
Wednesday	3
Thursday	4
Friday	5
Saturday	6
Sunday	7

The problem when using allocated codes is that the coding implies something about the difference between different classes. For example in the weekday example the coding implies that the difference between Monday and Tuesday is the same as the difference between Friday and Saturday. Using indicator variables instead of allocated code variable avoids this problem of inherent assumptions.

5.1.2 Regression diagnostics

There exist many methods for analyzing regression models. The list of methods used in this thesis is by no means exhaustive. In this thesis the analysis is done mostly by visual methods supplemented by some mathematical methods.

Total sum of squares is a measure of the variance in the observed values. The equation for it is

$$SSTO = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \boldsymbol{Y}^T\boldsymbol{Y} - \left(\frac{1}{n}\right)\boldsymbol{Y}^T\boldsymbol{J}\boldsymbol{Y} \tag{17}$$

Where Y_i is the *i* th observed value of the response variable
$\bar{Y}$ is the mean of the observed values
$\boldsymbol{J}$ is a matrix of appropriate size full of ones

Error sum of squares is a measure of how much the regression line deviates from observed values. The equation for it is

$$SSE = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} e_i^2 = \boldsymbol{e}^T\boldsymbol{e} = \boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{b}^T\boldsymbol{X}^T\boldsymbol{Y} \tag{18}$$

Where $\hat{Y}_i$ is the *i* th fitted value of the response variable
e_i is the value of *i* th residual
$\boldsymbol{b}$ is a vector of the estimated regression coefficients

Coefficient of multiple determination is a measure of how much of the variation in the observed values the regression model explains

$$R^2 = 1 - \frac{SSE}{SSTO} \tag{19}$$

With multiple predictor values the adjusted coefficient of multiple determination is often used. The issue is that when adding more predictor variables to the model R^2 cannot get smaller. In the adjusted version the formula is modified so that each sum of squares is divided by its associated degrees of freedom. With this modification the R^2 value can get smaller if the added variable does not decrease SSE enough to offset losing a degree of freedom. The equation for R^2-adjusted is

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \frac{n-1}{n-p} \cdot \frac{SSE}{SSTO} \tag{20}$$

Where n is the number of observations
p is the number of predictor variables plus one for the intercept term

The variance of residuals is estimated by error mean square which is defined as follows

$$MSE = \frac{SSE}{n-p} \tag{21}$$

The square root of MSE is called standard error which is an estimate of standard deviation.

The predictive capabilities of the model can be evaluated by making a forecast and then looking at mean absolute error and mean absolute percent error which are defined as follows

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \tag{22}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}_i\right|}{Y_i} \cdot 100\% \tag{23}$$

Prediction intervals can also be calculated. The 100(1 - α) % prediction intervals for a future observation is

$$\hat{Y} \pm t(\frac{\alpha}{2}, n-p) \cdot \sqrt{MSE[1 + \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0]} \tag{24}$$

Where $t(\frac{\alpha}{2}, n-p)$ is the 100(1 - α/2)th percentile of t-distribution with $n-p$ degrees of freedom
n is the amount of observations in the base data
p is the number of predictor variables plus one for the intercept term
$\boldsymbol{x}_0$ is a vector of predictor variable values from which the prediction is being made

Multicollinearity

When predictor variables are correlated with each other it is called intercorrelation or multicollinearity is said to exist among them. Often the term multicollinearity is reserved for situations when the correlation is high. If there is a perfect correlation between predictor variables then the columns in the **X** matrix are linearly dependent and as mentioned in chapter 5.1.1 this means that the unique estimators of regression coefficients cannot be calculated. In a model with high multicollinearity the interpretation of the regression coefficients is hard or impossible as the values can vary a lot when new data is introduced to the model.

A widely used formal method for investigating multicollinearity is the use of variance inflation factors

$$(VIF)_k = \frac{1}{1 - R_k^2} \qquad k = 1,2, \dots, p-1 \tag{25}$$

In the equation R_k^2 is the coefficient of determination where the k th predictor variable is acting as the response variable and other variables are used as predictor variables. As a rule of thumb a factor above 10 is considered to be an indication that multicollinearity is influencing the least squares estimates of regression coefficients disproportionately. A factor of 10 means that the other variables in the model can be used to explain 90% of the variation of the *k*th variable.

5.2 Base forecasting models

The forecasting model implemented in the application is not just a single model. The application includes a year-based model for predicting network losses and another year-based model for predicting loss percentages. In this context year-based means that it uses all the data in the defined range to form the regression models. In addition to the year-based models there are models for each individual month for both network loss and network loss percent prediction. For month-based models the regression model is formed by using only the data for the appropriate month. The analysis and investigation of different models was done mostly with the program R (R Core Team 2012).

The base data for all the models is the same. Data input for the application consists of hourly measurements of energy input to the network, energy output of the network and temperature in Jyväskylä. For temperature only a single measurement is used for the whole network. Some investigations were also done with additional measurement points around Elenia’s network but they did not seem improve the forecasting capabilities of the models and actually in some cases they got worse. From the energy input and energy output values the observed network losses and loss percentage are calculated for each hour. The temperature is used to calculate a 48 hour and a 24 hour rolling average where the temperature for each hour is the average of the previous 48 or 24 hours including the hour in question. Using temperature averages makes sense because for example heating loads do not react instantly to temperature. Instead there is a delay depending on the type of residence.

For this thesis the available data included the aforementioned variables for the year 2012. The models were calculated from the data in the time period from 3.1.2012 to 31.12.2012. As more data becomes available it is possible to change this period to include newer data and possibly drop older data out of the models.

The reason for keeping both the year-based and the month-based models is that the year-based model is less sensitive to data quality issues and exceptional weather. For example February in 2012 was quite cold with average temperature in Jyväskylä being -11.2 °C while in 2013 the average temperature was -3.9 °C with the long time average being -8.5 °C. On the other hand month-based models seem to be able to better follow the changes in losses. Another reason for year-based models is that the data quality of early 2012 is slightly in question so they act as a sanity check when making predictions.

5.2.1 Model variables

The predictor variables for all the models are the same except that month-based models do not have the indicator variables for different months. Also while the application is calculating the regression coefficients for each model it will try out both 48 and 24 hour temperature averages and settles on the one which gives a higher R^2-adjusted value. This temperature selection is done every time the models are updated. Table A.1 in Appendix A lists the variables used in the models.

The data is time series data. For the given hour the month and hour variables are 1 or 0 depending when the hour is. Similarly the value of Saturday and Holiday variables is 1 or 0 depending on which day the hour is. Holiday gets the value 1 when the day is either a Sunday or one of the following days when they do not fall on Saturday: New Year's Day, Epiphany, Good Friday, Easter Monday, First of May, Ascension Day, Midsummer's Eve, Independence Day, Christmas Eve, Christmas Day or Boxing Day.

Temperature

Temperature can be considered as the main variable of the model. Temperature is chosen because it has a large impact on the consumption of electrical energy especially in houses with direct electrical heating. Figure 5.1 illustrates the temperature dependency of transmission volume in Elenia's network. The temperature in the chart is Jyväskylä's temperature and distribution volume displays the total energy coming in to the network measured at the connection points. The temperature and distribution volume are rolling 24 hour average values to smooth the curves. The correlation between electricity consumption and temperature is displayed very well during the periods of cold weather.

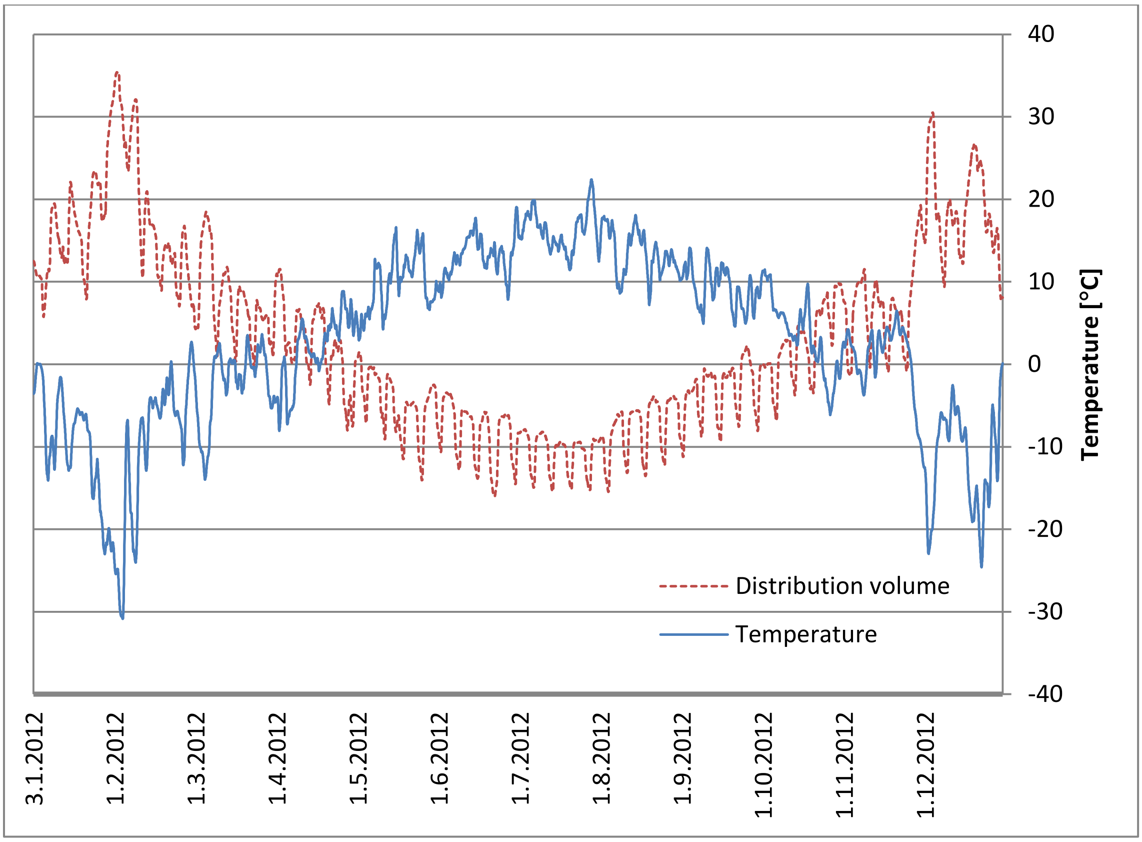

Figure 5.1. *Temperature dependency of distribution volume.*

Temperature forecasts can be expected to be accurate in the 4 to 7 day range. Most of the uncertainty comes from cloudiness which can have a noticeable effect on temperature. During winter cloud cover raises the temperatures and in summer it lowers the temperatures. When considering losses the effect of cloudiness is at its highest during high pressure in winter. (FMI 2013)

5.2.2 Discarded variables

Early on the use of a predictor variable which denoted the proportion of middle-voltage and high-voltage distribution from overall distribution volume was investigated. Large consumers are connected usually straight to 20 kV network and some very large consumers to 110 kV area network. It had been noted that when this proportion was higher the network losses were proportionally lower as the distribution at higher voltages meant less resistive losses. However the variable was eventually discarded as it didn't bring enough extra information to the model. It was found that the value of the proportion variable could be predicted from the other variables with an R^2-adjusted value of over 0.9. Another large problem with it was that it was practically impossible to forecast beforehand which made its use as a predictor variable questionable.

More hour-based variables were also investigated such as a variable for peak hours denoting hours from 8 to 20 on workdays. These variables were discarded as well for not bringing enough new information to the model. Also in an effort to lower the total amount of variables in the model the variables for Sunday and midweek holiday were combined to a single Holiday variable.

While constructing the models all the possible averages from 1 to 48 hours were investigated for temperatures. Some month-based models got the highest R^2-adjusted value with 48 hour or 39 hour averages for example while some others peaked at 22 or 25 hour averages. As a solution the forecasting application now compares 24 and 48 hour temperature averages when updating the models and chooses appropriately for each model every time the models are updated with new data. The values 24 and 48 were chosen instead of arbitrary values because they are easier to understand and more logical.

5.2.3 Year-based models

Year-based models include all the data in the chosen data range. The regression formula can be expressed as follows

$$\begin{aligned} Y = \; & b_0 + b_{Feb}X_{Feb} + b_{Mar}X_{Mar} + b_{Apr}X_{Apr} + b_{May}X_{May} + b_{Jun}X_{Jun} + \\ & b_{Jul}X_{Jul} + b_{Aug}X_{Aug} + b_{Sep}X_{Sep} + b_{Nov}X_{Nov} + b_{Dec}X_{Dec} + b_{H1}X_{H1} + \\ & b_{H2}X_{H2} + b_{H3}X_{H3} + b_{H4}X_{H4} + b_{H5}X_{H5} + b_{H6}X_{H6} + b_{H7}X_{H7} + \\ & b_{H8}X_{H8} + b_{H9}X_{H9} + b_{H10}X_{H10} + b_{H11}X_{H11} + b_{H12}X_{H12} + b_{H13}X_{H13} + \\ & b_{H14}X_{H14} + b_{H15}X_{H15} + b_{H16}X_{H16} + b_{H17}X_{H17} + b_{H18}X_{H18} + \\ & b_{H19}X_{H19} + b_{H20}X_{H20} + b_{H21}X_{H21} + b_{H22}X_{H22} + b_{H23}X_{H23} + \\ & b_{Sat}X_{Sat} + b_{Hol}X_{Hol} + b_{Temp}X_{Temp} + e \end{aligned} \tag{26}$$

The values for the regression coefficients are expressed in Table A.2 in Appendix A for both year-based models. The values for the loss percentage model are multiplied by 100 and thus expressed in percentage units. With 2012 base data for the loss model the temperature average used was the 24 hour average and for the loss percentage model the average used was the 48 hour average as they gave a better R^2-adjusted value as explained in chapter 4.2.1.

As can be seen from the table, some coefficients for the hour indicator variables are close to zero. This means that their statistical significance is low. These are kept in the model because they have a logical reason for being included and after updating the models with newer data the relative values might change. If the model had been reduced to arbitrary hour variables it might not make any sense after updating.

For the loss model the highest VIF value is 5.3 when the temperature average is being predicted from the other variables. This makes sense because temperature is quite dependent on the month of the year. For loss percent model the same VIF value is 5.9. The values are different because the loss percent model uses 48 hour temperature average and loss model uses 24 hour average. These values are still under 10 but they are starting to get in the problematic region when considering multicollinearity.

The residuals are graphed for the year-based loss model in Figure 5.2. Residuals are the difference between observed losses and fitted values. The observed losses are the network loss values that are obtained from loss calculations as explained in chapter 4.2. The true losses are not exactly known due to problems explained in chapter 4.1 however the observed losses can be assumed to be fairly close to true losses. The fitted values are given by the regression formula for the same data that was used as the basis of estimating the regression coefficients. The smaller the residuals are the better the regression model developed can explain the variations in the base data. Residuals are also used to analyze the developed models.

