LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
Department of Energy and Environmental Technology

Suvi Huovila

SHORT-TERM FORECASTING OF POWER DEMAND IN THE NORD POOL MARKET

Thesis for the degree of Master of Science in Technology

Supervisor: Professor Lasse Koskelainen

Instructor: Dr. (Tech.) Jukka Ruusunen

Espoo, 1. April 2003

LAPPEENRNTA UNIVERSITY OF TECHNOLOGY
Department of Energy and Environmental Technology

ABSTRACT OF MASTER'S THESIS

Author:	Huovila Suvi	Date: 1.4.2003
		81 pages , 62 figures, 4 appendices
Title of thesis:	**Short-term forecasting of power demand in the Nord Pool market** Master's thesis	
Department:	Department of Energy and Environmental Technology	Chair: Power Plant Engineering
Supervisor:	Professor Lasse Koskelainen	
Instructor:	Dr. (Tech.) Jukka Ruusunen	Fortum Power and Heat Ltd.

Abstract:

The short-term forecasting of power demand has been studied over a long period of time. The deregulated electricity market in the Nordic countries is used as input in the load forecasts.

The thesis begins with a survey of the literature. The behaviour of the load is studied. The suitability of temperature data for usage in load forecasts is evaluated. Load forecasts are made on an hourly basis for a forecast period of one week.

The thesis studies the availability and the quality of the load and the temperature data from the Nord Pool market. The properties and the qualities of the data have an impact on the construction of the hourly load forecasts. Two approaches to load forecasting are modelled.

The regression model and the autoregressive (ARX) model are tested. The least-square method was used for the estimation of the parameters.

The results show that the load and the temperature data has to be checked afterwards because of the poor quality of the online input data. The temperature is affecting the load in the winter but the temperature factor can be excluded in the summer. The regression model is more robust than the ARX model. The error term of the regression model can be modelled with the time-series model.

Keywords: Short-term load forecast, demand, load profile, temperature, specific load, regression model, autoregressive model (ARX)

LAPPEENRANNAN TEKNILLINEN YLIOPISTO
Energia- ja Ympäristötekniikan osasto

DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Huovila Suvi	Päiväys:	1.4.2003
		81 sivua, 62 kuvaa, 4 liitettä	
Työn nimi:	**Sähkönkulutuksen lyhyen aikavälin ennustaminen Nord Pool markkinalla** Diplomityö		
Osasto:	Energia- ja ympäristötekniikan osasto	Professuuri:	Voimalaitos-tekniikka
Työn valvoja:	Professori Lasse Koskelainen		
Työn ohjaaja:	TkT Jukka Ruusunen	Fortum Power and Heat Oy	

Tiivistelmä:

Sähkönkulutuksen lyhyen aikavälin ennustamista on tutkittu jo pitkään. Pohjoismaisien sähkömarkkinoiden vapautuminen on vaikuttanut sähkönkulutuksen ennustamiseen.

Aluksi työssä perehdyttiin aiheeseen liittyvään kirjallisuuteen. Sähkönkulutuksen käyttäytymistä tutkittiin eri aikoina. Lämpötila tilastojen käyttökelpoisuutta arvioitiin sähkönkulutusennustetta ajatellen. Kulutus ennusteet tehtiin tunneittain ja ennustejaksona käytettiin yhtä viikkoa.

Työssä tutkittiin sähkönkulutuksen- ja lämpötiladatan saatavuutta ja laatua Nord Poolin markkina-alueelta. Syötettävien tietojen ominaisuudet vaikuttavat tunnittaiseen sähkönkulutuksen ennustamiseen. Sähkönkulutuksen ennustamista varten mallinnettiin kaksi lähestymistapaa.

Testattavina malleina käytettiin regressiomallia ja autoregressiivistä mallia (autoregressive model, ARX). Mallien parametrit estimoitiin pienimmän neliösumman menetelmällä.

Tulokset osoittavat että kulutus- ja lämpötiladata on tarkastettava jälkikäteen koska reaaliaikaisen syötetietojen laatu on huonoa. Lämpötila vaikuttaa kulutukseen talvella, mutta se voidaan jättää huomiotta kesäkaudella. Regressiomalli on vakaampi kuin ARX malli. Regressiomallin virhetermi voidaan mallintaa aikasarjamallia hyväksikäyttäen.

Avainsanat: Sähkönkulutuksen ennustaminen, kulutus, kulutusprofiili, lämpötila, ominaiskulutus, regressiomalli, autoregressiivinen (ARX) malli

Ei lainata ennen: Työn sijaintipaikka:

Preface

I have prepared this thesis at Fortum Power and Heat Oy. I want to thank Vice President of Business Development and Analysis and my instructor Jukka Ruusunen, for providing the opportunity to do the thesis and from the advises during the process of my thesis. I would also like to thank my supervisor Lasse Koskelainen. Special thanks to all my colleagues, especially Mika Gillberg and Mikael Palmgren, at Fortum from the support they provided.

I want to thank my parents from all the efforts and support during my studies.

Suvi Huovila

Espoo, 1. April 2003

TABLE OF CONTENTS

1 INTRODUCTION

Forecasting power demand is very important for power companies. The power markets have been deregulated in the Nordic countries. Some changes in the structure of demand can be seen due to the trade cycle and the rate of growth of consumption.

Demand forecasting can be divided into very short-term, short-term, mid-term and long-term forecasting. In very short-term forecasting, forecasts will be made for a few minutes to half an hour with a time span of a few hours at the most. A short-term forecast is usually made with a time span of a few hours to a few weeks. Mid- and long-term forecasts are made for longer time spans. The time span can be a few weeks to decades. (Bunn, Farmer 1985)

Very short-term and short-term demand forecasts are applied to load distribution and production allocation of electricity in generating units. Mid-term and long-term forecasts are needed when bigger investments are planned. (Karanta, Ruusunen 1991)

This thesis concentrates on the short-term forecasting of demand in the Nord Pool market. There is generally a strong interest for short-term load forecasting. The deregulation of the Nordic power markets has created the need to forecast the demand more accurately. Demand forecasts are needed e.g. in order to forecast the spot price of electricity.

Demand forecasting has long traditions. (Bunn, Farmer 1985) The forecasting models have basically stayed the same but today there are more efficient computers and techniques to handle larger amounts of data. Artificial neural networks are examples of new model structures that have been used in forecasting power demand. (Chen at al.1992, Taylor and Buizza 2002)

This thesis studies the availability and quality of load and temperature data from the Nord Pool market. The properties of the data influence the making of hourly load forecasts. Two different mathematical approaches are tested.

Diagnostics of the chosen models are presented. Proposals on how to proceed in practice with the development and the implementation of the hourly load forecasts are made.

2 THE STRUCTURE OF THE THESIS

This thesis concentrates on studying the load forecasting for the Nord Pool areas. The Nord Pool areas are Norway, Sweden, Finland and Denmark. Norway is actually divided into three different price areas and Denmark into two price areas, but in this thesis they have been treaded as one area.

Two different load forecasting models are studied. The first one is based on the regression model and the second one is a stochastic time-series model. The properties of the forecasting models are compared when using the same input data. Input data consists of realised values for load and temperature, which are provided by Nord Pool and Finnish weather service Foreca Ltd.

Temperature is the external factor for the load forecasts in this thesis even if there are additional factors that influence the load. The goal for this thesis is to analyse the properties of the two models in short-term load forecasting.

The models are tested with different time spans. The time span for estimating the most suitable parameters of the models is also studied.

An important problem is to find out the temperature dependence of demand. Temperature dependence may vary during different seasons. It will be helpful to find the proper temperature dependence for the different seasons of the year and for the whole year. At the end of the thesis suggestions for developing the models will be presented.

3 THE NORD POOL MARKET

Electricity is an extraordinary product for trading purposes because it cannot be stored easily. In the power grid demand and production have to be in balance at all times.

The electricity market in Norway was opened up for competition in 1993. This means that power generation and supply were allowed business areas for every power company. Only transmission and distribution remained regulated monopolies. In Sweden the electricity market was deregulated in 1996, with the same model as in Norway. At the same time a Norwegian-Swedish spot market called Nord Pool was established.

The electricity market law in Finland was introduced in 1995. The deregulation of the market was implemented in 1998 when Finland became a member of Nord Pool and an independent price area on the Nord Pool Exchange. West-Denmark joined Nord Pool in 1999 and East-Denmark was integrated into the Exchange area in 2000. (Nord Pool 2003)

On the Nord Pool Exchange the players, who are mainly from Norway, Sweden, Finland and Denmark, can trade electricity. Nord Pool arranges the trade in electricity on physical and financial markets and also provides a clearing service. (Nord Pool 2003)

3.1 The physical market

The participants bid for day-ahead contracts for physical electricity deliveries in the Elspot market. The power contracts are traded daily. A trading day consists of 24 different hourly products, and bids are made separately for these products. The market operator clears each market by determining market clearing prices on the basis of received bids. The clearing process results in a series of contracts between Nord Pool and each of the market participants. A contract is related to a specific hour and consists of an amount of power (MW) and the price of that hour (NOK/MWh). Contracts are binding and when the clearing process is finished financial settlement will take place. Deviations between demand vs. supply in the spot market and actual demand vs. generation are priced in the balancing markets managed by the system operators. (Bergman et. al., 2001)

The bids are made in the form of price-quantity pairs for each hour. These show the quantities in MW that the participants are prepared to purchase or sell from the spot market at different prices.

Elspot closes at 12.00 am and there are at least 12 hours, and at the most 36 hours, between the time of trade and the time of delivery. Nord Pool balances supply and demand by comparing the supply and demand curves. A market clearing price in NOK/MWh is calculated for each hour. This price, the so-called "system price", is calculated on the assumption that there are no violations of transmission constraints. Nord Pool organises the bids and the power flows over the interconnectors. If the comparison between planned and feasible power flows imply that there are no violations of transmission constraints, the system price will coincide with the "area price" in all bidding areas. Otherwise, the bids will be treated separately for each transmission constrained area. Market clearing prices for each of these areas are calculated and area prices will be different in all or some of the bidding areas. (Nord Pool 2003)

The system price and area prices are published within two hours after the Elspot market has closed. Nord Pool keeps the bid information of each market participant confidential, but the market price and the aggregate volumes are publicly available. After Elspot has closed there is - for Finnish and Swedish players - the possibility to trade on Elbas. Elbas is the aftermarket for Elspot. That enables hour to hour trading in every day of the year. (Nord Pool 2003)

Elspot's price mechanism is used to regulate the flow of power when there are capacity restrictions in the grid. Thus, Elspot may be viewed as a combined energy and capacity market.

3.2 The financial market

Futures are listed for shorter delivery periods and forwards are listed for longer delivery periods. Settlement and delivery are carried out financially without any physical delivery of electricity. Market participants can use the financial market in their risk management activities. European options on forwards are also traded on Nord Pool. (Nord Pool 2003)

4 ELECTRICITY DEMAND

Society is nowadays highly electricity dependent. The residential, commercial and service sectors are heavy users of many electricity intensive systems. This is one reason for the growth of demand during the past decade. Additionally, the industrial sector is growing continuously.

In all the Nordic countries (except for Denmark) there is an electricity intensive industry sector. Some industries have their own power generation but the industry in general is still dependent on the market for electricity and in this way dependent on the electricity price.

Use of electrical space heating as a domestic heating system has been growing after the 1990's. One reason for this increase may have been the high price of oil. This demand area is highly dependent on price and its behaviour. There are electric boilers and pumps in use at district-heating utilities. These utilities have a possibility to react to changes in the electricity price. If the price is high the electricity boilers and pumps are only partially in use. (Electricity Market 2002)

The electricity market price level influences demand. The large players on the market are highly dependent on price. They have to react to fluctuations in the market price by controlling their own resources of electricity.

4.1 Factors influencing electricity demand

There are many factors that can influence electricity demand. In the Nordic countries the most important factor is temperature. A considerable part of the residential, commercial and service sectors' heating in these countries is implemented by electrical space heating. For this kind of heating the influence of temperature fluctuations is considerable.

There are also other weather factors that influence demand. These include wind speed, sun radiation, humidity etc. (Bunn, Farmer 1995). On the other hand, sun radiation is correlated to out-door temperature and thus partly included in the temperature related load forecasting models. (Karanta, Ruusunen 1988)

Electricity demand is highly dependent on human and economic activities. This can be seen as a rhythm in the electricity load. There is more demand during daytime and less during night. Weekdays and weekends have different rhythms due to weekends having lower demand. Mondays differ a little from other weekdays because Sunday still influences the load. Special days (Christmas, Easter etc.) have their own special rhythm.

4.2 Electricity demand and production in the Nordic countries

Total electricity consumption in the Nordic countries was 393 TWh in 2001. This figure covers all the electricity demand including electric boilers. The gross consumption was 386 TWh, which represents an increase of 2.5%, compared with 2000. (Nordel 2001)

4.2.1 Electricity demand in Norway

In 2001 total electricity consumption in Norway was 125 TWh. This represented an increase of 1.4% from 2000. The consumption in power intensive industries was 31.5 %, a reduction of 1.9%. Total industry sector consumption was 45%. Overall power consumption for electric boilers and pumped storage power was 5.9 TWh, a reduction of 11.9% from 2000. (Nordel 2001)

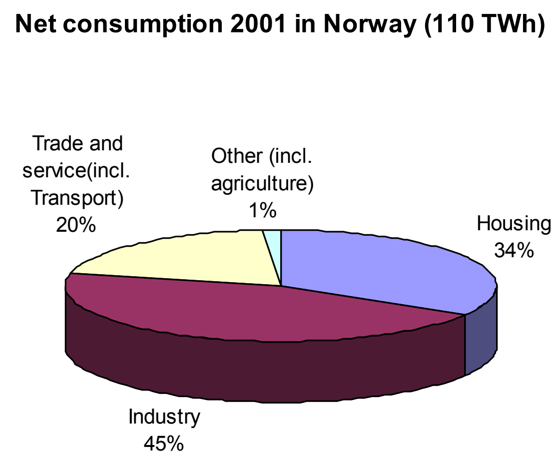

Figure 4.1 Norwegian net consumption (total consumption excluding electric boilers, pumped storage power and losses) by consumption sectors in 2001.

4.2.2 *Electricity demand in Sweden*

Total electricity consumption in Sweden was 151 TWh in 2001, an increase of 2.8% from the previous year. Demand in the entire industrial sector has fallen 3%, to a level of 43%. High electricity prices, a high dollar rate and the economic downturn towards the end of the year caused this reduction. (Nordel 2001)

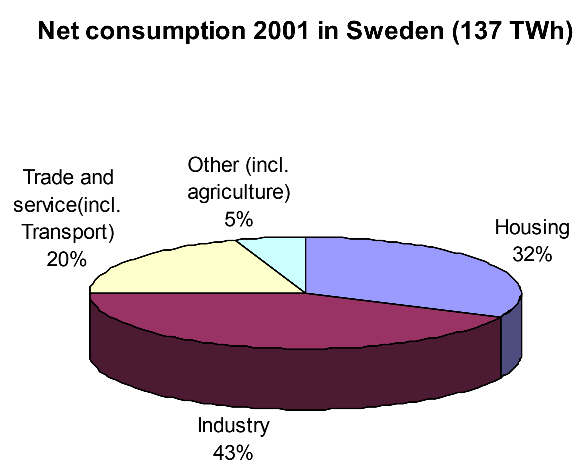

Figure 4.2 Swedish net consumption (total consumption excluding electric boilers, pumped storage power and losses) by consumption sectors in 2001.

4.2.3 Electricity demand in Finland

In 2001, a total of 82 TWh of electricity was used in Finland, which was over 3% more than the year before. Industry and the construction sector accounted for 57% of all electricity demand, households and agriculture for 25% and the service and public sectors for 17%. Consumption in the industry sector has fallen slightly while households and public sectors have seen increases of 10%. (Nordel 2001)

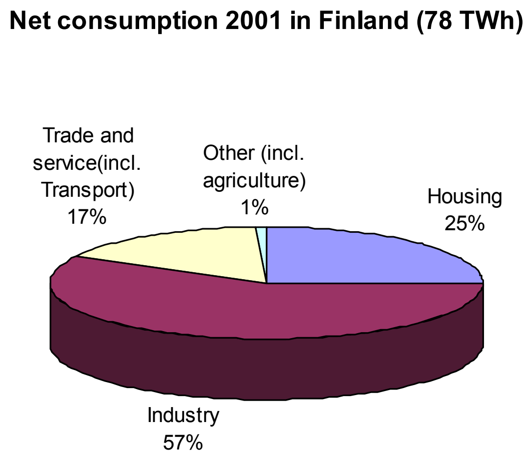

Figure 4.3 Finnish net consumption (total consumption excluding electric boilers, pumped storage power and losses) by consumption sectors in 2001.

4.2.4 Electricity demand in Denmark

Total electricity consumption in Denmark was 35 TWh in 2001. The increase from year 2000 was 1.5%. Housing, industry and the trade and service sector each accounted for approximately 30% of electricity consumption. (Nordel 2001)

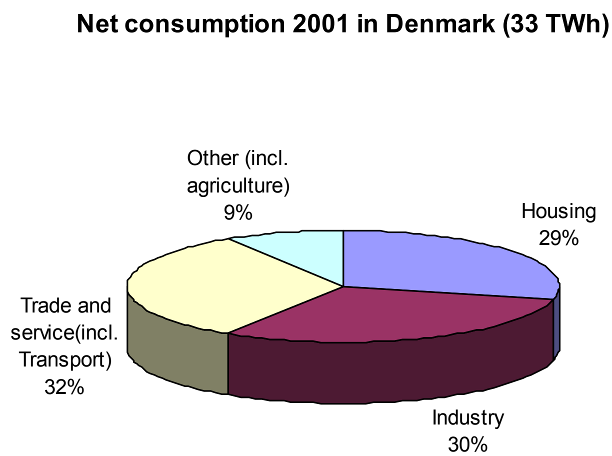

Figure 4.4 Danish net consumption (total consumption excluding electric boilers, pumped storage power and losses) by consumption sectors in 2001.

4.3 Electricity generation in the Nordic countries

Total electricity production in the Nordic countries was 387 TWh in 2001, an increase of 0.3% compared with 2000. The largest production source was hydropower (219 TWh), which represented 55% of overall production in 2001. Other production sources were nuclear power (91 TWh), other thermal power (78 TWh) and all other energy, including wind power (5 TWh).

