SHORT TERM LOAD FORECASTING WITH NEURAL NETWORKS

Oscar A. Garrido

Master's thesis
2012:E10

LUND UNIVERSITY

Faculty of Science
Centre for Mathematical Sciences
Mathematical Statistics

CENTRUM SCIENTIARUM MATHEMATICARUM

TYP AV DOKUMENT		DOKUMENTBETECKNING
☒ Examensarbete	☐ Kompendium	LUNFMS-3036-2012
☐ Delrapport	☐ Rapport	

INSTITUTION

Matematikcentrum. Matematisk statistik, Lunds universitet, Box 118, 221 00 LUND

FÖRFATTARE

OSCAR GARRIDO SUXS

DOKUMENTTITEL OCH UNDERTITEL

Abstract:
The purpose of this thesis is to get knowledge of which parameters and settings are suitable for making an accurate forecast of the electricity load. The focus will be on one day ahead forecasts. The importance of accurate electricity load forecasts are high since they are used for energy trading and production planning. For instance in the Nordic power exchange market they have trading of energy contracts for a day-ahead. The production of electricity is a complex routine which is very cost and time consuming. Non optimized planning could therefore have severe consequences on the budget. The approach is based on trying different input variables for a feed forward neural network. The study involves analysis of lagged variables, weather variables and deterministic time components for capturing the seasonality profile of the electricity load. The conclusions that can be drawn is that raw time variables are the best deterministic time components for modeling the electricity load profile.

NYCKELORD

DOKUMENTTITEL OCH UNDERTITEL - SVENSK ÖVERSÄTTNING AV UTLÄNDSK ORIGINALTITEL

Lastprognoser med neutrala nätverk

UTGIVNINGSDATUM		ANTAL SID	SPRÅK		
år 2011	mån 12		☐ svenska	☒ engelska	☐ annat

ÖVRIGA BIBLIOGRAFISKA UPPGIFTER	
	ISSN
	ISBN
	2012:E10

Signature ______________________ Date ____________

Abstract

The purpose of this thesis is to get knowledge of which parameters and settings are suitable for making an accurate forecast of the electricity load. The focus will be on one day ahead forecasts. The importance of accurate electricity load forecasts are high since they are used for energy trading and production planning. For instance in the Nordic power exchange market they have trading of energy contracts for a day-ahead. The production of electricity is a complex routine which is very cost and time consuming. Non optimized planning could therefore have severe consequences on the budget. The approach is based on trying different input variables for a feed forward neural network. The study involves analysis of lagged variables, weather variables and deterministic time components for capturing the seasonality profile of the electricity load. The conclusions that can be drawn is that raw time variables are the best deterministic time components for modeling the electricity load profile.

Contents

4

Acknowledgment

I wish to express my gratitude to all who have contributed to this Master's Thesis. First I would like to thank Erik Lindström for his help and guidance. I would also like to thank my supervisor Björn Malmström and the Energy Opticon staff for giving me the opportunity to research within the field. Finally I would like to thank my family and friends for being there for me with unconditional support.

Thank you

April 23, 2012

Chapter 1

Introduction

1.1 Purpose

This thesis covers investigation of artificial neural network models for short term electricity load forecasts. The purpose is to get knowledge of which parameters and settings are suitable for making an accurate forecast of the electricity load. The implementation takes in to account several weather variables and seasonal dependencies. The focus will be on one day ahead forecasts (24 hours ahead). The importance of accurate electricity load forecasts are high since they are used for energy trading and production planning. For instance in the Nordic power exchange market(Elspot) they have trading of energy contracts for a day-ahead. The production of electricity is a complex routine which is very cost and time consuming. Non optimized planning could therefore have severe consequences on the budget. The desired forecast may vary and is usually defined in three types of length.

- from one year and ahead (also referred as long term) is used for long term planning, maintenance and extension of production units.
- from one week up to one year ahead (also referred as medium term) is mainly used for planning and maintenance.
- from one hour to one week ahead (also referred as short term) is used for production planning and energy trading.

The focus of this paper is on short term forecast models.

1.2 Energy Opticon

Energy Opticon is one of the leading providers of software with economic production optimization and load forecasting for energy producing companies. The software tool is called Energy Optima and delivers systems with district heating forecasts, electricity load forecasts, production optimization, long term planning and electricity trading. Energy Opticon currently has customers in Sweden, Denmark, Finland, Germany, Netherlands, France and Switzerland [12].

1.3 Outline of the thesis

Chapter 2 covers the data description and analysis. In Chapter 3 the theory for neural networks is presented. In Chapter 4 the selected neural network models are trained and evaluated. Chapter 5 contains a conclusion of the model evaluation with some further suggestions in areas of research for a best fit model.

Chapter 2

Data Analysis

In this chapter the data for creating a forecast model of the electricity load demand is presented. In order to be able to identify the best fit model it's necessary to analyze which type of variables are able to capture the dependencies of the load. There exists a huge amount of articles and information around which variables are important for the electricity load forecast.

Lagged variables are the most commonly used input variables when creating a model for the electricity load demand [9]. The importance of the input variables may differ depending on the load consumption type. There are three types of main classes of electricity consumption types. These are the Industrial-, commercial- and domestic consumption. For the industrial electricity consumption the weather dependency may not have as much impact since the usage is usually highly effected by the human activity rather then the weather factor. On the other hand we have domestic electricity consumption which has a very strong seasonal weather dependency. For instance during the winter we can assume that the electricity consumption will be higher due to heating and lightning. For optimal modeling the best practice is to distinguish the different types of energy load classes [9]. In this case study there is no knowledge of which type of electricity load that is being handled, therefore we can not assume any special conditions. The data analysis will involve all obtained variables with focus on lagged variables. Further there will also be a study of the periodicity due to high seasonality.

2.1 Data Description

The acquired data consists of time series of the electricity load and weather observations that come from a country in Northern Europe. The data signals are electricity power load, air pressure, humidity, wind direction, wind speed, global radiation and outdoor temperature. All measurements have an hourly time resolution. The time line for the data is from 1st of January 2008 to 1st of November 2010.

All the acquired data observations are visualized in Figure 2.1. The figures indicate that there is a notable dependency between our target variable the electricity load and three weather variables. These are the outdoor temperature, global radiation and humidity. For instance when the temperature observations are in general low the electricity load goes up and vice versa. The dependencies for the other weather variables are analogue to the temperature example. An interesting thing will be to see if the usage of more weather variables can effect the performance accuracy of our models. Temperature and humidity are two highly correlated weather variables. In many cases the temperature and humidity may be computed to a single component referred to as temperature-humidity index or heat index which temps to define the human-perceived temperature[9]. The global radiation variable can be used as a way of estimating the heating of the ground and buildings. The global radiation variable may in certain cases cause false indications due to effects such as cloudiness therefore it is recommended to be used with caution. The observed variables are defined in Table 2.1.

Data variable	Notation	Unit
Electricity power load	f_t	MW
Temperature	T_t	°C
Global radiation	G_t	W/m^2
Wind speed	W_t	m/s
Wind direction	D_t	°
Humidity	H_t	%
Air pressure	P_t	hPa

Table 2.1: Data variables and their notation

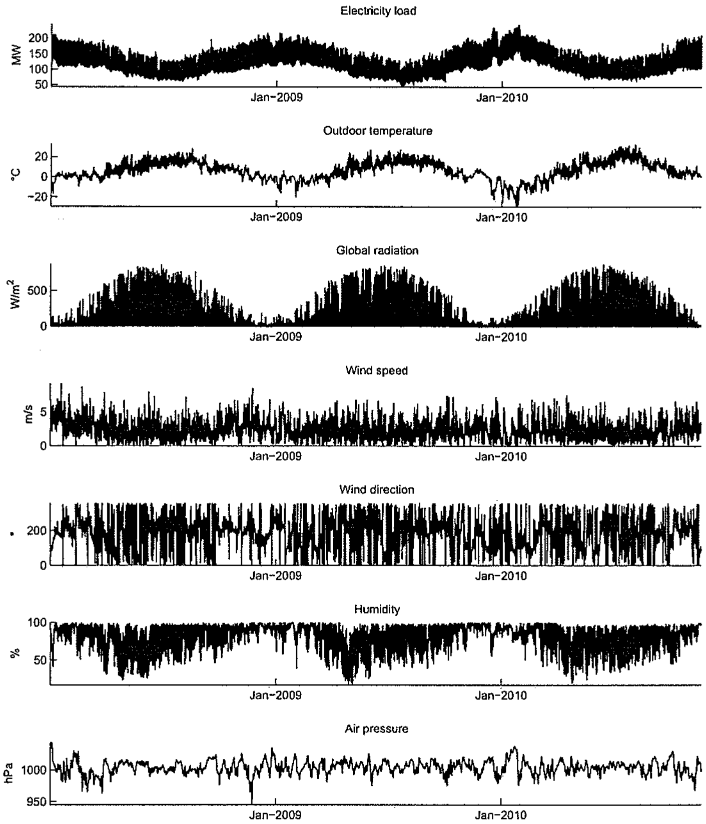

Figure 2.1: Time series of the energy load and all the obtained weather observations

2.2 Correlation Analysis

The relationship between the given observations are studied with linear correlation analysis. An easy approach is by visualizing two different variables against each other for different time steps. The most known seasonal components for electricity load are the daily, weekly and yearly dependency [9]. In Figure 2.2 four graphs are shown. The first is of the average daily electricity load against the actual load. The rest consist of lagged variables against the actual electricity load. The delayed variables show a strict linear dependency but the previous day average load seems to have some kind of conditional linear dependency that varies depending on other factors. To get a clear overview of the electricity load correlation between different time lags the ACF and PACF is computed. In Figure 2.3 the ACF and PACF is demonstrated. Both correlation functions clearly have a seasonal dependency specially the ACF.

In a classical time series approach a suggested method to decrease the seasonal dependency would be to differentiate the electricity load. A suggestion of differentiation is clearly 24 hours and 168 hours. Figure 2.4 illustrate the ACF and PACF of the electricity load differentiated with 24, 168 and also two times differentiated time series of electricity load. The differentiations do not remove the strong correlation in time lags. The two times differentiated time series seems to handle it best but still has strong correlation after 24 hours. The differentiated time series shows an unsatisfying model result since there are notable lag correlation for length larger then 24 hours.

The actual electricity load has been shown to be a highly seasonal component. In Figure 2.5 graphs for the actual electricity load are plotted against all weather variables. A first and second degree polynomial curve is estimated for each plot. Only the outdoor temperature shows significant linear dependency. The temperature, humidity and air pressure can be described by a second degree polynomial curve. The other variables are not easily described by polynomials but show clear non-linearity.

In Figure 2.5 the dependency of the electricity load and the weather variables were shown but no notable linear dependency could be identified. In order to identify the correlation between the electricity load and the weather variables the CCF is computed. In Figure 2.6 all the CCF's are shown. They all show a strong seasonal correlation. In other words this means that all our weather variables are correlated with the electricity load. Some of the weather variables such as temperature, global radiation and humidity have much stronger cross correlation with the electricity load than others.

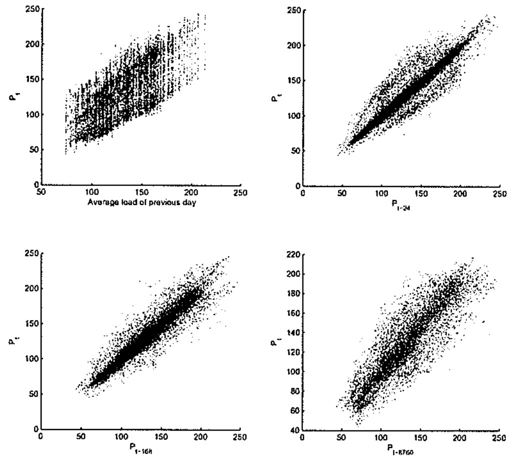

Figure 2.2: The dependency between the electricity load in different time steps.