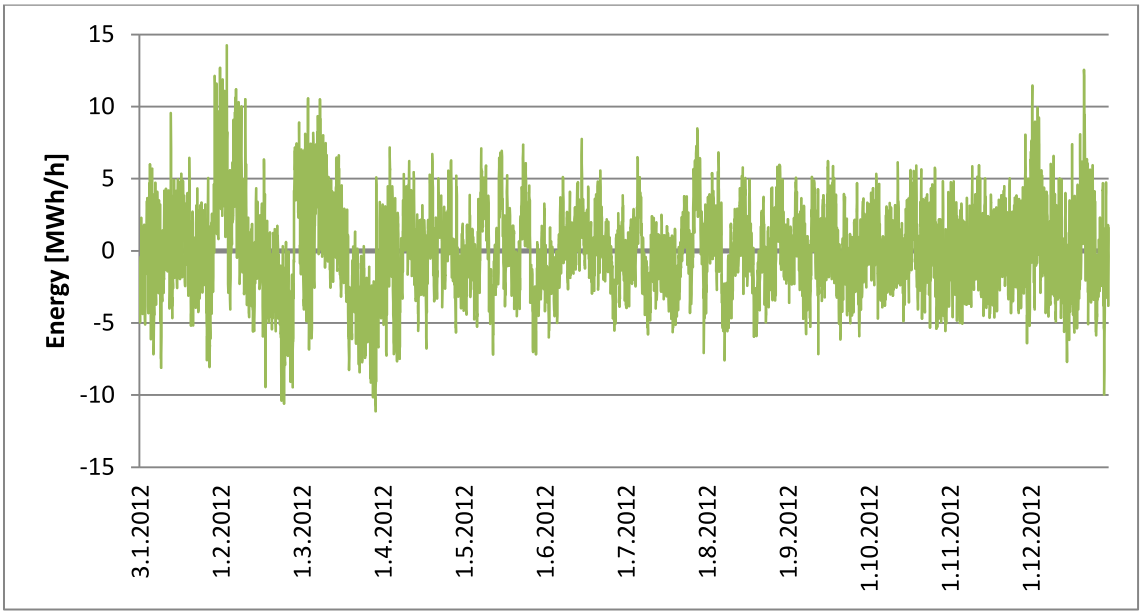

Figure 5.2. *Residuals for the year-based loss model.*

Residuals for February and March seem to display a clear pattern while December which was very cold as well looks better in comparison. As said in chapter 5.2 there are some doubts about the data quality of early 2012 which might also be one of the causes for the pattern. Once data for 2013 becomes available it can be investigated further. The second peak in the residuals for December is during the 22^{nd} and we can conjecture that this peak is caused by Christmas preparations. The reason for the first peak in December is not clear but it coincides with the temperature getting very cold for that time of the year with temperatures falling under -20 °C.

Figure 5.3 shows the residuals graphed for the year-based loss percent model. During March there is similar pattern in the residuals as there was with the loss model in Figure 5.2. During the summer the overall volumes are a lot smaller than during winter which might lead to the loss percentage being more volatile.

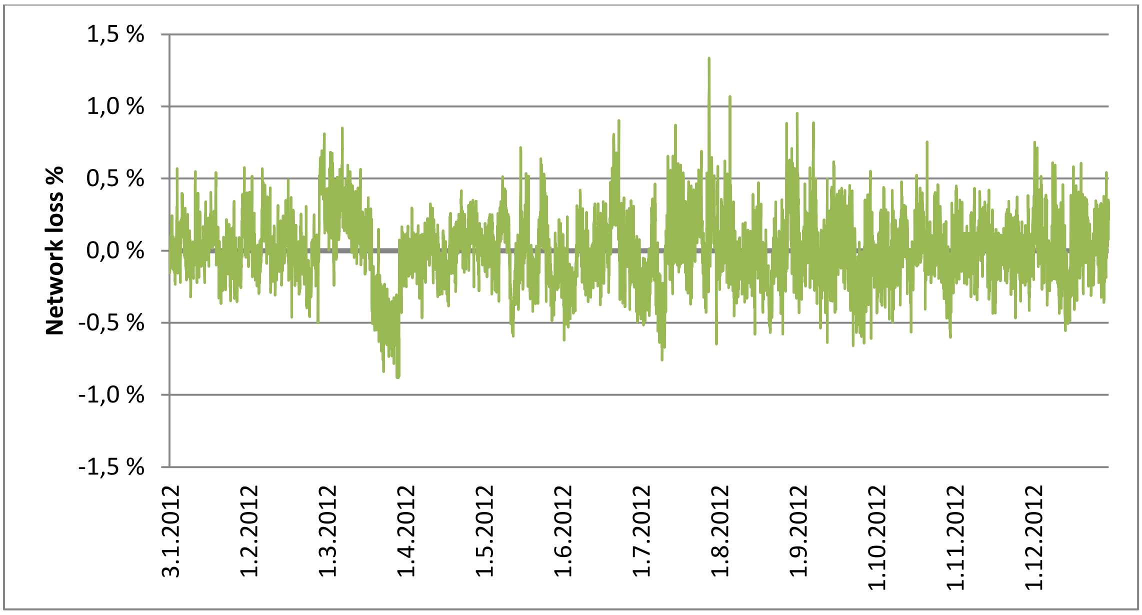

Figure 5.3. *Residuals for the year-based loss percent model.*

Table 5.2 shows the coefficients of determination for the year-based models. The values are fairly high. A high R^2-value means that the regression model can explain a high proportion of the variation in the base data. While having a high R^2-value does not guarantee that good predictions can be made with the model it is a good starting point.

Table 5.2. *Coefficients of determination for year-based models.*

Statistic	Losses as response variable	Loss percent as response variable
R^2	0.9347	0.8265
R^2-adjusted	0.9344	0.8257

Table 5.3 shows the statistics for the residuals of the year-based models. First quartile means that 25% of the values of the residuals are between the minimum value and the quartile value. Respectively third quartile means that 25% of the values are between it and maximum value. In other words the quartiles denote the range in which 50% of the values of the residuals fall between.

Table 5.3. *Statistics for the residuals of the year-based models.*

Residuals	Losses as response variable	Loss percent as response variable
Maximum	14.239	1.334 %
3rd quartile	1.796	0.134 %
Median	-0.057	-0.003 %
1st quartile	-1.995	-0.136 %
Minimum	-11.130	-0.883 %
Standard error	3.014	0.231 %

In Figure 5.4 the histogram for the residuals of the year-based loss model residuals is plotted. In the regression model (equation 12) an assumption is made that the error terms have a mean of zero, constant variance σ^2, are normally distributed and that the error terms are uncorrelated. The residuals of a sound regression model should exhibit similar qualities.

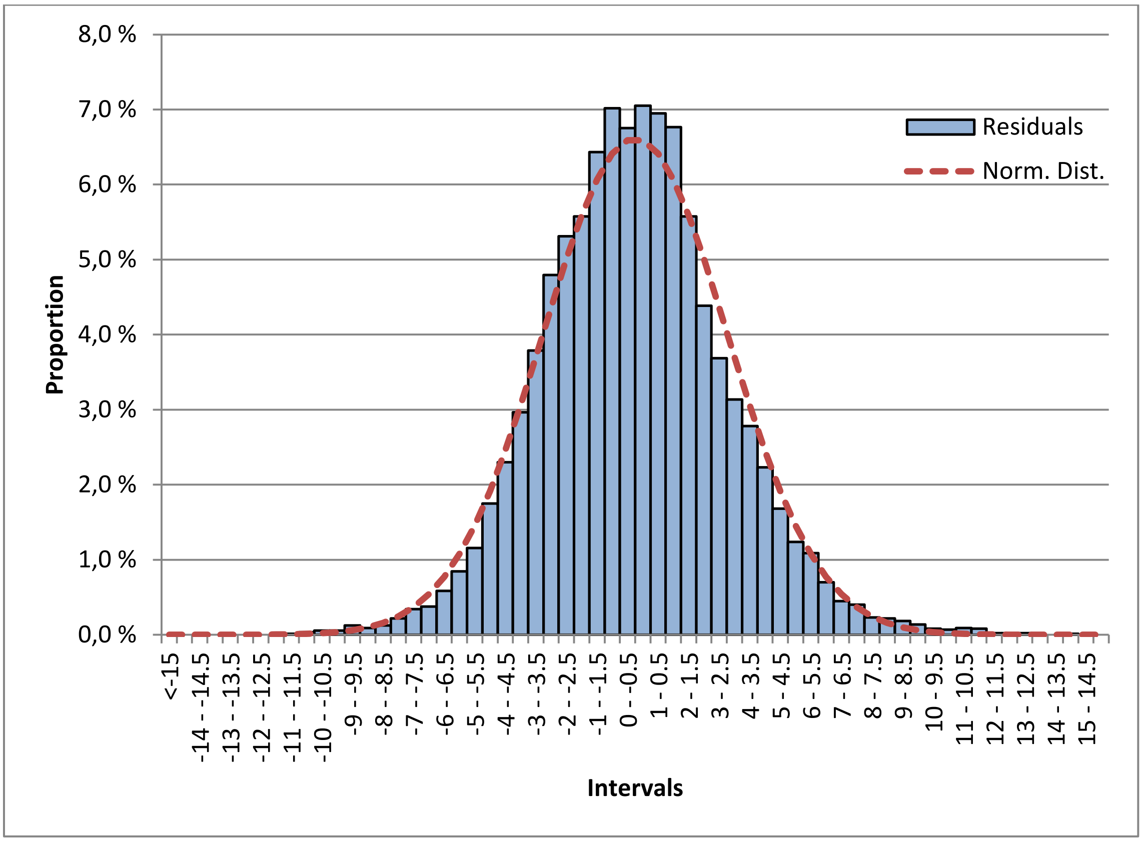

Figure 5.4. *Histogram of residuals together with a normal distribution plot for the year-based loss model.*

Each column is 0.5 MWh wide. For the normal distribution plot the standard error of 3.014 from Table 5.2 is used as the standard deviation. From the chart we can see that the residuals seem to be very close to normally distributed. This gives us assurance that the model is theoretically a proper model.

Similarly to the loss model the histogram for the year-based loss percent model is graphed in Figure 5.5.

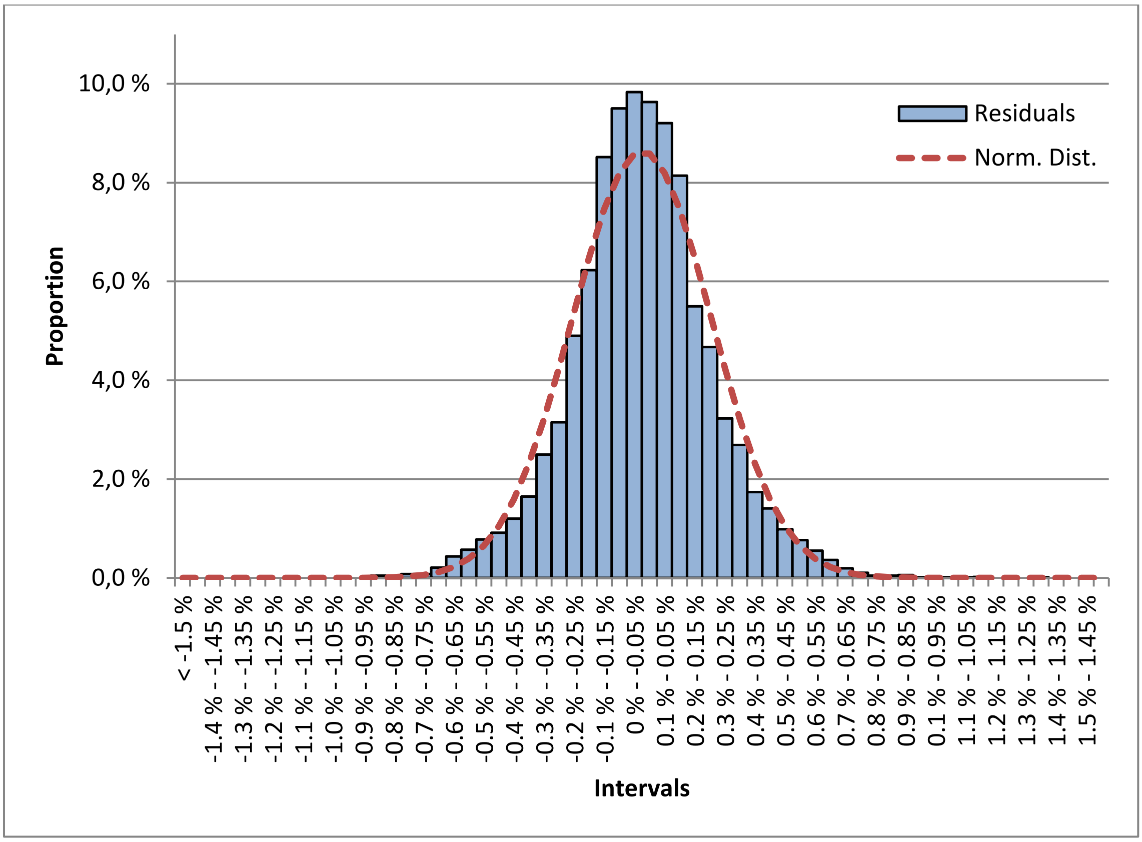

Figure 5.5. *Histogram of residuals together with a normal distribution plot for the year-based loss percent model.*

In this case the columns are 0.05 %-units wide and the normal distribution plot uses a standard deviation of 0.231 %-units. Again the residuals are fairly close to being normally distributed.

5.2.4 Month-based models

Month-based models use the same base data as the year-based models however there are independent models for each month. The regression formula is similar to equation 26 except that there are no variables for the months

$$
\begin{aligned}
Y = \; & b_0 + b_{H1}X_{H1} + b_{H2}X_{H2} + b_{H3}X_{H3} + b_{H4}X_{H4} + b_{H5}X_{H5} + b_{H6}X_{H6} + \\
& b_{H7}X_{H7} + b_{H8}X_{H8} + b_{H9}X_{H9} + b_{H10}X_{H10} + b_{H11}X_{H11} + b_{H12}X_{H12} + \\
& b_{H13}X_{H13} + b_{H14}X_{H14} + b_{H15}X_{H15} + b_{H16}X_{H16} + b_{H17}X_{H17} + \\
& b_{H18}X_{H18} + b_{H19}X_{H19} + b_{H20}X_{H20} + b_{H21}X_{H21} + b_{H22}X_{H22} + \\
& b_{H23}X_{H23} + b_{Sat}X_{Sat} + b_{Hol}X_{Hol} + b_{Temp}X_{Temp} + e
\end{aligned} \qquad (27)
$$

For the month-based models only January will be analyzed here. The regression coefficients are represented in Table A.3 in Appendix A. The loss model uses 24 hour temperature average and the loss percent model uses 48 hour temperature average with the 2012 data.

Similarly to the year-based model there are some coefficients that are close to zero which are kept in the model for consistency. For the month-based models the VIF values are under two for all the variables which is very good. In Figure 5.6 we can see the residuals for all the month-based loss models put together.

