Forecasting model of electricity demand in the Nordic countries

Tone Pedersen

3/19/2014

Abstract

A model implemented in order to describe the electricity demand on hourly basis for the Nordic countries. The objective of this project is to use the demand data simulated from the model as input data in the price forecast model, EMPS model, at Vattenfall. The time horizon is 5 years, 6 years including the current year. After different models tried out, the final model is described by fundamental and autoregressive time variant variables, an ARX model. The variable of temperature is described by historical data from 46 years which are used to create an idea of the outcome variation depending on the weather. Non parametric bootstrap of the residuals is used when adding noise to the simulation. The ARX parameters was estimated by prediction error method but a two-step estimation was also tried, by first estimating the fundamental parameters and then model the rest of the demand by an AR process. The second method was supposed to increase the weight on the fundamental variables. Results of the simulation Indicates of a realistic description of the electricity demand which is an improvement of the earlier demand input to the EMPS model but the difference is not always seen in the outcome of the price forecast. The results are discussed in Chapter 5.

Acknowledegements

I would like to thank my supervisor at Vattenfall, Anders Sjögren, and specially Roger Halldin at Vattenfall for all his help and guidance throughout this thesis. I would also like to thank Björn Wetterberg at Vattenfall and my examiner at Lund University Erik Lindström.
.

Contents

1. Introduction

1.1 Background

This master thesis is a project for Vattenfall AB and the business unit Asset optimization Trading, AOT . Vattenfall is one of the biggest energy companies in Europe where one of its main tasks is to produce and sell electricity to a profitable price.

In order to maximize the revenue it is important to have a good price prognosis. The MA/ seasonal planning department is the department that is in charge of the price prognosis of the next five year. The MA/Seasonal planning is using a forecast price model, the EMPS model, or EFI as it is called internal in Vattenfall. The EMPS model is taking into consideration input parameters as such consumption, transmission capacity water values (which are computed in the system), fuel prices, etc.

1.1.1 EMPS model

The main purpose of the EMPS model is to forecast the spot price in deregulated markets. The price is evaluated by considering estimation of the water values and the input data provided by the user.

The EMPS model of Vattenfall, called EFI, is a forecast of the next five years, six including the current year, electricity price with a weekly resolution.

The EMPS model has two steps in its procedure. The first part, called the strategy part, develops the water values. In the second part the system uses the water values and the input data from the user to simulate the output. The output consists of water values for water power stations, prognosis to the Nordic system price, the price areas, the power production and reservoir development.

EMPS is a model created to optimize and simulate the hydropower system. It takes the transmissions between different regional reservoirs and between bigger areas into consideration. The system uses the flexibility in the hydropower system to stabilize future uncertain inflows which are less steerable or non-steerable. The hydropower is optimized in relation to regional hydrological inflows, thermal generation and power demand.

The steering of the hydropower is done by computing water values for each region. A high water value indicates a lower water level/volume in the reservoir. The water value is computed by a stochastic dynamic programming where also the interaction between the areas are included. Optimal operational decisions are evaluated for each time step for thermal and hydro production. This is done by considering the water value of the aggregated regional subsystems. In each subsystem a more detailed plan of the distribution of the production is done according to the number of plants and reservoirs.

In the EFI model, the price is simulated with 46 different weather scenarios, actual historical weather years. The scenarios are used in blocks of five years to include an actual historical 5 years weather change.

The existing EMPS model simulates the 46 simulations 5 times per week which will expand to 168 times per week if the model will be transformed to an hourly based model. An update of the EFI model with 84 times 46 simulations per week is done but not jet implemented due to non-updated input data.

1.1.2 What is the need of a demand forecast on an hourly basis?

Over the years the renewable energy sources wind and solar power have rapidly increased. The sources are obviously only producing when there is wind or sun which is why these renewable sources cannot be a base energy source, a source to rely on. When wind or solar power or both are producing a lot of energy they also increase the supply on the market. The supply and demand curves of the electricity market will meet at a lower electricity price. If the wind and solar power would produce a constant amount of electricity during specified periods the electricity spot price would not be hard to forecast. This is where the time resolution becomes a problem for the model. The wind and solar power are only producing when there is wind or solar which is sometimes hard to forecast and can change quickly during a day. This results raises the volatility in the electricity price during the day. Because the price forecast is set on a weekly basis the daily high and low peak prices will not be caught in the model. The information that is lost makes the model output lose its momentum.

In order to catch the momentum during the day and improve the resolution of the output data, the input data must contain information on the same time scale basis as the output data that is on an hourly basis. One of the input data to the model is the electricity demand forecast for the next 5 years. The demand data is today on a monthly basis which is added together to annual data, the price model is then distributing the demand data over the year on an two hourly basis. The system has a daily, weekly and yearly profile of the consumption which is used when distributing the monthly consumption data.

1.1.3 Background of the electricity demand

Parts of the electricity consumption are static processes in form of the consumption of the households and the consumption of lighting. This part is relatively easy to model since it is not affected by unexpected events. The households consumption of electric heating during the winter is closely related to the temperature. When the temperature is very low then the electric heating stagnates, meaning it does not increase more. Electricity consumption versus temperature is deferring depending on if you are in southern, middle or in northern Europe. The difference is due to the use of air condition in southern and middle Europe. As the air condition uses as much electricity as the electric heating the electricity consumption is higher in southern and middle Europe than in northern Europe. In the north the temperature related electricity consumption could be omitted during summer.