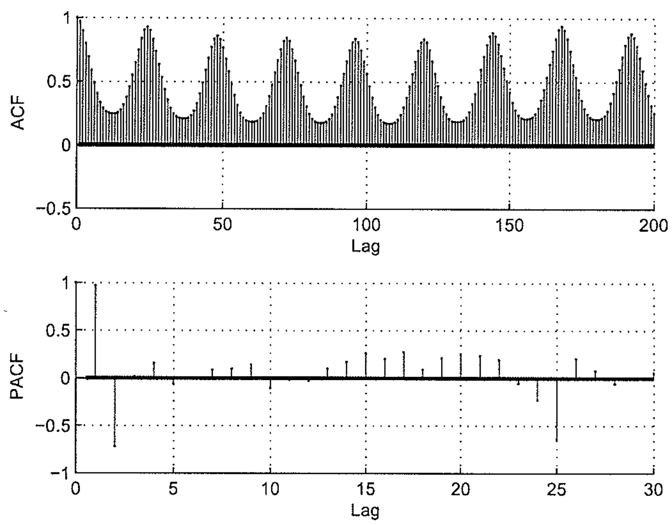

Figure 2.3: Correlation functions ACF and PACF of the the electricity load.

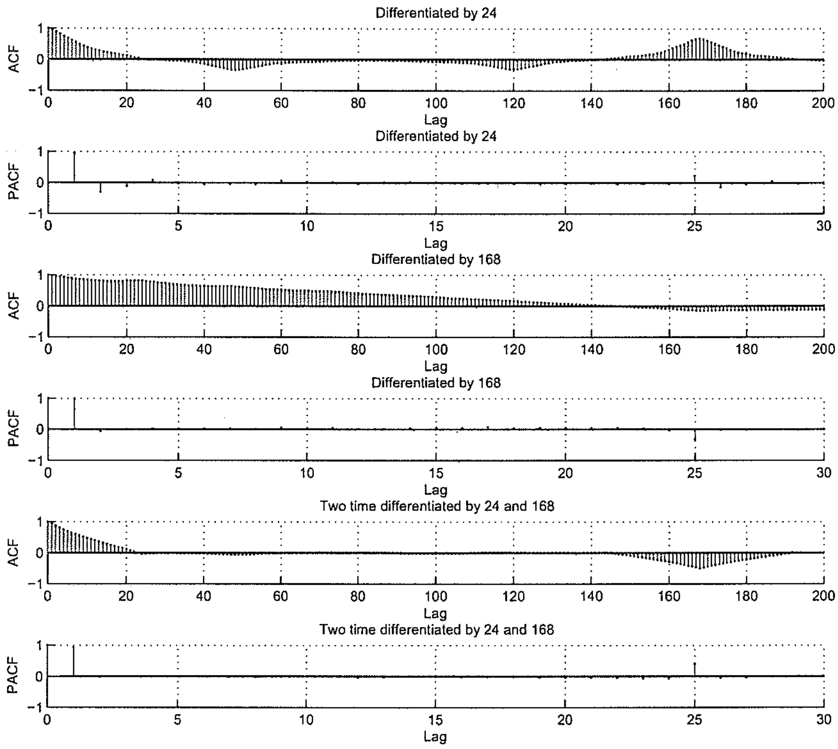

Figure 2.4: Correlation functions ACF and PACF of the differentiated electricity load.

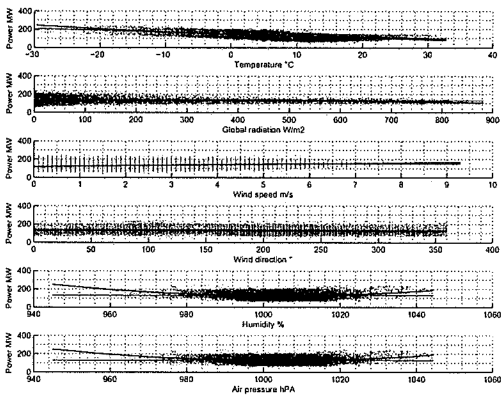

Figure 2.5: The dependency between the weather and the electricity load.

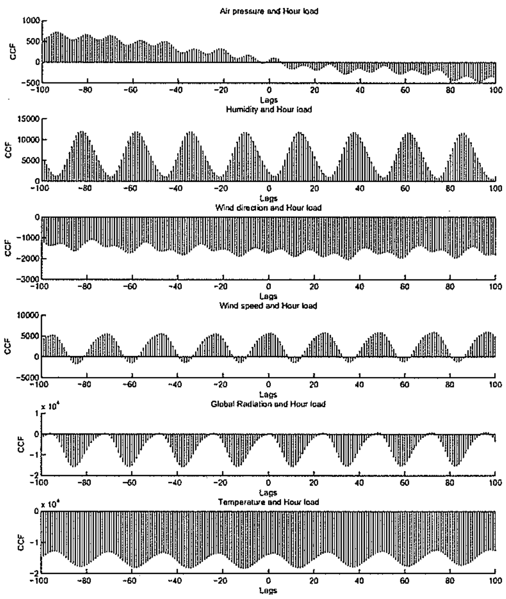

Figure 2.6: The cross correlation between the weather and the electricity load in different time steps.

2.3 Deterministic components

The focus has so far been on input variables that consist of delayed load and weather variables. Both types of variables showed strong seasonal correlation. There is no doubt that the seasonality variable is the most important input factor for the electricity load demand.

Figure 2.9 shows the daily electricity load profile of each day of the week for all obtained data. The daily load has indeed a special pattern for each day of the week. Except for the load usage the pattern of the load for each hour seems to remain throughout each day of the week. Therefore a deterministic variable such as hour of the day, day of the week, month or day of the year should definitely be considered.

Previous correlation analysis demonstrate a strong seasonal correlation between the daily load but also the load of the previous week at the same day and hour. In Figure 2.7 a regular week example of the electricity load time series is shown. The load has a strong oscillating effect that represents the daily load. The figures of a weekly load profile and the daily load profile concludes that there also is a periodicity in the weekly time length.

Fourier series is a decomposed periodic function. If we can identify the periodicity in the electricity load time series we may use these variables in the modeling. To identify the periodicity in our fluctuation we perform spectral analysis on the electricity load. In Figure 2.8 the power spectrum is shown. The figures show peaks for frequencies which in descending order correspond to the periods 24, 12, 168 and 84 hours. Hence the time series can be represented by a Fourier model. In classical time series analysis an approach would be to use these variables in our multiple regression model. In this case we will use them as input variables to see if the components may improve the accuracy of our network.

$$f_t = a_0 + \sum_{k=1}^{4} (a_k \cos(2\pi t f_k) + b_k \sin(2\pi t f_k)) \tag{2.1}$$

where $\{t \in [1, 168]\}_{\mathbb{Z}}$.

The analysis of the deterministic component concludes that there is a thorough seasonal pattern in all the analyzed variables. The need for deterministic time variables may therefore capture this seasonality which includes the hourly and daily effect. In Table 2.2 all the different time variables that will be used when doing model selection are presented. The suggested time variables are presented further in the next subsections of this chapter.

Data variable	Notation	Variable(s)	Type
Hour of the day	h_t	1	Raw
Weekday	d_t	1	Raw
Month	m_t	1	Raw
Day of the year	y_t	1	Raw
Hour of the day	I_{h_t}	24	Indicator
Weekday	I_{d_t}	7	Indicator
Hour of the day	g_{h_t}	2	Fourier
Weekday	g_{d_t}	2	Fourier
Day of the year	g_{y_t}	2	Fourier
Hour of the day	a_t	7	Combined
Weekday	b_t	24	Combined

Table 2.2: Indicator time variables.

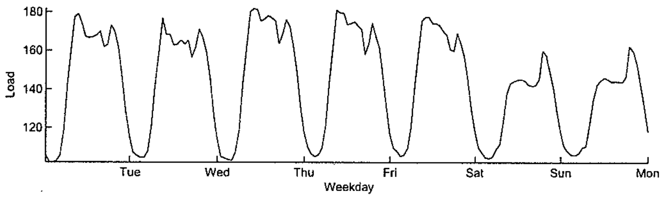

Figure 2.7: Example of the electricity load a regular week.

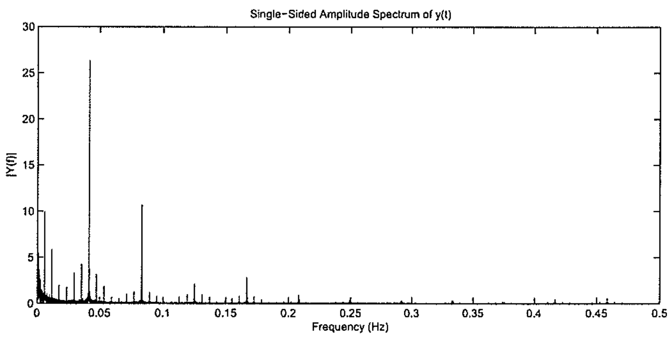

Figure 2.8: Power spectrum of the electricity load.

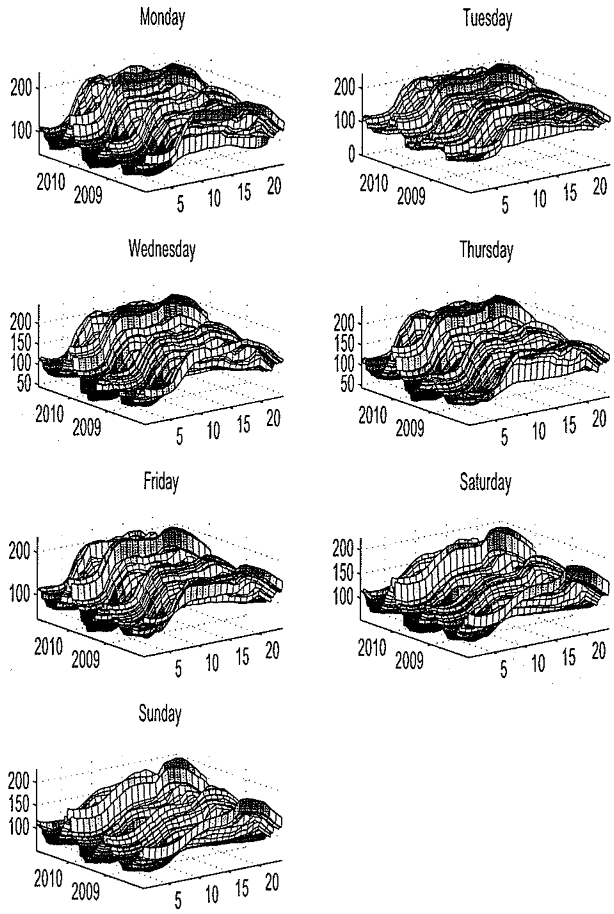

Figure 2.9: The daily load for all series.

2.3.1 Raw time variables

Raw time variables are defined as integer values in a certain range. The defined variables are hour of the day (1-24), day of the week (1-7), day of the year(1-365) and month of the year(1-12). The variables are defined below. In Figure 2.10 a example of a regular week is plotted for each variable.

This yields

$$h_t = \text{hour of the day at time } t \tag{2.2}$$

$$d_t = \text{day of the week at time } t \tag{2.3}$$

$$y_t = \text{day of the year at time } t \tag{2.4}$$

$$m_t = \text{month of the year at time } t \tag{2.5}$$

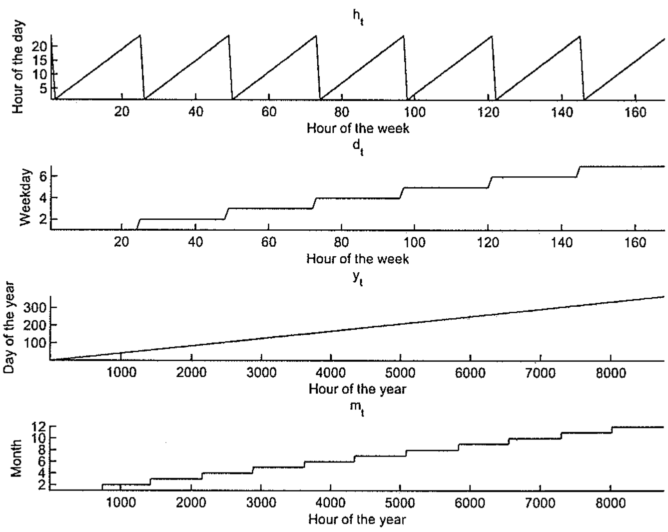

Figure 2.10: The plots demonstrate the deterministic values of the raw input variables for a regular week.