Power trading on the Nord Pool market totalled 21 TWh. Trades made with Germany, Russia and Poland totalled 19 TWh. During the year Sweden was the largest exporter and Finland the largest importer. (Nordel 2001)

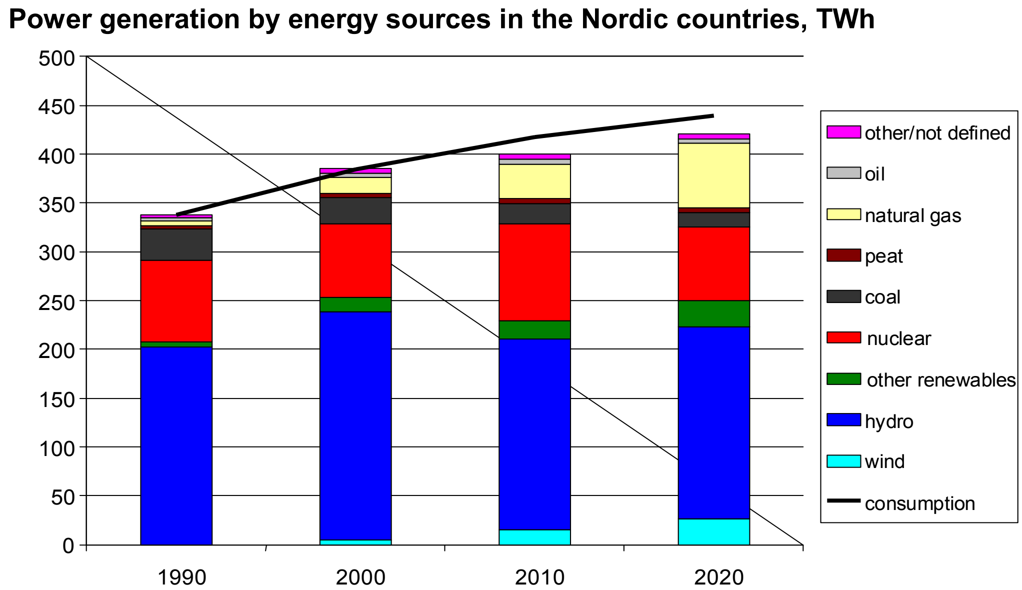

Figure 4.5 Total electricity generation in the Nordic countries and forecasts for 2010 and 2020. Sources: Eurelectric, Nordel

5 DATA FOR STUDYING THE LOAD FORECAST MODELS

The input data for the electricity load forecast models consists of the realised load and the realised temperature data. The idea is to evaluate the models based on how accurate load forecasts they produce, and compare the results with already realised load values. Nord Pool provides the load data and the temperature data is coming from the Finnish weather service Foreca Ltd. Temperature and load data are available for all Nordic countries from 1.1.2001 to 31.1.2003.

5.1 Electricity demand data

The demand curve consists of hourly demand values. The demand curves for one year are presented in figures 5.1-5.3 for each country. The curves begin at 1.1.2001 0:00. The seasonal variation can be seen in the figures. In the winter the demand is almost twice as high as in summer.

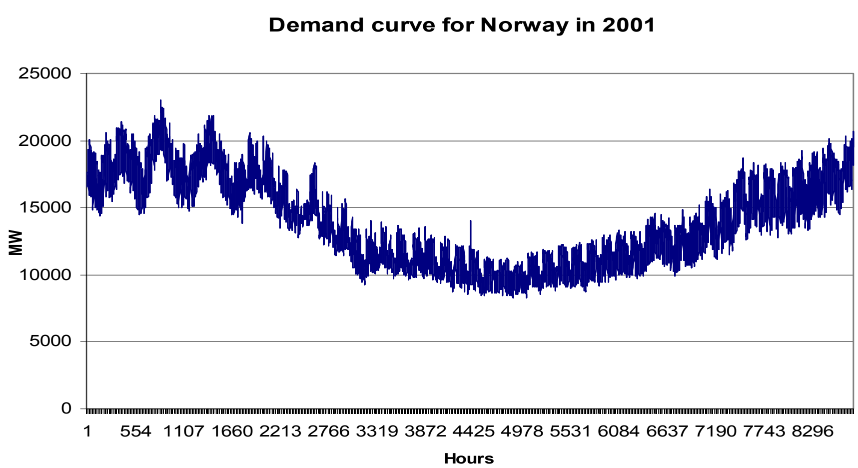

Figure 5.1 Electricity demand curve in Norway 2001.

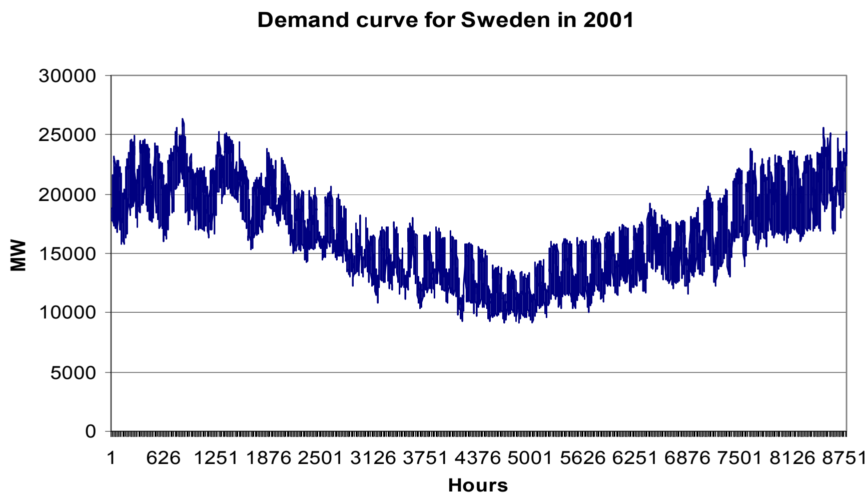

Figure 5.2. Electricity demand curve in Sweden 2001.

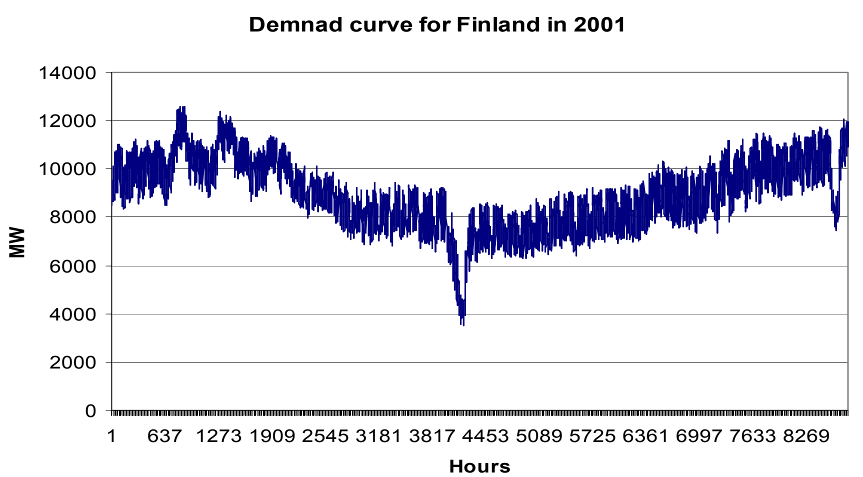

Figure 5.3 Electricity demand curve in Finland 2001.

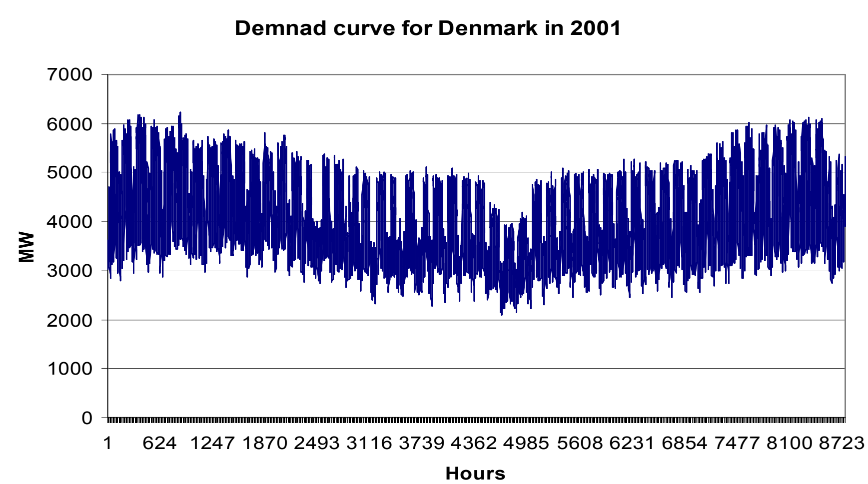

Figure 5.4 Electricity demand curve in Denmark 2001.

There are some downward peaks especially in the Finnish demand curve. These represent Midsummer in the middle of June and Christmas at the end of the year. The demands on these occasions are at a very low level. Midsummer also affects demand in Sweden but this is not so observable in the yearly demand curve. Christmas diminishes demand in every Nordic country.

If a high pressure covers Scandinavia during winter, the temperatures can be extremely low. This will affect the demand upwards and cause hourly peak loads. If cold weather is prevailing at Christmas, the low peaks of demand will be evened out.

The demand curve for Denmark is much more stable than for the other Nordic countries. Denmark has a different climate than Norway, Sweden and Finland. The structure of heating is also different than in the other Nordic countries.

The autocorrelation function tells how much correlation there is (and by implication how much interdependency there is) between neighbouring data points in the time-series. (Pindyck, Rubinfeld 1991) The autocorrelation functions pictured in figures 5.5-5.8, show that the hourly load time-series have also shorter cyclical effects. The peaks are arising every 24 hours. The weekly rhythm is slightly observable every 168 hours.

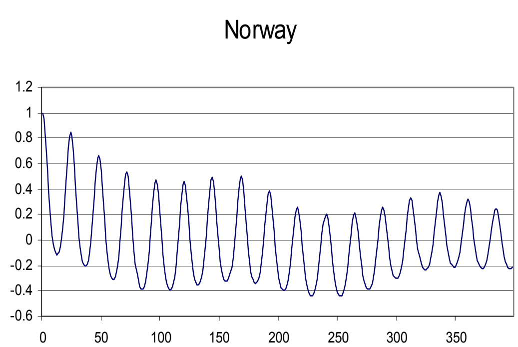

Figure 5.5 The sample autocorrelation function of the demand time-series for Norway.

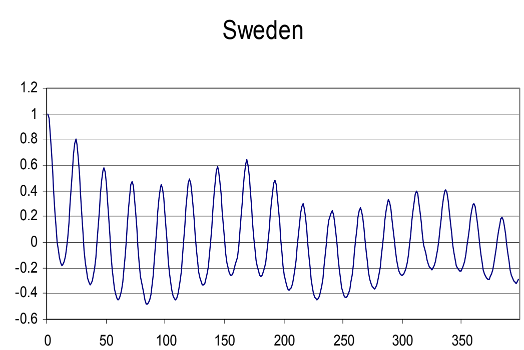

Figure 5.6 The sample autocorrelation function of the demand time-series for Sweden.

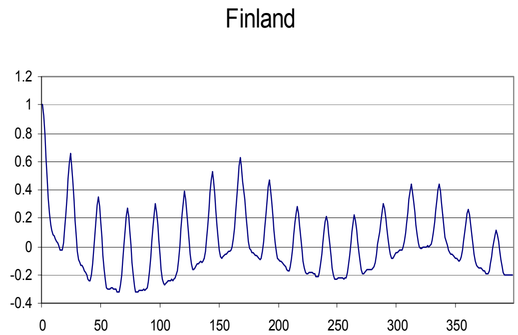

Figure 5.7 The sample autocorrelation function of the demand time-series for Finland.

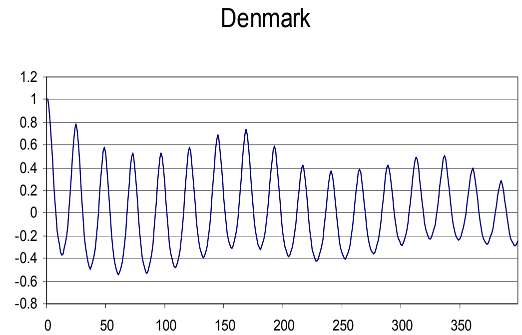

Figure 5.8 The sample autocorrelation function of the demand time-series for Denmark.

The weekly rhythm is due to the division between working days and weekends. On working days the load is higher than on Saturdays and Sundays. As an example, figure 5.9 presents the load in Finland for two weeks in January 2001 in a daily and weekly pattern. The first day and hour of the curve is Monday and its first hour. The shape of the demand curve for two weeks in the other Nordic countries is similar to the Finnish demand curve.

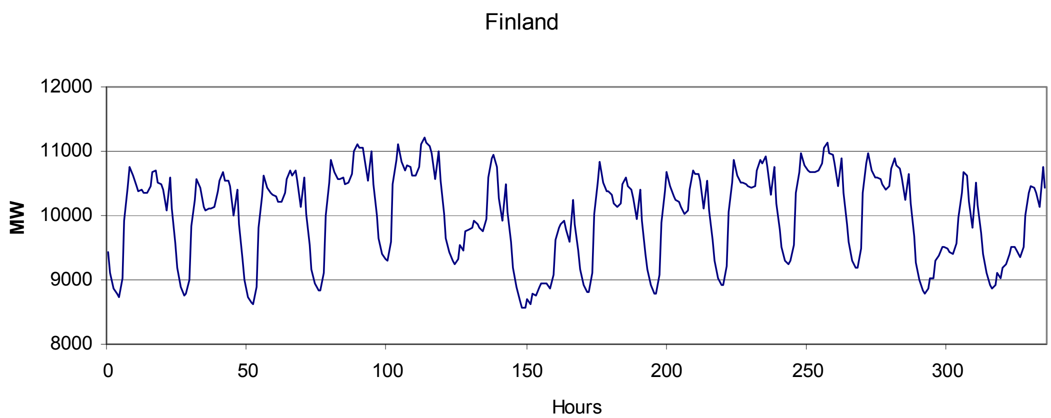

Figure 5.9 The electricity demand curve in Finland 8.1.2001 0:00 - 21.1.2001 23:00.

On the other hand, people’s behaviour during the day impacts the daily rhythm. Most people sleep during nights and consequently the demand is lower. People's daily activities such as working, watching television etc., increase demand. Figures 5.10-5.13 show daily demand patterns in every Nordic country on Wednesday January 10th 2001, Saturday January 13th 2001 and Sunday January 14th 2001.

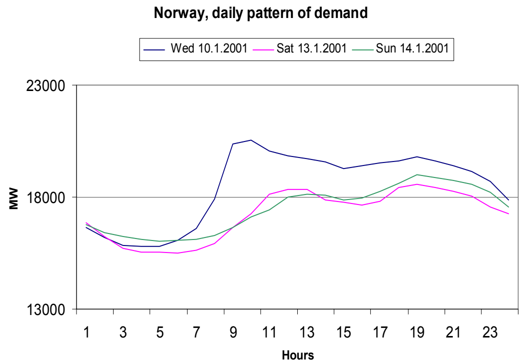

Figure 5.10 The daily pattern of demand. in Norway.

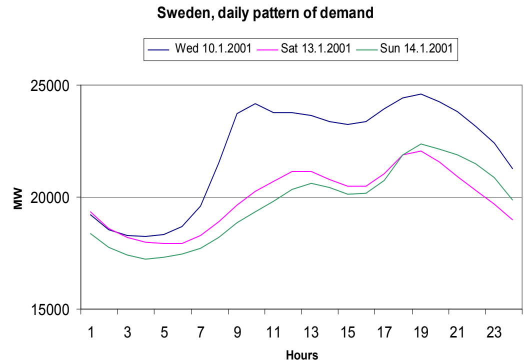

Figure 5.11 The daily pattern of demand in Sweden.

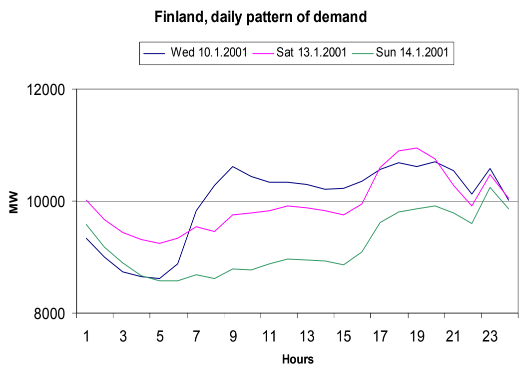

Figure 5.12 The daily pattern of demand. in Finland.

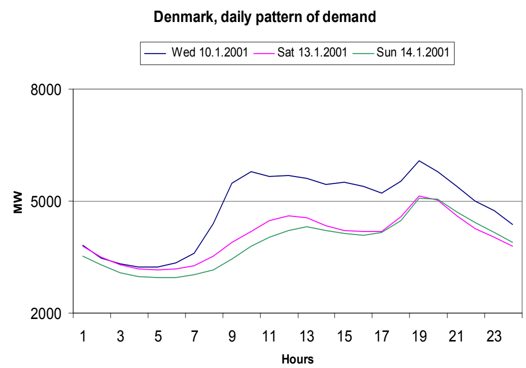

Figure 5.13 The daily pattern of demand in Denmark.

The figures show there are two peaks in demand on Wednesday. The first peak follows from the beginning of the working day and the other peak, slightly lower, follows from the leisure time of the evening. In the Finnish demand there is one further peak between 10 p.m. and midnight. Electrical space heating is connected at that time of the evening because Finnish electricity companies use lower electricity tariffs for so-called night electricity.

The daily pattern is, of course, a little bit different for each weekday. Mondays and Fridays are slightly influenced by the proximity of the weekend, but they are still working days. (Park et al. 1991) In load forecasting they are considered in the same way as Tuesdays, Wednesdays and Thursdays. Saturdays and Sundays have a different rhythm. Figures 5.10-5.13 show examples of the demand profile for a Saturday and a Sunday in the Nordic countries.

The demand profiles for Saturdays and Sundays in Sweden and Denmark follow the same shape as in Norway. The shape of the demand profile in Finland is a bit different compared to the Norwegian shape, especially on Saturdays between 4 p.m. and 9 p.m. This can be explained by the Finnish sauna culture.

In load forecasting special days, for example Christmas, Easter and Independence Day, create problems, especially when the special days occur on weekdays. Therefore these special days require special treatment. Figures 5.14-5.15 show demand curves for Good Friday and Christmas Day in Norway.

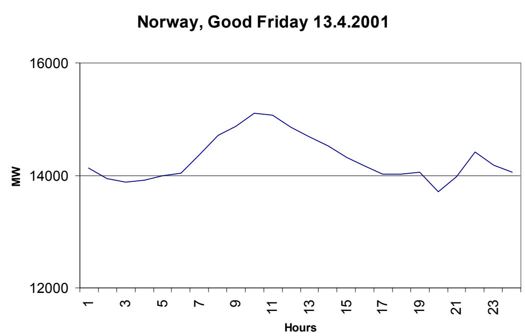

Figure 5.14 Good Friday demand in Norway 2001.