Another part of the electricity consumption is the electricity consumed by the industry. The industry looks very stable on a daily basis but looking at a long run perspective the industry production varies with the economic cycle. If the economic situation is falling the situation in the country will worsen and some industries will have difficulties to survive, these might therefore decrease the production and in worst case shut down their industry. This part is hard to model since it needs to be observed from many perspectives and deeper investigation is needed to detect the industries that will disappear. Because of this it was decided to use the already existing demand forecast on a monthly basis where investigation and previous knowledge is creating the forecast.

The model should partly be based on analyses of these factors and analyses of other possible contributing factors.

1.1.3.1 Previous models used within electricity demand

The electricity demand, is a quite well investigated area where most of the models have a time horizon of either intraday, one week, one year or long term as 10 years ahead. Speaking of general modeling of electricity demand, the traditional techniques of forecasting are regression, multiple regression, exponential smoothing and iterative least squares techniques (Singh, Ibraheem, Khatoon, & Muazzam, 2013). The range of models are varying between manual methods that has been tested operationally and formal mathematical approaches. (BOFELLI & MURRAY, 2001)

The most popular model is the multiple regression model which is describing the demand by a number of factors that are affecting the electricity demand in different ways (Singh, Ibraheem, Khatoon, & Muazzam, 2013). The interest of load forecasts is typically aimed to the hourly quantity of the total system load. (Alfares & Nazeeruddin, 2002)

The development of the traditional methods has modified the previous techniques by keep track of the environmental changes and update the parameters during the forecast along with the changes. The most popular model among the modified technique is the stochastic time series methods. The time series models are looking for internal structures such as seasonal trends and autocorrelation.

All models tried out are more or less imprecise and uncertain due to the fact that there are unknown or totally random variables. Instead of using hard computing, trying to find the exact solution, soft computing solution has over the last few decades been used. The soft computing is using the environment of approximation rather than being exact. (Singh, Ibraheem, Khatoon, & Muazzam, 2013)

Looking at the chronological order, the ARMAX model is found straight after the stochastic time series. As many of the popular models are best suited to short term load forecasting this model has also been tried out on long term where the result was compared with regression methods. (Alfares & Nazeeruddin, 2002)

Another technique that has been used within the electricity demand model is the dynamic factor model which can be modified in several ways. (Mestekemper, 2011) It is briefly explained as reducing the dimension of the original set of data which gives e.g. better parameter estimation. The method is mostly used when the time horizon is short. See appendix B for further reading.

1.2 Purpose

The main purpose of the model is to describe the electricity consumption during the days depending on what weekday it is, if it is a public holiday or an expected vacation day. The model should also be able to observe how the electricity consumption is varying depending on the temperature input. The model should also be able to change the pattern while the input data is being updated and explain future year's consumption.

The idea of the project is to improve the model of the consumption data used for the EMPS model with an hourly resolution by systemizing it and increasing the quality and reliability of the model, compared to today's monthly consumption data. The outcome of the project is an implemented model where the output is forecasted hourly consumption for the next five years. The main focus in the model is to create a normal consumption year and then use the monthly forecasted consumption, from the consumption model that is used for the EMPS model today, to profile the future years. The model should be implemented and calibrated for each of the Nordic countries separately.

The path of the project includes investigations of different models and to find a suitable definition of the demand. Different variables were tested to conclude what is affecting the electricity consumption. The analysis of the observed consumption is investigated mostly by time series analysis and several time series models have been investigated for a possible fit. Static models have also been tried and by then combinations of stochastic processes based on regression models was also included.

In order to handle the noise added to the model I used block bootstrap and tried different ways to identify the periods of the more significant noise.

The simulation was tested out by different methods, e.g. prediction with different time horizons. The final method was to use prediction with one time step as time horizon and then add noise. This was executed for each time point and continued until the timeline of the simulation was outlined.

1.3 Limitations

Included in the specification when I started the project was to be able to see the trends in the price if the electricity consumption was distributed differently during the day. The goal was to see the change in price if the consumers was affected by the information available in order to consume the electricity when the price was lower during the night.

There is no trend seen when looking at historical data which means that the data has to be created by modification of the real data. A possible way of create this data is to investigate the behavior within peoples habits and what would be a possible future scenario of the consumption. This investigation would need a lot of information analyzed and this was too time consuming.

Another aspect of the project that had to be removed from the scope was to modify the economic growth of the year more precisely. The idea was to identify the industrial consumption since the factories is the group which is most affected of an economic national change. If the economic growth goes down a lot of factories has to decrease their production and worst of all shut down. And the other way around if the economic situation will improve. In order to identify the industrial consumption in the northern countries a lot of data had to be found to identify the factories. The simplified solution was to modify all the consumption which still gives a better way to describe the consumption than the outcome data used before.

1.4 Result

The main result of this master thesis shows that the model developed captures the annually, weekly and daily trends. It also follows the changes in the temperature very good. With the noise added and with 46 weather scenarios as temperature input data the 95% confidence interval gives has a relatively good spread. The validation of the spread is done by calculating the amount of observed demand hours which is outside the confidence interval which is slightly above 5%. 5% would be approved as it might be in the quantiles of the 95% confidence interval.

Another objective was to use the demand model to simulate input data to the EMPS model. The result showed that the more specific hourly model improved the variation from day to night during the summer period. There is hardly any significant changes in the spring period, some days have more variations from day to night others are the same as with the old demand data.

1.5 Outline

This thesis is structured into 6 sections. It starts with an introduction about the EMPS model and why this project was set up. Section 2 is handling the theory needed to know to understand the modeling part and some of the result section. The theory can be read with varied carefulness depending on the mathematical background. This section is mostly describing and defining methods used within time series analysis but also describes bootstrap which is used when simulating the model. The modeling part, section 4, is guiding the reader through the way taken to arrive to the ARX model which was chosen. Section 4 is also showing the method used for simulating the model. The modeling section is followed by the results which is mostly analyzing the result of the demand model. The section concludes with result from the EMPS model, if the change from demand on annually basis to hourly basis had an impact on the price and how the noise added to the demand data was seen in the price. The last section is discussing the results and also give some tips of future work.