2.3.2 Indicator time variables

Indicator variables are also known as dummy variables which can be defined either as one or zero. The handled time variables are for the hour of the week and the day of the week. Since indicator variables can only hold one categoric effect the need for multiple input variables are needed for capturing the chosen time variables. The indicator variables for the hour of the day consist of 24 variables, one for each hour of the day. The indicator variables for the day of the week are 7, one for each day of the week. In Figure 2.11 two graphs illustrate the values for two of the indicator values. The first is for the 16th indicator variable for the hour of the day and the latter is the Wednesday indicator variable for the day of the week.

This yields

$$I_{h_t} = (I_{1,h_t}, I_{2,h_t}, I_{3,h_t}, ..., I_{24,h_t}) \tag{2.6}$$

$$I_{j,h_t} = \begin{cases} 1 & j = t \\ 0 & j \neq t \end{cases} \tag{2.7}$$

$$I_{d_t} = (I_{1,d_t}, I_{2,d_t}, I_{3,d_t}, ..., I_{7,d_t}) \tag{2.8}$$

$$I_{j,h_t} = \begin{cases} 1 & j = d_t \\ 0 & j \neq d_t \end{cases} \tag{2.9}$$

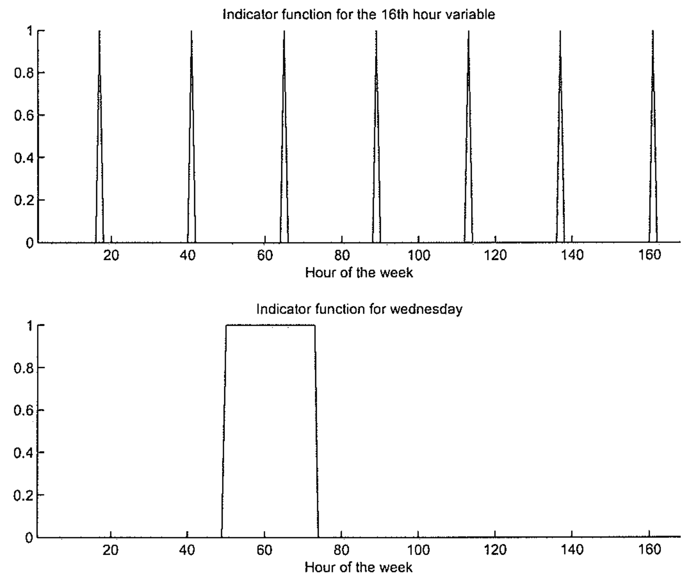

Figure 2.11: The first plot shows how the indicator variable for the 16th hour looks like for a general week. The second plot shows the indicator variable for Wednesday during a regular week.

2.3.3 Fourier series time variables

Previous identification through the power spectrum showed strong periodicity for the time periods 24, 168, 12 and 8760 hours. Fourier series are used in classical time series for it's ability to capture periodicity for strongly seasonal time series. In this study the chosen Fourier variables are for the time periods of 24, 168 and 8760 hours. Each Fourier series component consists of two variables. The Fourier series variables are defined as trigonometrical functions of sinus and cousins where the period length varies. In Figure 2.12 examples of the chosen variables are shown.

This yields

$$g_{h_t} = (\cos(2\pi t/24), \sin(2\pi t/24)) \tag{2.10}$$

where t is the hour of the day

$$g_{d_t} = (\cos(2\pi t/168), \sin(2\pi t/168)) \tag{2.11}$$

where t is the hour of the week

$$g_{y_t} = (\cos(2\pi t/8760), \sin(2\pi t/8760)) \tag{2.12}$$

where t is the hour of the year

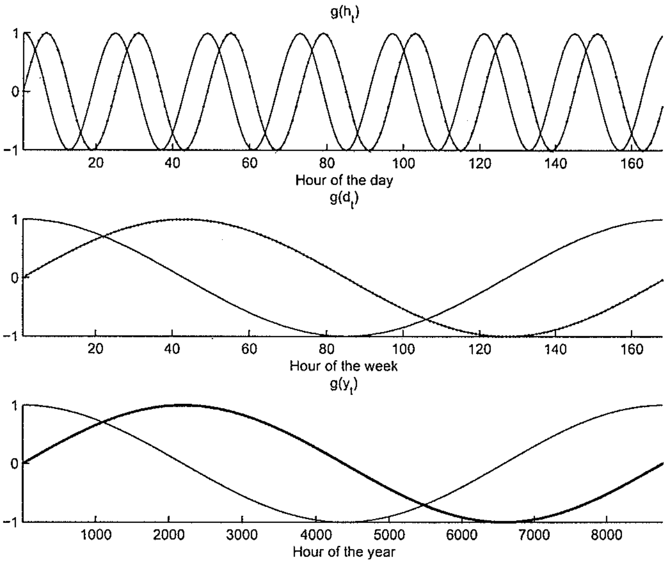

Figure 2.12: Each plot has two variables the regular line is the cosine variable and the dotted line is the sine variable. The first plot shows values for the daily periodicity of a regular week. The second plot is for the weekly periodicity. The third plot shows the yearly periodicity during one year.

2.3.4 Combined time variables

Combined time variables consist of combinations of our previous time variables. Since the input variables can have a multi category property there is the possibility to add several categorical effects in to the same variable. The merged properties are the hour of the day and day of the week component. The chosen combination is between the raw and the indicator time variables. This leads to two new multi categorical variables. The first is a combination of the indicator for the actual day of the week against the actual raw value for which hour of the day. The combination consist of 7 variables where the defined values are from 0-24. The second combination is of the indicator for the actual hour of the day against the actual raw value for which day of the week. The combination consists of 24 variables where the defined values are from 0-7. In Figure 2.13 examples of the combined time variables are shown for a regular week. The first figure shows the Wednesday variable variable for out combined variable $a_{3,t}$. The second figure shows the 16th variable for our combined variable $b_{16,t}$.

This yields

$$a_t = I_{d_t} \cdot h_t \tag{2.13}$$

$$a_t = (a_{1,t}, a_{2,t}, ..., a_{7,t}) \tag{2.14}$$

$$b_t = I_{h_t} \cdot d_t \tag{2.15}$$

$$b_t = (b_{1,t}, b_{2,t}, ..., b_{24,t}) \tag{2.16}$$

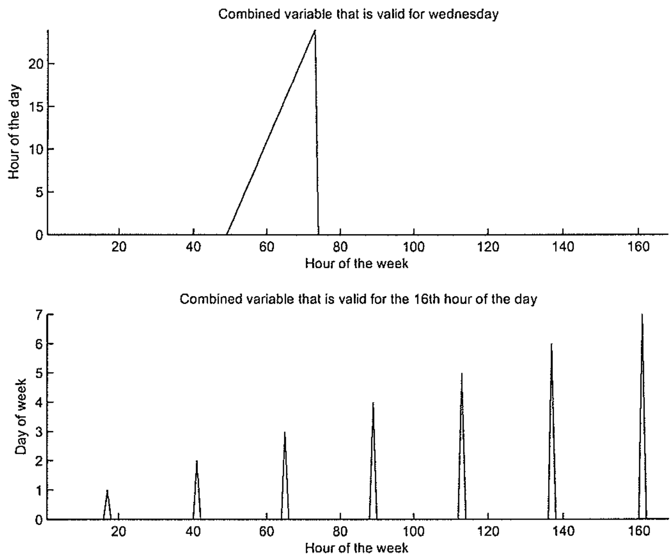

Figure 2.13: An example of a regular week.

2.4 Example of a linear model

A linear model estimate of an electricity load is made with the following input matrix.

$$X_t = \Big(T_t, H_t, h_t, d_t, I_{\text{wd}\in\{0,1\}}, \bar{T}_t, f_{t-24}, f_{t-168}\Big) \tag{2.17}$$

Each variable consist of a column vector with N rows corresponding to the amount of measurement observations.
The electricity load model is

$$f_t = BX_t + e_t \tag{2.18}$$

where e_t is white noise.

We estimate the models parameters in the vector B with least squares. The result is presented in the table beneath. In Figure 2.14 we clearly see that the residuals are highly correlated after 24 lags. The estimated model is not satisfying due correlated and non-gaussian residuals. The result in MAPE is slightly below 5 percent for the linear model.

MODEL	In sample			Out sample		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
(2.18)	6.0392	4.9331	8.3593	6.8598	4.9830	9.5717

Table 2.3: Performance results of the linear model M(2.18) is measured using MAE, MAPE and RMSE. The estimated parameters of the model are made with the in sample data. The out of sample data is used to test the model with unknown data.

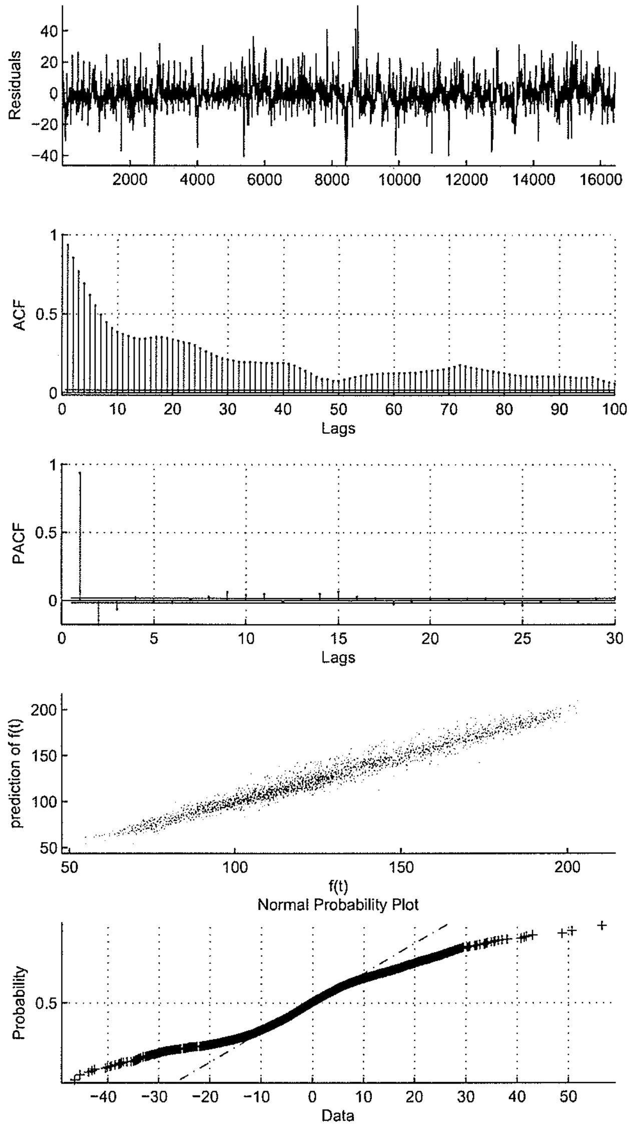

Figure 2.14: The figure consist of residuals from the in sample data for the model 2.18

Chapter 3

Neural Networks

3.1 Background

A neural network is a circuit of biological neurons. A neuron can be connected to several other neurons. The connection is called synapses and is built from axons to dendrites. The human brain is estimated to contain around 10-100 billion neurons and around 10000 billion connections, with each neuron linked with around 10000 synapses.

Artificial neural networks (ANN) is a mathematical model that is inspired by the biological neural networks. Although in modern usage the term neural networks(NN) is referred to the mathematical model artificial neural networks. Today neural network models are used in research areas such as statistics, signal processing, neurophysiology, informatics, meteorology, engineering, business forecasting and so on. The applications are amongst other for prediction/forecasting, pattern recognition, adaptive control such as speech recognition, explosives detection at airports, rainfall prediction, credit card fraud detection and electrical load/demand forecasting. Neural networks is a non-parametric black box model which means that no priori assumption is made. Solely the inputs and the network settings effect the output.