Figure 5.15 Christmas Day demand in Norway 2001.

The demand curves for Midsummer and for a normal summer Sunday in Finland are shown in figure 5.16.

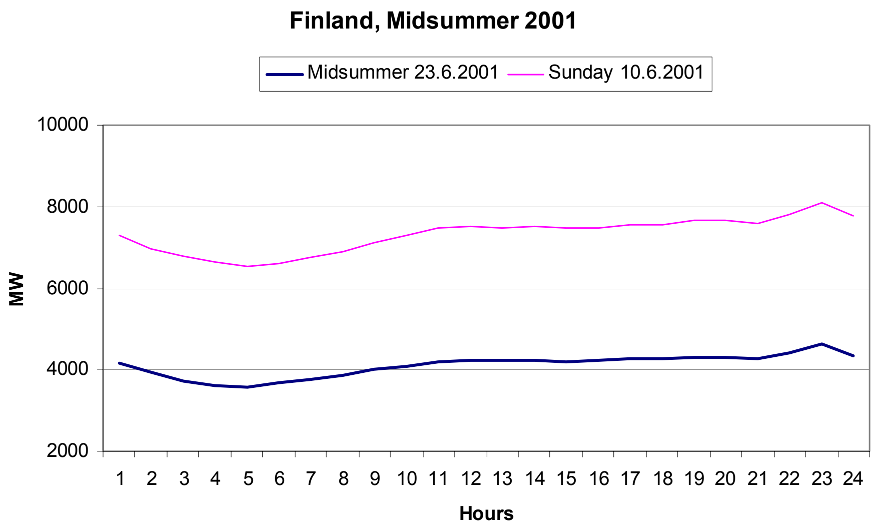

Figure 5.16 Demand curves for Midsummer and a normal summer Sunday in Finland 2001.

5.1.1 The quality of the load data

Nord Pool receives the load data from the system operators. The load data is published in real time by Nord Pool. Real time data is partly an estimate from actual demand. The system operators publish corrected data with some delay. The corrections needed to the load values can be significant. The Nord Pool's load data and the system operators' load data for Denmark are shown in figure 5.17.

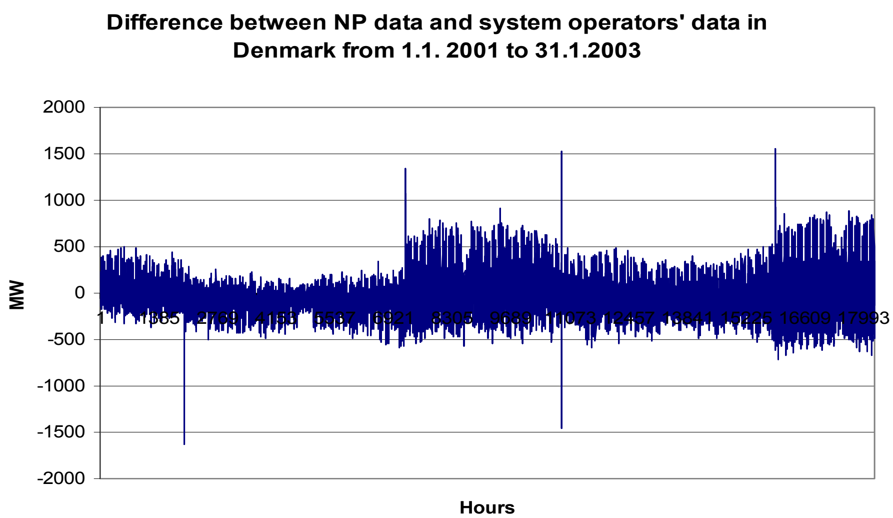

Figure 5.17 Difference between Nord Pool's data and the system operators' data in Denmark from 1.1.2001 to 31.1.2003.

5.2 Temperature data

Temperature observation points are located all around the Nordic countries. In this thesis four different observation points for each country (except for Denmark) have been chosen. For Denmark there is only one observation point. The weighted temperatures for each country were calculated from the data originating from these observation points.

Temperature data for the Nordic countries (except for Denmark) are available from the beginning of 2001 in 3-6 hour intervals. For Denmark temperature values are available only on a daily basis. The temperature intervals are linearised to construct hourly values. There are some missing measurement values, but these are ignored because missing values are impossible to get afterwards.

Suitable temperature weights have already been constructed for the Nordic countries. These weights are used in the calculations of this thesis. The temperature weights are used to calculate the weighted temperature for each country.

5.2.1 Temperature data for Sweden

SMHI temperature observation points are located in Malmö, Stockholm, Frösön and Gothenburg. Stockholm and Gothenburg are used as observation points both by SMHI and Foreca. The best alternative observation points for Malmö and Frösön have to be determined.

There are three observation points (Kalmar, Ängelholm, Jönköping) geographically close to Malmö. Correlation coefficients for these points – as compared to Malmö - were calculated. The best result was achieved for Kalmar, even if Ängelholm is geographically closer. The measured temperatures for Ängelholm have more missing values than the measurements for the other two locations. 2/3 of the correlation coefficients were better for Kalmar. Jönköping has the most accurate values, but it is situated in the inland whereas Malmö lies on the coast. Therefore Jönköping’s values would give misleading information. For Kalmar the temperature weight is the same as for Malmö.

An alternative temperature measuring point for Frösön is also established. There are three points that could be considered; Umeå, Borlänge and Luleå. The same calculations as for Malmö were made. The correlation coefficients for Umeå were the best, as well as the quality of the temperature data. Luleå is geographically too far away from Frösön. There are minor difficulties with the use of Umeå's values because Umeå is located on the coast whereas Frösön is located in the inland. Despite of Umeå's location, Umeå was chosen as the alternative for Frösön. The same temperature weight for Umeå as for Frösön is used.

5.2.2 Temperature data for Finland

Foreca provides time-series of temperatures in Finland. The temperature values are linearised for the time intervals used. For Finland the measuring points are Helsinki-Vantaa, Tampere, Joensuu and Rovaniemi.

There has earlier been more observation points than four and due to that the weights are allocated to more than four locations. As a consequence of this, the weights have to be adapted for this thesis.

Tampere's weight was updated, Lappeenranta was left out and Helsinki-Vantaa's weight was increased. With these changes more weight was given to the south and the middle of Finland. In this way the weights correspond more accurately the areas where the main part of electricity demand occurs.

5.2.3 Temperature data in Norway

The temperature time-series were linearised between measured temperatures. There were some missing measurement values (e.g. 3.1.2002 08:00 - 4.1.2002 03:00) but this was solved by the linearising between missing values. Foreca and SMHI use the same four locations as observation points. The temperature weight rates are the same as the ones used by SMHI. The observation points are Bergen, Oslo, Trondheim and Tromsö.

5.2.4 Temperature data for Denmark

Temperature data for Denmark is only supplied on a daily basis by SMHI. In Denmark only one observation point is available. There are many missing values in the measurement data. E.g. temperature values are 0 for March 2002. The missing values were replaced with the forecasted values.

The daily temperature time-series are divided into hourly values. The same daily average temperature is used for every hour of the day. The temperature weights for Denmark are the ones used for the observation points Odense and Kalmar. Even if Kalmar is in Sweden, the location of Kalmar is very close to Själland. Själland does not have an observation point of its own. More than half of the weight rate is based on the values for Odense, the rest being based on the values for Kalmar.

6 FORECASTING OF POWER DEMAND

6.1 General approach to forecasting the power demand

Already in 1970 Box and Jenkins published a book about a mathematical approach to analysing and forecasting time-series (Box and Jenkins 1971). The currently used load forecasting techniques are e.g. Box-Jenkins approaches, exponential methods, regression analysis methods, state space methods and artificial neural network approaches. (Bunn, Farmer 1985)

Bunn and Farmer divide the load forecasting methods into two types. The first one is an univariate method that forecasts the load with the help of realised loads and the time of the day. Often, the load is modelled as a sum of a standard load curve and a residual. The second type is the multivariate model, which includes the exogenous input e.g. temperature. (Bunn and Farmer 1985)

Other possible classifications are deterministic and stochastic, parametric methods and artificial intelligence methods etc. (Charytoniuk, Niebrzydowski 1998).

Deterministic models are providing only the forecast values, but not the measure for the forecasting error. The stochastic models give the forecast as the expectation of the identified stochastic process. They allow calculations of statistical properties of the forecasting error, which rely on the assumptions made in the model. (Pauli Murto 1998)

7 MODELLING

7.1 The regression model

The multiple linear regression model is based on the assumption that the load can be divided into two components: the specific load and the component that depends on the temperature. Changes in temperature are taking place rather slowly. Furthermore, the temperature values are correlated with a fairly long lag (e.g. 12 hours). Thus, it is difficult to differentiate from the sequential temperature values the temperature value that influences the demand of a specific hour. (Karanta, Ruusunen 1995)

7.1.1 The structure of the model

The error associated with a forecasting procedure can be derived from a combination of four distinct sources. The random nature of the additive error process in a linear regression model guarantees that forecasts will deviate from true values, even if the model is specified correctly and its parameter values are known. Secondly, the process of estimating the regression parameters introduces error because estimated parameter values are random variables that may deviate from the true parameter values. Thirdly, when a conditional forecast is made, errors are introduced when forecasts are made for the values of the explanatory variables in the period in which the forecast is made. Finally, errors may be introduced because the model specification may not be an accurate representation of the true model. (Pindyck, Rubinfeld 1991)

In this thesis the influence of temperature on a linear one parameter model is analysed. It is assumed that the demand depends linearly on the average of the previous hours out-door temperature values. (Räsänen, Ruusunen 1992)

The regression model can be presented as:

$$y_t = \sum_{i=1}^{N} \alpha_i x_{i,t} + \sum_{j=1}^{M} \beta_j T_{t-j-d} + \varepsilon_t \qquad (1)$$

where

y_t= the t:th realised demand

α_i = the i:th specific demand

N = number of specific demand hours

$x_{i,t}$= the 0/1-sign variable ($x_{i,t}$=1 when i=t, otherwise $x_{i,t}$=0)

β_i = temperature parameter

T_t = the t:th temperature

d = lag of temperature

M = number of temperature parameters

ε_t= a zero mean, constant variance, random error term.

The error term occurs because of the simplification of reality and the stochastic nature of demand. Actually, several omitted variables related to demand might be included in the error term. If these omitted effects are small it can be assumed that the error term is random. Also the source of errors can be associated with the collection and measurement of the data. Data will frequently be difficult to measure. The relationship of the equation is presented as a stochastic one, based on the error terms that are presented.

The important assumptions of the two-variable linear regression model are:

1. The relationship between y and T is linear, as described in Equation (1).
2. The T_t's are nonstochastic variables whose values are fixed.
3. a. The error term has zero expected value and constant variance for all observations; that is, $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma^2$.

   b. The random variables ε_t are statistically independent. Thus, $E(\varepsilon_i \varepsilon_j) = 0$, for $i \neq j$.

The list of the assumptions constitutes the classical linear regression model. Equation (1) is often termed the specification of the model. It is presumed that y is related to T, rather than vice versa. The assumption that the T's are fixed is equivalent to the assumption that the independent variable in question is controlled. It means that its value can be changed in accordance with experimental objectives. (Pindyck, Rubinfeld 1991)

The hourly load forecast in this thesis is presented as:

$$y_t = \sum_{i=1}^{168} \alpha_i x_{i,t} + \beta\, T_{Z,t} + \varepsilon_t \tag{2}$$

where α_i (i= 1...168) are the specific demand parameters for the different hours of the week and $T_{Z,t}$ is the 12 hour average of the temperature.

7.1.2 Parameter estimation

The least-square method is used in parameter estimation. The parameters are α_i (i=1...N) and β which represent specific demand at the i:th hour and the temperature coefficient, respectively.

The least-square method is computationally simple and it penalises large errors relatively more than small errors. The least-squares criterion is defined as the "line of best fit", a line which minimises the sum of the squared deviations of the points of the graph from the points of the straight line.

The least-squared criterion can be restated formally as:

$$Minimize \sum_{t=1}^{N} \left(y_t - \hat{y}_t\right)^2 \tag{3}$$

where $\hat{y}_t = \alpha_i + \beta T_t$ represents the equation for a straight line with intercept α and slope β. y_t is the actual value of y for observation t, corresponding to the value of T, while n is the number of observations.

The least-squares solution for the slope and intercept are:

$$\hat{\beta} = \frac{\sum_{all} y_i T_i - \frac{1}{N_1}\sum_1 y_i \sum_1 T_i - \ldots - \frac{1}{N_n}\sum_n y_i \sum_n T_i}{\sum_{all} T_i^2 - \frac{1}{N_1}\left(\sum_1 T_i\right)^2 - \ldots - \frac{1}{N_n}\left(\sum_n T_i\right)^2} \tag{4}$$

and

$$\hat{\alpha}_k = \frac{1}{N_k}\left(\sum_k y_i - \beta \sum_k T_i\right),\ k \in [1, N] \tag{5}$$

Here in this case the sub-indexes of the sums describe specific demand hours. E.g. $\sum_k y_i$ is the sum of the observations from the first hour of a rhythm period. (Gillberg 1997)

If the temperature parameter is left out, the estimation contains only the estimation of constants that are obtained directly as the averages of the load from the corresponding hours. (Gillberg 1997)

7.1.3 The use of the regression model

The simplicity of the regression model is a credit of the model. The regression model is only estimating a constant when the temperature factor is excluded.

The regression model is constructed from the specific load and the temperature factor. The specific load is estimated at the same time as the parameter for temperature. The specific demand is not the realised value but the estimation of the realised values of the load. The specific load parameter is estimated for every hour of the load forecast period.

The realised load values for the estimation period (15.1 - 11.2.2001) in Finland are presented in figure 7.1.

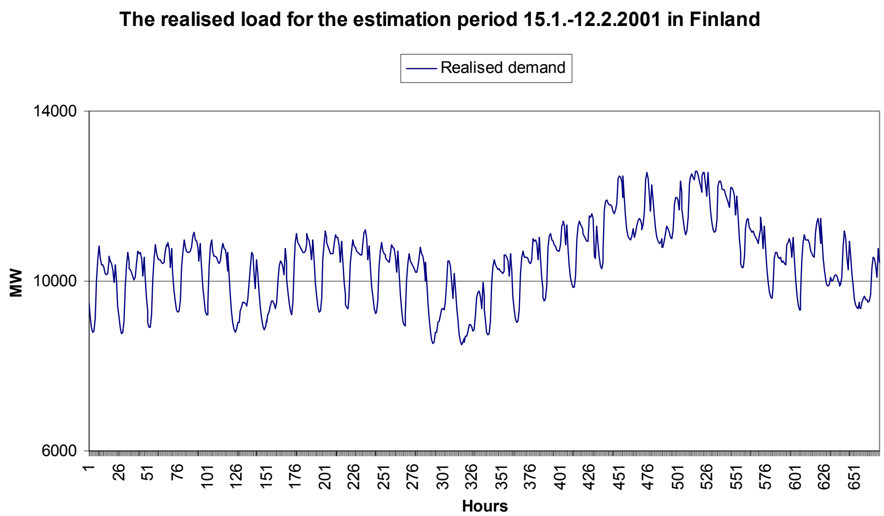

Figure 7.1 The realised load for the estimation period in Finland.

One specific demand parameter is estimated for every hour of the week from the 4 weeks data set. The specific load parameters for the load forecast, which begins on 11.2.2001 0:00 with a time span of 168 hours, are shown in figure 7.2.

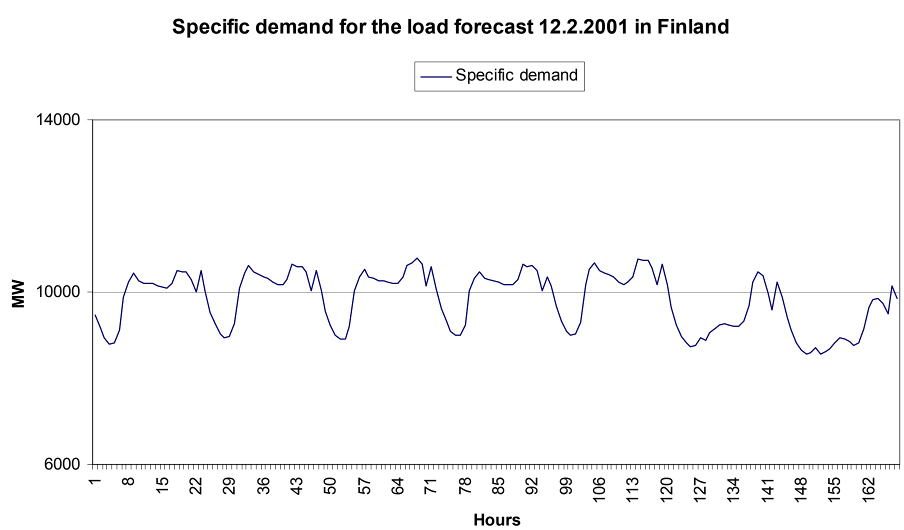

Figure 7.2 Specific demand for the load forecast on 12.2.2001 in Finland.

The parameter for the temperature factor is estimated from the same period as the specific demand. The temperature values are 12 hour moving average temperatures for the whole data set. Only one temperature parameter is estimated for the whole forecasting period. The temperature dependence of demand and temperature for the load forecast period is shown in figure 7.3.

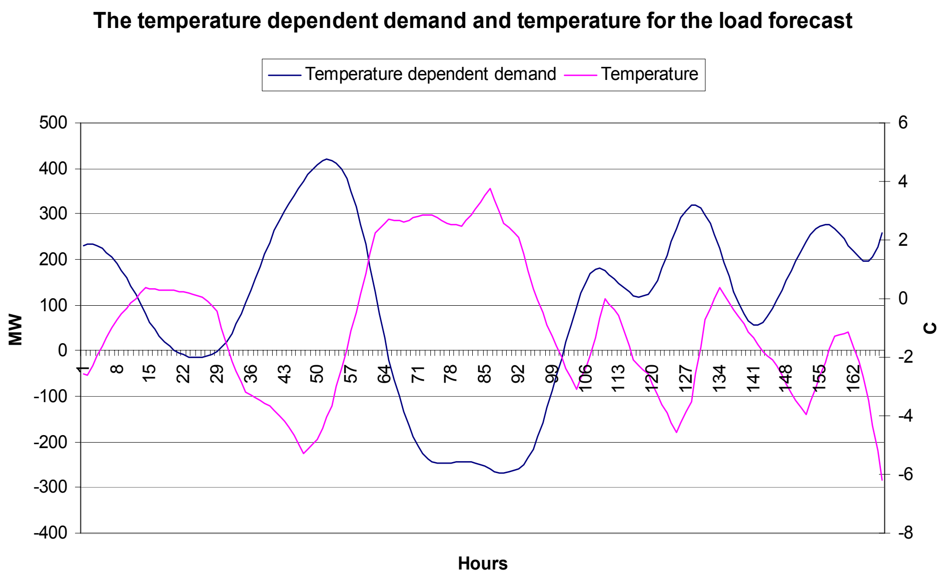

Figure 7.3 The temperature dependent demand and temperature for the load forecast in Finland.