2. Theory

This part contains the theory behind the tools that has been used to conclude the best model solution. The final model is structured by a time series model with deterministic input variables that are effecting the electricity consumption in a one way relation. Meaning e.g. the electricity demand would change if the temperature changed but the temperature would not be effected if the electricity consumption changed.

2.1 Time series analysis

A time series $\{x_t, t = 0 \pm 1, \ldots\}$ is a realization of a stochastic process $\{X_t, t = 0, \pm 1 \ldots\}$.

DEFINITION 1 Stochastic process

The process $\{X_t, t = 0, \pm 1 \ldots\}$ is a family of random variables $\{X(t)\}$ where t belongs to an index set.

Time series analysis are statistical methods, e.g. time series models, that are often used to model physical events as stochastic processes.
The stochastic process has two arguments$\{X(t, \omega), \omega \epsilon \Omega\}$, $X(t,\cdot)$ is a random variable for a fixed t and ω is the sample space on the set of all possible time series, Ω, that can be generated by the process.

The stochastic process described must have an ordered sequence of observations. The ordered sequence of observations is the order through time at equally spaced time intervals. (Madsen, Time Series Analysis, 2008)

The model used can be seen as tools often used when it is hard to describe all patterns by deterministic variables or there are no fundamental parameters at hand. The time series models finds the underlying forces by observing the correlation between previous and current data points. The linear time series models are constructed by either autoregressive parameters or a moving average parameters, or a combination of these two. The autoregressive parameters is looking after a lagged correlation between the current data point and previous data points. The moving average is looking after the correlation in the residuals that is deviating from a mean of all data points.

Two linear processes that are the base for time series models are the Moving average process, MA process and the autoregressive process, AR process which are defined as follow

DEFINITION 2 The MA(q) process

The process $\{Y_t\}$ given by

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \qquad (2.1)$$

Where $\{\varepsilon_t\}$ is white noise, is called a *Moving Average process* of order *q.* In short it is denoted an MA(q) *process.*

DEFINITION 3 The AR(p) process

The process $\{Y_t\}$ given by

$$Y_t + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} = \varepsilon_t, \tag{2.2}$$

Where $\{\varepsilon_t\}$ is uncorrelated white noise, is called an *autoregressive process* of order *p* (or an *AR(p) process).*

(Madsen, Time Series Analysis, 2008)

2.1.2 Time series models

In the following part, the theory of the models I tested will be described shortly just to be sure that reader understand section 3, the modeling part.

2.1.2.3 ARMA

Auto regressive moving average process has the following equation

DEFINITION 4 The ARMA(*p,q)* process

The process $\{Y_t\}$ given by

$$Y_t + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_t + \cdots + \theta_q \varepsilon_{t-q}, \tag{2.3}$$

Where $\{\varepsilon_t\}$ is white noise is called an *ARMA(p,q) process.*

2.1.2.2 SARMA

SARMA, Seasonal Autoregressive moving average process, removes the seasonal pattern from the data sequence before fitting the data to the ARMA model.

DEFINITION 5 Multiplicative $(p, q) \times (P, Q)_s$ Seasonal model

The process $\{Y_t\}$ is said to follow a multiplicative $(p, q) \times (P, Q)_s$ *seasonal model* if

$$\varphi(B)\Phi(B^s)\nabla_s^D Y_t = \theta(B)\Theta(B^s)\varepsilon_t \tag{2.4}$$

Where $\{\varepsilon_t\}$ is white noise φ and θ are polynomials of order p and q, respectively, and Φ and Θ are polynomials of order P and Q, which have all roots inside the unit circle.

The method of seasonal adjustment is used to capture the underlying trend which becomes more distinct when removing the seasonal trend. Long term forecasting is less suitable for seasonal models as they are adapt for non-stationary data which is hard to forecast in long term. More reading will follow in section 3.1.1.

2.1.2.1 SARIMA

The SARIMA, seasonal autoregressive integrated moving average process differentiates the data by both considering the seasonal periodic pattern of order D and differentiating afterwards the data by order d to become stationary. After filtering the data by the seasonal differentiation the model structure will be easier to identify by ACF and PACF. The equation of SARIMA looks like,

DEFINITION 6 Multiplicative $(p, d, q) \times (P, D, Q)_s$ Seasonal model

The process $\{Y_t\}$ is said to follow a multiplicative $(p, d, q) \times (P, D, Q)_s$ *seasonal model* if

$$\varphi(B)\Phi(B^s)\nabla^d\nabla_s^D Y_t = \theta(B)\Theta(B^s)\varepsilon_t \qquad (2.5)$$

Where $\{\varepsilon_t\}$ is white noise φ and θ are polynomials of order p and q, respectively, and Φ and Θ are polynomials of order P and Q, which have all roots inside the unit circle. (Jakobsson)

This seasonal method is hard to use when the focus of the model should be on the seasonal pattern and not only to find the underlying factors. This is the same issue as in 2.1.2.2.

The other problem with SARIMA is that it is integrated one time which means in this case that it has been differentiated. The differentiation also makes it difficult to go backwards since it is only the difference between the data points that is used when the model is done.

2.1.1 Identification of the model and model order

The identification of the model is based on the stochastic process found in the data. The data modeled is one or more time series. To be able to describe a stochastic process with a time series model the process has to be stationary.