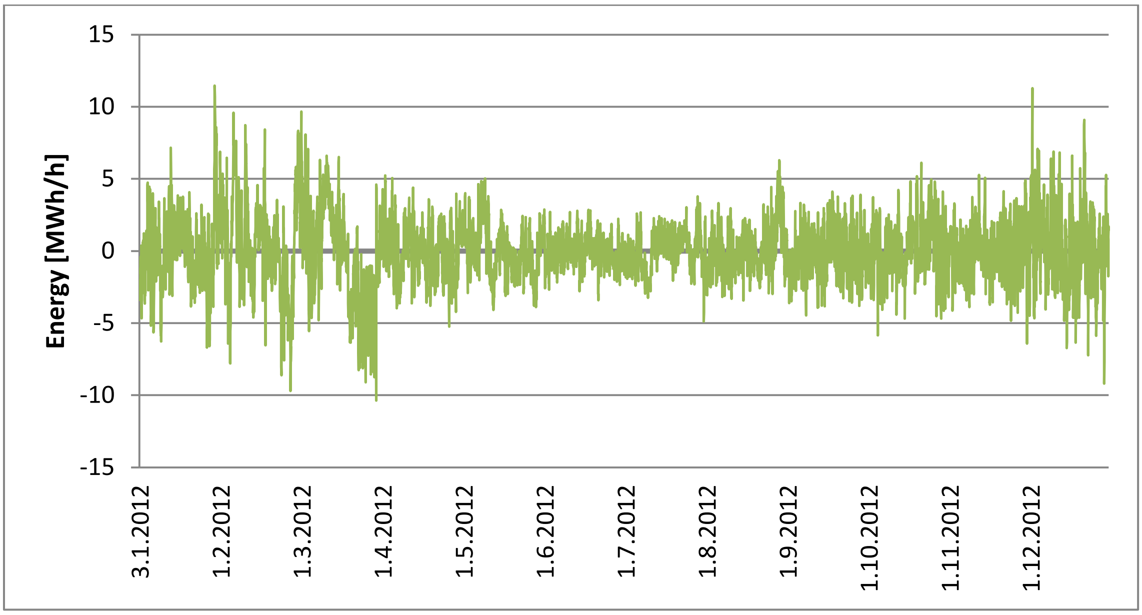

Figure 5.6. *The residuals for the month-based loss models.*

Comparing the statistics for residuals is harder because there is a different month-based model for each month while the year-based model covers the whole year. Visually we can see when compared to the year-based model in Figure 5.2 that overall the residuals have less variance in them. However we can also see that the early 2012 has similar issues with the patterns in residuals.

In Figure 5.7 we can see the residuals for all the month-based loss percent models.

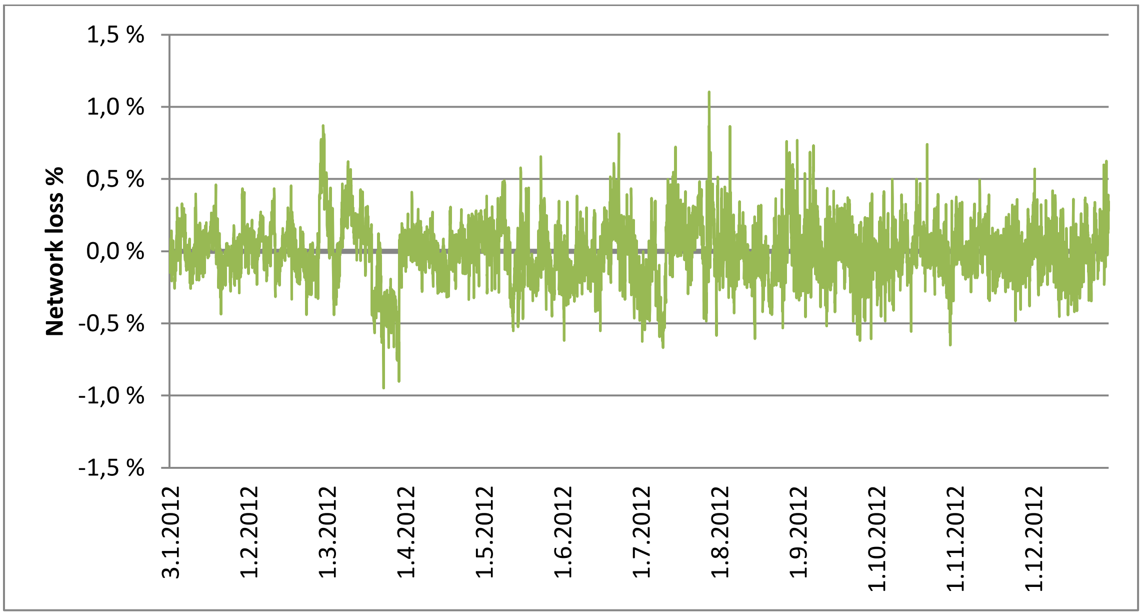

Figure 5.7. *The residuals for the month-based loss percent models.*

Overall the graph and residuals look fairly similar to the year-based model in Figure 5.3 including the pattern in March.

In Table 5.4 we can see the coefficients of determination for the month-based models of January. While the R^2-values are lower than they are for year-based models it doesn't necessarily mean that the models are worse for prediction. The year-based models have more variables in them which increases the R^2-value and with a lot of data points also usually the R^2-adjusted value.

Table 5.4. *Coefficients of determination for month-based models of January.*

Statistic	**Losses as response variable**	**Loss percent as response variable**
R^2	0.8738	0.7761
R^2-adjusted	0.8689	0.7674

In Table 5.5 the statistics for residuals for the January models are displayed. While the R^2-adjusted values are not as high as for the year-based models in Table 5.2 the residual distributions statistics look slightly better although as discussed above the data sets for the models are not equal.

Table 5.5. *Statistics for the residuals of month-based models of January.*

Residuals	**Losses as response variable**	**Loss percent as response variable**
Maximum	11.460	0.460 %
3rd quartile	1.400	0.089 %
Median	-0.032	-0.016 %
1st quartile	-1.747	-0.097 %
Minimum	-6.687	-0.436 %
Standard error	2.592	0.144 %

In Figure 5.8 we can see the residual histogram of the month-based loss model for January plotted together with a normal distribution curve with the standard error of 2.592 from Table 5.5. The residuals seem to be reasonable close to being normally distributed although not quite as nicely as for the year-based model in Figure 5.4. Sample size for the month-based model is naturally a lot smaller which plays a part.

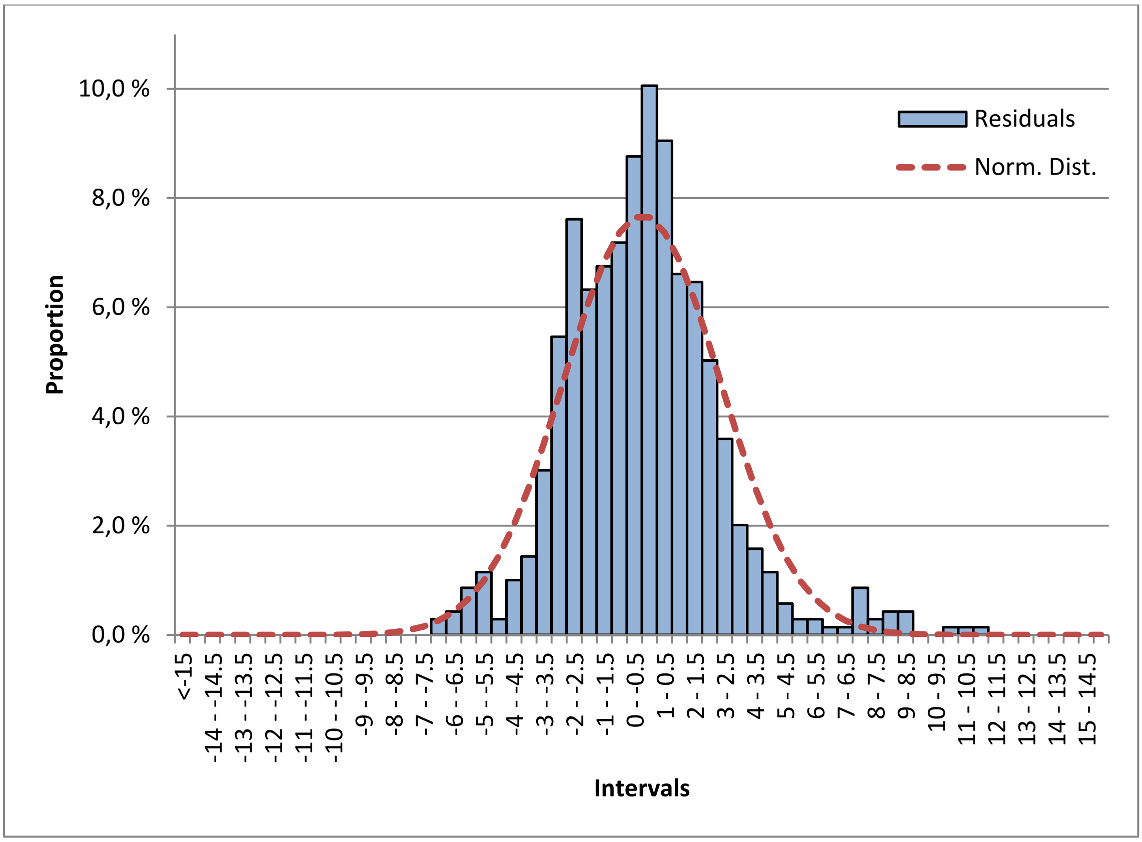

Figure 5.8. *Histogram of residuals together with a normal distribution plot for the month-based loss model of January.*

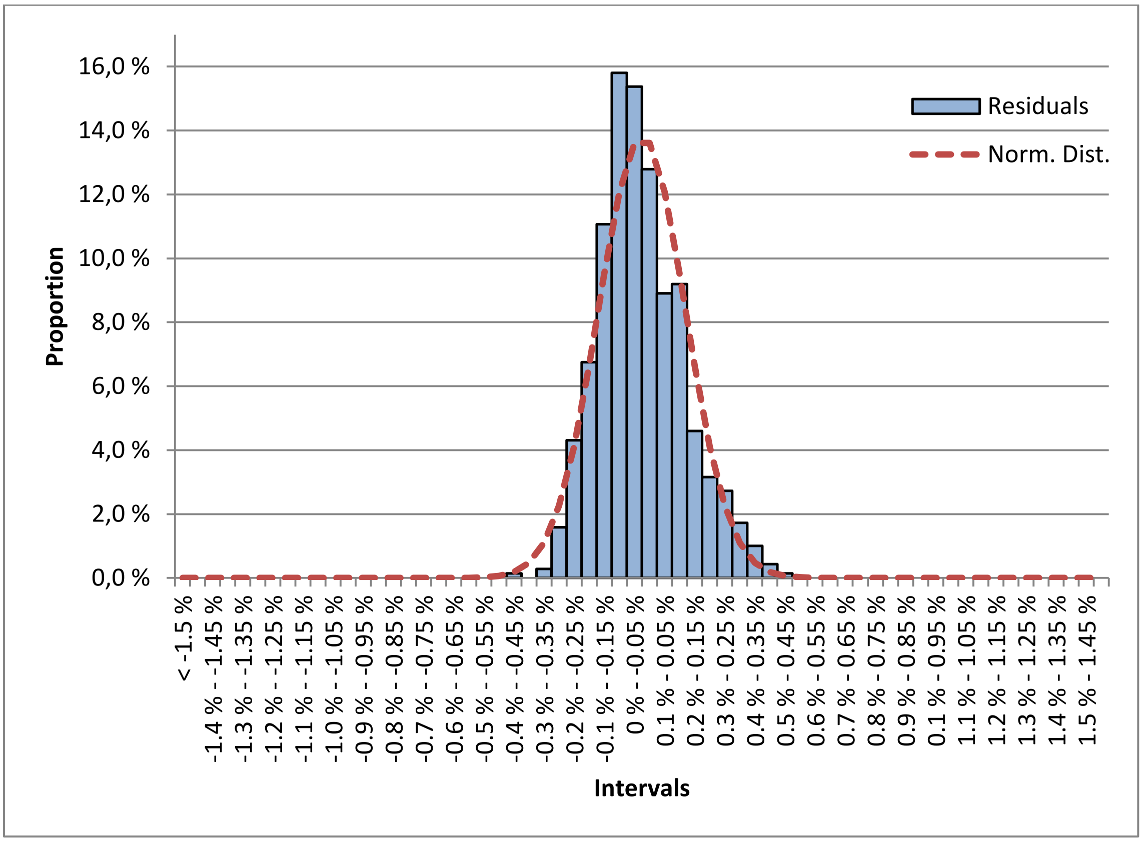

Figure 5.9. *Histogram of residuals together with a normal distribution plot for the month-based loss percent model of January.*

In Figure 5.9 we can see the residual histogram of the month-based loss percent model for January plotted together with a normal distribution curve with the appropriate standard error from Table 5.5. Again the residuals are close to being normally distributed.

5.3 Volume-based monthly loss forecasting

The application has a feature that uses network loss and network loss percent models to forecast losses for a given total distribution volume for a given month (see chapter 4.2.1). The forecast is done by first producing a base forecast for network losses and for network loss percentages for each hour of the given target month. Then the total distribution volume is estimated in the following way

$$E_{estimated} = \sum_{\forall h} \frac{l_h}{p_h} \tag{28}$$

Where $E_{estimated}$ is the estimated distribution volume
h is an hour of the target month
l_h is the forecasted loss for the hour h
p_h is the forecasted loss percent for the hour h (divided by 100)

After the estimated volume is calculated then the forecasted losses for each hour are multiplied by a factor k given by

$$k = \frac{E_{target}}{E_{estimated}} \tag{29}$$

Estimating the total distribution volume from these multiplied losses will equal the desired target volume. The application will calculate the forecasts using all the four combinations of year-based and month-based loss and loss percent models for comparison purposes.

6 EVALUATING THE NEW FORECASTING MODELS WITH JANUARY 2013 DATA

In this section we evaluate the forecasting models constructed in chapter 5 by using two different methods. First we calculate the coefficients for the models using data for whole 2012 plus 2013 January and then with January 2013 data combined with year 2012 data without January. These coefficients are compared to each other and the coefficients obtained in chapter 5 with the data for 2012. The second method used to evaluate the models is to forecast January 2013 losses based on the observed temperatures and comparing predicted losses to the actual observed losses.

6.1 Model coefficient comparisons

6.1.1 Loss model

The first model to be looked at is the loss model. In Figure 6.1 there are the values of coefficients graphed for all the variables in the year-based model with the three different data sets. We can see that the values for the month variables seem to change visibly but the other variables stay fairly constant. Considering most of the data for the year-based model stays the same the changes seem surprisingly large.

The intercept term of the regression model is omitted from figures Figure 6.1 and Figure 6.2 to keep the vertical scales more appropriate. The intercept terms for the year-based and month-based models with different data sets are displayed in Table 6.1. We can see that the intercept term has changed in magnitude by approximately the same amount as the month variable coefficients in Figure 6.1 have changed. The relative values between month variables seem to be fairly similar.