The specific demand and the temperature dependent demand form the load forecast for every hour of the week. The specific demand, the temperature dependent demand and the load forecast are presented in figure 7.4.

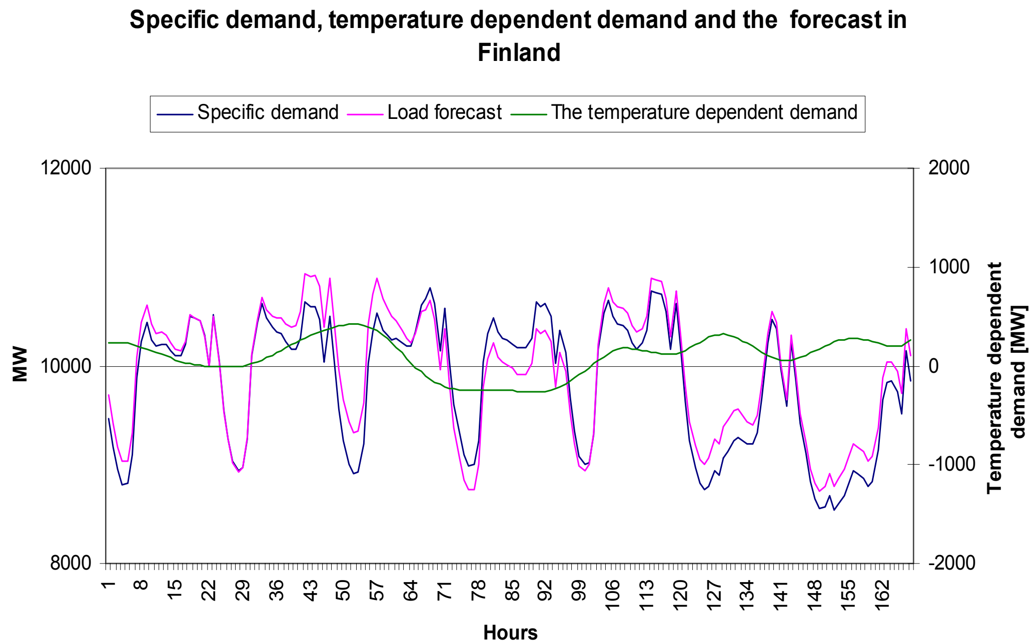

Figure 7.4 Specific demand, temperature dependent demand and the load forecast for Finland from the beginning of 11.2.2001.

The regression model estimates the parameters and makes the load forecasts for 168 hours. The load forecast is completed when the specific demand factors and the temperature dependent demand are added. The realised load and the load forecast are not exactly the same because of the error term in the regression model. The error term is caused by e.g. the model structure, the quality of the data or random parameters. The 168 hours load forecast and the realised load for Finland from the beginning of 12.2.2001 are presented in figure 7.5.

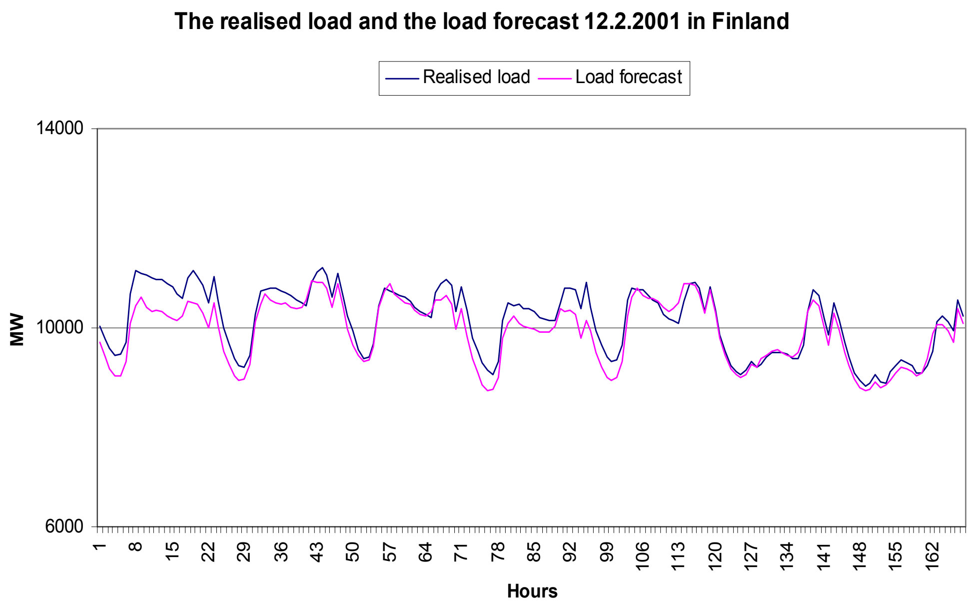

Figure 7.5 The realised load and the load forecast for Finland from the beginning of 12.2.2001.

7.2 The stochastic time-series model

In this thesis an autoregressive model with an exogenous variable (ARX) is used as a stochastic time-series model. In the ARX model the process depends on a weighted sum of its past values, exogenous variables and a random disturbance term.(Söderström, Stoica 1989) In this process the time series are assumed to be stationary and invariant with respect to time. (Pindyck, Rubinfeld 1991)

7.2.1 The structure of the stochastic time-series model

The autoregressive process can be cyclical, depending on the numerical values of the parameters. Some information about the order of the autoregressive process can be obtained from the oscillatory behaviour of the sample's autocorrelation function. (Pindyck, Rubinfeld 1991)

The ARX model is a second-order autoregressive model. In a second-order model each observation is highly correlated with those surrounding it, resulting in a discernible overall up-and-down pattern. An autoregressive model in itself eliminates the correlation of the time-series. (Pindyck, Rubinfeld 1991)

The results of the autocorrealation function with lag of 24 hours are presented in figures 5.5.-5.8. When the first differentiation is calculated the autocorrelation function diminishes rather quickly with increasing lag. The weekly rhythm is still present in the autocorrelation function. The autocorrelation function for the load in Norway after the first differentiation is presented in figure 7.6.

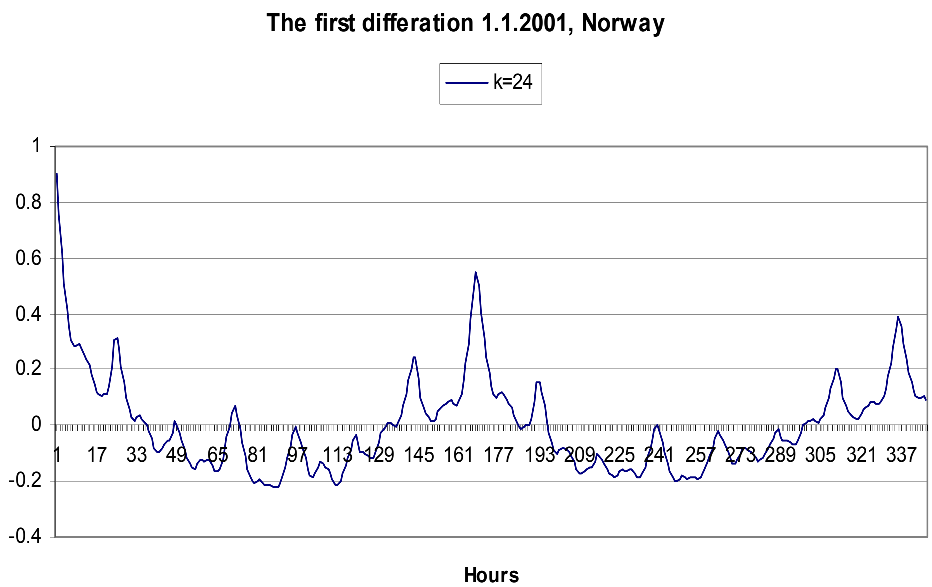

Figure 7.6 Autocorrelation function after 24-hour differentiation for the load in Norway.

The ARX model is as follows:

$$y_t = \sum_{\tau=1}^{l} a_{t,\tau} y_{t-\tau} + \sum_{\tau=1}^{m} b_\tau T_{t-\tau-d} + \varepsilon_t \tag{6}$$

where

y_t = the t:th realised demand
T_t = the t:th temperature
τ = the weekly difference factor
ε_t = a random disturbance error term, zero mean
a_τ (τ= 1...l) is the load parameter
b_τ (τ= 1...m) is the temperature parameter.

From this general model structure the following form can be derived:

$$
\begin{aligned}
y_t - y_{t-168} &= a_1(y_{t-24} - y_{t-192}) + a_2(y_{t-48} - y_{t-216}) \\
&\quad + b_0(T_t - T_{t-168}) + b_1(T_{t-24} - T_{t-192}) + b_2(T_{t-48} - T_{t-216}) + \varepsilon_t
\end{aligned}
\tag{7}
$$

=>

$$
\begin{aligned}
y_t &= a_1(y_{t-24} - y_{t-192}) + a_2(y_{t-48} - y_{t-216}) + b_0(T_t - T_{t-168}) \\
&\quad + b_1(T_{t-24} - T_{t-192}) + b_2(T_{t-48} - T_{t-216}) + y_{t-168} + \varepsilon_t
\end{aligned}
\tag{8}
$$

This structure reflects the daily and weekly rhythms of the demand, combined with the temperature dependency.

7.2.2 Parameter estimation

Linear estimation of parameters (Pindyck, Rubinfeld 1991) is used for solving the parameters that give the best fit to the measurements. The assumptions listed in chapter 7.1 should hold here as well. However, it is clear that the y-variables on the right hand side do not fulfil assumption 2. Because $y_{t-\tau}$ is measured with an error, the problems arise with the least-square method. The parameters will be biased and inconsistent. The problem of errors in measurement of regression variables is quite important, and yet econometricians do not have much to offer in the way of useful solutions. (Pindyck, Rubinfeld 1991)

7.2.3 The use of the stochastic time-series model

A time-series model accounts for patterns in the past movements of a variable and uses that information to forecast its future movements. A time-series model is usually used when little is known about the determinants of the variable of primary concern and sufficiently large amounts of data is available. (Pindyck, Rubinfeld 1991)

In this thesis the ARX model uses the previous load data and temperature for the load forecasting. The difference in the hourly load between 168 hours is calculated. The model takes into account two load factors that have a difference of 24 hours.

The temperature index is calculated for each Nordic country. Realised temperatures are weighted for each observation point separately. The weighted temperatures are added into one country specific temperature index.

The ARX model contains three temperature factors. The first temperature factor contains the difference in the temperature index between the previous hour and the hour 168 hours previously. The next two temperature factors have the same differences, but are calculated from the preceding 24 and 48 hour values.

The estimation of the parameters is made from the load and temperature differences from the previous four weeks period. One parameter for every factor has been estimated for the whole forecasting period. The ARX model is using these five parameters in the load forecast for every hour. 168 hourly load forecasts are made separately.

The ARX model calculates the change of the load and adds this change to the corresponding realised load of the previous week. The model uses the realised load as basis, and due to that the model is sensitive to measurement errors.

The load forecast is made from the beginning of 12.2.2001 with a 168 hours time span. The four weeks' realised load, the load forecast and the temperature indices for the forecast period are shown in figure 7.7.

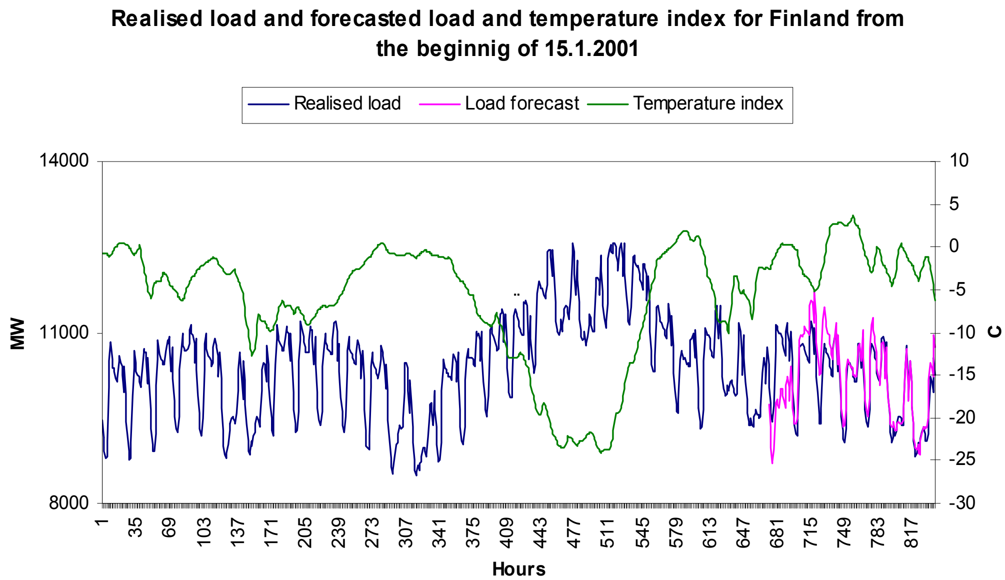

Figure 7.7 Realised load, forecasted load and realised temperature in Finland from the beginning of 15.1.2001.

The figure shows the sensitivity of the ARX model to the measured data. The realised load begins to grow after two weeks and affects the load forecasts of the following weeks. The temperature indices show that the weather has been colder in the third week. The ARX model is not handling the beginning of the load forecast very well.

8 MODEL VALIDATION

In this thesis the goal is to compare the regression model and the stochastic time-series model in forecasting the electricity demand on the Nord Pool market. To compare the models, a plan for testing and for result evaluation has to be made.

The tests are made in an offline process. The existing data for both models are the realised loads and temperatures from the beginning of 2001. The models are using hourly values and thus the forecasts are made for every following hour. The models have been tentatively tested and the maximum forecasting period seems to be two weeks.

There are some limitations to forecasting a two weeks period when considering an online forecasting process. The temperature forecast is provided every day for a ten-day period by the Finnish weather service. To respect these limitations and to find a meaningful test period, a seven days forecast period was chosen. That way every weekday is taken into account.

The forecasts are made for three different time periods. The periods chosen were the ones used by Nord Pool in their financial market; Winter 1 (1.1.-30.4.), Winter 2 (1.10.-31.12.) and Summer (1.5.-30.9.). These periods are used in every evaluation of the forecast errors.

8.1 Evaluation methods

Forecast error measures methods can be used for comparing the models. They are easy to understand and computationally efficient. Forecast error measures can be categorised into standard and relative errors. Standard errors are expressed in the same unit as the forecast. This can induce problems in comparing the different periods of time and accuracy across the time series. Relative errors do not have this kind of problems. They are dimensionless and the quality of the forecast is easier to understand. (Sanders 1997)

In the literature a commonly used method for measuring the electricity load forecast error is the Mean Absolute Percentage Error (MAPE) or standard deviation. (Räsänen, Ruusunen 1996, Papalexopoulos, Hesterberg 1990, Moghram, Rahman1989).

8.1.1 *Absolute error of the load forecast*

To evaluate generally the forecast errors in the regression model and in the time-series model, the absolute error of both models have been calculated. (Sanders 1997)

The absolute error is defined as:

$$e_t = |y_t - \hat{y}_t| \tag{9}$$

where
e_t = the t:th absolute forecast error
y_t = the t:th measured load value
$\hat{y}_t$ = the t:th load forecast value.

8.1.2 *Histograms of the absolute errors*

In this thesis number of load forecasts is relatively high. The distribution of the forecast errors is illustrative for a large set of values. The distribution of the forecast errors can be observed with the histogram. E.g. the histograms can be seen in figures 9.1-9.4.

The histogram is based on the frequency table, which contains the error categories. The error categories take into account how many values belong to each category. The number of values of each category comprises the histogram of the load forecast errors. (Tilastomatematiikan opintomoniste 1996)

The data set for a histogram is formed from the error values that correspond to the 28th and 34th hour from the beginning of the forecast period. The 28th and the 34th hours are chosen in order to get an impression of the forecast accuracy for the following day. The 28th hour is between 03:00-04:00 at night and the lowest demand typically occurs at this time. The 34th hour is between 09:00-10:00 in the morning and it is the highest demand period of the day. These histograms are made for Norway, Sweden, Finland and Denmark. The histograms for Norway are presented as an example.

8.1.3 Mean absolute percentage error

In this thesis the mean absolute percentage error has been chosen for the evaluation of the accuracy of the models. Because MAPE is unit-free measure, it is also easy to compare the result between the regression model and the stochastic time-series model.

MAPE is defined as: (Sanders 1997)

$$e_{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{|y_t - \hat{y}_t|}{y_t}\right]*100\%, \tag{10}$$

where

N = number of the forecast values

y_t= the t:th realised load value

$\hat{y}_t$ = the t:th forecast load value

Approximately 200 forecasted values are used, depending on the length of the season, to calculate one MAPE value. The electricity load forecasts are calculated with both models and always for the same test period. In the beginning, the parameters of both models are estimated from the previous four weeks. The estimation period of the parameters can also be a different time period. The most important issue that has to be evaluated is how accurate forecasts the models can give for the next day and for the next week.

8.1.4 Evaluating the mean absolute percentage error

The forecasts start at 0:00, the same input data is used for all hours until 24:00 and the new forecasts are made after 24 hours for the next seven days period. All the load forecasts and the errors are calculated separately for Winter 1, Winter 2 and Summer. All the seven days periods are placed on the same level, so that the all the forecast hours are equally far away from the starting point of the forecast. MAPE is then calculated for every hour of the seven days period.

MAPE is calculated for the whole data set for the time periods Winter 1, Winter 2 and Summer. MAPE is evaluated with errors that are less than 15% of the realised load for one hour. From these MAPE values one can roughly say which model is better.

9 ANALYSIS OF THE TEST RESULTS

The load forecasts are made for both models using the load and temperature data from the beginning of 2001. The input data is available from the beginning of 2001 until 31.1.2003. The load forecasts are made until 31.1.2003. The estimation of parameters in both models requires input data from a four weeks period before the load forecast itself begins. January of 2001 is then used for estimating the parameters, and the load forecasts begin on 1.2.2001. From a statistical point of view, the January 2003 input data is included into the forecast period to produce equal time periods for the two years.