DEFINITION 7 weak stationary

A process $\{X(t)\}$ I said to be *weakly stationary* of *order k* if all the *first k* moments are invariant to changes in time. A weakly stationary process of order 2 is simply called *weakly stationary.* (Madsen, Time Series Analysis, 2008)

One of the primary tools for time series analysis is the estimation of the correlation.

2.1.1.1 ACF – autocorrelation function

DEFINITION 8 Autocovariance function

The autocovariance function is given by

$$\gamma_{XX}(t_1, t_2) = \gamma(t_1, t_2) = Cov[X(t_1), X(t_2)] = E\big[\big(X(t_1) - \mu(t_1)\big)\big(X(t_2) - \mu(t_2)\big)\big]$$

And the autocorrelation function is given by

$$\rho_{XX}(t_1,t_2) = \rho(t_1,t_2) = \frac{\gamma_{XX}(t_1,t_2)}{\sqrt{\sigma^2(t_1)+\sigma^2(t_2)}} \qquad (2.6)$$

The time series is from start at least or is integrated to become weak stationary before continuing the modeling. If the time series is stationary the autocorrelation function will be a function of the time difference $\tau = t_2 - t_1$,

$$\rho(\tau) = \frac{\gamma_{XX}(\tau)}{\gamma_{XX}(0)} = \frac{\gamma_{XX}(\tau)}{\sigma_X^2} \qquad (2.7)$$

The difference from the earlier equation is that the variance is the same no matter of where in the process you are. For example $\gamma_{XX}(t_1 - t_2) = \sigma_{t_1}\sigma_{t_2} = \sigma_X^2$ due to the variance should be equal in all time steps.

To investigate which model structure that is appropriate to describe the data the autocorrelation function, ACF is a good way to start with.

As defined above the autocorrelation function is a description of the relation between the covariance of two data points and the variance of the process. The output is the correlation between the data from different time steps in the process.

To identify if the process is a pure AR process, autoregressive process, or a pure MA process, moving average process the following theorem will show what to be observant of.

THEOREM 1 Property for AR processes

For an AR(p) process *it holds*

$$E[\widehat{\varphi_{kk}}] \cong 0 \qquad (2.8)$$

$$Var[\widehat{\varphi_{kk}}] \cong \frac{1}{N} \qquad (2.9)$$

k=p+1,p+2,...

where N is the number of observations in the stationary time series and φ_{kk} is the PACF.

The ACF of the AR process has the characteristics as a damped exponential and/or a sine function.

THEOREM 2 Property for MA processes

For a MA(q) process *it holds*

$$E[\widehat{\rho(k)}] \cong 0 \qquad (2.10)$$

$$Var[\widehat{\rho(k)}] \cong \frac{1}{N}\left[1 + 2\left(\hat{\rho}^2(1) + \cdots + \hat{\rho}^2(q)\right)\right] \qquad (2.11)$$

k=q+1,q+2,...

where N is the number of observations in the stationary time series and $\rho(k)$ is the ACF.

The orders of the pure processes can be observed by looking at the estimated variables $\widehat{\rho(k)}, (k > q)$ and $\widehat{\varphi_{kk}}, (k > p)$, which should be approximately normally distributed. For example the $\widehat{\varphi_{kk}}$ is approximately zero after the order p since there are no correlation in the current time step and the delays further on.

If the process is mixed, there are both MA and AR processes within the stochastic process, the model order is far more difficult to discover where the trial and error method is useful. Different orders are tried out until no significant lags are seen. (Madsen, Time Series Analysis, 2008)

The right model order is found when the parameters of the Sample Autocorrelation Function, SACF, no longer are significant except of k=0 which will always be 1 since it is the variance divided by itself. SACF is the autocorrelation function of one sample. (Madsen, Time Series Analysis, 2008)

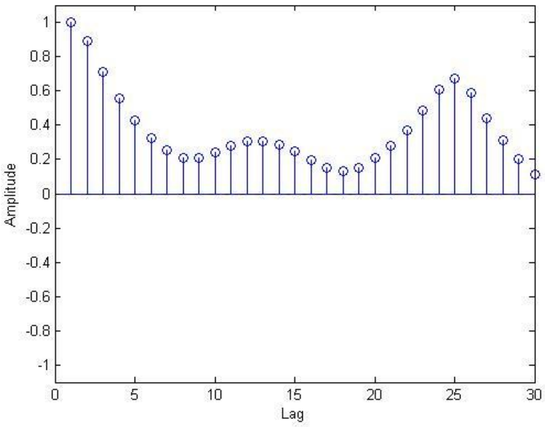


Figure 1 The autocorrelation function when there is a correlation between current data point and the data point in time point 25. This is a pure AR process since it has the form a sine function.

When analyzing the residuals after a simulation theorem 3 can be used to see if the simulation is simulating correct.

DEFINITION 9 The inverse autocorrelation function

The inverse autocorrelation function (IACF) for the process, *$\{Y_t\}$*, is found as the autocorrelation at lag k is denoted $\rho i(k)$.

THEOREM 3 Inverse Autocorrelation function for AR processes

For an AR(p) process it holds that

$$\rho i(k) \neq 0, k \leq p, \tag{2.12}$$

$$\rho i(k) = 0, k > p.$$

PROOF Follows from the fact the at the AR(p) process, $\varphi(B)Y_t = \varepsilon_t$, can be written as $Z_t = \varphi(B)\varepsilon_t$ if Y_t is stationary.

2.1.1.2 Optimization of the number of parameters and model order

The following methods are used in order to validate the model order and also to see if the number of parameters is optimized.

The loss function is using the residual sum of squares, RSS, and is comparing the RSS for different number of parameters in order to see when the number of parameters is optimized. The RSS is written as

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.13}$$

The RSS gives information of the proportion of the variation explained by the model compared to the total explained variation in y. (Madsen, Time Series Analysis, 2008) (page 34)

The optimized number of parameters is seen when the model is not being improved by additive parameters.