Table 6.1. *Intercept terms for loss models based on different data sets.*

January data set	Year-based model	Month-based model - January
2012	37.948	36.941
2013	35.779	34.118
2012+2013	36.655	35.177

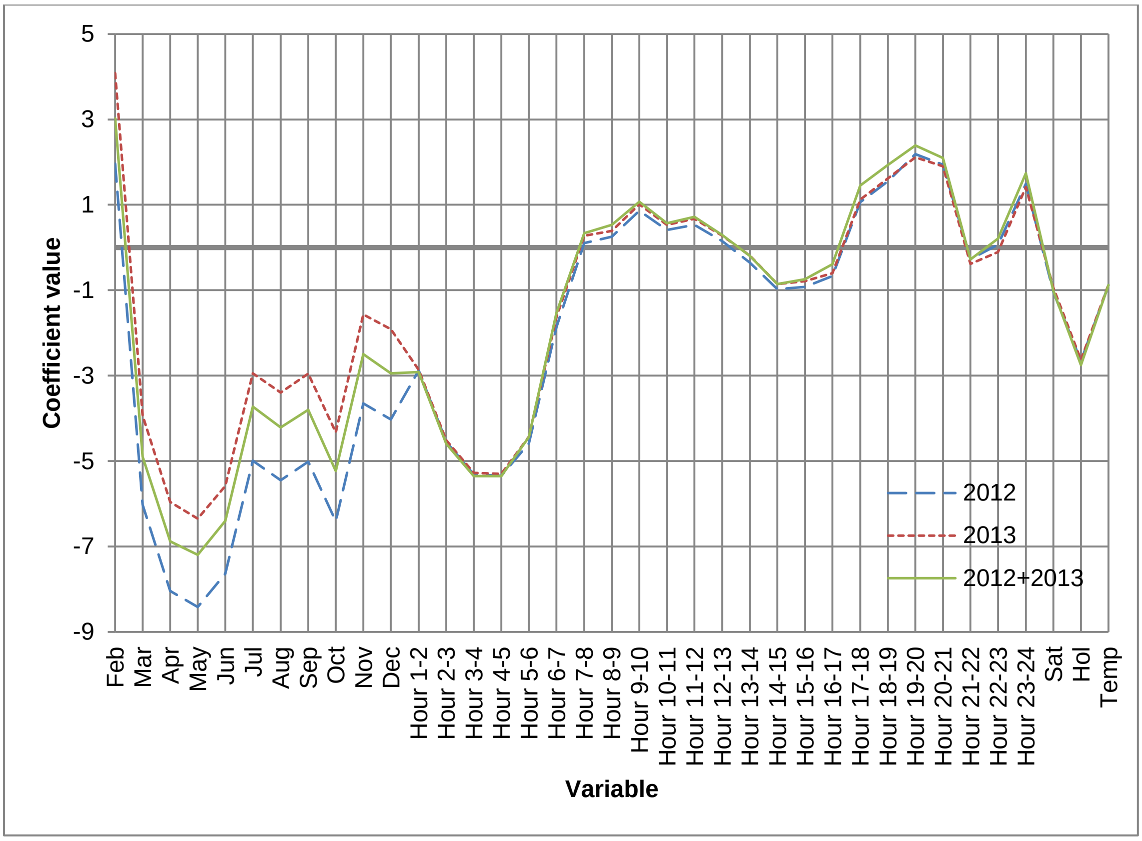

Figure 6.1. *Year-based loss model coefficients with different data sets.*

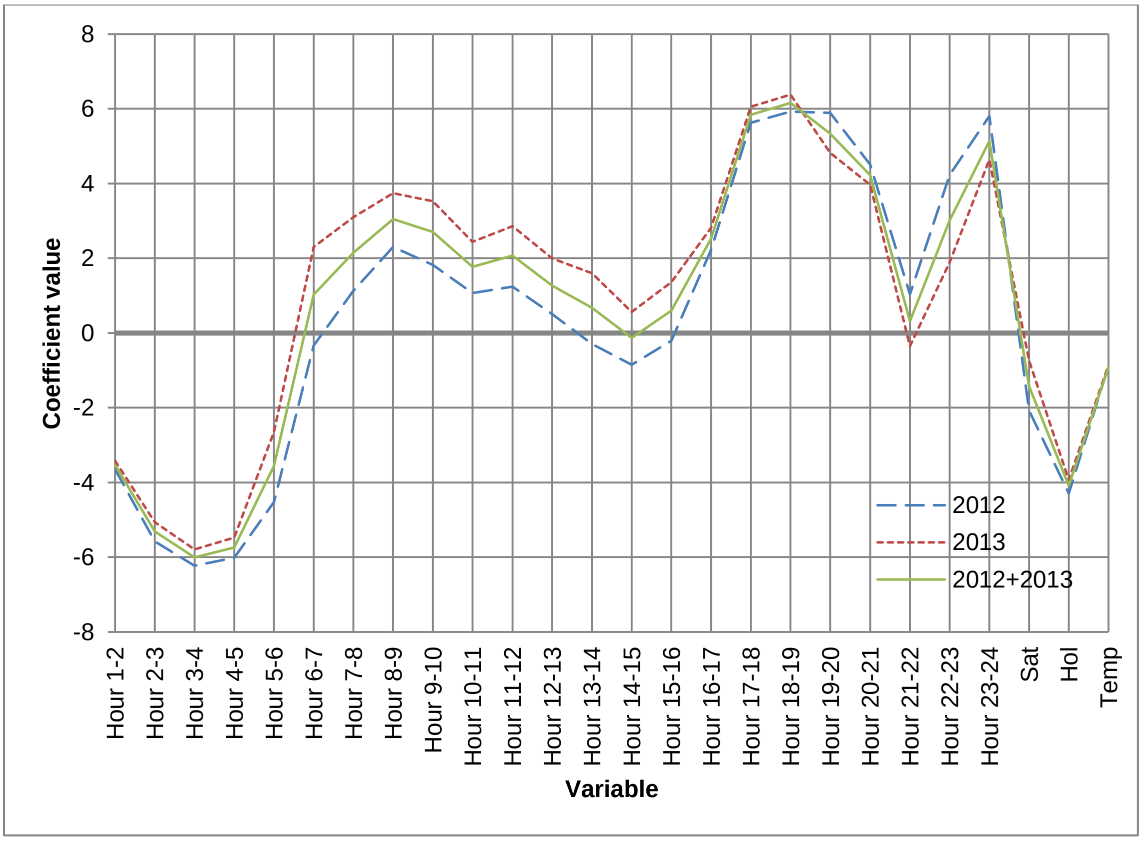

Figure 6.2. *Month-based loss model coefficients for January model with different data sets.*

In Figure 6.2 we can see the coefficient values for the month-based model for January with different data sets. Overall the coefficients and their relative values seem to be fairly similar even with completely different data sets. This gives further confidence in that the model might be able to produce sensible forecasts.

6.1.2 Loss percent model

In Figure 6.3 we can see the coefficients for the year-based loss percent models with different data sets. Similarly to the year-based loss model we can see changes in the values of the month variable magnitudes and intercept term magnitudes. The intercept terms for the loss percent models are shown in Table 6.2.

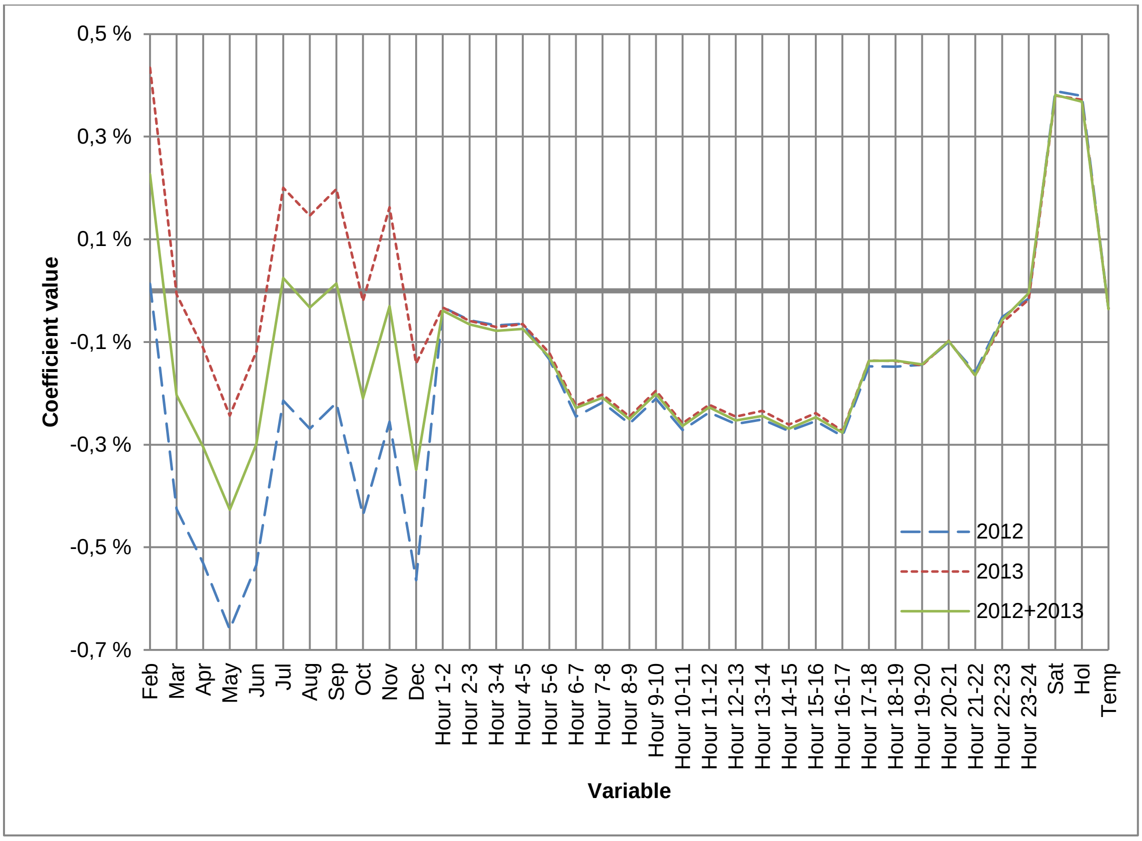

Figure 6.3. *Year-based loss percent model coefficients with different data sets.*

Table 6.2. *Intercept terms for loss percent models with different data sets.*

January data set	**Year-based model**	**Month-based model - January**
2012	4.269 %	4.367 %
2013	3.846 %	3.888 %
2012+2013	4.044 %	4.077 %

In Figure 6.4 we can see the coefficient values for the month-based models for January with different data sets.

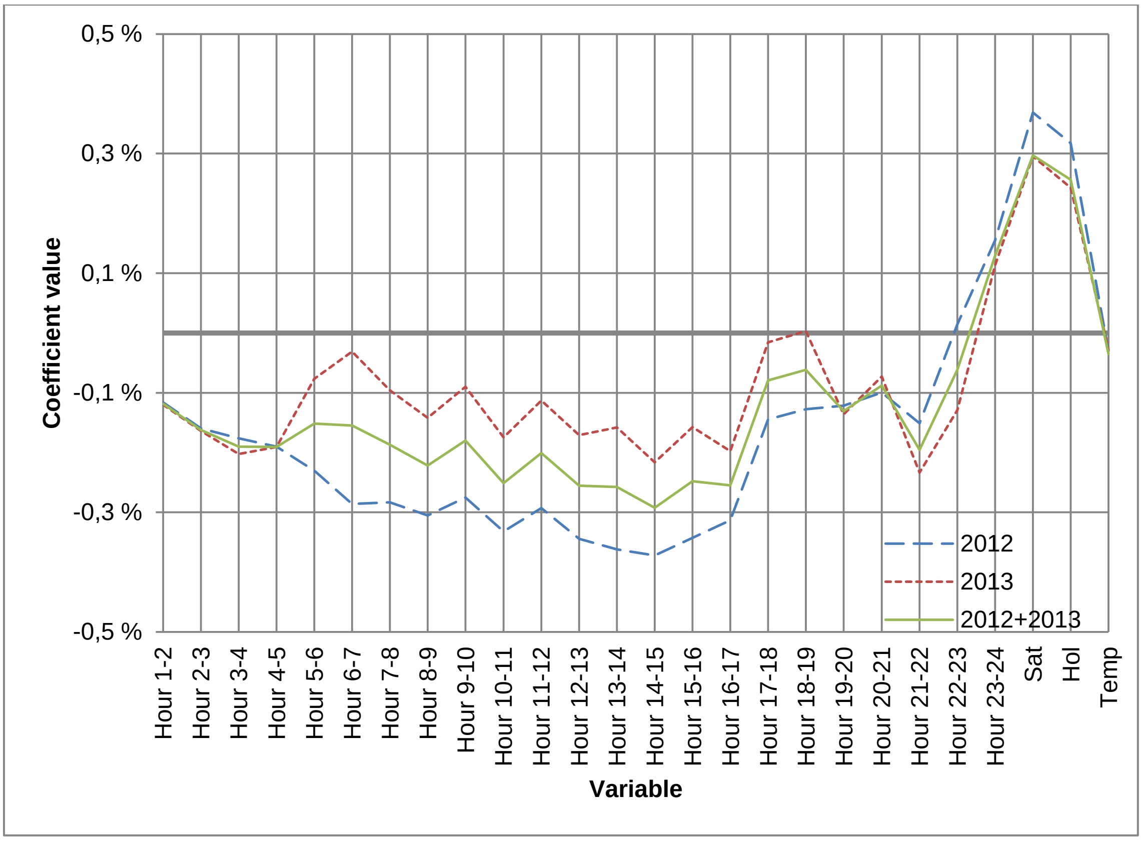

Figure 6.4. *Month-based loss percent model coefficients for January with different data sets.*

The coefficients seem to be fairly similar again although during the day there seems to be a bit of a difference.

6.2 Forecasting January 2013 losses

In this section we look at a forecast made with the application based on 2012 data. Temperature used for making the forecast was the observed temperature of January 2013. Naturally when making a real forecast the temperature estimate won't be so exact but nonetheless it is useful to investigate how well the model can forecast in an ideal situation. Uncertainties of the temperature forecasts are a separate issue that needs to be taken into consideration when making forecasts.