Test results are produced for all the Nordic countries with the same methods that have been specified in the previous chapter. There are 730 different load forecasts which all begin at midnight. Each day the load forecasts are made for every day in periods of seven days.

9.1 Results of the regression model

Evaluating the error compared to realised demand tests the regression model. The forecast errors have been dealt with by using all the methods mentioned in the previous chapter.

The model is tested for longer forecast periods to compare the parameter estimation periods and also to apply it to the study of the temperature dependence of demand in chapter 10.

9.1.1 Distribution of the load forecast error values for the regression model

In the regression model the construction of the load forecast is comprised of the specific load value and the temperature factor. The specific demand is estimated from the load values of the previous four weeks. The present values of the realised demand do not affect the rhythm of the load as much as has been estimated.

The distribution of the load forecast is defined in chapter 8. The histograms are drawn for the 28th and the 34th hours of the load forecasts. Every season has its own histogram and the load forecast error points are taken from every load forecast. The example figures are from the Norwegian load forecast because the most electricity dependent demand is in Norway.

In the load forecasts, errors can occur due to different reasons. One of the reasons can be the accuracy of temperature data. Also the quality of realised load data can be affecting the forecasted demand. The measured load values that the model is forecasting can be different from real demand.

The model uses the Finnish calendar with Finnish special days. The load pattern for Sunday is used for the special days. Errors in the load forecasts occur because all the special days in Finland are not necessarily such in the other Nordic countries. Furthermore, the actual load patterns for Easter, Christmas and Midsummer differ from the load pattern for Sunday. This creates additional errors in the forecasted values but in this case the periods around the special days are excluded from the analysis.

The forecast errors should be zero mean. In the results the forecast errors are not zero mean. The forecast errors deviate more to the positive side of the histogram, as can be seen for Winter 1 in figures 9.1-9.2. The regression model forecasts the hourly values as too high when compared to the realised load values. The parameters are estimated from the previous four weeks. The demand dependence of temperature may have been higher during this period and the model has estimated the specific demand as too high.

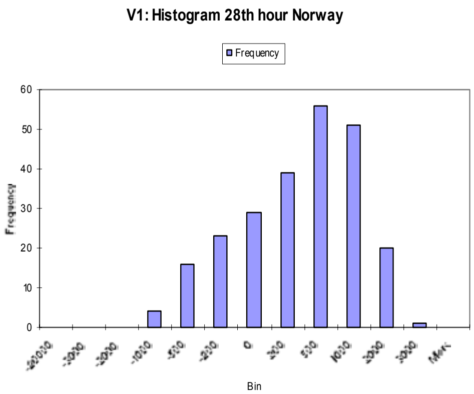

Figure 9.1 Histogram 28th hour from the beginning of the load forecast in Winter 1

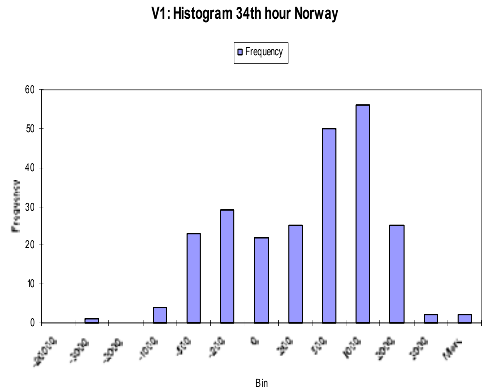

Figure 9.2 Histogram 34th hour from the beginning of the load forecast in Winter 1

In the summer the forecast errors are more diffused than in the winter. The regression model takes into account the temperature effect also in the summer. As it is known the demand is not highly temperature dependent in the summer and thus the regression model overparametrizises. The industrial load diminishes in the summer due to maintenance of plants, which can been seen from the specific load parameters. Figures 9.3-9.4 present the histograms for the summer.

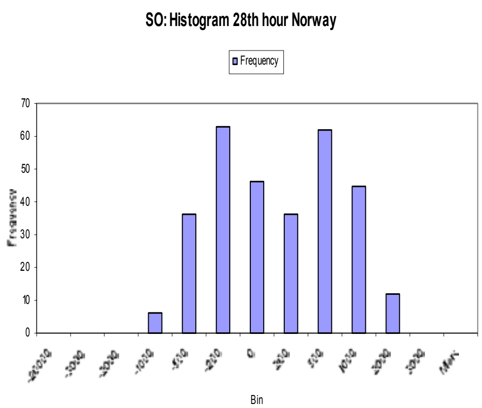

Figure 9.3 Histogram 28th hour from the beginning of the load forecast in Summer.

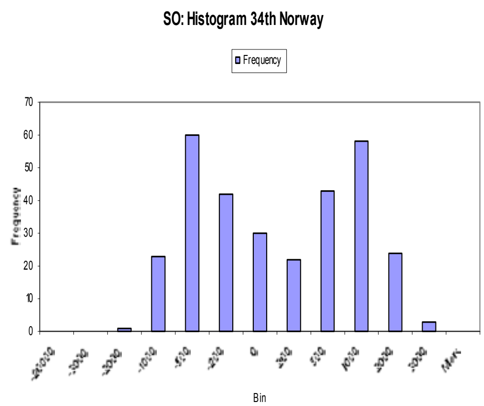

Figure 9.4 Histogram 34th hour from the beginning of the load forecast in Summer.

During Winter 2 the forecast errors are emphasised more to the left side of the histogram. This means that the model is forecasting the load as too low when compared to realised values. The regression model estimates the parameter somewhat lower, which can be derived from the lower demand in the estimating period than in the forecasting period. The temperature dependence of demand is not correcting the forecasted load enough. The histograms for Winter 2 are shown in figures 9.5-9.6.

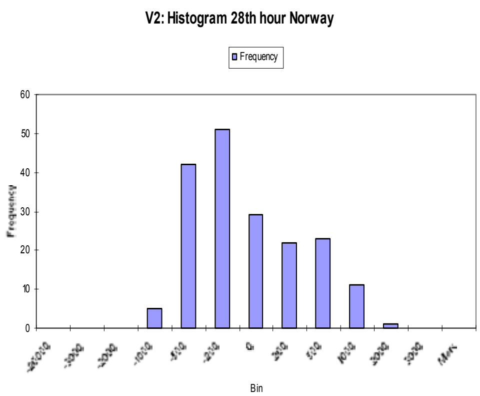

Figure 9.5 Histogram 28th hour from the beginning of the load forecast in Winter 2.

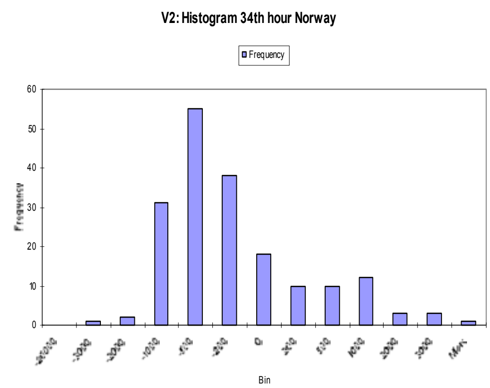

Figure 9.6 Histogram 34th hour from the beginning of the load forecast in Winter 2.

The numeric values for the forecast error distributions are shown in appendix 1. For comparability the average demand of all demand that occurs at 03:00 and 09:00 is calculated. The bins are compared to these average demands. This is one reference point when checking how much the forecast errors deviate from real demand. The results show that 60-90% of the forecast errors are within –1000 MW and 1000 MW. This is under 10% of average demand. Percentages are presented in appendix 2.

9.1.2 Mean absolute percentage error analysis of the regression model

The regression model has been evaluated with MAPE. The results are analysed on an hourly basis. All the load forecasts are constructed to begin at the same time and thus every hour of each forecast is equally far from the starting point. Every forecast period is 168 hours and they are constructed for every season.

MAPE emphasises the relative error compared to realised load values of all the data points analysed. From this behaviour the forecast errors can be observed. Approximately 200 data points are used, depending on the length of the season, to calculate one MAPE value.

The regression model uses the Finnish calendar with Finnish special day. The other Nordic countries' calendars are not used. On a special day the model is using the load pattern for Sunday.

MAPE is calculated so that errors larger than 15%, compared to the realised demand, are left out. There are some errors that are too big because of the special days or because the parameters are over-estimated. Easter, Midsummer and Christmas also induce changes in the demand that are not caused by temperature, and the current model is unable to handle these situations. MAPE values for all seasons are presented in figures 9.7-9.9.

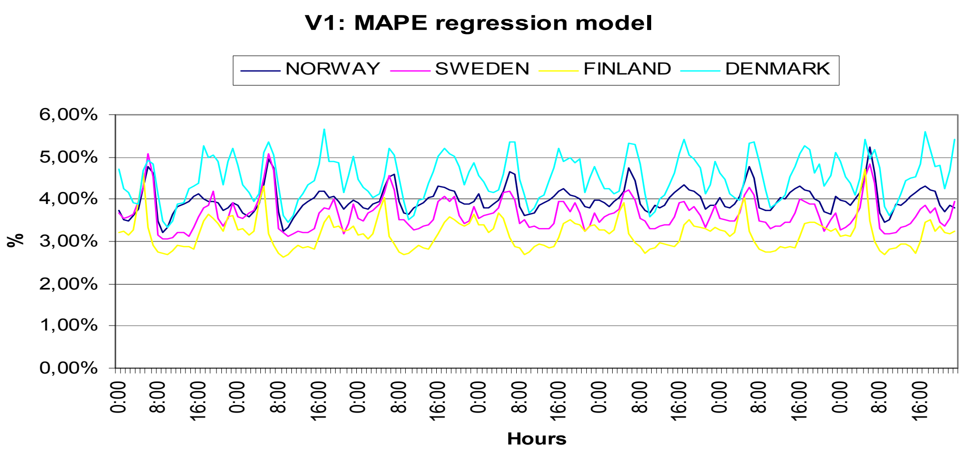

Figure 9.7 MAPE for all Nordic countries on an hourly basis in Winter 1.

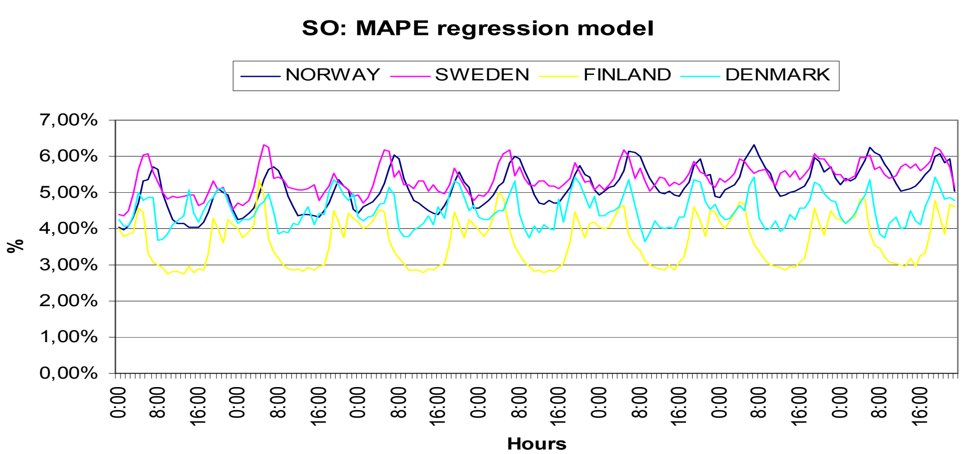

Figure 9.8 MAPE for all Nordic countries on an hourly basis in Summer.

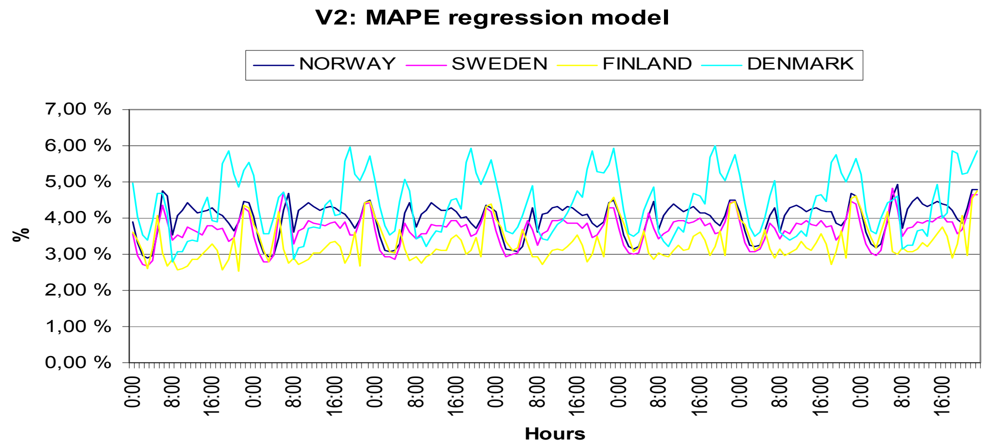

Figure 9.9 MAPE for all Nordic countries on an hourly basis in Winter 2.

MAPE follows a part of the daily rhythm, which can be seen from the figures above. Temperature is a slow variable and that can effect the daily rhythm of MAPE. During Winter 1 and Winter 2 the wide temperature changes can produce difficulties in forecasting. This temperature dynamics of the model should be developed.

For both winter seasons MAPE is relatively high for Denmark compared to the other Nordic countries. In Denmark the temperature does not affect the demand as much as in the other Nordic countries. The weather in the winter is also milder. From this point of view the model is overparametrizing the coefficients for Denmark.

In the summer the temperature does not affect the electricity demand as much as in the winter. Even if the regression model takes the temperature into account, the MAPE values are almost at the same level as in the winter.

MAPE values vary between 2.5% and 6% in all the Nordic countries and for all the seasons. This can be considered a fairly good result considering that the input data may not be the most accurate.

The amount of out of bound values is high for every season in Denmark. Temperature dependence of demand has little significance in forecasting demand. The regression model estimates the parameter for the forecasts as too high and the load forecast results are much more unstable than for the other Nordic countries.

One of the reasons for the forecast errors that are over 15% is special days. The special days cause the structural change to the load that is not dependent from temperature. The regression model is using the Finnish calendar, and this causes more errors to the results for the other Nordic countries. In every season with a longer special day period the demand is lower than on a normal Sunday. The diminished industrial load affects the demand and the model does not take this kind of changes into account.

9.2 Results of the stochastic time-series model

The program that is based on the ARX model is implemented into Excel. The program is designed so that it forecasts the load on an hourly basis. Because of the limitations of the software, the model is only used for hourly load forecasting and the parameters can be estimated from the previous four weeks. The forecast errors are analysed in the same way as in chapter 8.

9.2.1 Distribution of the load forecast error values for the stochastic time-series model

The ARX model is using the latest realised load values as a basis for the load forecast. The temperature factors are then correcting the previous load upwards or downwards. The forecast errors are calculated for all seasons and the histograms are presented in the same way as in the regression model. Both the 28th and the 34th hour of the load forecasts are evaluated. In the ARX model the forecasts are affected by the same inaccuracies of the input data as in the regression model.

The ARX model does not include the special day calendar. In the original daily ARX model the special days are corrected manually. The special days, which appear on weekdays, are not corrected with the Sunday load pattern. In the time-series model these kinds of errors repeat themselves in the later forecasts and also in the parameter estimation. Despite of this shortage the model is tested and this is taken into account in the analyses.

In Winter 1 the forecast errors for Norway are settled more to the right side of the histogram. This implies that the parameters are estimated as too high. Easter is also messing up the forecast model and is affecting almost a month in the forecasts. The histograms for Winter 1 are presented in figures 9.10-9.11.

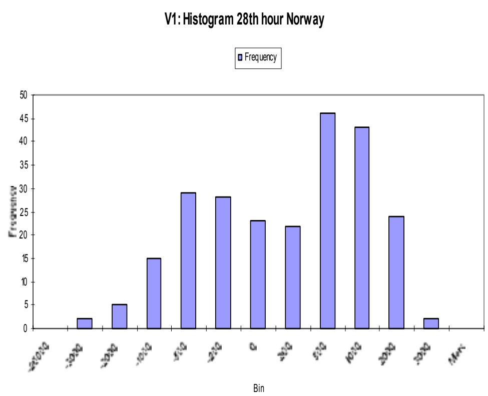

Figure 9.10 Histogram 28th hour from the beginning of the load forecast in Winter 1.

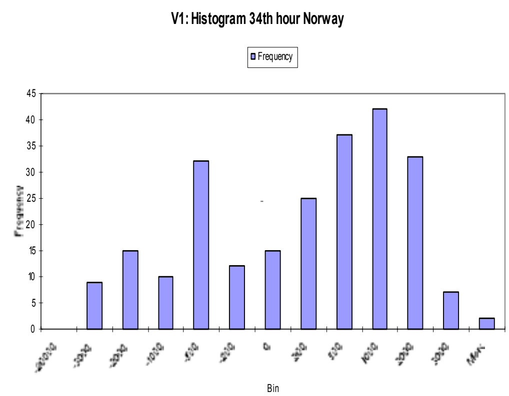

Figure 9.11 Histogram 34th hour from the Beginning of the load forecast in Winter 1.

In the Summer the forecast errors are quit close to zero mean. Temperature is not affecting the load and the model has estimated the parameters at the right level. There are errors that are more than –3000 MW. The errors begin to grow exponentially, and during Midsummer in Finland the model could not calculate any of the forecasts for the time periods 21.6-23.6.2001 and 20.6-22.6.2002. The histograms are shown in figures 9.12-9.13.

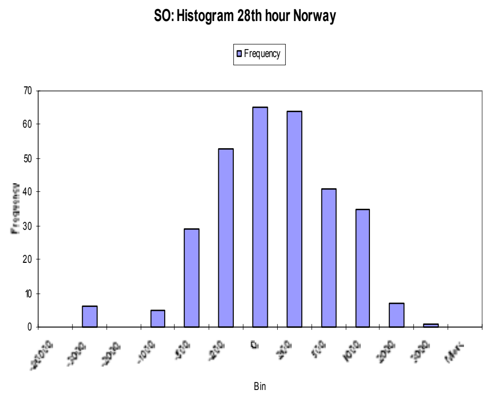

Figure 9.12 Histogram 28th hour of the load forecast in Summer.

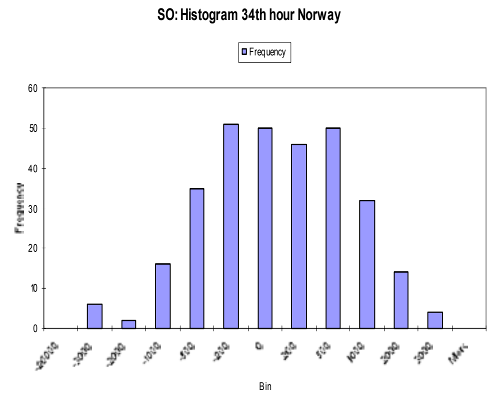

Figure 9.13 Histogram 34th hour of the load load forecast in Summer.