The loss function is written as

$$S(\theta_{\underline{i}}) = \sum_{i=1}^{N} \varepsilon_t^2 \, (\theta_{\underline{i}}) \tag{2.14}$$

Where the index $\underline{i}$ is the number of parameters. The loss function seeks its minimum for the least number of parameters. The expression holds that when the model is extended with one more parameter, from $\theta_{\underline{i}}$ to $\theta_{\underline{i}+1}$, then $S(\theta_{\underline{i}+1}) < S(\theta_{\underline{i}})$. The gain of including on more parameter will be less as the number of parameters increase and the loss function curve will stagnate. (Madsen, Time Series Analysis, 2008)

Another choice when a leak of data points is the issue is Akaike´s information criteria, AIC, is an option. The AIC measures the quality of the model and the information lost when describing the observed data.

$$AIC = -2log(max. likelihood) + 2n_i \tag{2.15}$$

where n_i is the number of estimated model parameters. AIC is choosing the model order when it is minimized.

The maximum likelihood is a way of estimating parameters. The estimation of the parameters are optimized when the maximum of the probability of the estimated parameter values is reached. The formula of maximum likelihood is,

$$\hat{\theta} = \underset{\theta}{\arg\max}\, P(y; \theta)$$

Where P is the probability function of the parameter θ and $\hat{\theta}$ is the parameter estimated when the P is maximized.

Final prediction error is closely related to and has the following equation,

$$FPE(d) = V_n(\theta_n)\left(\frac{1+{}^{d}/_{n}}{1-{}^{d}/_{n}}\right) \tag{2.16}$$

Where d is the number of estimated parameters, n is the number of values in the input data set and $V_n(\theta_n)$ is the loss function of the estimated parameters θ_n.

For each model order tested an estimation of the FPE will be computed and the order which gives the minimum FPE will be chosen.

The FPE gives an approximation of the prediction error in the future. (Akaike, 1969)

2.2 Regression models

A classical regression model is describing a static relation between one dependent parameter Y_t and one or more independent parameters $X_{1t}, X_{2t}, \dots, X_{pt.}$ The regression model differ from the time series analysis by instead of using the time as an index, the variables are known for each time, t, and simple calculating one time step at the time. In the time series analysis the observations are modeled by the pattern over time and not by each time step. The regression model is written as follow

$$Y_t = f(\boldsymbol{X}_t, t;\ \boldsymbol{\theta}) +\ \varepsilon_t \tag{2.17}$$

Where $f(\boldsymbol{X}_t, t;\ \boldsymbol{\theta})$ is a known function with known variables, $\boldsymbol{X}_t$, at time t but with unknown parameters, $\boldsymbol{\theta}$. E.g. of the function is the general linear regression model

DEFINITION 10 General linear model

The general linear model (GML) is a regression model with the following model structure

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t \tag{2.18}$$

Where $\boldsymbol{x}_t = (x_{1t}, \dots, x_{pt})^T$ is a known vector and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$are the unknown parameters. ε_t is a random variable with mean $E[\varepsilon_t] = 0$ and covariance $Cov\left[\varepsilon_{t_i}, \varepsilon_{t_j}\right] = \sigma^2 \boldsymbol{\Sigma}_{ij}$.

The random variable ε_t is assumed to be independent of $\boldsymbol{x}_t$ since it is supposed to be the part of the observed data that was randomly around zero and impossible to model.

This way of modeling the data observed has advantages when all input variables are known and when different scenarios is wanted, scenarios meaning different outcome of the same variable e.g. different temperature scenarios. Different scenarios can be seen by modifying the input data or try different variables in order to see what is affecting the output data. With the aim of finding which parameters that optimize the model, hypothetical tests can be done. Briefly explained, the hypothetical test is checking the significance of the variable added, if it is adding value to the model or if it is only by chance adding value and would then be rejected, not included in the model.

Simulation done by a regression model will always give the same output except from the noise added.

2.3 Combining time series with deterministic modeling

The time series models are good in describing the data by non-fundamental factors and finding the underlying forces. The deterministic model is good when the dynamics in the data have to be included and scenarios are wanted. If the regression model is not good enough, due to all variables are not known, it is a good thing to combine these two ways of describing the observed data. The regression model will extend the time series model in terms of input exogenous parameters.

2.3.1 ARX

ARX, Autoregressive exogenous process, is an autoregressive time series model with exogenous input parameters. The known parameters will describe the model and a filter created by the correlation between the previous output data points will fulfill the model.

$$\varphi(B)Y_t = \omega(B)u_t + \varepsilon_t \tag{2.19}$$

Which can be expressed as,

$$Y_t + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} = u_t + \omega_1 u_t + \cdots + \omega_s u_{t-s} + \varepsilon_t$$

(Söderström & Stoica, 1989)

Where u_t is the external variable at time t. B is the backshift operator which is going to be defined and explained later.

2.3.2 ARMAX

ARMAX, Autoregressive moving average exogenous process, is an extension of an ARX where moving average parameters are added.

$$\varphi(B)Y_t = \omega(B)u_t + \theta(B)\varepsilon_t \tag{2.20}$$

Which can be expressed as,

$$\begin{aligned} Y_t + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} \\ = u_t + \omega_1 u_t + \cdots + \omega_s u_{t-s} + \varepsilon_t + \theta_1 \varepsilon_t + \cdots + \theta_q \varepsilon_{t-q}, \end{aligned}$$

2.4 Parameter estimation

There are several ways of estimating the parameters of a model. One way is by least squares estimates, LS. The LS estimation is aiming at estimating the parameters $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ such that the $f(\boldsymbol{x}_t; \widehat{\boldsymbol{\theta}})$ is describing the observations as good as possible. The LS method finds the parameters optimized when the residuals have the least square, $\sum[y_t - f(\boldsymbol{x}_t; \boldsymbol{\theta})]^2$.