3.2 Mathematical Description

The neural network task is to basically insert inputs into the network to match any given output. The output is normally referred as the target value. The network structure consists of three type of layers with nodes. In the input layer the nodes correspond the amount of variables which will be inserted. The hidden layer has arbitrary number of nodes depending on the type of process and number of inputs. Also the hidden layer may consist of several layers. Finally the number of nodes in the output layer is equal to the number of target variables. We demonstrate a simple feed forward neural network in Figure 3.1.

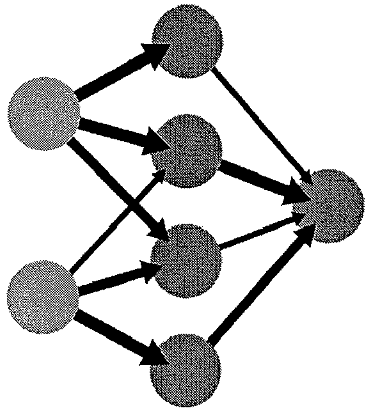

Figure 3.1: This figure consists of a simple feed forward neural network with one hidden layer. There are two inputs three nodes in the hidden layer and one output node.

3.2.1 Mathematical description

Artificial neurons

An artificial neuron is a mathematical function that is a unit of the neural network. It is defined as follows. Assume there are m inputs which are $x_1,...,x_m$. We denote the j:th node weight as $w_{j1},...,w_{jm}$. Further the bias weight is denoted as w_{j0} ans has x_0 equal to one. This yields

$$u_j = \varphi\left(\sum_{i=0}^{m} w_{ji}x_i\right) \tag{3.1}$$

Where $\varphi(\cdot)$ is the neurons activation function also known as the transfer function that is chosen depending on which properties the network should have. In general this is if there should be a linear or non-linear dependency. In this case we there is a strive to capture non-linearity so the chosen transfer function is the hyperbolic tangent sigmoid function. This is defined as

$$u_j = \varphi(\xi) = \tanh(\xi) = \frac{e^{2\xi} - 1}{e^{2\xi} + 1} \tag{3.2}$$

Network layers

As described earlier the layers consist of the input layer, the hidden layers and the output layer. If the variable u_j describes one node in an arbitrary layer k

then it can be denoted as

$$u_j^k = \varphi \left(\sum_{i=0}^{m} w_{ji}^k x_i^k \right) \tag{3.3}$$

If the layer M is the output layer and there is only one output the function is

$$y = \varphi \left(u^{M-1} \right) = \varphi \left(\sum_{i=0}^{m} w_{ji}^{M-1} x_i^{M-1} \right) \tag{3.4}$$

3.2.2 Training of neural network

To find optimal weights for our network functions we need to optimize the functions by minimizing the error. This performance (cost) function is in this case measured by the MSE which is defined as follows

$$\text{MSE} = \sum_{i}^{n} (e_i)^2 = \sum_{i}^{n} (y_i - \hat{y_i})^2 \tag{3.5}$$

There are different optimization methods available for minizirning the cost function. In this study the used optimization method is Levenberg-Marquardt(LM) which is a robust method for approximating functions.

Let the Jacobian of e_i be denoted J_i then the LM algorithm searches for the solution of p_k to the equation

$$\left(J_k^T J_k + \lambda_k I \right) p_k = -J_k^T e_k \tag{3.6}$$

where λ_k are nonegative scalars and I is the identity matrix.

Since the computations are made backwards through the network they are referred to as backpropagation algorithms The Levenberg-Marquardt optimization function is used as a complement to the chosen training function Bayesian Regulation. The Bayesian Regulation is a training function which generalizes how many weights and errors should be fitted to the network [10], [11].

3.3 Neural Network Setup

The architechture which is studied in this paper is feedforward neural network with a hyperbolic sigmoid transfer function. The training function is Bayesian Regulation with Levenberg-Marquardt optimization. Since we are using Bayesian Regulation the number of nodes are automatically adjusted. The number of hidden layers is set to a fixed number of 20.

Chapter 4

Models and Results

In this chapter the models that have been studied are presented an evaluated. The models are all based on an feed-forward neural network model which estimates the function of the electricity load (target variable). The study starts of with models with lagged electricity load variables and then testing of different kinds of deterministic components. Finally the weather factors are tried out. The performance is measured in MAE, MAPE and RMSE. The quality of the model is measured through correlations analysis and normality check of the residuals. The dependency of prediction against actual load is also taken in to account. The models are evaluated with both in sample and out sample data. The training set consists of data observations 2008-01-01 to 2009-10-31. The out of sample evaluation is made with data from 2009-11-01 to 2010-11-01. The out-sample data consists only of observations and not of forecast.

4.1 Lagged variables

This section presents our first set of models which are based solely on time lagged electricity load as input. In this set case the expected results are not a best fit model but to get an estimate of the importance of the lagged variables. In the assumption of having a model based on lagged variables we dismiss the importance of variables such as weather or the studied deterministic components. These models are solely based on lagged variables of 24, 84 and 168 hours. Model M(4.3) is also including the previous days average load.

$$f_t = f\left(f_{t-24}, f_{t-168}\right) + e_t \tag{4.1}$$

$$f_t = f\left(f_{t-24}, f_{t-84}, f_{t-168}\right) + e_t \tag{4.2}$$

$$f_t = f\left(f_{t-24}, f_{t-168}, \bar{f}_t\right) + e_t \tag{4.3}$$

where $\bar{f}_t$ is equal to previous day's average load and e_t is white noise.

4.1.1 Results

The first model M(4.1) has only two input variables. The 24 and 168 hours lagged electricity load variable. The quality results of the model are presented in Figure 4.1. The ACF shows a clearly strong dependency for time steps above 100 hours. This indicates that the residual still has strong correlation and that the model is not satisfying. The PACF shows only high dependency for one lag which can be described by an AR(1) model. The ACF and PACF demonstrates that the residuals are not uncorrelated and by viewing the normal probability plot we can also see that it's not normally distributed. The dependency plot shows that the predicted electricity load is more precise for loads around 75 and 125 MW.

The results for models M(4.2) and M(4.3) are almost analogue. For model M(4.2) the ACF is a bit lower but not remarkably. Table 4.1 shows the forecast accuracy results of the models. The conclusion for this set of models is that the M(4.2) with lagged variables 24, 168 and 84 is the best performing model of this set.

MODEL	In sample			Out sample		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
(4.1)	5.6112	4.5498	8.3089	7.2261	5.1170	11.3989
(4.2)	5.4947	**4.4514**	8.1162	6.8681	**4.8818**	10.5112
(4.3)	5.6193	4.5548	8.3201	7.5822	5.2893	13.7058

Table 4.1: Results of the performance of the models in this section.

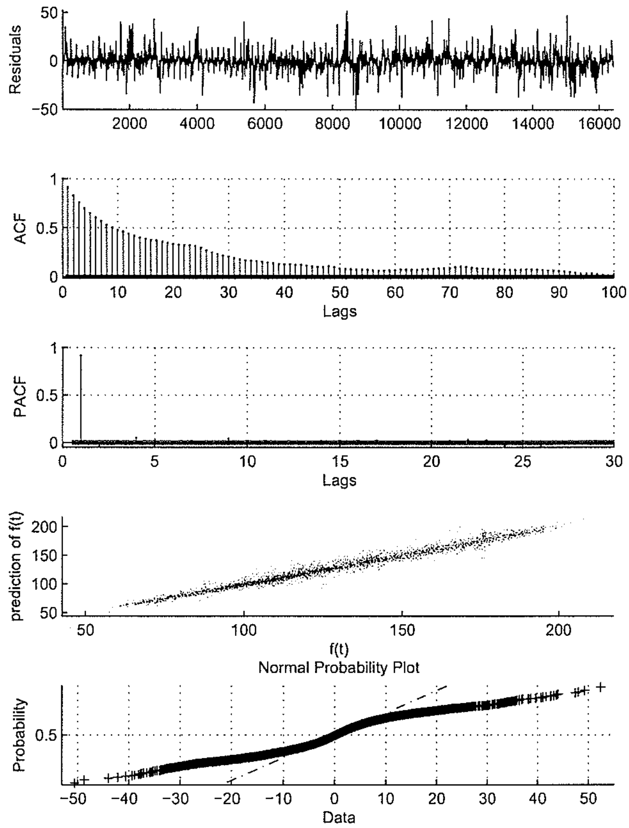

Figure 4.1: The figure consist of residuals from the in sample data for the model 4.1.

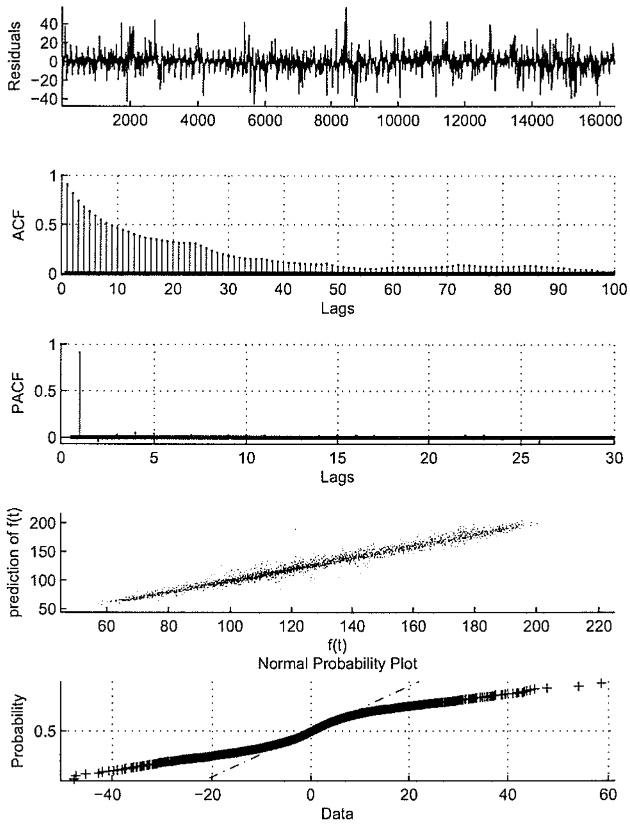

Figure 4.2: The figure consist of residuals from the in sample data for the model 4.2.

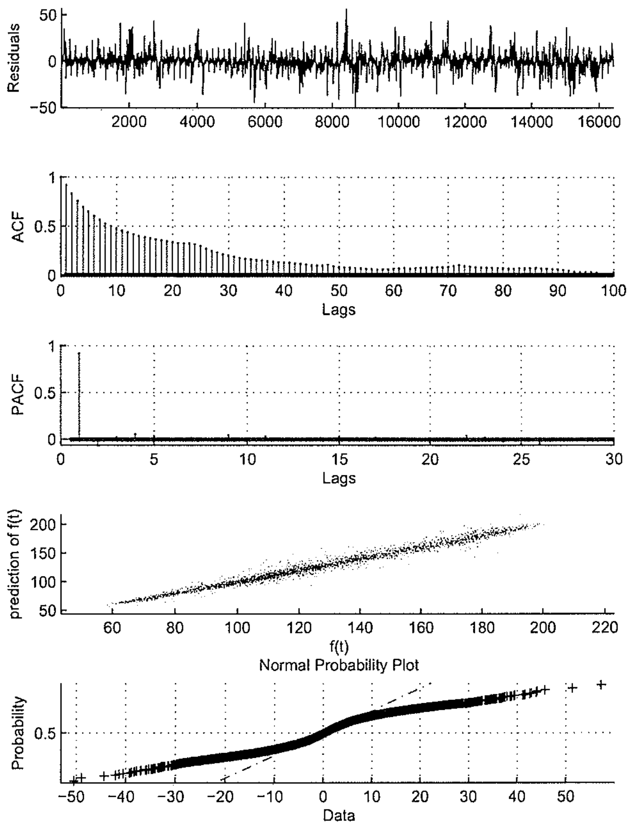

Figure 4.3: The figure consist of residuals from the in sample data for the model 4.3.

4.2 Lagged and Raw time variables

The set of models in this section all have the lagged electricity load of 24 and 168 hours as input variables. All models also have different combinations of raw time variables. The raw time variables consist of the "hour of the day", "day of the week", "day of the year" and "month of the year". The assumption of these models is that there should be a notable connection between the specific raw time variable and the electricity load. This should intentionally try to capture the load profile of the electricity load.