In Winter 2 the forecast errors are more diffused than in any other season. In this case Christmas is causing the larger errors. Demand is highly dependent on temperature in the winter but the model cannot correct for the lower demand during Christmas. The demand of industry and trade and the service sectors is lower than normal. Winter 2 has the same problem as Summer, the load forecast is growing exponentially and the model could not calculate any of the load forecasts for 25.12-26.12.2001 and 25.12-26.12.2002. The histograms are shown in figures 9.14-9.15.

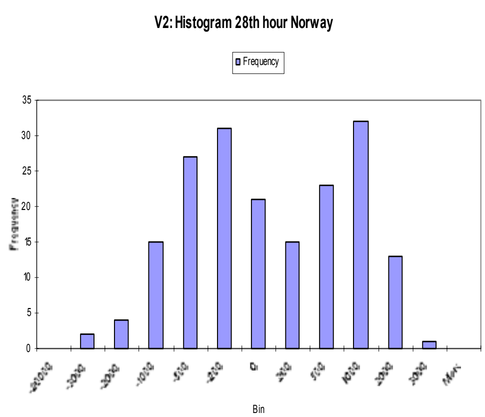

Figure 9.14 Histogram 28th hour of forecast in Winter 2.

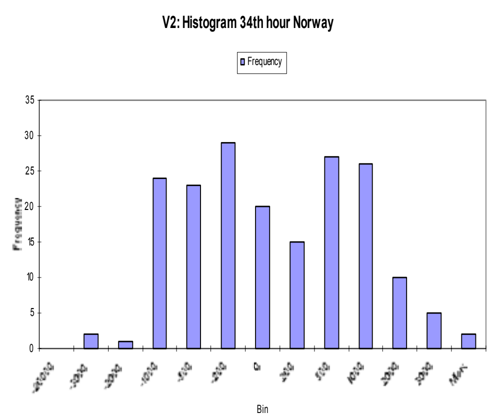

Figure 9.15 Histogram 34th hour of the load forecast in Winter 2.

The numerical values of the histograms are presented in appendix 3. As for the regression model, the average demands for the same hours are calculated. The results show that 60-80% of the forecast errors are between –1000 MW and 1000 MW. This is under 10% of average demand. The percentage is shown in appendix 4.

9.2.2 Mean absolute percentage error analysis of the stochastic time series model

MAPE is used to evaluate the forecast errors of the ARX model. The analyses have been made in the same way as in the regression model. The cross section for every hour has been taken from the group of forecast errors.

Problems with forecasting demand in Denmark appeared in the ARX model. The ARX model uses the previous load values as a prediction. When the temperature factor was taken into account, the ARX model overparametrizised the load forecast. The results were too far from the realised values, and it was decided to use only the AR model for Denmark. This means that the temperature factors were left out. This proves that demand is not very temperature dependent in Denmark.

All the forecast errors that are higher than 15% of realised load have been left out from the hourly MAPE values. The special days are causing higher MAPE values for all Nordic countries and for all seasons. MAPE values on an hourly basis are presented in figures 9.16-9.18.

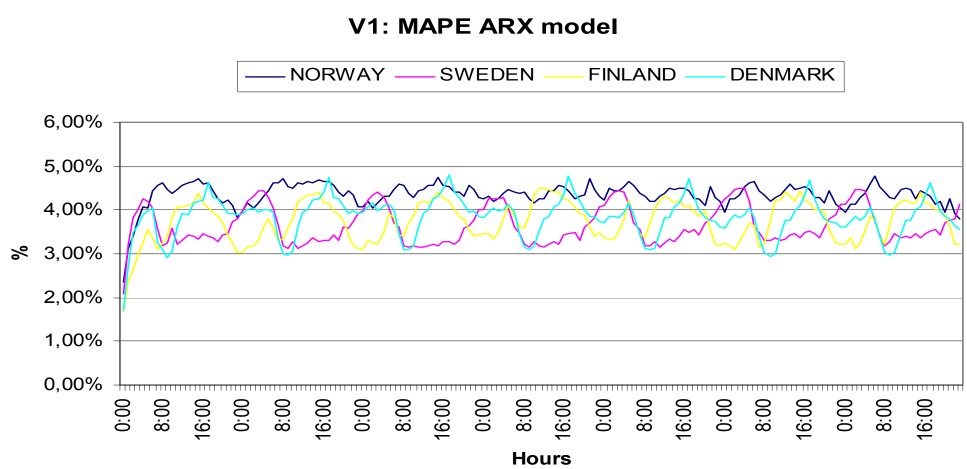

Figure 9.16 MAPE for all Nordic countries on hourly basis in Winter 1.

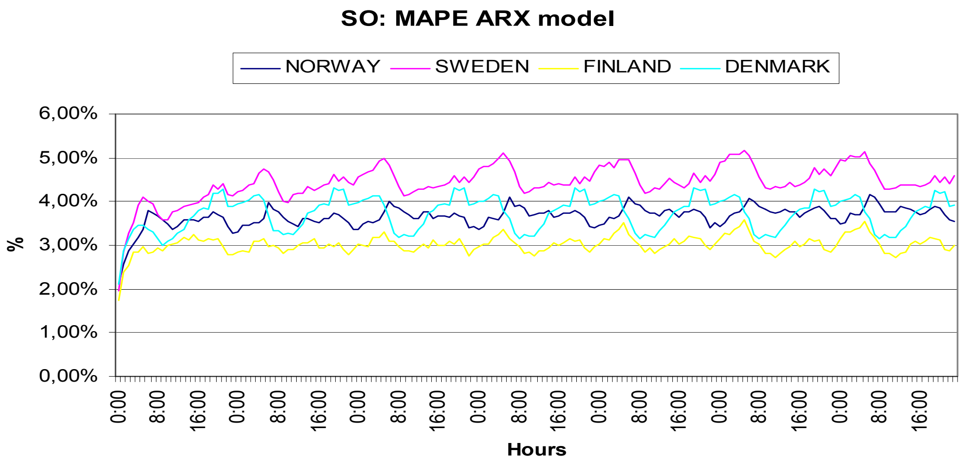

Figure 9.17 MAPE for all Nordic countries on hourly basis in Summer.

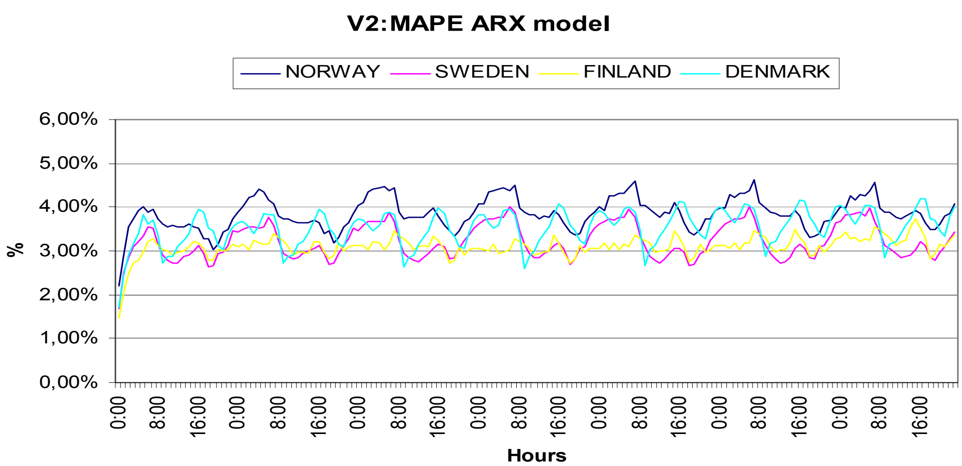

Figure 9.18 MAPE for all Nordic countries on hourly basis in Winter 2.

MAPE values show that the ARX model forecasts the first few hours of the forecast fairly well. After that the MAPE values begin to fluctuate, following some daily load pattern. The MAPE values are slightly increasing towards the end of the forecast period. In the ARX model the forecast error is repeated when the forecast period is extended.

In Winter 1 and Winter 2 the MAPE values are more stable for Norway than for the other Nordic countries. It can be assumed that the demand reacts to temperature changes quit quickly. MAPE values have been fairly stable for all Nordic countries also in the summer.

The MAPE values for ARX show that the model is managing the load forecasting well when there are no special situations. The MAPE values are varying between 1.5% and 5.5%.

One of the error sources is that the ARX model does not take the special days into account. Every season has some special days. Because the special days are not considered in any way, the forecast errors repeatedly appear in the forecasts. The special days are also affecting the estimation of parameters. The special days and the growth of the industrial load in the autumn are difficult to take into account with the structure of the ARX model.

Another reason is that the quality of the input data is poor. The ARX model is fairly sensitive to input data. Missing measuring values in the load data affects the load forecasts because the model is using the previous realised values as basis. The poor quality of the temperature data is also a probable cause for the high forecast errors.

The relative amount of errors is bigger for Denmark than for the other Nordic countries. A different structure of the time-series model is used to forecast the load for Denmark.

The number of errors is fairly high in the summer due to Midsummer in Finland. Industrial load around Midsummer is below half of the normal level. This is influencing almost all of the forecasts for the late summer. Also in Winter 2 the number of errors is high. The reason is Christmas and diminished industrial load.

9.3 Evaluating the results of the models

The load forecasts are made with the same input data in both the regression model and the ARX model. It enables the comparison of the models. The results on an hourly basis have been analysed in sections 9.1 and 9.2. The results on hourly basis show that the models have different qualities. The regression model is more stable when the whole data set is considered and when the number of errors is taken into account.

The ARX model forecasts normal situations (e.g. without any special days distracting the load forecast) fairly well. The periods around the special days are excluded from the analysis. The models are compared with each other by the MAPE values that are calculated from all the load forecasts for every season in every Nordic country. Figures 9.19-9.21 present the MAPE values for whole seasons in the Nordic countries.

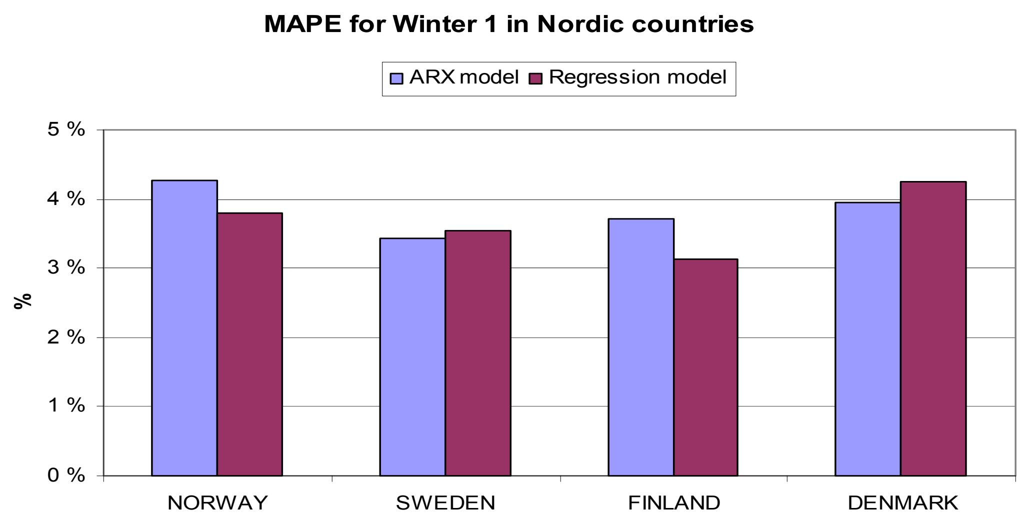

Figure 9.19 MAPE values for Winter 1 season in Nordic countries.

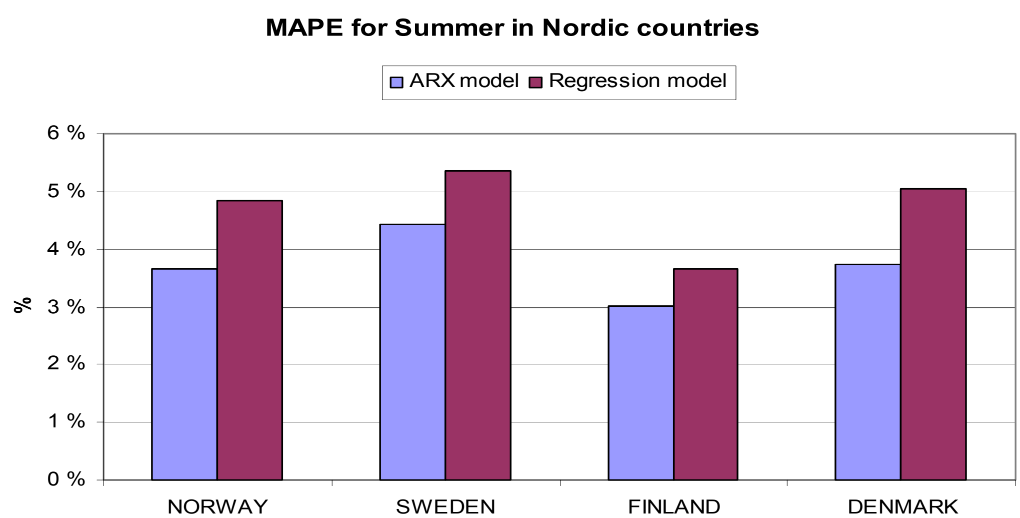

Figure 9.20 MAPE values for Summer season in Nordic countries.

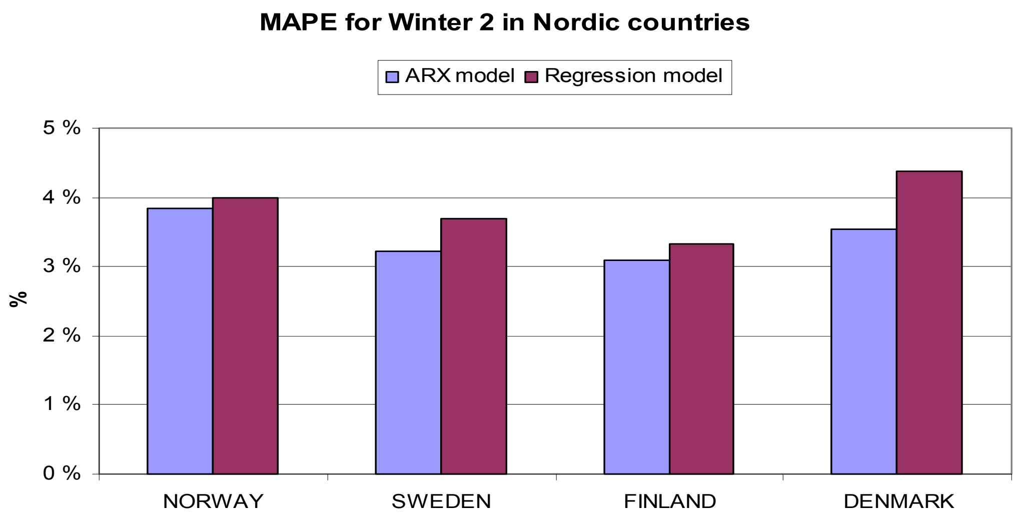

Figure 9.21 MAPE values for Winter 1 season in Nordic countries.

The results show that for Winter 1, and for Norway, the regression model is better than the ARX model. For the other seasons it seems that the ARX model is better. In the summer the ARX model is better at tracking the diminishing demand than the regression model. The difference between the MAPE values for the two models in Winter 2 is rather small.

In Sweden the ARX model is better than the regression model in every season. In Finland the regression model is better in Winter 1, but the ARX model is better in the other 2 seasons. In Denmark the ARX model is better than the regression model in every season.

The MAPE values show that the ARX model is better for any Nordic country in Summer. The difference is also much smaller than in both Winter seasons. Especially in Denmark, the larger error values are higher in the regression model than in the ARX model. This is achieved by the modification of the ARX model for Denmark. From these results one can conclude that in Summer the ARX model is better than the regression model in every Nordic country.

Overall, the preliminary conclusions that can be made, based on the previous results, is that the regression model is a better option to use for forecasting the electricity load in all the Nordic countries, except for Denmark. In the winter, the forecasts are more reliable in a changing situation with the regression model than with the ARX model. The ARX model is recommended to be used for forecasting the load in the summer.

In Denmark the temperature dependence of demand is so insignificant that it is better to exclude it from the model. If this is done, the ARX model achieves better results. It was not possible to test the forecasting ability of the regression model without the temperature factor.

To forecast the load with any of the models it is primarily important to check the quality of the input data. The quality of the load data is very important in the ARX model because it uses the previous realised values as basis for forecasting the electricity load. The regression model is more robust and not so sensitive to the input data.

10 USAGE OF THE MODELS

In this chapter the usage of the models is considered. The preliminary analyses have shown that the regression model is better to use for the winter in every Nordic country, except for Denmark. The ARX model is suitable for forecasting the electricity load in Denmark. When the temperature dependence is removed, the ARX model is suitable for forecasting the summer demand in the other Nordic countries as well.

The models can also be used for other purposes. The regression model is applied to studying the temperature dependence in the Nordic countries. The regression model is also briefly tested for estimating the parameters in longer periods and for the ability to forecast the electricity load for the whole summer.

The ARX model is restricted to estimating the parameters from the previous four weeks. Due to this reason the ARX model is tested only to forecast the electricity load without the temperature factors in Denmark.

10.1 Other applications for the regression model

10.1.1 Temperature dependence of demand

Out-door temperature is the most important factor that influences electricity demand in the Nordic countries, especially in the winter. To provide a suitable temperature dependence complicates the unlinearity of dependence. In this thesis the temperature dependence of demand is studied with the help of a regression model. The temperature dependence of demand is produced for the entire yearly profile and for all Nordic countries.

This thesis has been concentrating on evaluating the models' ability to forecast future demand. Examining the temperature dependence of the load has not been that important. Thus the temperature dependence is only shown in one case.

The temperature dependence of demand is assumed to be linear. The parameters of the model have been estimated from the same time period as from which the temperature dependence is produced. In this case the input data is gathered on an hourly basis from the whole year 2001. The specific demand for 2001 is calculated. The realised loads and the specific loads are compared to each other, as well as zero temperature to realised temperature. A scatter chart is applied to these data series and a linear regression line is fitted to the data points. The temperature dependence of demand in the Nordic countries is presented in figures 10.1-10.4.