DEFINITION 11 LS estimates

The *Least Squares (unweighted) estimates* are found from

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} S(\boldsymbol{\theta}),$$

Where

$$S(\boldsymbol{\theta}) = \sum_{\mathbf{1}}^{\boldsymbol{N}} \sum[y_t - f(\boldsymbol{x}_t; \boldsymbol{\theta})]^2 . = \sum_{\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{\varepsilon}_{\boldsymbol{t}}^{\mathbf{2}}(\boldsymbol{\theta}) \tag{2.21}$$

i.e. $\widehat{\boldsymbol{\theta}}$ is the $\boldsymbol{\theta}$ that minimizes the sum of squared residuals.

The term unweighted is used if the variance of the residuals is constant. The residuals might have a larger variance where correlation might occur, if that is the case weighted least squares estimations are made.

The variance of the parameters is used when calculating the confidence interval of the parameters which is an important observation to see if the parameter is significant or not.

$$Var[\widehat{\boldsymbol{\theta}}] = 2\hat{\sigma}^2 \left.\left[\frac{\partial^2}{\partial^2 \boldsymbol{\theta}} S(\boldsymbol{\theta})\right]^{-1}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \tag{2.22}$$

Where $\hat{\sigma}^2 = S(\widehat{\boldsymbol{\theta}})/(N-p)$.

(Madsen, Time series Analysis, 2008)

2.4.1 Prediction error model

An extension of the LS model is the prediction error model which is a more complex model where a minimization of the prediction error is implemented. Given a model the parameters $\boldsymbol{\theta}$ are calculated by

$$\widehat{\boldsymbol{\theta}} = arg \min_{\theta} \{S(\boldsymbol{\theta}) = \sum_{t=t_0}^{N} \varepsilon_t^2 \, (\boldsymbol{\theta})\}. \tag{2.23}$$

where

$$\varepsilon_t(\boldsymbol{\theta}) = Y_t - \hat{Y}_{t|t-1}(\boldsymbol{\theta}) \tag{2.24}$$

The conditioned estimated output is calculated by

$$\hat{Y}_{t|t-1}(\boldsymbol{\theta}) = E[Y_{\boldsymbol{t}}|Y_{\boldsymbol{t-1}}, \boldsymbol{\theta}] \tag{2.25}$$

As understood by the name of the method one is estimating the parameters by observing when the expected output of the model, with condition on the last output in t-1, is as close as possible to the observation. In other words the parameters are set when the prediction error one step ahead is in its minimum.

The problem is how to calculate the $E[Y_{\boldsymbol{t}}|Y_{\boldsymbol{t-1}}, \boldsymbol{\theta}]$ which is demonstrated below for a model with deterministic input, the ARX model.

To be able to understand the derivation of the conditional mean, knowledge of z-transformation and the backwards shifting operator is very useful.

The backward shifting operator is using the z transform to turn the time series difference equation to be convergent.

DEFINITION 12 The z-transform

For a sequence $\{x_t\}$ the z-transform of $\{x_t\}$ is defined as the complex function

$$Z(\{x_t\}) = X(z) = \sum_{t=-\infty}^{t=\infty} x_t z^{-t} \tag{2.26}$$

The z-transform is defined for the complex variables z for which the Laurent[1] series converges.

DEFINITION 13 Backward shifting operator

The backward shifting operator z^{-1} is defined as

[1] The series described in the definition.

$Z(\{x_{t-1}\}) = \sum_{t=-\infty}^{t=\infty} x_{t-1}\, z^{-t} = z^{-1} \sum_{t=-\infty}^{t=\infty} x_{t-1}\, z^{-(t-1)} = z^{-1}X(z) = z^{-1}Z(\{x_t\})$ (2.27)

The advantages with the z-transform and by then the backward shifting operator is that convolution in the time domain is equal to multiplication in the z domain. It is simpler to work with multiplication than convolution.

An example of how the backwards shifting operator, B, is used,

$$BY_t = Y_{t-1}$$

And if an autoregressive process of order p is described with an backward shifting operator it is written

$$\varphi(B) = (1 + \varphi_1 B + \cdots + \varphi_P B^P)$$

Where the polynomial of B indicates the order of the model and the backward shift operator is often expressed by the z-transform,

$$\varphi(B)Y_t = (1 + \varphi_1 B + \cdots + \varphi_P B^p)Y_t = (1 + \varphi_1 z^{-1} + \cdots + \varphi_P z^{-p})Y(Z)$$

$$= (1 + \varphi_1 Y_{t-1} + \cdots + \varphi_P Y_{t-p}).$$

For a time series model with deterministic input the prediction error would be,

$$Y_{\boldsymbol{t}} = H_1(B)u_t + H_2(B)\varepsilon_t \tag{2.28}$$

The $\{u_t\}$ is the deterministic variable. Here $H_1(B)$ and $H_2(B)$ are rational transfer operators where operator is the backward shifting operator. The transfer functions H_1 and H_2 is transforming the variables from time domain to z domain.