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t\right) + e_t \tag{4.4}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, y_t\right) + e_t \tag{4.5}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, m_t\right) + e_t \tag{4.6}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, y_t, m_t\right) + e_t \tag{4.7}$$

where e_t is white noise.

4.2.1 Results

Model M(4.4) has the raw time variables "hour of the day" and "day of the week". In Figure 4.4 it's results are shown. The residual ACF is unlike previous set of models uncorrelated for lags after 24. Although the PACF still shows dependency at time lag 1 and 25 the correlation has decreased. The graphs demonstrate that there are still correlation in the residuals but still not as much as in previous set. Models M(4.5), M(4.6) and M(4.7) have similar results and patterns but the model M(4.4) is the best performing model for out of sample data. Model M(4.6) has least correlation and performs best for the in sample data. The PACF for this model has no dependency after the first lag. The day of month variable seems to capture varying correlation for the 24 hours lag.

MODEL	In sample			Out sample		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
(4.4)	3.8312	3.1127	5.8001	4.6125	**3.3641**	6.8162
(4.5)	3.7727	3.0955	5.5303	4.8236	3.4833	6.8933
(4.6)	3.5551	**2.8911**	5.2364	4.6907	3.3718	6.8262
(4.7)	3.6487	2.9821	5.2146	4.8619	3.5197	6.9597

Table 4.2: Results of the performance of the models in this section.

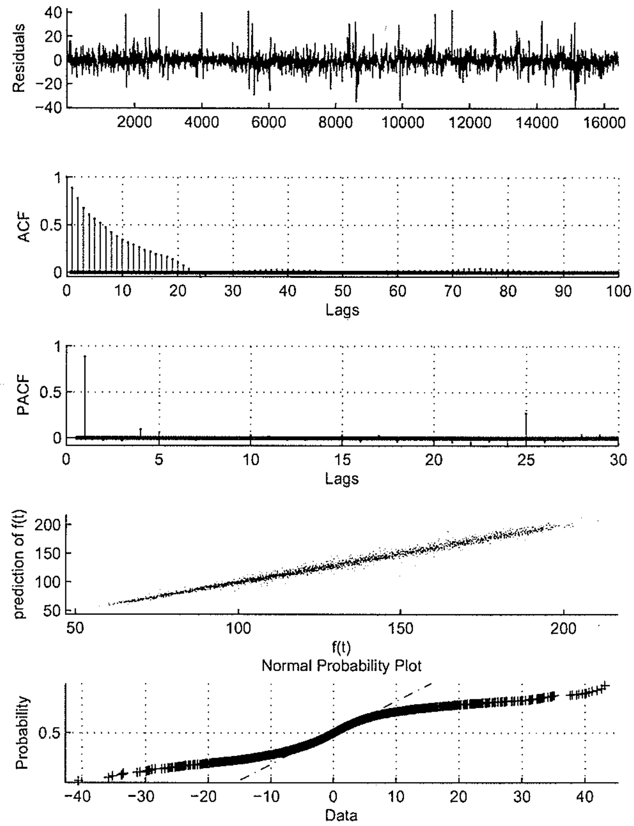

Figure 4.4: The figure consist of residuals from the in sample data for the model 4.4.

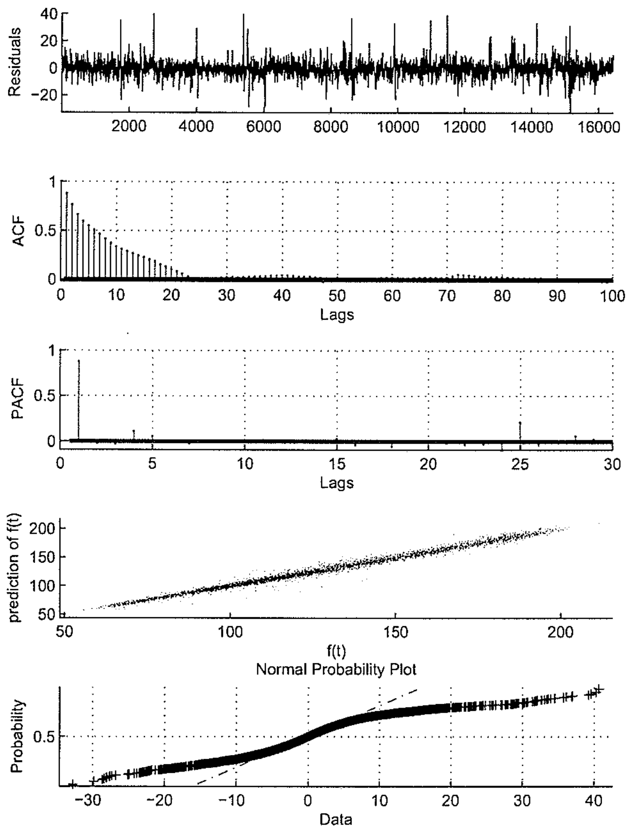

Figure 4.5: The figure consist of residuals from the in sample data for the model 4.5.

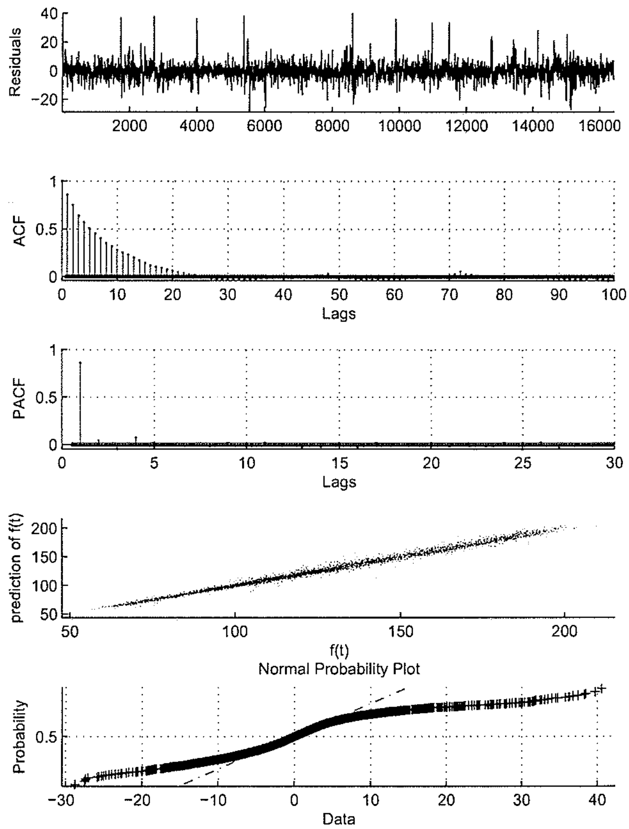

Figure 4.6: The figure consist of residuals from the in sample data for the model 4.6.

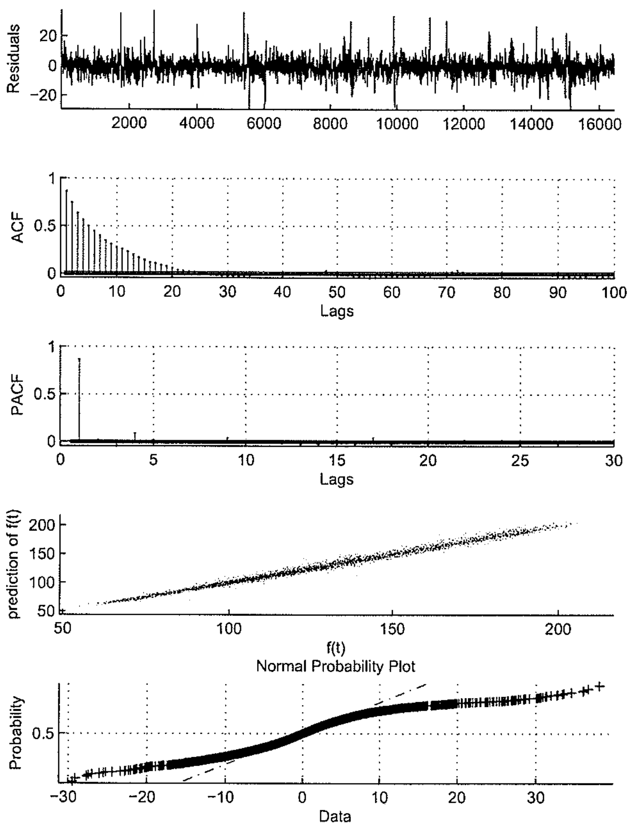

Figure 4.7: The figure consist of residuals from the in sample data for the model 4.7.

4.3 Lagged and Indicator time variables

This set of models handle indicator variables of the "day of the week" and the "hour of the day". The assumption is that there exists a individual dependency of the "hour of the day" or the "day of the week" for the whole data set. If this is true then the model will successfully capture the dependency. Each indicator variable consist of several inputs. The day of the week has 7 inputs and the hour of the day has 24 inputs as described in Section 2.

$$f_t = f\left(f_{t-168}, f_{t-24}, I_{h_t}\right) + e_t \tag{4.8}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, I_{d_t}\right) + e_t \tag{4.9}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, I_{h_t}, I_{d_t}\right) + e_t \tag{4.10}$$

where e_t is white noise.

4.3.1 Results

The first model M(4.8) performs the worst in this set. Figure 4.8 shows that the model has highly correlated residuals like in the previous set of lagged variable models. This means that the hourly indicator dependency is not effective as an input for our electricity load model. Model M(4.9) consist of daily indicators which performs much better than model M(4.8). Hence the result concludes that the daily pattern has a much higher importance as an indicator then the hourly load for the electricity load throughout the year. The performance is better but still the residuals are more correlated than for the previous set of models. Model M(4.9) has the best performance but has correlation similar to model M(4.9). Overall the indicator variables have a worse performance in comparison to the set of models with raw time variables.

MODEL	In sample			Out sample		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
(4.8)	5.6112	4.5498	8.3089	7.2261	5.1170	11.3989
(4.9)	4.2901	3.4977	6.3935	5.6265	3.9705	8.8310
(4.10)	3.5599	**2.8817**	5.3970	5.0240	**3.5841**	7.5456

Table 4.3: Results of the performance of the models in this section.

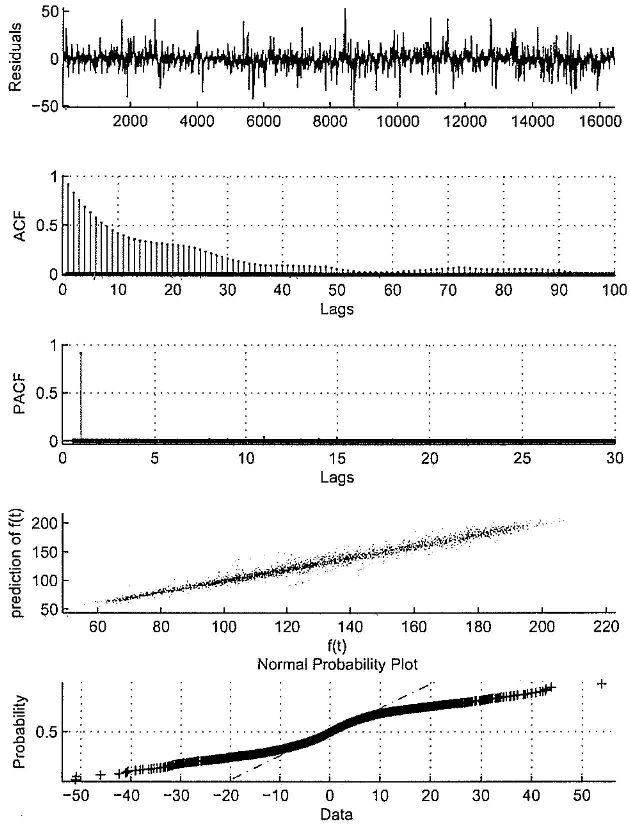

Figure 4.8: The figure consist of residuals from the in sample data for the model 4.8.