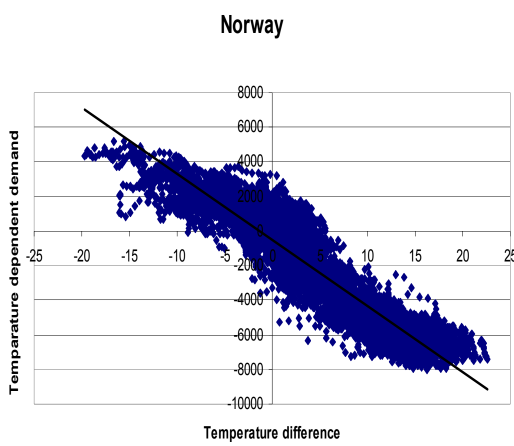

Figure 10.1 The temperature dependence in Norway.

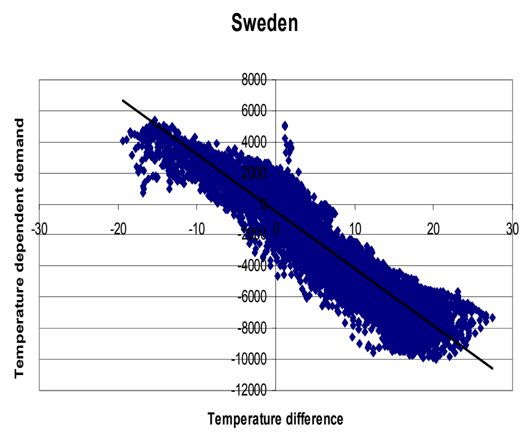

Figure 10.2 The temperature dependence in Sweden.

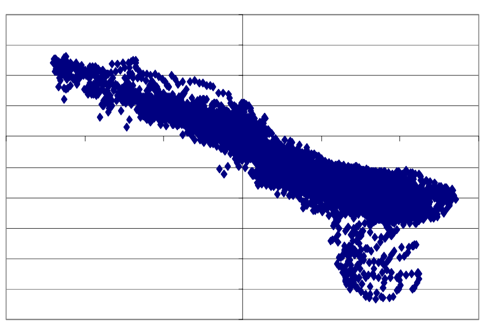

Figure 10.3 The temperature dependence in Finland.

Figure 10.4 The temperature dependence in Denmark.

The temperature dependence of demand is not linear as can be seen in the figures above. Temperature is quite a slow variable. A specific hour's temperature is not necessarily affecting the same hour's load, but can have an effect several hours later.

Dispersion of the data points can be caused by other factors that influence the demand. The weather and load measures are not as accurate as they are assumed to be and the result can be seen from the wideness of the scattered line. (Bunn and Farmer)

E.g. the demand of industry is not highly temperature dependent but corresponds to total electricity demand. The larger error values are mostly caused by changes in the electricity load of industry. Larger values also arise when special days are diminishing demand.

It is known that temperature dependence of the load is not linear, thus a nonlinear model could better describe the behaviour of dependence. The temperature dependence of the third degree would also be possible to calculate in some cases.

In this case the temperature dependence of demand is calculated for the whole year. One possibility is creating a separate linear regression line for all seasons. It would take into account the changes in temperature and their effect on the electricity load more accurately.

10.1.2 Parameters estimation period and the load forecast over summer

One of the problems in forecasting the load is a model's ability to handle structural changes in the load. The growth of demand in the autumn is difficult to handle with the help of previous loads.

Parameters for the load forecast models can be estimated for different time periods. Previously a four weeks time period has been used for forecasting the power demand. The effect of estimating the regression model parameters from the previous year is studied. Summer 2002 is used as forecasting period. The parameters are estimated from the summer 2001. The growth rate of demand during the year has been taken into account.

A correction for the Midsummer period is made in the load forecast for Finland. In Finland the electricity load of industry diminishes heavily a few days before Midsummer. The industrial load partly returns after two weeks. The correction rates are calculated from the load during the Midsummer period and a two weeks period before this Midsummer period. The correction of the Midsummer 2002 period compared with the realised load is presented in figure 10.5.

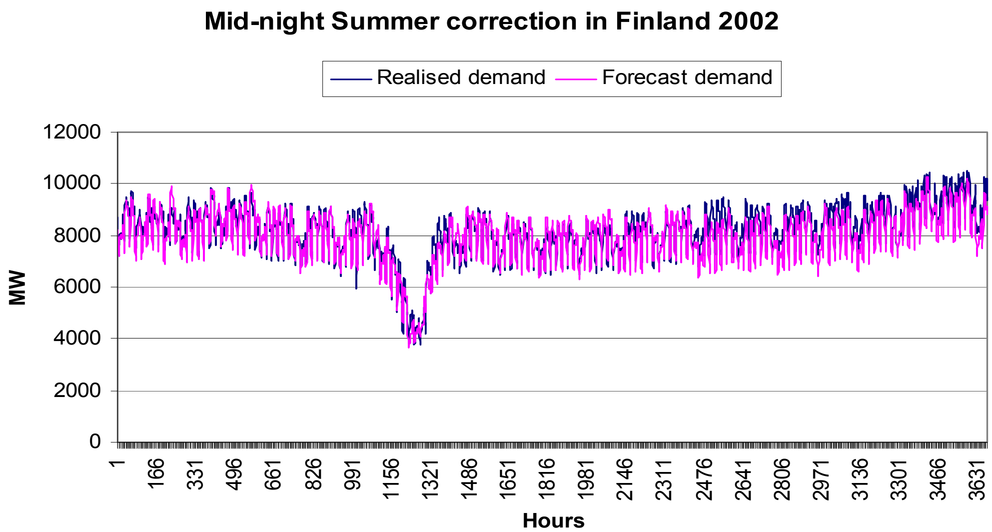

Figure 10.5 Forecasted and realised demand for Summer 2002.

The parameter estimation from the year before has benefits because the rhythm of the load for the whole season remains equal each year. It takes into account the rhythm that includes e.g. the special days and the holiday period in the summer. On the other hand, estimating the parameters from the year before does not consider the latest changes in the growth of demand. But this can be handled by using a short-term time-series model as an error structure in the regression model.

The forecast of the electricity load is made for the summer. One remark should be made. Although it is possible to estimate the overall demand profile for the whole summer, on the basis of the hourly forecast it becomes quite uncertain after a couple of weeks. MAPE that are calculated for the summer 2002 are shown in the figure 10.6.

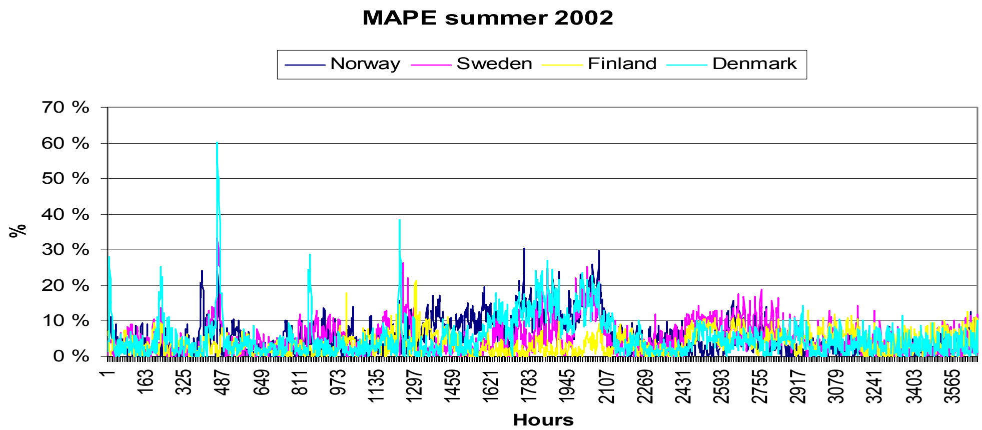

Figure 10.6 MAPE for every hour in summer 2002.

MAPE is also calculated from the whole summer 2002. MAPE for the Finnish load forecast is corrected for the Midsummer period. This correction improves the load forecast to a great extent. The load forecasts for the other Nordic countries are not corrected for the Midsummer period. This obviously affects the comparability of the loads of the different countries. The MAPE values for the whole summer 2002 with and without the corrected load forecasts is presented in figure 10.7.

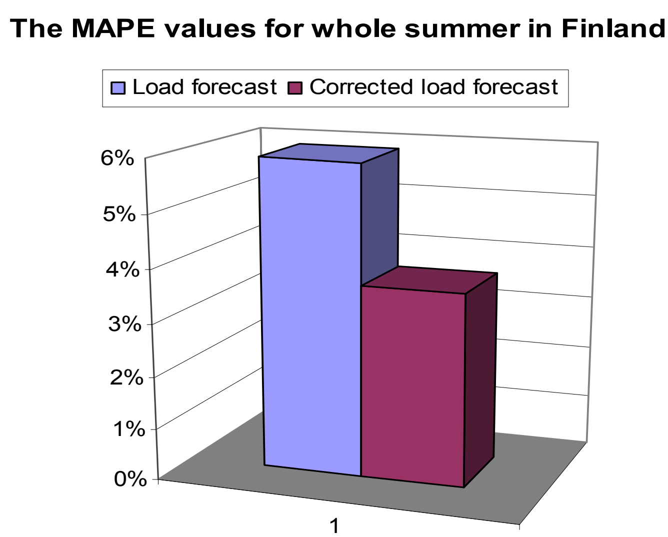

Figure 10.7 MAPE for the whole summer period of 2002 before and after corrected load forecasts in Finland.

10.2 Other applications of the time series model

The ARX model is tested for Denmark both with temperature factors and without them. The ARX model is easy to modify to exclude the temperature factors. The AR model might be a suitable model also for the other Nordic countries in the summer. Only the result of the load forecast with and without the temperature factors in Denmark are presented in figure 10.8.

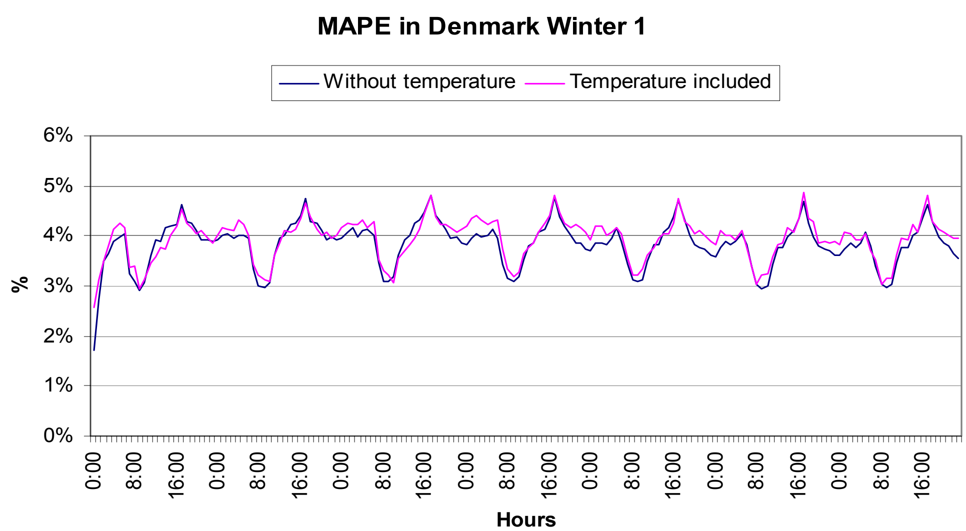

Figure 10.8 MAPE values in Denmark with and without the temperature factors in the Winter 1 season.

The real difference occurs in the amount of larger errors because there is a significant difference between the forecast that is made with the temperature factors and the one made without them. The errors of the load forecasts are more stable without the temperature factor and this proves that the AR model is better at forecasting the load in Denmark.

The special day calendar must be implemented into the AR model if it is to be used to forecast the load. One solution might be that the model should use the special days of previous years.

11 CONCLUSIONS AND PROPOSALS FOR FURTHER DEVELOPMENT

In this thesis two different electricity load models have been analysed. Both models are handled separately. The results show that the development of these models should continue. Some suggestions for achieving better load forecasts for the Nord Pool market are presented in this chapter.

11.1 Data improvements

Nord Pool provides the load data. The system operators in the Nordic countries supply the data to Nord Pool. The load data is published real time. The real time data may be partly inaccurate and the system operators correct the realised data afterwards. The data used in the load forecasts should be the corrected data.

Missing measurements are also a problem. Missing data in the real time load values can occur. Correcting the realised load data with the system operators' data can solve this problem.

The missing temperature data is much more complicated to correct afterwards. Foreca supplies the temperature data every 3-6 hours in real time. Getting missing data afterwards is usually difficult. One solution is to use the supplied forecasted temperature values. Another solution is to use the average temperatures from the historical temperature data.

11.2 Improvements to the temperature dynamics and the functional form

The AR model is better when the load is forecasted for Denmark. This model does not use temperature values at all. Temperature dependence of demand for Denmark has been handled in section 10.1.1. The structure of demand in Denmark differs from the other Nordic countries. The industrial sector is not a heavy user of electricity.

The analyses show that the temperature dependence of demand for the summer is not playing the big part in the load forecasts. The regression model and the ARX model overparametrized the load forecasts when the temperature factors were included. The AR model may be the better solution for the summer.

In this thesis the temperature dependence of demand is considered as linear. In reality the dependence is not linear. A piecewise linearity could describe the behaviour of demand better. One solution might also be that the central part of the temperature dependence could be linear while the boundaries could be curved.

11.3 Suggested approaches in the load forecasting

The regression model can use the realised load data from the previous year to form the basic rhythm of the load and the temperature dependence. The basic rhythm of the load and the temperature dependence will be used to forecast the following years' load. To be able to forecast the load from two factors, one has to include the increase of the load from the previous year.

A preliminary study is made to forecast the load with the basic rhythm of the load and the temperature dependence. The parameters are estimated from the period 11.6 - 9.7.2001. The realised load values for Finland from the estimation period are presented in figure 11.1.

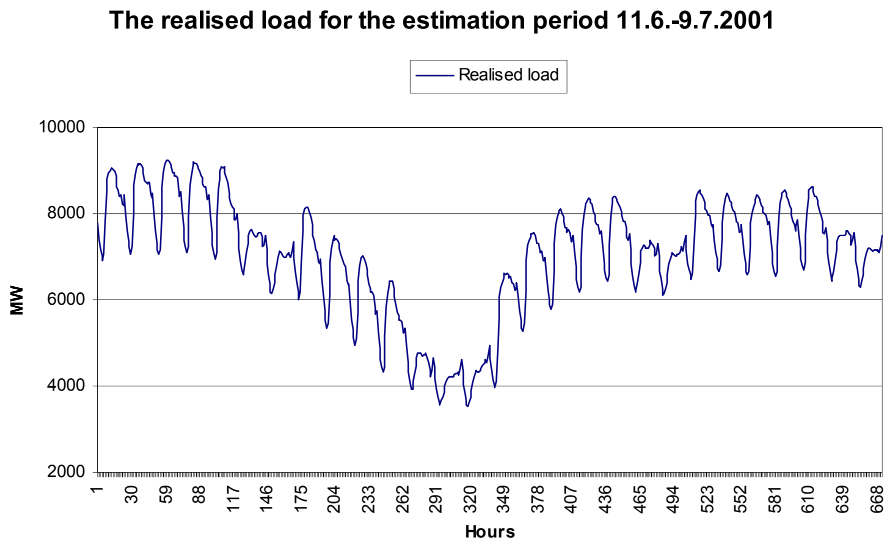

Figure 11.1 The realised load for the estimation period for Finland.

The specific demand is estimated from the same time period for the previous year. The estimation period is four weeks. The specific demand is corrected with the relative growth of demand from the previous year. The Midsummer correction in the load forecast is also taken into account in the same way as in section 10.1.1. The specific demand and the Midsummer corrected specific demand are presented in figure 11.2.

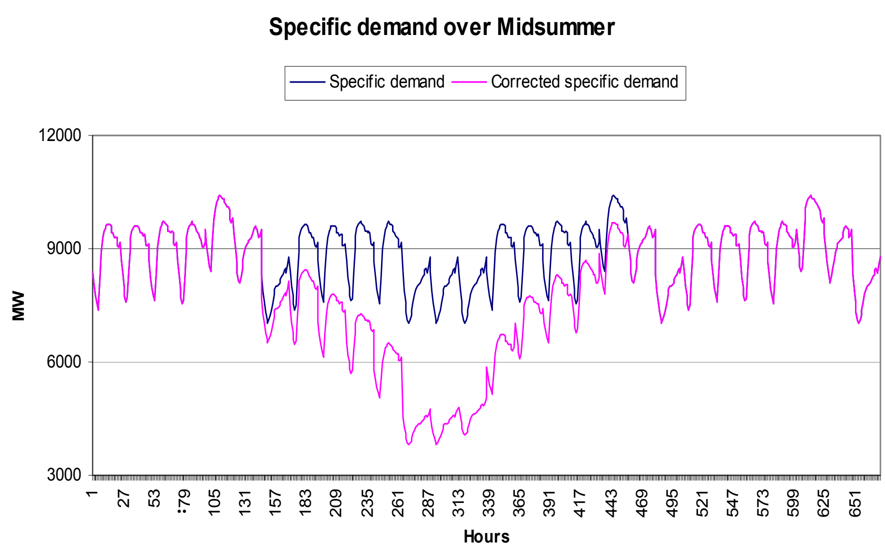

Figure 11.2 The specific load and the Midsummer corrected specific demand 10.6.-8.7.2002.

The load is not very dependent on temperature in the summer. Other structural changes have more affect on the load. The temperature index and the realised load are presented in figure 11.3.

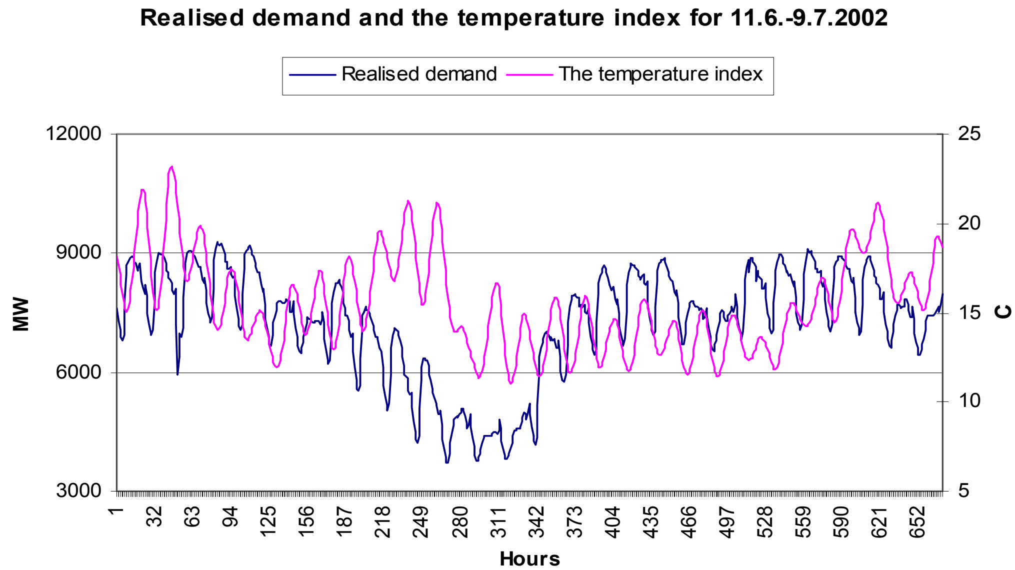

Figure 11.3 The realised demand and the temperature index for the 11.6.-9.7.2002

The load forecast is made for 11.6 - 9.7.2002. The realised load and the forecasted load with and without the Midsummer correction are presented in figure 11.4.