6.2.1 Loss forecast

In Figure 6.5 there are displayed the observed losses, predicted losses made with month-based loss model and the prediction error between predicted and observed losses. Generally speaking the predicted losses seem to follow the trends in the observed losses fairly well. Figure 6.6 shows the same data for a forecast made with the year-based model. It follows the general trend fairly well but the changes are much more moderate. Year-based model has much more trouble following the peaks and lows of the observed

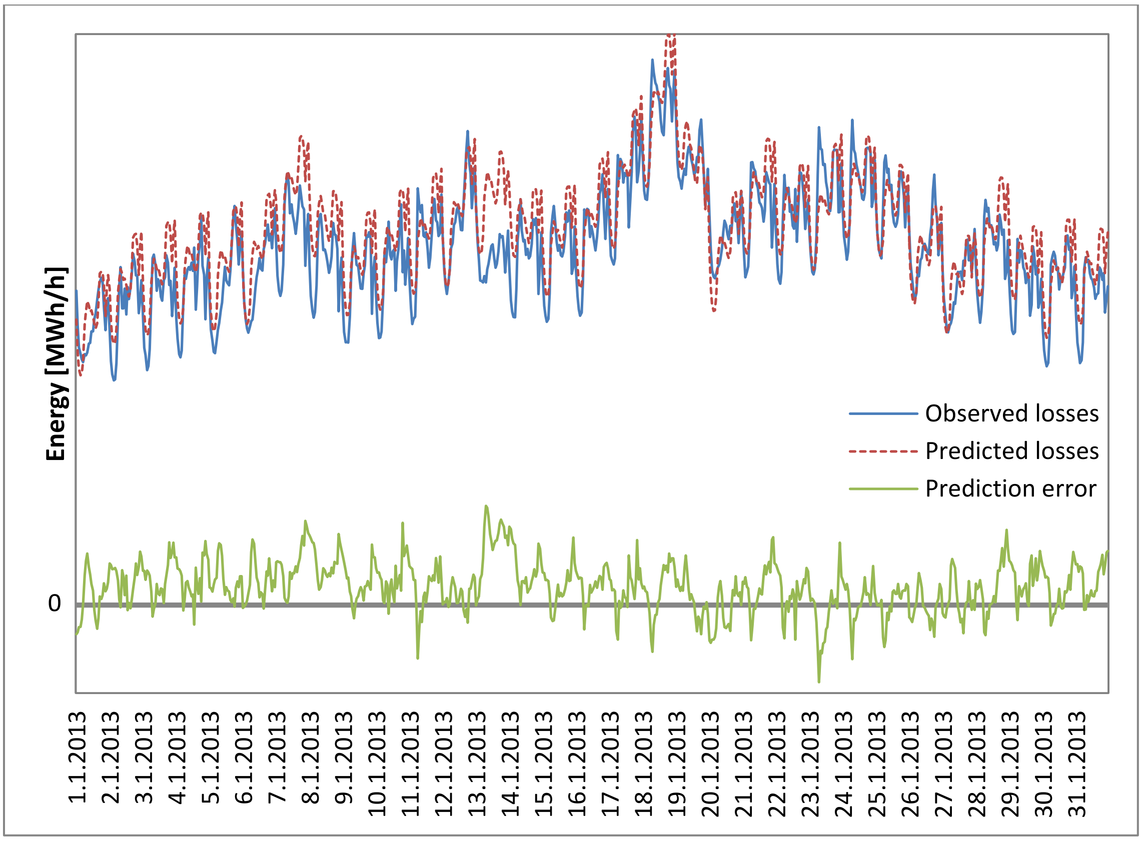

Figure 6.5. *Predicted and observed losses for January using the month-based model.*

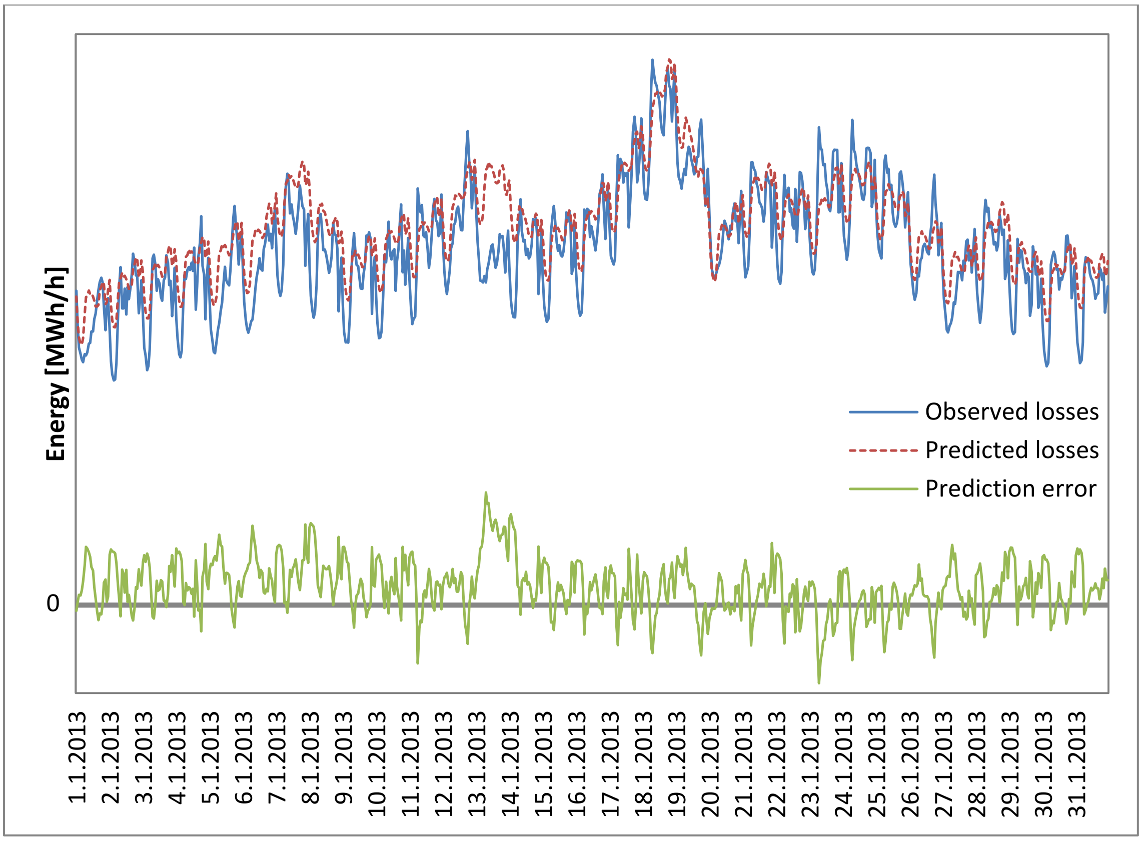

Figure 6.6. *Predicted and observed losses for January using the year-based model.*

losses. In Appendix A there are charts for each week for the month-based forecast. In these charts we can more clearly see the relation between observed and predicted losses. Overall the predicted losses seem to follow the hourly changes and overall magnitude fairly well.

The sums of the predicted losses for the month were 32 GWh for both models while the observed losses were 30.4 GWh. The month-based model for January is based on January 2012 data. The overall losses in the data for January were about 33 GWh. Some data quality issues have been identified and improved upon in the MDMS over time and it has been estimated that the losses for January 2012 might be too high by around 1-2 GWh in the data (Halkilahti 2013). This is meaningful because as was mentioned in chapter 5.1 in the regression model the fitted values sum up to the observed values if the values for the predictor variables are the same. Thus if the base data has higher losses then also the predicted values will be higher. Mean of the prediction error in both forecasts is around 2.1 MWh per hour which would add up to around 1.5 GWh over the month.

Table 6.3 shows the mean absolute error and mean absolute percent error for the prediction errors of the loss predictions in addition to 95% prediction intervals.

Table 6.3*. Statistics for January loss prediction errors.*

	Year-based loss model	**Month-based loss model**
MAE	2.95 MWh	2.84 MWh
MAPE	7.77 %	7.29 %
95% prediction intervals	±5.9 MWh	±5.2 MWh

The error statistics are fairly similar for both models. Month-based model seems to be slightly more accurate despite the data quality issues mentioned earlier. The prediction intervals are fairly large and the data issues discussed above mean that they are not particularly useful.

6.2.2 Loss percent forecast

Figure 6.7 and Figure 6.8 show the observed and predicted loss percent along with the prediction error calculated with month-based and year-based loss percent models. The predicted values have a similar overall trend but the observed values are much more volatile. Similarly to the loss prediction in chapter 6.2.1 the year-based model has smaller changes while maintaining similar average value as the month-based model.

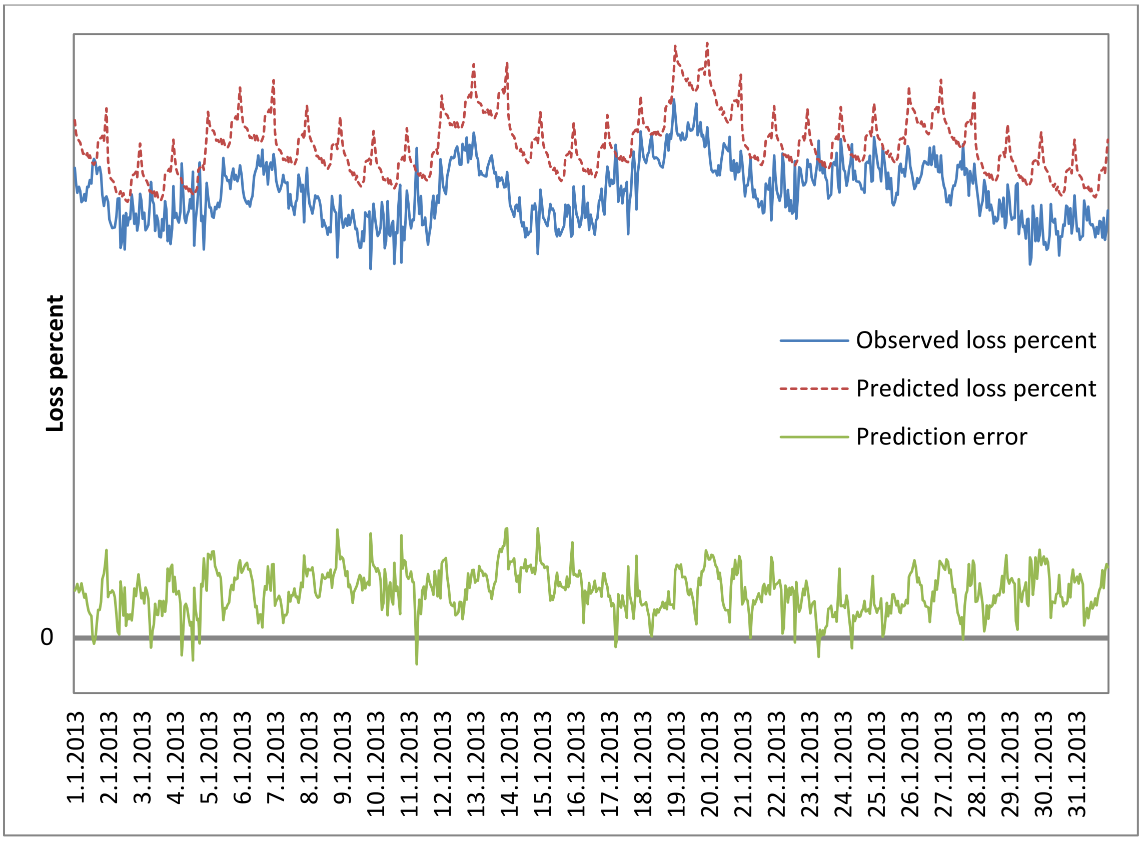

Figure 6.7. *Forecasted loss percentages for January 2013 using the month-based model.*

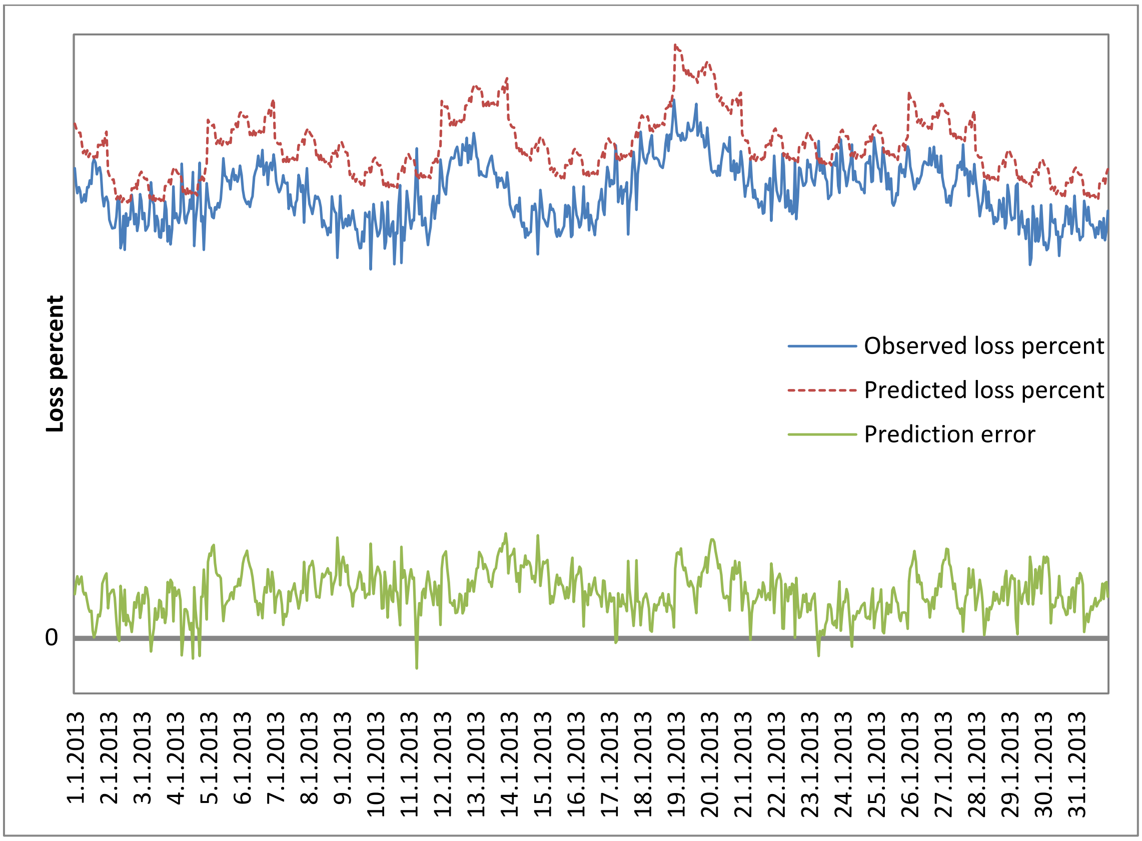

Figure 6.8. *Forecasted loss percentages for January 2013 using the year-based model.*

As discussed in chapter 6.2.1 the losses for January 2012 are slightly higher than they should be. Same goes for the loss percent. The observed loss percent for January 2012 was 0.5 %-units higher than the observed loss percent for January 2013. Table 6.4 shows the mean absolute error and mean absolute percent error for the prediction errors of the loss percent predictions in addition to 95% prediction intervals.

Table 6.4. *Statistics for January loss percent prediction errors.*

	Year-based loss model	**Month-based loss model**
MAE	0.42 %-units	0.43 %-units
MAPE	10.59 %	10.85 %
95% prediction intervals	±0.45 %-units	±0.29 %-units

MAPE is displayed in percentages and the other two in percentage units. Again both models seem to be fairly similar but now the predictions made with the year-based model are slightly better. However the year-based models prediction intervals are larger.

7 INVESTIGATING THE USABILITY OF THE WEEKLY PRODUCT ON THE FINANCIAL MARKET

In this section we take a look at the financial products. First we have a brief overview of hedging in general. Then we investigate the viability of the week future products. With the help of the developed forecasting application one could forecast the network losses for the following week.

7.1 Hedging of network loss procurement in general

When thinking about hedging policies it is good to keep in mind that the profitability of a forward or a future product depends only on the price of the product and the average Elspot system price during the delivery period of the product. The actual losses and their procurement from Elspot market or from a supplier is a different issue. It can be shown with a simple calculation as follows

$$C_{Spot} = \sum_{t=1}^{n} P(t)L(t) \tag{30}$$

Where C_{Spot} is the cost of buying all the loss energy from Elspot
$P(t)$ is the Elspot system price for hour *t*
$L(t)$ are the losses at hour *t*
t is an hour during the delivery period (t = 1…n)

With a financial product for volume *V* acquired for price *D* the total price is

$$\begin{aligned} C_{Total} &= \sum_{t=1}^{n} \{DV + [L(t) - V]P(t)\} \\ &= \underbrace{\sum_{t=1}^{n} P(t)L(t)}_{=C_{Spot}} - V\left[\sum_{t=1}^{n} P(t) - Dn\right] \end{aligned} \tag{31}$$

From the equation above we can deduce that

$$C_{Spot} = C_{Total} \Leftrightarrow D = \frac{1}{n}\sum_{t=1}^{n} P(t) = P_{Average} \tag{32}$$

By combining equations 31 and 32 we can see that if the price *D* of the financial product is higher than $P_{Average}$ then the total costs are higher than having no hedging product at all. Respectively if the price D is lower than $P_{Average}$ then the total cost to acquire the loss energy is lower. Same can be shown for CfDs by replacing the spot price with the difference between the area price and system price which by definition is the price of the CfD.