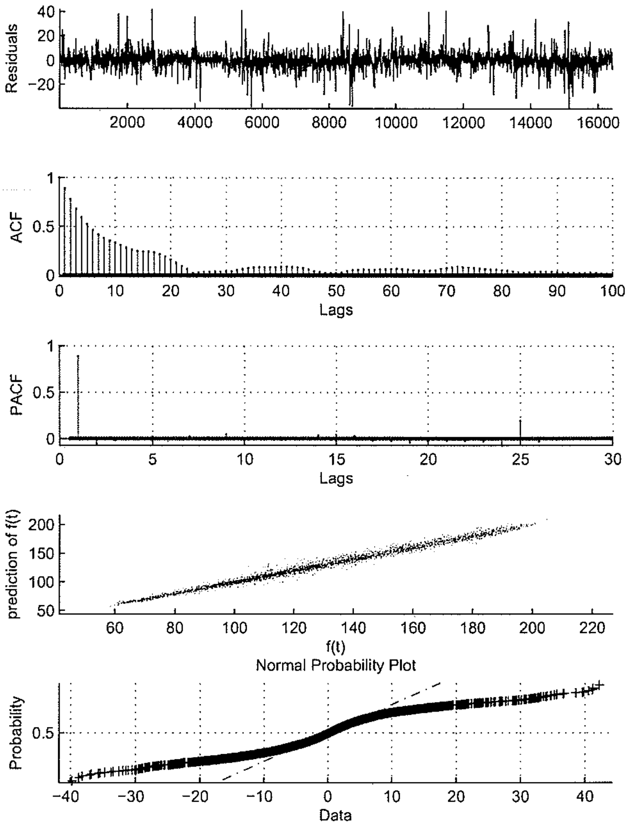

Figure 4.9: The figure consist of residuals from the in sample data for the model 4.9.

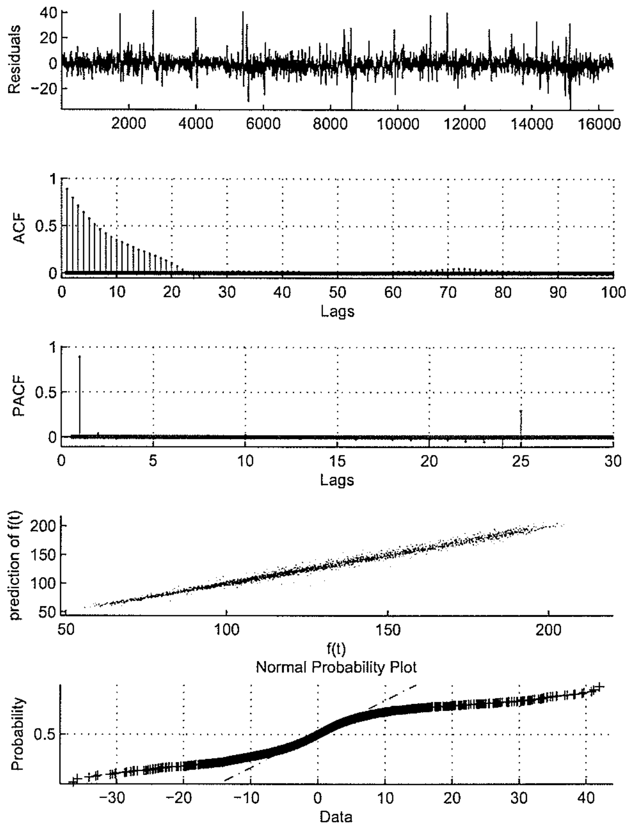

Figure 4.10: The figure consist of residuals from the in sample data for the model 4.10.

4.4 Lagged and Fourier time variables

This set of models uses Fourier series of periods corresponding to 24, 168 and 8760 hours as input variable. Each Fourier period consist of two variables. The testing of Fourier series as inputs is interesting since they are in general used to model linear periodic time series.

$$f_t = f\left(f_{t-168}, f_{t-24}, g_{h_t}\right) + e_t \tag{4.11}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, g_{d_t}\right) + e_t \tag{4.12}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, g_{y_t}\right) + e_t \tag{4.13}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, g_{h_t}, g_{d_t}\right) + e_t \tag{4.14}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, g_{h_t}, g_{d_t}, g_{y_t}\right) + e_t \tag{4.15}$$

where e_t is white noise.

4.4.1 Results

Models M(4.11) and M(4.13) have the worst performance in this set. Both have highly correlated residuals. Model M(4.11) uses the period of 24 hours and Model M(4.12) uses the period of 8760 hours (approximately one year). These input variables are unfortunately not capturing the load patterns in a effective way. The ACF of the residuals for both models indicate that there's still notable lag dependency after 24 hours. Model M(4.12) is one of the better performing models. The model periodicity corresponds to the weekly(168 hour) dependency. In Figure 4.12 the residuals ACF show small correlation after 24 lags. The PACF shows dependency for 1 and the 25th lag. The conclusion is that importance of the day of the week periodicity is higher than the daily and the yearly. The best performing out of sample model is the M(4.14) which is a combination of the weekly and daily period. The best performing in sample model is the M(4.15) which has the weekly, daily and yearly periodicities as input variables. The residuals correlation function for this model is similar to M(4.14).

MODEL	In sample			Out sample		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
(4.11)	5.3267	4.3004	7.8910	6.8578	4.7961	10.7472
(4.12)	3.7287	3.0382	5.6597	4.8632	3.5008	7.5900
(4.13)	4.8936	4.0122	6.9241	6.6418	4.7809	9.2924
(4.14)	3.6933	2.9970	5.6277	4.6778	**3.3856**	7.0056
(4.15)	3.5868	**2.9455**	5.2322	4.7791	3.4550	6.9832

Table 4.4: Results of the performance of the models in this section.

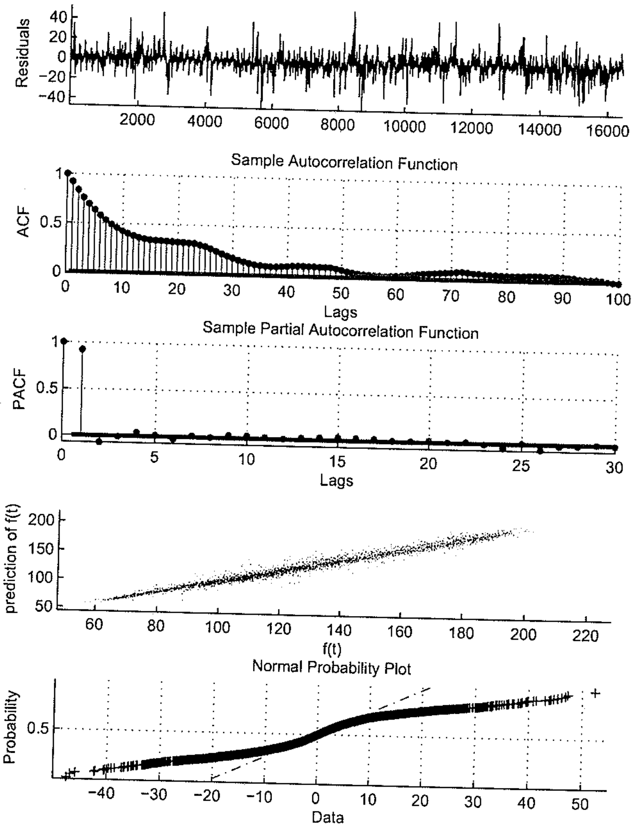

Figure 4.11: The figure consist of residuals from the in sample data for the model 4.11.

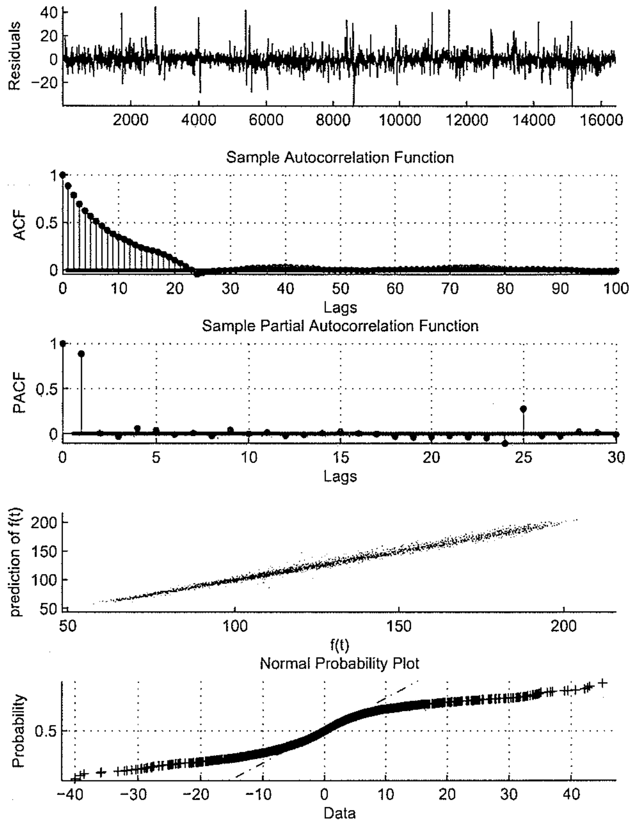

Figure 4.12: The figure consist of residuals from the in sample data for the model 4.12.

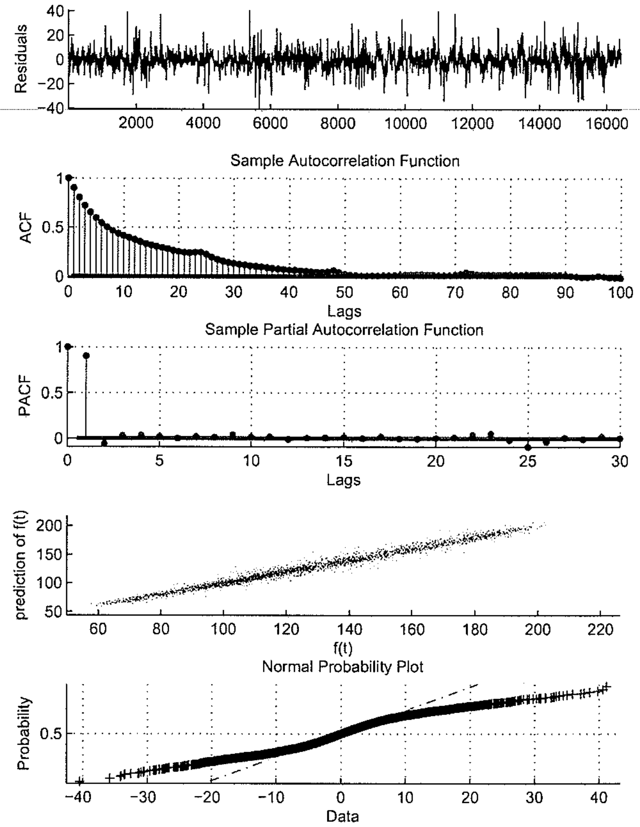

Figure 4.13: The figure consist of residuals from the in sample data for the model 4.13.

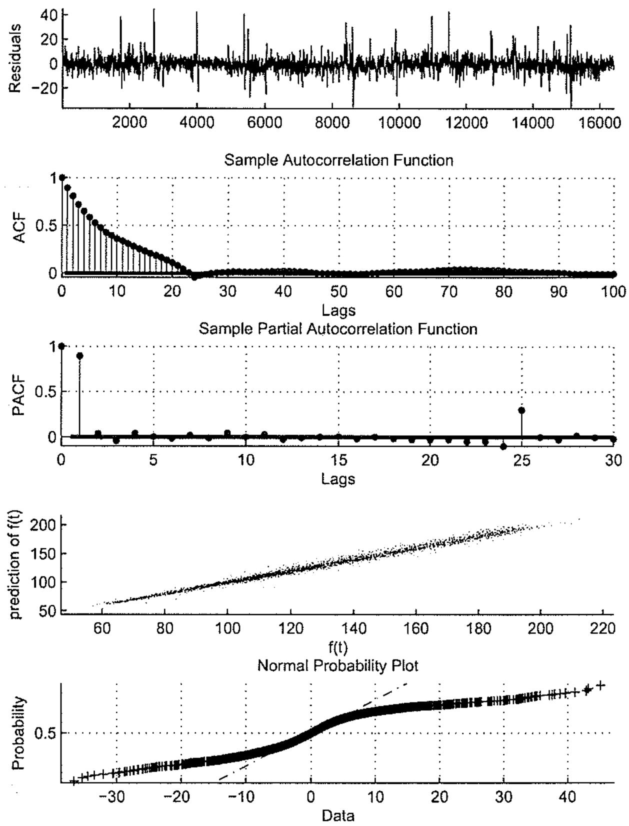

Figure 4.14: The figure consist of residuals from the in sample data for the model 4.14.