For an ARX model the equation would be

$$\varphi(B)Y_t = \omega(B)u_t + \varepsilon_t \tag{2.29}$$

And the rational transfer function would be

$$H_1(B) = \varphi^{-1}(B)\omega(B) \tag{2.30}$$

$$H_2(B) = \varphi^{-1}(B) \tag{2.31}$$

The conditional mean, $E[Y_{\boldsymbol{t}}|Y_{\boldsymbol{t-1}}, \boldsymbol{\theta}]$ for the ARX model following the formula (24) and keeping the rational transfer operators is demonstrated beneath. The goal of the derivation is to achieve an expression where the observed Y_t is included and from there have an expression of the prediction error. The parameters are chosen in order that the minimum of the prediction error is allocated.

$$\hat{Y}_{t|t-1}(\boldsymbol{\varphi}) = E[Y_{\boldsymbol{t}}|Y_{\boldsymbol{t-1}}, \boldsymbol{\varphi}] = H_1(B)u_t + H_2(B)\varepsilon_t - \varepsilon_t = H_1(B)u_t +$$

$$[H_2(B) - 1]\varepsilon_t = H_1(B)u_t + [H_2(B) - 1]\varepsilon_t \tag{2.32}$$

$$[\varepsilon_t(\boldsymbol{\theta}) = Y_t - \hat{Y}_{t|t-1}(\boldsymbol{\theta}) = {H_2}^{-1}(B)(Y_t - H_1(B)u_t)]$$

$$\hat{Y}_{t|t-1}(\boldsymbol{\varphi}) = H_1(B)u_t + [H_2(B) - 1]{H_2}^{-1}(B)(Y_t - H_1(B)u_t) = H_1(B)u_t + \left[1 - {H_2}^{-1}(B)\right][Y_t - H_1(B)u_t]$$

$$= Y_t(\mathbf{1} - {H_2}^{-1}(B)) + {H_2}^{-1}(B)H_1(B)u_t \qquad (2.33)$$

With initial conditions given e.g. $\hat{Y}_{t|t-1} = 0$, $Y_t = 0$ and $u_t = 0$ for $t \leq 0$ it is possible to calculate the prediction error.

The equation for the ARX model is

$$\hat{Y}_{t|t-1}(\boldsymbol{\varphi}) = Y_t\left(1 - \varphi(B)\right) + \varphi(B)\varphi^{-1}(B)\omega(B)u_t = Y_t\left(1 - \varphi(B)\right) + \omega(B)u_t \qquad (2.34)$$

(Madsen, Time Series Analysis, 2008)

With expression (2.34) it is possible to calculate the prediction error, formula (2.24).

It can happen that the PEM model in the software program does not find the right minimum when optimizing the parameters. The minimum error is being found by the loss function which can have a shape that contains local minimum.

The software also assumes that the parameter optimization worked well and calculates the variance from this assumption.

In order to check that the right minimum was chosen different initial values can be tested and if the same value is set to the optimized parameter when the initial values were in different parts of the function it is clear that the global minimum was found. (Söderström & Stoica, 1989).

2.4.1.2 Model validation

The next step in modeling is to evaluate if the estimated model is describing the observation in an adequate way? There are a number of methods available but none of the methods can be used by itself say that if the model is good or bad. Several methods have to be used and analyzed to give different aspects of the model. In this section some of the methods will be presented.

2.4.1.2.1 Cross validation

One of the most common checks of the model is cross validation which is using the model on a dataset that was not included when estimating the model. The method is used to evaluate the accuracy of the model in practice by comparing the output is compared to observed data. (Madsen, Time Series Analysis, 2008)

2.4.1.2.2 Residual analysis

The aim when estimating a model is to describe the observed data so well that the remaining residuals will only be white noise.

By just observing the plot of the residuals, $\{\varepsilon_t\}$, it will be revealed if there are outliers and non-stationary. If the residuals are white noise it will be seen in the autocorrelation function, ACF explained earlier in section 2.1.1.1 and the output $\rho(\tau)$

$$\widehat{\rho_\varepsilon}(\tau) \in_{approx.} \mathrm{N}(0, \frac{1}{N})$$

The output, $\hat{\rho}_\varepsilon(\tau)$, will be approximately zero except for the τ =0 which will be $\hat{\rho}_\varepsilon(0) = 1$. (Madsen, Time Series Analysis, 2008)

2.5 Simulation and prediction

In this section prediction and simulation is going to be explained and the difference between them.

2.5.1 Prediction

An important theorem to explain prediction is

THEOREM 4

Let Y be a random variable with mean E[Y] then the minimum of $E[(Y-a)^2]$ *is obtained for a=E[Y].*

PROOF

$$\begin{aligned} E[(Y-a)^2] &= E[(Y-E[Y]+E[Y]-a)^2] \\ &= E[(Y-E[Y])^2] + (E[Y]-a)^2 + 2E\big[Y-E[Y]\big](E[Y]-a) \\ &= Var[Y] + (E[Y]-a)^2 \geq Var[Y] \end{aligned}$$

Equal sign in the last step is achieved if $E[Y] = a$ and the proof is followed.

An even more important theorem is the following

THEOREM 5 Optimal prediction

$$\min_g E[Y-g(X))^2|X=x] = E[(Y-g^*(x))^2|X=x]$$

Where $g^*(x) = E[Y|X=x]$.

The proof follows the proof of theorem 3.

This theorem says that the minimum of the expected value of the squared prediction error is when the expected value is described by the conditioned expected value. By then the prediction is optimized.

The confidence interval is one of the big difference between the simulation and prediction. The prediction will be discovered to have a much wider interval within a few samples comparing to the simulation. This will be discussed later.

(Madsen, Time Series Analysis, 2008)

Depending on the time horizon ℓ the prediction will be more or less accurate. As ℓ is increasing the prediction accuracy is deteriorating.