Naturally when acquiring the financial products the future Elspot prices are not known. Elspot prices can be volatile and difficult to predict. Figure 7.1 displays the monthly system average prices and Finland area prices since 2000. Even on monthly scales there are large spikes during some winters but not during every winter. Over time the separation between Finland's area price and Elspot system price has become more volatile as well.

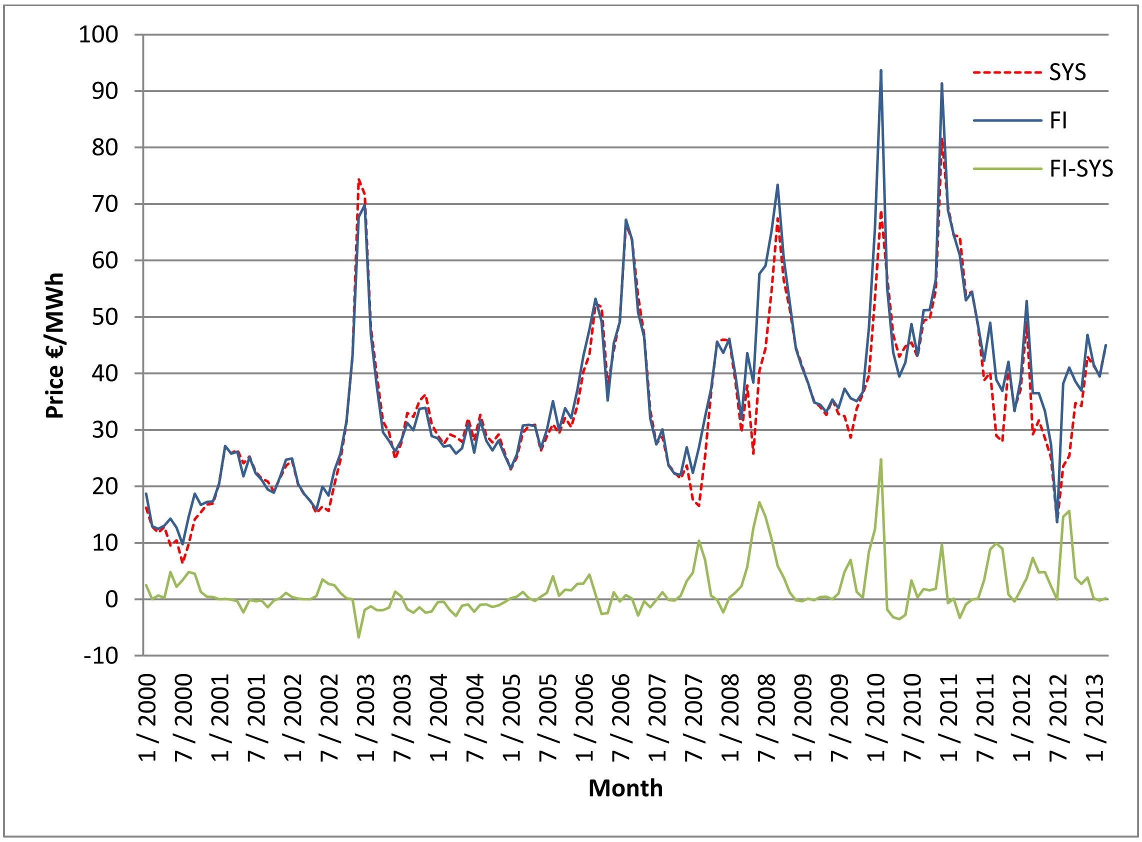

Figure 7.1. *Monthly Elspot system prices, Finland area prices and the differences since 2000.*

There are many factors influencing the exchange prices: for example temperature, hydrologic balance, emissions trading, fuel prices, functioning of the market and general economic state (Karjalainen 2006). These factors combined with practically non-existing elasticity of demand mean that the price of electricity can be very volatile. Volatile prices combined with network loss volumes that can vary by a factor of five from an hour during a summer night to peak hour during a cold winter day propose a challenge for DSOs.

Forecasting losses is the pre-requisite for making appropriate decisions on protection levels. Once loss volumes are predicted a complete hedge can be attained by first hedging the required volume with forward contracts, then hedging against the area price risk by acquiring CfDs for the same time period and volume and finally by procuring the volume of electric energy from the spot market (NASDAQ 2012).

One of the most used basic hedging strategies is to distribute the acquisition of financial products over time according to pre-set hedging levels (Karjalainen 2006). For example Fingrid starts the price hedging of network losses five years before and distributes it over time. Hedging products are evenly purchased according to forecasts so that the following year is fully hedged in the autumn. (Fingrid 2013)

In Finland network losses vary significantly over the duration of a year. The amount of year forwards needs to be tailored so that during the summer when losses are at their lowest the volume risks are not too high. Usually the Elspot prices are at their lowest during this time as well. During July 2012 the average system price was down to 13.7 €/MWh. Similarly the amount of quarter and month forward products needs to be scaled appropriately. Figure 7.2 shows an example of basic hedging with different financial products during a winter week where the temperature is significantly colder during a part of the week. Equivalent amounts of CfD products should be acquired for a full hedge.

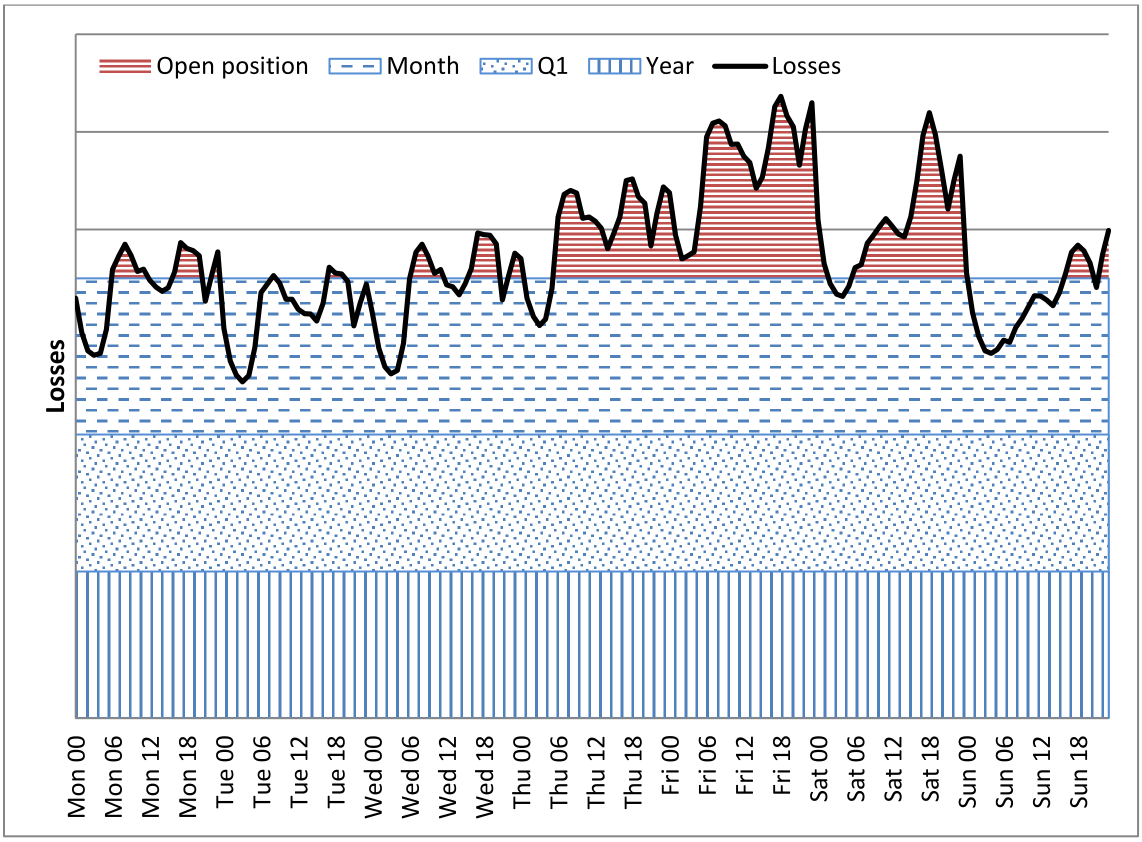

Figure 7.2. *Network loss hedges during a hypothetical winter week.*

Year, quarter and month products form the base of the hedging. Profile risk imposed by varying losses can be seen from the figure. The hedging products are constant while the losses can vary a lot with time. Open position shows the times when the losses are not fully hedged and are exposed to price risk. At other times there is a situation of over hedging which also causes a price risk. Week and day futures provide a possibility for more fine-grained hedging. However the shortest duration for a CfD contract available is a month.

7.2 Weekly products during winter 2012-2013

In this section we investigate the week future products from November to March during winter 2012-2013. With the help of the forecasting tool and accurate weather forecasts the losses for the following week could be estimated. In case of cold weather and higher than expected losses some additional hedging could be done. Week products were chosen for this investigation because weather forecasts become increasingly unreliable after about a week as mentioned in chapter 5.2.1. Table C.1 in Appendix C shows the price data for the weeks during the winter in question.

Table 7.1 shows the calculations based on the price data for a few different scenarios.

Table 7.1. *Week future calculations (unit is €). Profit is positive.*

Week	Case A	Case B	Case C	Case D
44	-3108.00	-2696.40		
45	-1789.20	-1058.40		
46	-16.80	411.60		
47	-2281.44	-1990.80		
48	789.60	1402.80	789.60	1402.80
49	2103.36	2217.60	2103.36	2217.60
50	1209.60	-2284.80	1209.60	-2284.80
51	1352.40	1881.60	1352.40	1881.60
52	579.60	991.20	579.60	991.20
1	-982.80	-1024.80		
2	1512.00	1411.20		
3	-3894.24	-4258.80	-3894.24	-4258.80
4	-3281.04	-2226.00	-3281.04	-2226.00
5	-1221.36	50.40		
6	-1318.80	-1587.60		
7	-105.84	420.00		
8	-38.64	-109.20		
9	949.20	798.00		
10	-277.20	-352.80	-277.20	-352.80
11	1780.80	1260.00	1780.80	1260.00
12	2931.60	1839.60	2931.60	1839.60
13	2847.35	1920.50	2847.35	1920.50
Total	-2259.85	-2985.10	6141.83	2390.90

The total amount bought in each scenario for each week is 5 MW. In case A it is assumed that week futures are bought throughout the final trading week so that they are acquired for the average closing price of the weekly future product for each week ($P_{Average}$ column in Table C.1). In case B the week futures are assumed to be bought during the last trading day for the final closing price each week (P_{Final} column in Table C.1). Case C is similar to case A except that week futures are only bought if the week's average temperature will be below the long time average temperature. Case D is similar to case B with the same restriction as in case C.

In the total row at bottom we can see that categorically buying week futures each week would not have been profitable in either case while buying only when the weather was cold it would have been profitable. It is interesting that it would have been profitable to acquire week futures even during the final trading day when the forecast for the next week can be expected to be reasonably accurate at least when it comes to temperatures. However the profit or loss in each case is only few thousand euros for a reasonable amount of work over the winter especially when trading every day and checking temperature forecasts.

The temperature used is the temperature in Jyväskylä which is located fairly middle of Southern Finland and can be expected to reasonably portray the overall weather in Southern Finland where most of the Finnish consumption is located. The temperatures over the whole Nordic area should be taken into account as well when deciding whether to acquire week futures or not when anticipating cold weather.

8 CONCLUSIONS

The main goal of this thesis was to develop an application for forecasting network losses. With the use of multiple linear regression and Excel VBA programming this goal was achieved in the form of a spreadsheet application. The implemented loss and loss percentage models were analyzed and some predictions were made with the application. The network loss data for the models was hourly measurement data for year 2012 that was aggregated from AMR meters. In addition to loss data the hourly temperatures in Jyväskylä were used.

The loss models and loss percentage models were analyzed by looking at the distributions of residuals. The residuals were fairly close to being normally distributed which means that the basic assumption of linear regression holds and we can conclude that the models used are applicable regression models. In addition January 2013 data was used to calculate new regression coefficients and these matched fairly well to the coefficients obtained with 2012 data. With the models validated a forecast for January 2013 was investigated. When taking some data quality issues into consideration the month-based loss model seemed to be able to predict the network losses with reasonable accuracy. Network loss percent forecast was able to follow with a similar trend but the observed loss percent was much more volatile. When more data becomes available the predictive capabilities can be further investigated.

Overall the month-based loss model seems to be able to reasonably accurately forecast network losses given accurate temperature forecasts. With an accurate forecast the model can be used to predict the following week’s network losses or by utilizing the built-in average temperatures it can be used to forecast month volumes. The network loss percentages seem too volatile to be predicted with such a simple model but the general trend of the loss percentages can be predicted. However the loss percentage prediction seems even more sensitive to issues with the base data.

Main issue with the models used is that there are quite a lot of variables, especially in the year-based models. Also the use of the hour variables in the models mean that the basic hour profile is the same for every day. The values just get scaled depending on temperature and other variables. This can be clearly distinguished from the charts in Appendix A. From experience we know that hour profiles for workdays are quite different from weekends and other special days. One possible option would be to develop models to forecast daily energy loss amounts and then derive the hourly distribution by other means. In addition other statistical methods such as polynomial regression, nonlinear regression and neural networks to mention a few could be investigated. Another possible venue to investigate would be to divide Elenia’s distribution network into different areas for prediction as the network covers a fairly large geographical area and the weather conditions can vary significantly in different portions. However EnergyIP does not currently make this division possible.

REFERENCES

Aalto, J., Karlsson, P., Kaukoranta, J-P., Pirinen, P., Ruuhela, R. & Simola H. Climatological statistics of Finland 1981-2010. Reports 2012:1. Finnish Meteorological Institute, Helsinki

ABB. 2010. Liquid filled transformers. IEC standard small and medium, rated power<2500 kVA, HV ≤36 kV. Brochure. [accessed on 2.1.2013]. Available at: http://www.abb.com

Aro, M., Elovaara, J., Karttunen, M., Nousiainen, K. & Palva, V. 2003. Suurjännitetekniikka. Helsinki, Otatieto. 520 p. (in Finnish).

EEA. 2013. Energy efficiency agreements. Webpage. [accessed on 23.4.2013]. Available at: http://www.energiatehokkuussopimukset.fi/en/

Elovaara, J. & Haarla, L. 2011a. Sähköverkot I: Järjestelmätekniikka ja sähköverkon laskenta. Helsinki, Otatieto. 520 p. (in Finnish).

Elovaara, J. & Haarla, L. 2011b. Sähköverkot II: Verkon suunnittelu, järjestelmät ja laitteet. Helsinki, Otatieto. 550 p. (in Finnish).

EMV. 2010. Liittymismaksujen ja siirtohäviöihin kuluvan energian hankinnasta syntyneiden kustannusten käsittely valvontamallissa. Consultation report by Pöyry Management Consulting Oy. (in Finnish).