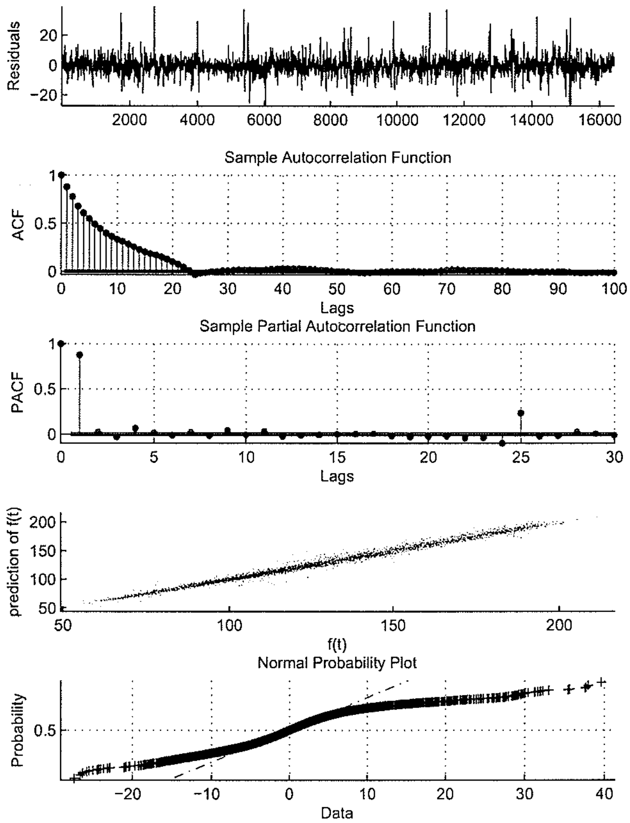

Figure 4.15: The figure consist of residuals from the in sample data for the model 4.15.

4.5 Lagged and combined time variables

This set of models consist of a merge between raw time variables and indicator time variables. By inserting more information to each input variable we may possibly be able to capture more of the time periodicity. The input variables are described in the data analysis chapter.

$$f_t = f\left(f_{t-168}, f_{t-24}, a_t\right) + e_t \tag{4.16}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, b_t\right) + e_t \tag{4.17}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, a_t, b_t\right) + e_t \tag{4.18}$$

where e_t is white noise.

4.5.1 Results

Model M(4.16) has one variable per weekday as input. The objective of this time variable is to capture the daily load profile. Model M(4.17) has one variable per hour of the day as input. The objective of this variables are to capture the hourly load profile. Between the models M(4.16) and M(4.17) the best results are given by the first combination. Model M(4.18) is the best performing model for this set both for "in sample" and "out sample" data. Surprisingly the combined variables do not perform better then the best raw and Fourier time variables.

MODEL	In sample			Out sample		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
(4.16)	3.7190	3.0252	5.6004	5.0657	3.6133	7.8109
(4.17)	5.0383	4.0196	7.4520	6.1047	4.3177	9.1268
(4.18)	3.6397	**2.9514**	5.4925	4.9131	**3.5198**	7.4189

Table 4.5: Results of the performance of the models in this section.

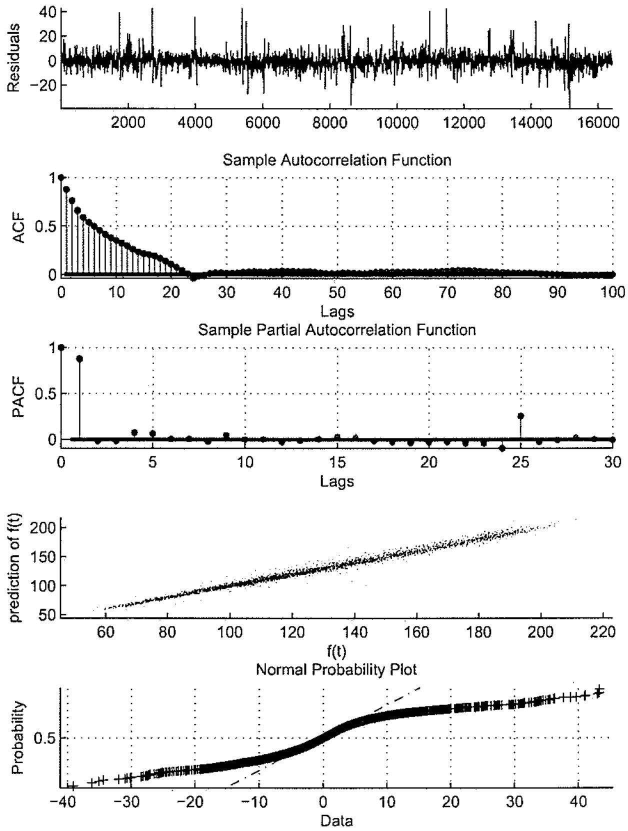

Figure 4.16: The figure consist of residuals from the in sample data for the model 4.16.

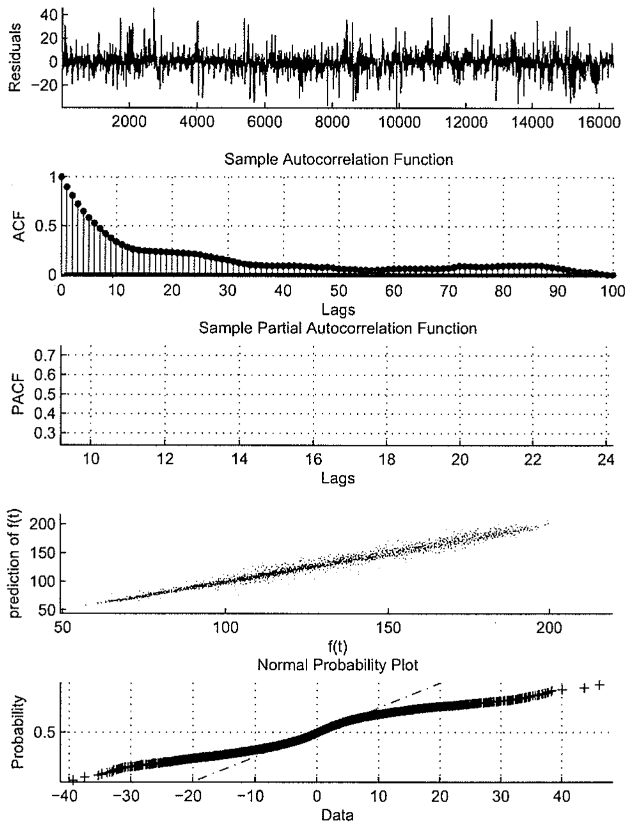

Figure 4.17: The figure consist of residuals from the in sample data for the model 4.17.

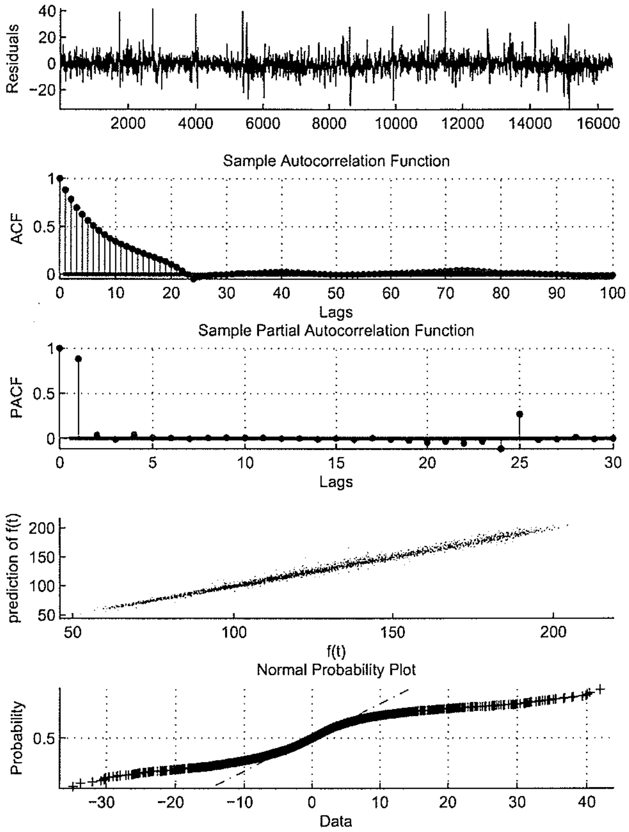

Figure 4.18: The figure consist of residuals from the in sample data for the model 4.18.

4.6 Best performance models without weather variables

The two best performing models are models M(4.4) and M(4.14). Both models consist of networks with input variables for "day of the week" and "hour of the day". In this set of models a comparison of these two models are made. Also a merged model of these two are made. The best performing model should be used when adding weather variables in the next model set.

$$f_t = f\left(f_{t-168}, f_{t-84}, f_{t-24}, h_t, d_t\right) + e_t \tag{4.19}$$

$$f_t = f\left(f_{t-168}, f_{t-84}, f_{t-24}, g_{h_t}, g_{d_t}\right) + e_t \tag{4.20}$$

$$f_t = f\left(f_{t-168}, f_{t-84}, f_{t-24}, h_t, d_t, g_{h_t}, g_{d_t}\right) + e_t \tag{4.21}$$

where e_t is white noise.

4.6.1 Results

All the models perform slightly well. The ACF for all models is low after 24 lags and the PACF has only two strong lag dependencies. The best model in the set is Model M(4.19). Although the difference in accuracy between M(4.19) and the model in the set of raw time variables M(4.4) is minimal.

MODEL	In sample			Out sample		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
(4.19)	3.7888	3.0850	5.7397	4.6238	**3.3530**	6.8181
(4.20)	3.6839	2.9984	5.5673	4.7639	3.4296	6.9923
(4.21)	3.6778	**2.9877**	5.4985	4.8005	3.4431	7.1331

Table 4.6: Results of the performance of the models in this section.

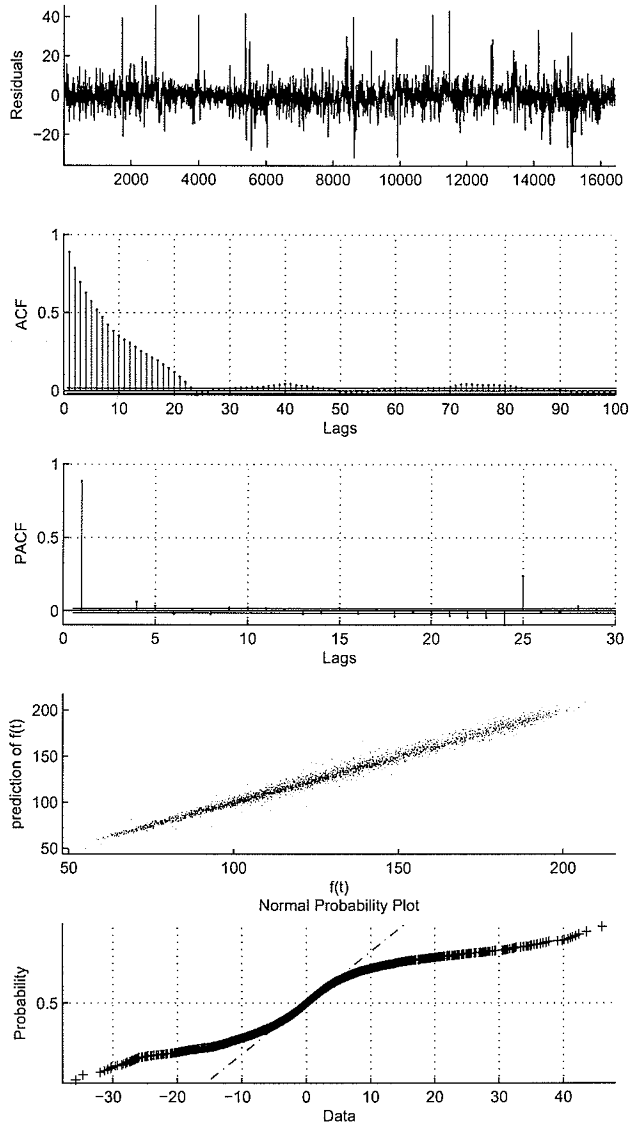

Figure 4.19: The figure consist of residuals from the in sample data for the model 4.19.