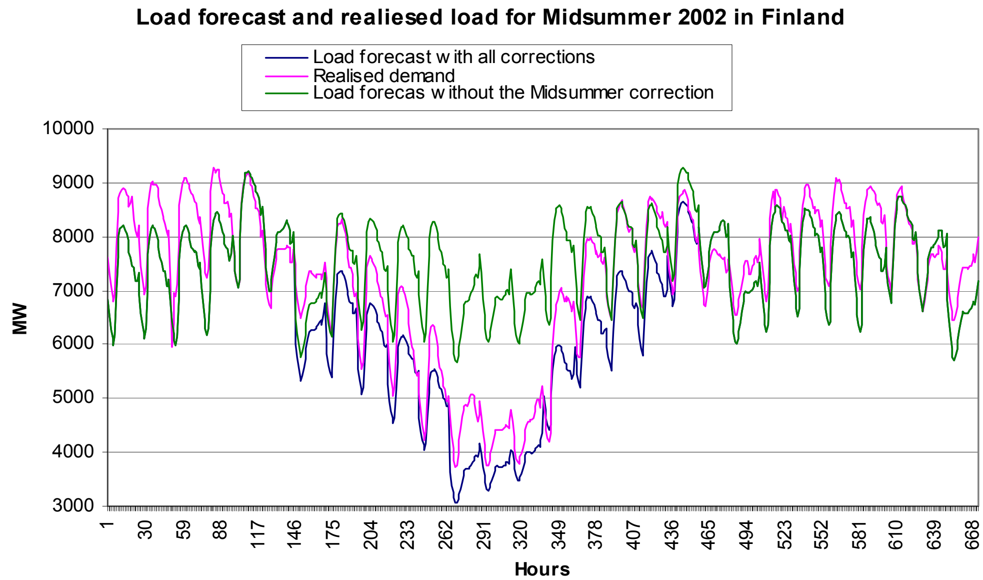

Figure 11.4 The realised load and the load forecast with and without the Midsummer correction 2002 in Finland.

The figure shows that the load forecasts without the Midsummer correction are following the realised load to some extent. The corrected forecasts for Midsummer are much closer to the realised values and are rather easy to produce. The correction rates are calculated from the load during the Midsummer period and a two weeks period before the Midsummer period from the previous year. This approach might be useful in further studies.

The special days cause difficulties because the load pattern for these days differs from the Sunday load pattern. The demand on special days is much lower than demand on a normal Sunday. One solution could be the usage of the specific special day model. The model can be used in addition to the original model. The specific special day model can be used when special days occur. In the literature specific special day models have been studied. The approaches in these models can be modified to satisfy the needs.

The time-series model can be used to improve the short-term forecasting ability. The regression model is a more stable model for load forecasting. The error term of the regression model can be modelled with the time-series model. The ARMA-process may be a proper solution and it has already been examined in the previous studies. The solution combines the good properties of the models; accuracy and robustness. (Karanta, Ruusunen 1995)

11.4 Conclusions

The time-series model uses the previous realised load values to forecast the load. The model takes into account the latest changes that occurs in demand. The time-series model forecasts the load more accurately for the first few hours of the forecast. On the other hand, the time-series model is very sensitive to measuring errors. The errors repeat themselves as the forecast proceeds.

The regression model creates the specific load for the forecast. The estimated specific load is stable and the measuring errors are less disturbing than in the time-series model. The load forecasts are more accurate for a longer time span when using the regression model than when using the time-series model. The time-series model is observed to start deviating when the time span exceeds ten days. The regression model is more robust than the time-series model.

12 SUMMARY

Hourly load forecasting is a key activity for the players in the physical power market. In this thesis the availability and the quality of the hourly load and the temperature data for the Nord Pool market have been analysed. The realised electricity load data is provided by Nord Pool and the emperature data by Foreca. The electricity load forecast is studied for the Nord Pool market area.

The factors that can influence electricity demand are electricity usage of the industry and e.g. temperature, wind speed, sun radiation and humidity. In this thesis only the effect of temperature is taken into account. A considerable part of the residential, commercial and service sectors' heating in these countries is electrical space heating. Electrical space heating is highly influenced by the temperature fluctuation.

The realised load values are hourly averages. The load data is published real time by Nord Pool. Real time data is partly an estimate from the actual demand. The system operators publish corrected data with some delay. The temperature data is linearised for every hour from the realised values published every 3-6 hours. Temperatures are weighted into one temperature index for each of the Nordic countries.

The load profile is studied for each Nordic country. The yearly load profile and the weekly and hourly rhythm of the loads are presented. The Midsummer in the middle of June and Christmas at the end of the year produce the largest downward peaks, especially in the Finnish demand profile. The weekly rhythm consists of the weekday pattern and the weekend pattern. The demand during weekends is lower than during weekdays. People's behaviour during the day impacts on the daily rhythm. The electrical space heating in Finland is connected late in the evening, which also affects the daily load rhythm.

In this thesis a regression model and one of the Box-Jenkins time-series approaches (as a so-called autoregressive model) have been compared. The regression model theory is based on the specific load that is the characteristic rhythm of the load and temperature fluctuation. The ARX model is based on the previous load and the temperature fluctuation. Both models are implemented on an hourly basis. The parameters of the models are estimated with the least-squares method.

The influence of temperature is treated as a linear one parameter model. It is assumed that demand depends linearly on the average of the previous hours out-door temperature values. The load forecasts are made for both models using load and temperature data from the beginning of 2001. The estimation of parameters in both models requires input data from a four weeks period before the load forecast itself begins.

The number of load forecasts is relatively high in this study. The load forecast are evaluated by the distribution of the forecast errors, which are shown in histograms.

The ARX model shows that the temperature dependence of demand in Denmark is not very important. The model was modified for Denmark by removing the temperature factors. The results were better without the temperature factors.

MAPE is chosen for the evaluation of the model accuracy. MAPE is a unit-free measure. Therefore it is easy to compare the results between the regression model and the stochastic time-series model.

MAPE showed that the ARX model is better for all Nordic countries in the summer. The difference is also much smaller than in both the winter seasons. Especially for Denmark, the larger error values are higher for the regression model than for the ARX model. One can conclude that the ARX model is better than the regression model for every Nordic country in the summer. The analysis showed that the regression model performs better in the winter for all other Nordic countries except for Denmark.

The models were also studied for other purposes. The regression model was applied to examine the temperature dependence in the Nordic countries. The regression model was briefly tested for estimating the parameters of the longer period and for the ability to forecast the electricity load for the whole summer season.

The temperature dependence of the load is not linear, thus a nonlinear model could better describe the behaviour of dependence. The temperature dependence of the shape of the curve function can be applied.

The load forecast for the summer in Finland was studied. The parameters were estimated from the year before. The regression model can use the realised load data from the previous year to form the basic rhythm of the load and the temperature dependence. The Midsummer correction and the growth of demand from the previous year were taken into account in the load forecast. MAPE for the Finnish load showed that the correction improves the load forecast to a great extent.

The development of the regression model and the ARX model should continue. Some suggestions for load forecast improvements were presented. The load and temperature data have to be corrected and checked afterwards. The temperature dynamics and functional forms have to be considered and the possibilities to use other functional forms for the temperature dependence could be studied.

It might be better to estimate the parameters from the previous year. The basic rhythm of the load and the temperature dependence is used to forecast the following year's load. The time-series model can be used to improve the short-term forecasting ability. The error term of the regression model can be modelled with the time-series model. For the special days it would be better to use the specific special day model.

LIST OF SOURCES

Anon., 2002, Electricity Market 2002. Swedish Energy Agency, Eskilstuna.

Anon., 2001. Nordel Annual Report 2001. Nordel. Oslo.

Bergman L,. Brunekreeft, G,. Doyle, C., von der Fehr, N.-M., Newbery, D,. Pollitt, M., Régibeau, P., and Vaitilingam, R., 2001 , A European Market for Electricity? Centre for Economic Policy Research, London.

Box, G. E. P, Jenkins, G. M.. 1971 Time series analysis, forecasting and control, Holden-Day, San Fransisco.

Bunn, D. W., Farmer, E. D., 1985. Comparative Models for Electrical Load Forecasting. John Wiley & Sons, Belfast.

Charytoniuk W., Niebrzydowski J. 1998. Confidence interval construction for load forecast. Electricity Power Systems Research Vol 48, June 1998, pp. 97-103.

Chen, Shin-Tzo, Yu, David C., Moghaddamjo, A.R. 1992. Weather sensitive short-term load forecasting using nonfully conected artificial neural network. IEEE Transactions on Power Systems, Vol 7, No3, August 1992, pp. 1098-1105.

Eurelectric, 2001a. Statistics and prospects for the European electricity sector (1980-1999, 2000-2020). EURPROG Network of Experts. Ref: 2001-2745-0002. Brussels, November 2001.

Gillberg, Mika. 1997. Sähkönkulutuksen ennustaminen poikkeamasähkökaupaassa. Sovelletun matematiikan erikoistyö, Helsinki University of Technology. Espoo.

Karanta, Ilkka, Ruusunen, Jukka. 1991. A model for electricity demand forecasting in communal electric utility. Research report A40, Systems Analysis Laboratory, Helsinki University of Technology. Espoo.

Karanta, Ilkka, Ruusunen, Jukka. 1988. Sähkön ja kaukolämmön kulutuksen lyhyen aikavälin ennustamisjärjestelmä energialaitokselle. Research Reports B13, Systems Analysis Laboratory, Helsinki University of Technology. Helsinki.

Moghram, Ibrahim, Rahman, Saifur. 1989. Analysis and evaluation of five short-term load forecasting techniques. IEEE Transactions on Power Systems, Vol. 4, No. 4, October 1989, pp. 1485-1491.

Murto, Pauli. 1998. Neural network models for short-term load forecasting. Master's Thesis, Department of Engineering Physics and Mathematics, Helsinki University of Technology. Espoo.

Nord Pool (2003) http://www.nordpool.no/organisation/index.html 9.1.2003

Papalexopoulos, A. D., Hesterberg, T. C., 1990. A regression-based approach to short-term system load forecasting. IEEE Transactions on Power Systems, Vol. 5, No. 4, Novenber 1990, pp. 1535-1550.

J.H. Park, Y.M. Park, K.Y. Lee, 1991, Composite modelling for adaptive short-term load forecasting, IEEE Transactions on Power Systems, Vol. 6, No. 2, May 1991, pp. 450-457.

Pindyck, R. S., Rubinfeld, D. L., 1991. Econometric models & economic forecasts, McGraw-Hill, Singapore.

Räsänen, Mika, Ruusunen, Jukka. 1992. Verkoston tilan seuranta mittauksilla ja kuormitusmalleilla. Research Reports B17. Systems Analysis Laboratory, Helsinki University of Technology. Espoo.

Räsänen, Mika, Ruusunen, Jukka, Hämäläinen, Raimo P. 1995. Customer level analysis of dynamic pricing experiments using consumption-pattern models. Energy Vol. 20, No. 9, pp. 897-906

Räsänen, Mika, Ruusunen, Jukka. 1996. Object-Oriented Modeling software for Electric Load Analysis and Simulation. Simulation 66:5, May 1996, pp. 275-288.

Sanders, Nada R., 1997. Measuring forecast accuracy: some practical suggestions. Production and Inventory Management Journal, First Quarter 1997, pp. 43-46.

Söderström, Torsten, Stoica, Petre 1989. System Identification, Prentice Hall, Hertfordshire.

Taylor, James W., Buizza, Robero, 2002. Neural network load forecasting with weather ensemble predictions. IEEE Transactions on Power Systems, Vol 17, No. 3, August 2002, ppl 626-632.

Tilastomatematiikan opintomoniste, 1996, Lappeenrannan Teknillinen Korkeakoulu, Tietotekniikan osasto. Lappeenranta.

APPENDIX 1

Numeric values of histograms at 28th and 34th hour of the beginning of the forecast for regression model.

Winter 1 (1.1.-30.4.)

28th hour

Bin	Frequency
-20000	0
-3000	1
-2000	0
-1000	4
-500	23
-200	29
0	22
200	25
500	50
1000	56
2000	25
3000	2
More	2

34th hour

Bin	Frequency
-20000	0
-3000	0
-2000	0
-1000	4
-500	16
-200	23
0	29
200	39
500	56
1000	51
2000	20
3000	1
More	0

Summer (1.5.-30.9.)

28th hour

Bin	Frequency
-20000	0
-3000	0
-2000	0
-1000	6
-500	36
-200	63
0	46
200	36
500	62
1000	45
2000	12
3000	0
More	0

34th hour

Bin	Frequency
-20000	0
-3000	0
-2000	1
-1000	23
-500	60
-200	42
0	30
200	22
500	43
1000	58
2000	24
3000	3
More	0

Winter 2 (1.10.-31.12.)

28th hour

Bin	Frequency
-20000	0
-3000	0
-2000	0
-1000	5
-500	42
-200	51
0	29
200	22
500	23
1000	11
2000	1
3000	0
More	0

34th hour

Bin	Frequency
-20000	0
-3000	1
-2000	2
-1000	31
-500	55
-200	38
0	18
200	10
500	10
1000	12
2000	3
3000	3
More	1

APPENDIX 2

Forecast errors in regression model that occurs at 3 o'clock and 9 o'clock compared to the average demand

Winter 1

28th hour

Error in MW	Norway
-20000	130 %
-3000	20 %
-2000	13 %
-1000	7 %
-500	3 %
-200	1 %
0	0 %
200	1 %
500	3 %
1000	7 %
2000	13 %
3000	20 %

34th hour

Error in MW	Norway
-20000	111 %
-3000	17 %
-2000	11 %
-1000	6 %
-500	3 %
-200	1 %
0	0 %
200	1 %
500	3 %
1000	6 %
2000	11 %
3000	17 %

Summer

28th hour

Error in MW	Norway
-20000	207 %
-3000	31 %
-2000	21 %
-1000	10 %
-500	5 %
-200	2 %
0	0 %
200	2 %
500	5 %
1000	10 %
2000	21 %
3000	31 %

34th hour

Error in MW	Norway
-20000	168 %
-3000	25 %
-2000	17 %
-1000	8 %
-500	4 %
-200	2 %
0	0 %
200	2 %
500	4 %
1000	8 %
2000	17 %
3000	25 %

Winter 2

28th hour

Error in MW	Norway
-20000	153 %
-3000	23 %
-2000	15 %
-1000	8 %
-500	4 %
-200	2 %
0	0 %
200	2 %
500	4 %
1000	8 %
2000	15 %
3000	23 %

34th hour

Error in MW	Norway
-20000	123 %
-3000	19 %
-2000	12 %
-1000	6 %
-500	3 %
-200	1 %
0	0 %
200	1 %
500	3 %
1000	6 %
2000	12 %
3000	19 %

APPENDIX 3

Numeric values of histograms at 28th and 34th hour of the beginning of the forecast for ARX model.

Winter 1 (1.1.-30.4.)

28th hour

Bin	Frequency
-20000	0
-3000	2
-2000	5
-1000	15
-500	29
-200	28
0	23
200	22
500	46
1000	43
2000	24
3000	2
More	0

34th hour

Bin	Frequency
-20000	0
-3000	9
-2000	15
-1000	10
-500	32
-200	12
0	15
200	25
500	37
1000	42
2000	33
3000	7
More	2

Summer (1.5.-30.9.)

28th hour

Bin	Frequency
-20000	0
-3000	6
-2000	0
-1000	5
-500	29
-200	53
0	65
200	64
500	41
1000	35
2000	7
3000	1
More	0

34th hour

Bin	Frequency
-20000	0
-3000	6
-2000	2
-1000	16
-500	35
-200	51
0	50
200	46
500	50
1000	32
2000	14
3000	4
More	0

Winter 2 (1.10.-31-12.)

28th hour

Bin	Frequency
-20000	0
-3000	2
-2000	4
-1000	15
-500	27
-200	31
0	21
200	15
500	23
1000	32
2000	13
3000	1
More	0

34th hour

Bin	Frequency
-20000	0
-3000	2
-2000	1
-1000	24
-500	23
-200	29
0	20
200	15
500	27
1000	26
2000	10
3000	5
More	2

APPENDIX 4

Forecast errors in ARX model that occurs at 3 o'clock and 9 o'clock compared to the average demand

Winter 1

28th hour

Error in MW	Norway
-20000	130 %
-3000	20 %
-2000	13 %
-1000	7 %
-500	3 %
-200	1 %
0	0 %
200	1 %
500	3 %
1000	7 %
2000	13 %
3000	20 %

34th hour

Error in MW	Norway
-20000	111 %
-3000	17 %
-2000	11 %
-1000	6 %
-500	3 %
-200	1 %
0	0 %
200	1 %
500	3 %
1000	6 %
2000	11 %
3000	17 %

Summer

28th hour

Error in MW	Norway
-20000	207 %
-3000	31 %
-2000	21 %
-1000	10 %
-500	5 %
-200	2 %
0	0 %
200	2 %
500	5 %
1000	10 %
2000	21 %
3000	31 %

34th hour

Error in MW	Norway
-20000	168 %
-3000	25 %
-2000	17 %
-1000	8 %
-500	4 %
-200	2 %
0	0 %
200	2 %
500	4 %
1000	8 %
2000	17 %
3000	25 %

Winter 2

28th hour

Error in MW	Norway
-20000	153 %
-3000	23 %
-2000	15 %
-1000	8 %
-500	4 %
-200	2 %
0	0 %
200	2 %
500	4 %
1000	8 %
2000	15 %
3000	23 %

34th hour

Error in MW	Norway
-20000	123 %
-3000	19 %
-2000	12 %
-1000	6 %
-500	3 %
-200	1 %
0	0 %
200	1 %
500	3 %
1000	6 %
2000	12 %
3000	19 %