EMV. 2011. Sähkön jakeluverkkotoiminnan ja suurjännitteisen jakeluverkkotoiminnan hinnoittelun kohtuullisuuden valvontamenetelmien suuntaviivat vuosille 2012 – 2015. Document. (in Finnish).

EMV. 2012. Sähköverkkotoiminnan tunnusluvut vuodelta 2011. Statistics compilation. [accessed on 4.1.2013]. Available at: http://www.energiamarkkinavirasto.fi/ (in Finnish).

Energiateollisuus. 2006. Sähkön kulutushuiput tammikuussa 2006. Report.

EU. 2012. Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency.

Fingrid. 2010. Fingrid Oy customer magazine 1/2010. (in Finnish).

Fingrid. 2012. Vuosikertomus 2011. Annual report. (in Finnish).

Fingrid. 2013. Loss energy. Webpage. [accessed on 2.5.2013]. Available at: http://www.fingrid.fi/en/powersystem/loss%20energy/Pages/default.aspx

FMI. 2013. Finnish Meteorological Institute. Webpage. [accessed on 15.4.2013]. Available at: http://ilmatieteenlaitos.fi/ (in Finnish).

Government. 2010. Valtioneuvoston periaatepäätös energiatehokkuustoimenpiteistä. Decision in principle. (in Finnish).

Government. 2013a. Hallituksen esitys eduskunnalle sähkö- ja maakaasumarkkinoita koskevaksi lainsäädännöksi. Legislative proposal. (in Finnish).

Government. 2013b. National Energy and Climate Strategy. Government Report to Parliament on 20 March 2013.

Halkilahti, M. 2013. Process Developer, Elenia Oy. Tampere. Various conversations during 2012-2013.

Itäpää, Aila. 1979. Sähkönjakeluverkon teho- ja energiahäviöt suunnitteluun vaikuttavana tekijänä. Master of Science Thesis. Tampere University of Technology. 109 p. (in Finnish).

Järvinen, M. 2013. Network loss forecasting tool guide. Internal document.

Karjalainen R-M. 2006. Sähkökaupan riskit ja riskienhallinta. Master of Science Thesis. Lappeenranta University of Technology. 98 p. (in Finnish).

Kinnunen, M. 2002. Sähkönjakeluverkon tuntihäviöiden ja markkinaperusteisen häviökustannuksen mallinnus. Licentiate Thesis. Helsinki University of Technology. 156 p. (in Finnish).

Koivisto, M. 2010. Tuntimittausdatan käyttö sähkökuorman ennustamisessa. Master of Science Thesis. Aalto University. 87 p. (in Finnish).

Kuisma, K. 2008. Sähköverkon häviöiden mallintaminen ja häviösähkön hankinta. Master of Science Thesis. Tampere University of Technology. 85 p. (in Finnish).

Kutner, M., Li, W., Nachtsheim, C. & Neter, J. 2005. Applied Linear Statistical Models. 5th edition. New York, McGraw-Hill. 1396 p.

L 17.3.1995/386. Sähkömarkkinalaki. (in Finnish).

Laininen, P. 2000. Tilastollisen analyysin perusteet. 5th edition. Helsinki, Otatieto. 281 p. (in Finnish).

Mutanen, A., Repo, S. & Järventausta, P. 2011a. Customer Classification and Load Profiling Based on AMR Measurements. 21st International Conference on Electricity Distribution, Frankfurt, 6-9 June. CIRED. 4p

Mutanen, A., Ruska, M., Repo, S. & Järventausta, P. 2011b. Customer Classification and Load Profiling Method for Distribution Systems. IEEE Transactions on Power Delivery, vol. 26, No. 3. July 2011.

Mäkelä, P. 2011. New business and process development opportunities utilizing meter data management system. Master of Science Thesis. Tampere University of Technology. 78 p. (in Finnish).

NASDAQ. 2012. Trade at NASDAQ OMX Commodities Europe's Financial Market. Document.

Nord Pool Spot. 2011. Annual Report 2011.

Nord Pool Spot. 2012a. The Nordic Electricity Exchange and the Nordic Model for a Liberalized Electricity Market. Document.

Nord Pool Spot. 2012b. Webpage. [accessed on 21.12.2012]. Available at: http://www.nordpoolspot.com/

NordREG. 2010. NordREG memo on price peaks in the Nord Pool Spot market

Nousiainen, K. 2007. Magneettipiirit ja muuntajat. Study materials for course SVT-3300 Muuntajat ja sähkökoneet. Tampere University of Technology. (in Finnish).

Paloposki, Juho. 1999. Taajamasähköverkon jakeluhäviöiden vähentäminen. Master of Science Thesis. Helsinki University of Technology. 68 p. (in Finnish).

Partanen, J., Viljainen, S., Lassila, J., Honkapuro, S., Tahvanainen, K., Karjalainen, R., Annala, S. & Makkonen, M. 2012. Sähkömarkkinat. Course material for Electricity Markets course. Lappeenranta University of Technology. [accessed on 21.12.2012]. Available at: https://noppa.lut.fi/noppa/opintojakso/bl20a0400/materiaali (in Finnish).

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Seppälä, A., Etula, J.-H., Forsman, S. & Högberg, J. 2011. Real Time Purchase and Settlement of Distribution Losses. 21st International Conference on Electricity Distribution, Frankfurt, 6-9 June. CIRED. 5 p.

Seppälä, A. & Trygg, P. 2011. Sähkönjakeluverkon häviösäästöjen laskennan ja raportoinnin kehittäminen ja yhdenmukaistaminen. Report. (in Finnish).

Sievi, A. 2013. Service Manager, Energy Data Management, Elenia Oy. Tampere. Discussions Spring 2013.

Tyynismaa, P. 2003. Sähkönjakeluverkon häviöiden määrittäminen. Master of Science Thesis. Helsinki University of Technology. 71 p. (in Finnish).

APPENDIX A: REGRESSION VARIABLES AND COEFFICIENTS

Table A.1. *Predictor variables used in the models.*

Variable	Variable type	Symbol
February	Indicator	X_{Feb}
March	Indicator	X_{Mar}
April	Indicator	X_{Apr}
May	Indicator	X_{May}
June	Indicator	X_{Jun}
July	Indicator	X_{Jul}
August	Indicator	X_{Aug}
September	Indicator	X_{Sep}
November	Indicator	X_{Nov}
December	Indicator	X_{Dec}
Hour 1-2	Indicator	X_{H1}
Hour 2-3	Indicator	X_{H2}
Hour 3-4	Indicator	X_{H3}
Hour 4-5	Indicator	X_{H4}
Hour 5-6	Indicator	X_{H5}
Hour 6-7	Indicator	X_{H6}
Hour 7-8	Indicator	X_{H7}
Hour 8-9	Indicator	X_{H8}
Hour 9-10	Indicator	X_{H9}
Hour 10-11	Indicator	X_{H10}
Hour 11-12	Indicator	X_{H11}
Hour 12-13	Indicator	X_{H12}
Hour 13-14	Indicator	X_{H13}
Hour 14-15	Indicator	X_{H14}
Hour 15-16	Indicator	X_{H15}
Hour 16-17	Indicator	X_{H16}
Hour 17-18	Indicator	X_{H17}
Hour 18-19	Indicator	X_{H18}
Hour 19-20	Indicator	X_{H19}
Hour 20-21	Indicator	X_{H20}
Hour 21-22	Indicator	X_{H21}
Hour 22-23	Indicator	X_{H22}
Hour 23-24	Indicator	X_{H23}
Saturday	Indicator	X_{Sat}
Holiday	Indicator	X_{Hol}
Temperature average	Quantitative	X_{Temp}

Table A.2. *Regression coefficients for the year-based models.*

Coefficient	Losses as response variable	Std. Error	Loss percent as response variable	Std. Error
b_0	37.948	0.204	4.269 %	0.0157 %
b_{Feb}	1.960	0.162	0.013 %	0.0125 %
b_{Mar}	-6.028	0.168	-0.426 %	0.0128 %
b_{Apr}	-8.043	0.177	-0.531 %	0.0136 %
b_{May}	-8.412	0.208	-0.660 %	0.0162 %
b_{Jun}	-7.634	0.223	-0.534 %	0.0175 %
b_{Jul}	-4.988	0.242	-0.214 %	0.0190 %
b_{Aug}	-5.449	0.228	-0.269 %	0.0180 %
b_{Sep}	-5.010	0.208	-0.219 %	0.0163 %
b_{Oct}	-6.403	0.185	-0.437 %	0.0144 %
b_{Nov}	-3.651	0.177	-0.255 %	0.0137 %
b_{Dec}	-4.024	0.160	-0.564 %	0.0123 %
b_{H1}	-2.875	0.223	-0.032 %	0.0171 %
b_{H2}	-4.550	0.223	-0.058 %	0.0171 %
b_{H3}	-5.310	0.223	-0.068 %	0.0171 %
b_{H4}	-5.340	0.223	-0.064 %	0.0171 %
b_{H5}	-4.584	0.223	-0.134 %	0.0171 %
b_{H6}	-1.863	0.223	-0.246 %	0.0171 %
b_{H7}	0.103	0.223	-0.218 %	0.0171 %
b_{H8}	0.254	0.223	-0.259 %	0.0171 %
b_{H9}	0.858	0.223	-0.211 %	0.0171 %
b_{H10}	0.412	0.223	-0.271 %	0.0171 %
b_{H11}	0.535	0.223	-0.237 %	0.0171 %
b_{H12}	0.149	0.223	-0.260 %	0.0171 %
b_{H13}	-0.349	0.223	-0.251 %	0.0171 %
b_{H14}	-0.977	0.223	-0.274 %	0.0171 %
b_{H15}	-0.923	0.223	-0.254 %	0.0171 %
b_{H16}	-0.666	0.223	-0.283 %	0.0171 %
b_{H17}	1.061	0.223	-0.147 %	0.0171 %
b_{H18}	1.553	0.223	-0.148 %	0.0171 %
b_{H19}	2.184	0.223	-0.144 %	0.0171 %
b_{H20}	1.938	0.223	-0.100 %	0.0171 %
b_{H21}	-0.269	0.223	-0.159 %	0.0171 %
b_{H22}	0.069	0.223	-0.051 %	0.0171 %
b_{H23}	1.490	0.223	-0.015 %	0.0171 %
b_{Sat}	-1.044	0.094	0.389 %	0.0072 %
b_{Hol}	-2.657	0.088	0.380 %	0.0067 %
b_{Temp}	-0.874	0.007	-0.035 %	0.00057 %

Table A.3. *Regression coefficients for the month-based models of January.*

Coefficient	**Losses as response variable**	**Std. Error**	**Loss percent as response variable**	**Std. Error**
b_0	36.941	0.506	4.367 %	0.0283 %
b_{H1}	-3.648	0.681	-0.116 %	0.0377 %
b_{H2}	-5.573	0.681	-0.160 %	0.0377 %
b_{H3}	-6.228	0.681	-0.176 %	0.0377 %
b_{H4}	-6.013	0.681	-0.190 %	0.0377 %
b_{H5}	-4.527	0.681	-0.230 %	0.0377 %
b_{H6}	-0.335	0.681	-0.286 %	0.0377 %
b_{H7}	1.131	0.681	-0.283 %	0.0377 %
b_{H8}	2.304	0.681	-0.305 %	0.0377 %
b_{H9}	1.821	0.681	-0.275 %	0.0377 %
b_{H10}	1.065	0.681	-0.332 %	0.0377 %
b_{H11}	1.237	0.681	-0.293 %	0.0377 %
b_{H12}	0.502	0.681	-0.344 %	0.0377 %
b_{H13}	-0.295	0.681	-0.362 %	0.0377 %
b_{H14}	-0.853	0.681	-0.372 %	0.0377 %
b_{H15}	-0.210	0.681	-0.343 %	0.0377 %
b_{H16}	2.223	0.681	-0.313 %	0.0377 %
b_{H17}	5.624	0.681	-0.145 %	0.0377 %
b_{H18}	5.924	0.681	-0.128 %	0.0377 %
b_{H19}	5.894	0.681	-0.121 %	0.0377 %
b_{H20}	4.504	0.681	-0.100 %	0.0377 %
b_{H21}	1.037	0.681	-0.151 %	0.0377 %
b_{H22}	4.230	0.681	0.014 %	0.0377 %
b_{H23}	5.803	0.681	0.155 %	0.0377 %
b_{Sat}	-2.075	0.290	0.369 %	0.0162 %
b_{Hol}	-4.290	0.266	0.317 %	0.0147 %
b_{Temp}	-0.881	0.016	-0.030 %	0.00095 %