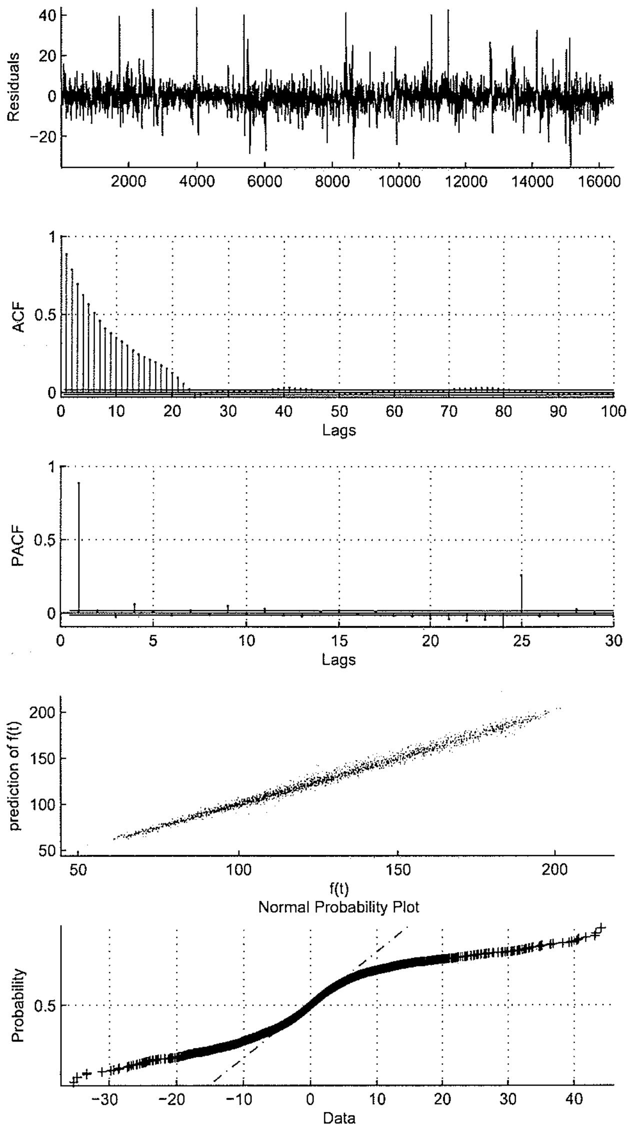

Figure 4.20: The figure consist of residuals from the in sample data for the model 4.20.

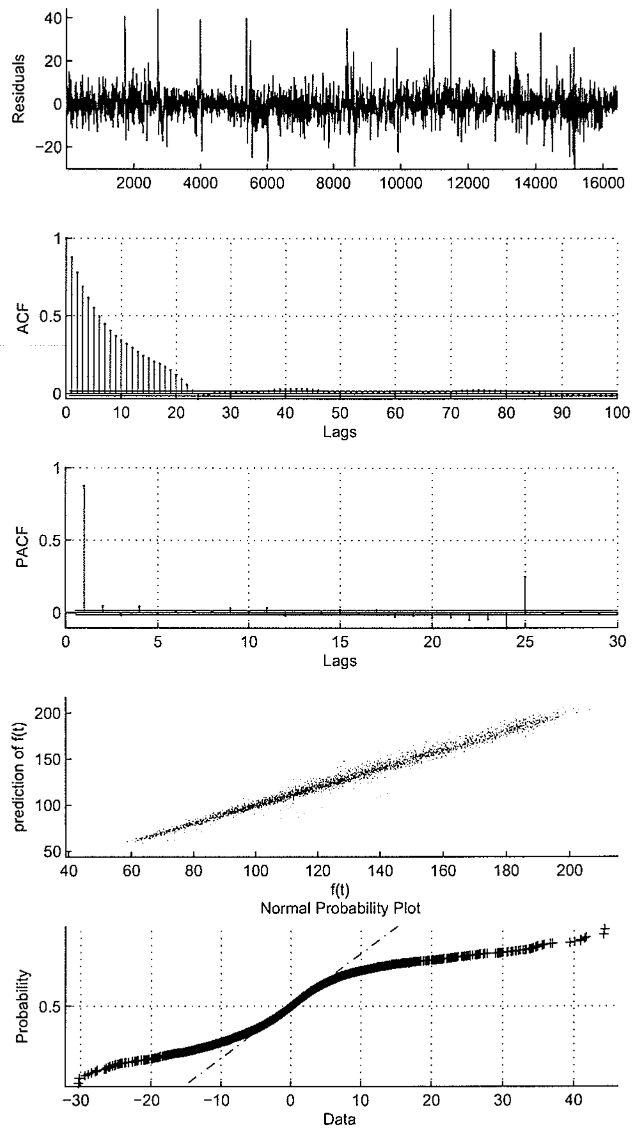

Figure 4.21: The figure consist of residuals from the in sample data for the model 4.21.

4.7 Models with weather variables

In this section we try out different weather input variables to see if a good model can be improved by inserting weather factors. The same input variables as for the model M(4.4) are used for all the models in this set. Since the cross correlation for all the weather variables have seasonal dependency such as with the lagged electricity load all the weather variables are tested with actual hour, 24 and 168 hour delay.

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, T_t, T_{t-24}T_{t-168}\right) + e_t \tag{4.22}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, \bar{G}_t\right) + e_t \tag{4.23}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, \bar{W}_t\right) + e_t \tag{4.24}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, D_t, D_{t-24}D_{t-168}\right) + e_t \tag{4.25}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, H_t, H_{t-24}H_{t-168}\right) + e_t \tag{4.26}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, P_t, P_{t-24}P_{t-168}\right) + e_t \tag{4.27}$$

$$f_t = f\left(f_{t-168}, f_{t-24}, h_t, d_t, T_t, T_{t-24}T_{t-168}, \bar{W}_t, \bar{G}_t\right) + e_t \tag{4.28}$$

where e_t is white noise.

4.7.1 Results

The result indicates that the M(4.22) with the outdoor temperature as input has the best performance compared to all the other models. As demonstrated in data analysis the result shows that the temperature input is the best fitting variable to the electricity load model. Most of the weather variables should be able to have a positive impact on the model. Trimming of them in forms of moving averages, other delays or other kinds of transformations may be needed to find clearer dependencies. In this case we are satisfied with the model M(4.22) that has the best performance in our study.

MODEL	In sample			Out sample		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
(4.22)	3.4253	2.8427	5.2021	4.3166	**3.1819**	6.4106
(4.23)	3.6782	3.0126	5.4889	4.6427	3.3883	6.7113
(4.24)	3.8102	3.1061	5.7557	4.5651	3.3284	6.7780
(4.25)	3.9175	3.2133	5.7565	4.9121	3.5767	7.1201
(4.26)	3.6258	2.9799	5.3978	4.6936	3.4446	6.7735
(4.27)	3.7445	3.0840	5.5395	4.8374	3.5293	7.2558
(4.28)	3.7445	3.0840	5.5395	4.8374	3.5293	7.2558
(4.28)	3.2904	**2.7465**	5.0304	4.3237	3.2625	6.3741

Table 4.7: Results of the performance of the models in this section.

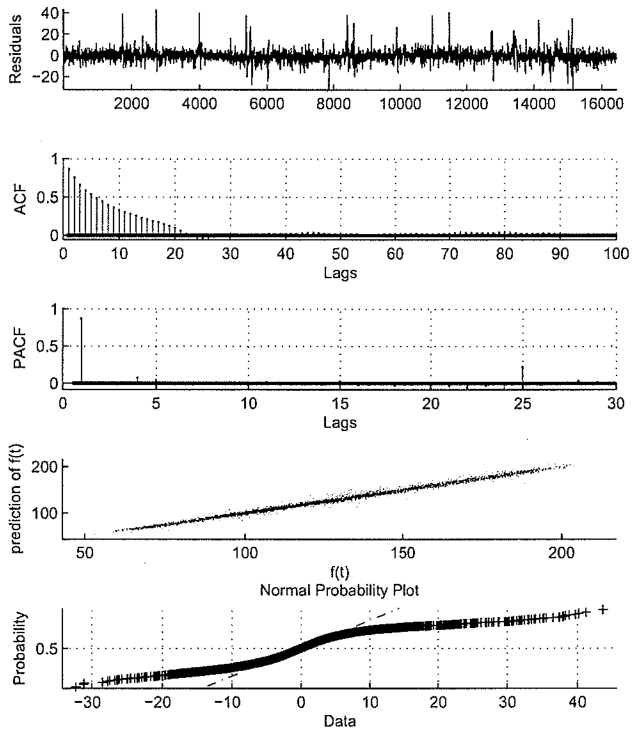

Figure 4.22: The figure consist of residuals from the in sample data for the model 4.22.

Chapter 5

Conclusion

The purpose of this study was to analyze which parameters produce the best fitting electricity load demand forecasts. Accurate electricity load forecasts are highly important for production planning and electricity trading. To identify which variables were the most essential the linear dependency was studied for delayed electricity load and weather variables. The highly seasonal impact of the electricity load was identified and further studied. The conclusion is that the time deterministic components are except for lagged load the most important factors for modeling the electricity load. The results demonstrate that usage of to many input variables may cause over fitting which causes increased errors. The best performing deterministic components are the raw time variables which also have the least used variables. The weather variables have a minor impact on the accuracy and should in general be chosen carefully. This clearly also has to do with which type of electricity load type it is. In some cases it may not even be needed. In our test only the temperature model M(4.22) and the wind model M(4.24) performed better then the best model without weather. The best fit model in the study was M(4.22) which had 2.84% MAE for in sample data and 3.18% MAE for out of sample data.

Some further interesting research would be to analyze the weather dependencies further. The usage of recurrent neural networks that model the moving average error would be interesting to test. The public holiday effect was not handled in this paper. To increase accuracy this should also be handled. To combine a neural network model with a linear model such as an ARIMA model would be interesting since there's already lots of studies around linear electricity load models.

Bibliography

[1] Hastie T., Tibshirani R. and Friedman J. (2001) The Elements of Statistical Learning, Springer.

[2] Madsen H. and Holst J. (2000) Modelling Non-Linear and Non-Stationary Time Series, IMM DTU

[3] Madsen H. (2007) Time Series Analysis, Chapman & Hall\CRC.

[4] Blom G. and Holmquist B. (1997) Statistikteori med tillämpningar, Studentlitteratur, Lund

[5] Lindgren G. and Rootzn H. (2005) Stationära stokastiska processer, KFS AB

[6] Hayking S. (1998) Neural Networks A Comprehensive Foundation (2nd Edition)

[7] Fan J. and Yao Q. (2005) Nonlinear Time Series: Nonparametric and Parametric Methods

[8] E.Gonzlez-Romera, M.A. Jaramillo-Morn, D.Carmona-Fernndez (2008) Monthly electric energy demand forecasting with neural networks

[9] Eugene A. Feinberg, Dora Genethliou (2004) Chapter 12 Load Forecasting, in Applied Mathematics For Restructured Electric Power Systems: Optimization, Control and Computational Intelligence

[10] MacKay J.C D. (1991) Bayesian Interpolation, Computation and Neural Systems, CalTech 139-74

[11] http://www.mathworks.se/help/toolbox/nnet/ref/trainbr.html

[12] Energy Opticon webpage with main presentation. http://www.opticon.se

[13] Deoras A. (2010) Electricity Load and Price Forecasting Webinar Case Study http://www.mathworks.se/matlabcentral/fileexchange/28684-electricity-load-and-price-forecasting-webinar-case-study

Master's Theses in Mathematical Sciences 2012:E10
ISSN 1404-6342

LUNFMS-3036-2012

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/