APPENDIX B: WEEKLY CHARTS FOR MONTH-BASED LOSS FORECAST FOR JANUARY 2013

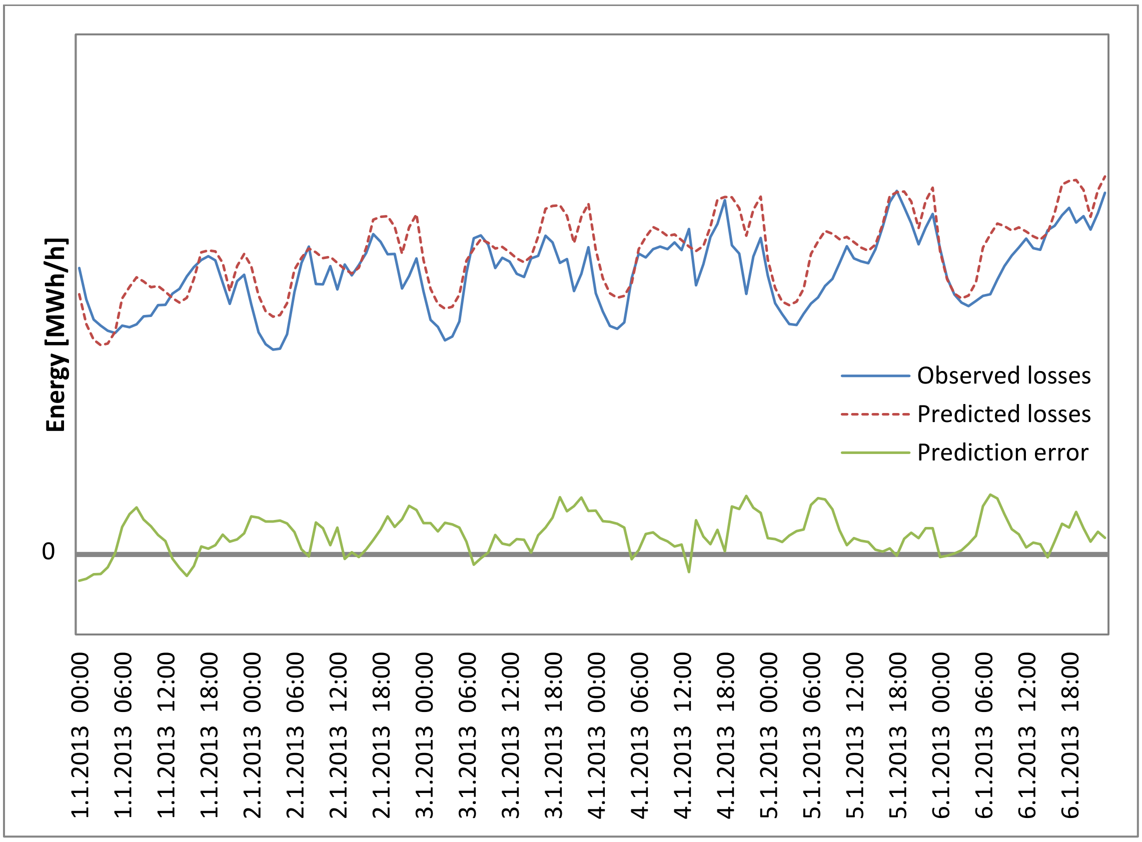

Figure B.1. *Month-based loss forecast for week 1.*

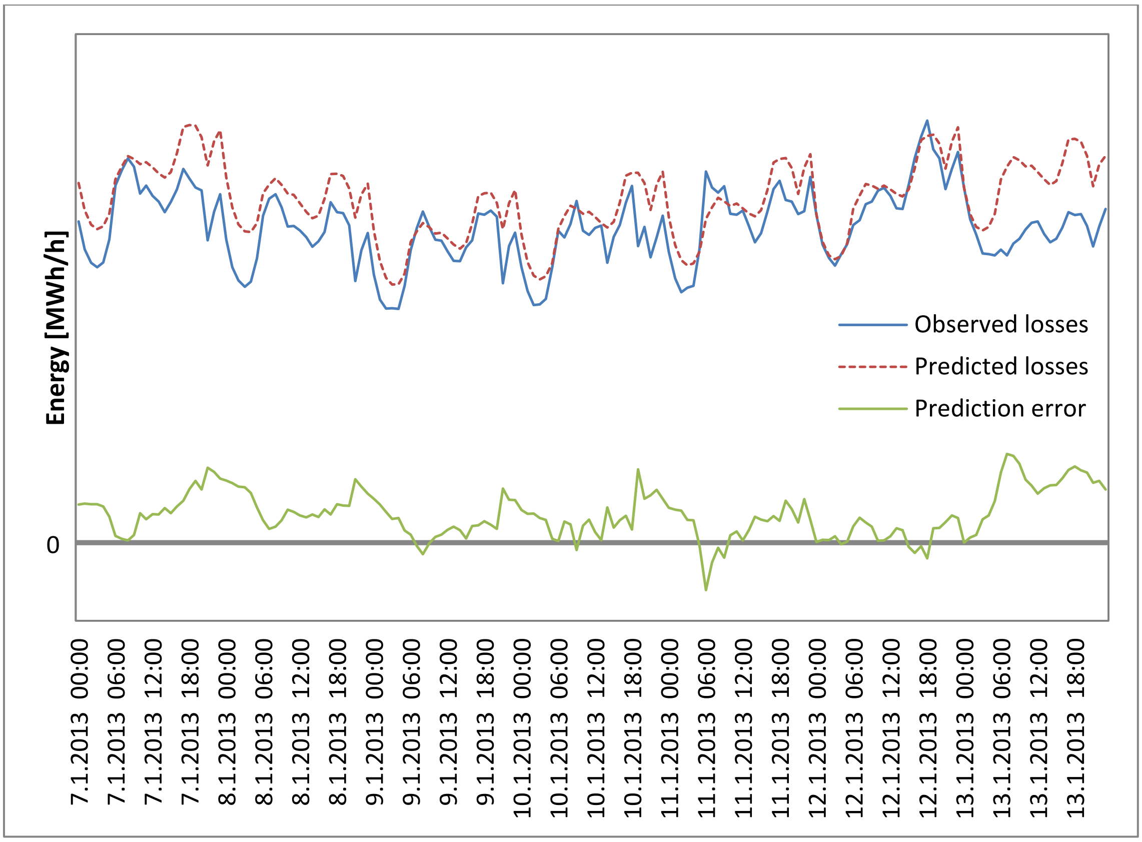

Figure B.2. *Month-based loss forecast for week 2.*

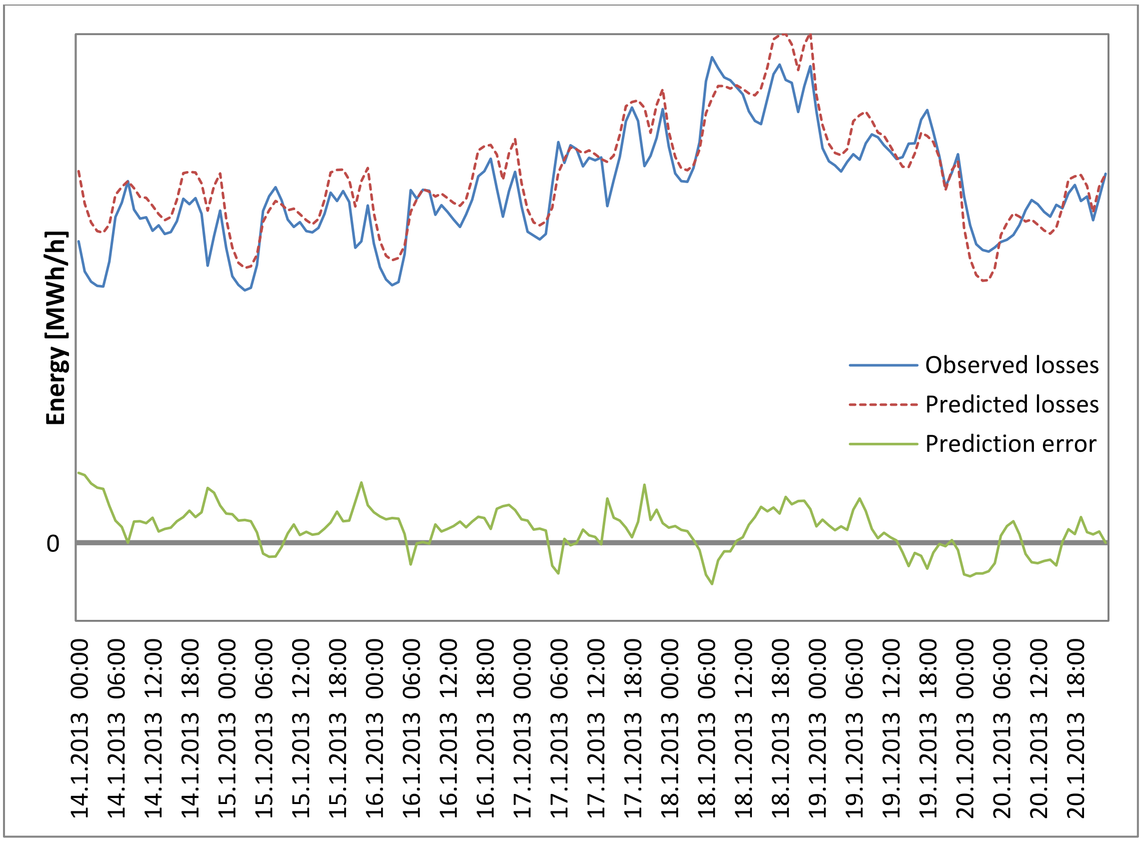

Figure B.3. *Month-based loss forecast for week 3.*

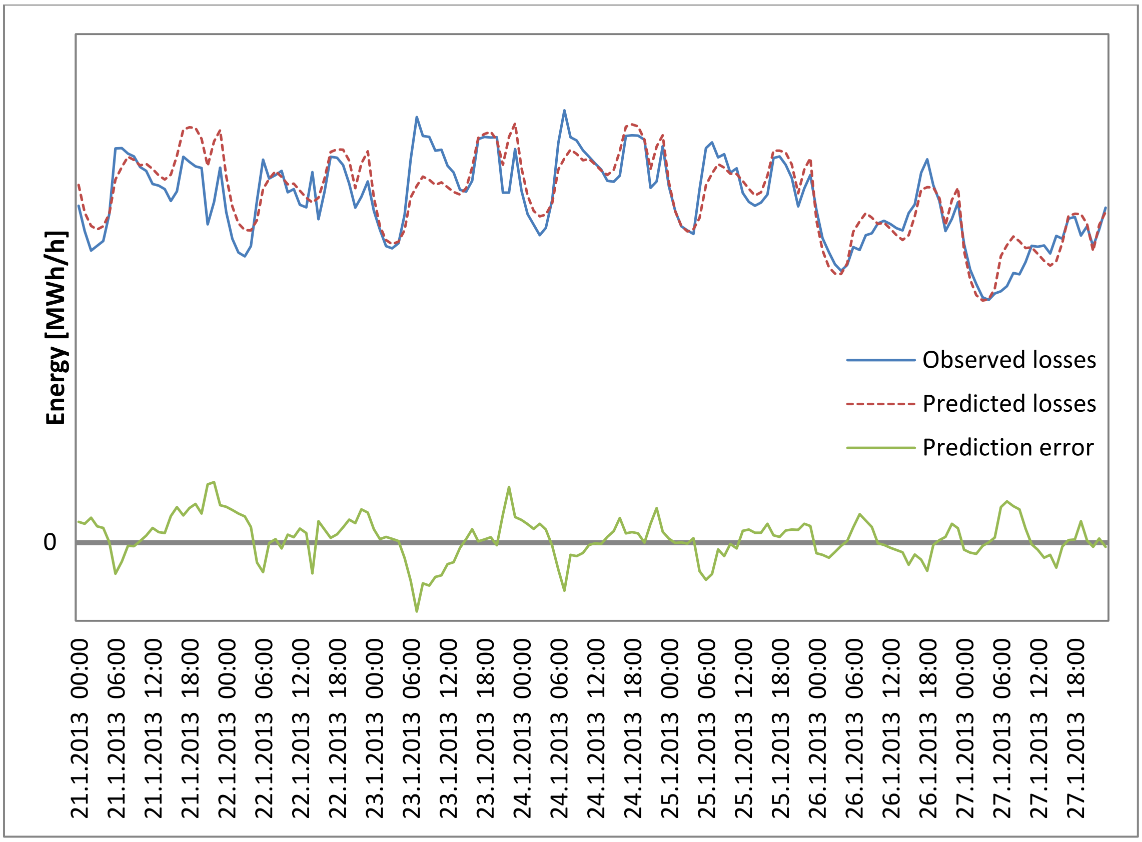

Figure B.4. *Month-based loss forecast for week 4.*

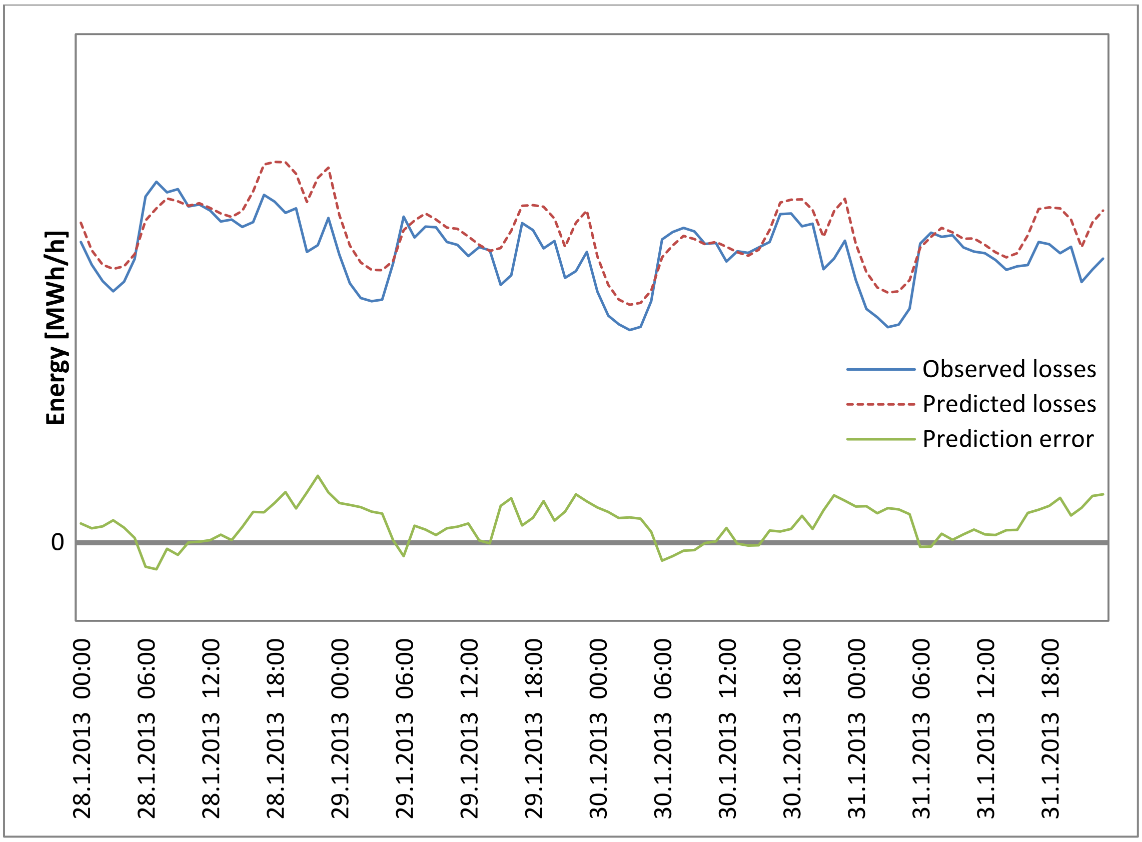

Figure B.5. *Month-based loss forecast for week 5.*

APPENDIX C: DATA FOR WEEK FUTURES

In Table C.1 P_{System} is the average Elspot system price for the week, P_{Final} is the final closing price of the week future product, $P_{Average}$ is the average price of the closing prices during final trading week for the week future product, T is the average temperature in Jyväskylä during the week and the final column is the difference between long time average temperature and the temperature in Jyväskylä during the week. The long time average temperature is a monthly average and there are two values when the week is split between months. The average temperatures are from Climatological statistics of Finland 1981-2010 (Aalto et al. 2012).

Table C.1. *Price data for week future products.*

Week	P_{System} (€/MWh)	P_{Final} (€/MWh)	$P_{Average}$ (€/MWh)	T (°C)	$T_{Ave} - T$ (°C)
44	35.19	38.40	38.89	1.1	-2.5 / 3.1
45	34.29	35.55	36.42	-0.3	1.7
46	34.29	33.80	34.31	2.6	4.6
47	32.63	35.00	35.35	4.1	6.1
48	37.27	35.60	36.33	-6.8	-4.8 / -0.6
49	48.39	45.75	45.89	-13.4	-7.2
50	48.98	51.70	47.54	-6.7	-0.5
51	41.89	39.65	40.28	-16.7	-10.5
52	35.33	34.15	34.64	-11.1	-4.9
1	34.38	35.60	35.55	-2.6	3.6 / 5.7
2	40.06	38.38	38.26	-8.2	0.1
3	46.43	51.50	51.07	-11.8	-3.5
4	46.10	48.75	50.01	-8.5	-0.2
5	36.76	36.70	38.21	-2.8	5.5 / 5.7
6	38.66	40.55	40.23	-6.8	1.7
7	40.95	40.45	41.08	-2.9	5.6
8	40.12	40.25	40.17	-4.3	4.2
9	40.05	39.10	38.92	-4.1	4.4 / -0.3
10	40.98	41.40	41.31	-10.2	-6.4
11	46.70	45.20	44.58	-12.5	-8.7
12	46.94	44.75	43.45	-8.8	-5.0
13	47.10	44.70	43.59	-6.2	-2.4