After each prediction step made, a confidence interval is calculated and a mean of the confidence interval is found where the next step in the prediction has it's starting point. This means that for each step in the prediction the confidence interval will grow because of more uncertainties in the trend discovered. When the confidence interval grows the starting point of the next step in the prediction will be more unsecure which affects the whole prediction path of the next step. The prediction path is described in the picture.

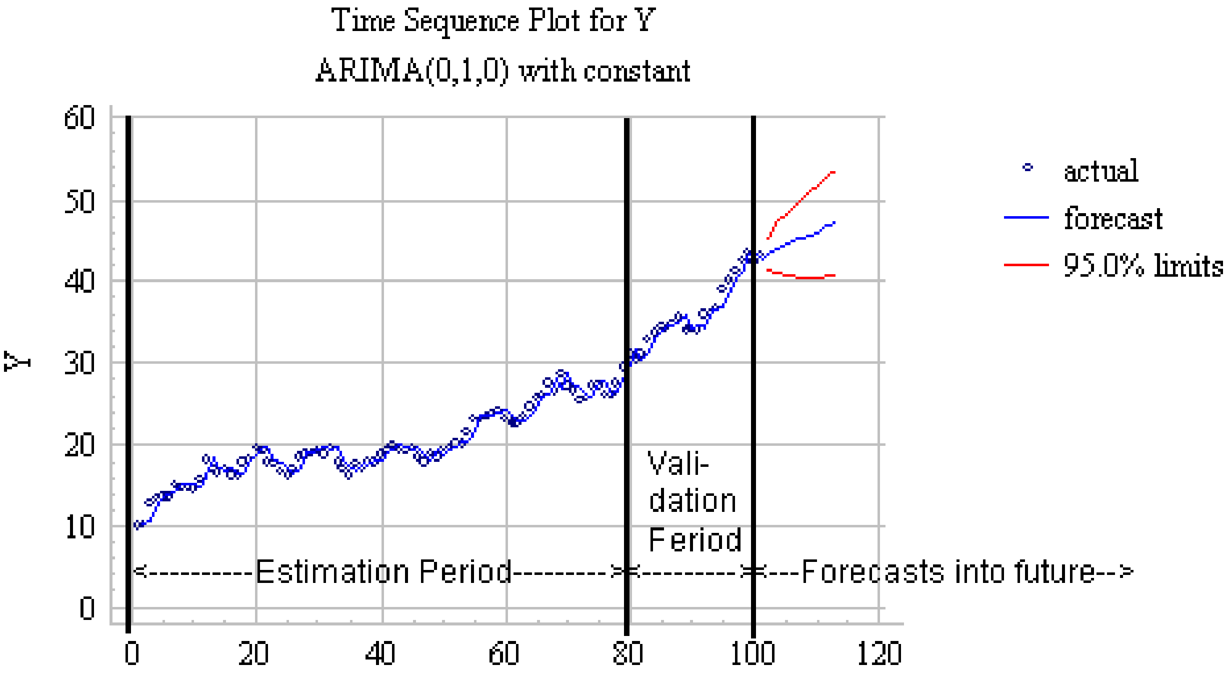


Figure 2.1 *Example of a prediction path.* **(University of Baltimore)**

2.5.2 Simulation

The simulation is used to reconstruct scenarios form historical data and to estimate the robustness of the algorithm so it can simulate the oscillations that are observed. The simulation needs to include the trend and variation that was not described by the time variant variables. Time variant parameters are often used in order to describe the trend in the latest periods (Brown, Katz, & Murphy, 1984)

If the simulation model is a deterministic model, in other words a static model where only if the parameters are updated the output will change. Otherwise the simulation will always be the same despite from the noise added. E.g. in the simulation of a regression model there will be a white noise or bootstrapped noise added.

If the predictor in equation 2.24 is used the prediction would in 5 year time be a linear combination of noise plus the last observed value.

$$Y_{t+(5*365*24)|t} = {\alpha_1}^{5*365*24} Y_t + \alpha_1^{5*365*24-1} \varepsilon_t + \alpha_1^{5*365*24-2} \varepsilon_{t+1} + \cdots + \alpha_1 \varepsilon_{t+5*365*24}$$

As seen is the impact of the autoregressive variables expressed by the noise. If the noise is bootstrapped there is a risk of a misleading confidence interval which would be fine if the noise were normal distributed. If the If a the model is simulated the autoregressive variables will be expressed by the last simulated outputs. The confidence bounds are measured by simulations of the noise. If the noise is simulated enough times, the confidence bounds will give an equally good prediction of the coming 5 years as the prediction with normal distributed noise. The method of the simulation of this model will be explained later in section 3.

2.6 Bootstrap

There are two kinds of bootstrap, parametric bootstrap and non-parametric bootstrap. Parametric bootstrap is when you want to estimate a parameter and want to know how accurate the estimation of the parameter is.

I will only explain the non-parametric bootstrap in this chapter since that is used in the model.

The bootstrap method that is going to be used in the model is also called residual resampling. The residual resampling can simply be explained by taking the residuals are seen as a set where noise is drawn and added to the simulation output of the model. The residual drawn is put back and the set is recovered.

This method, residual resampling is used when there is uncertainty of the distribution fitted for the residuals. The best way of sampling the noise is by describe it with the right distribution. It is hard to find a known distribution as the data often include outliers or are distributed in another way. The set of residuals do often include outliers which make them not a good fit to the normal distribution. Instead the student t-distribution would make a better fit where the tales are wider due to the outliers. The problem with the t-distribution is when the degrees of freedom are low, it can be an issue estimating the variance.

If the distribution is a bad fit to the residuals it will be misleading and seen in the simulation as well but on the other hand if the distribution is well fitted the confidence interval will be the most narrow of all the bootstrap methods used. What is said is that if the distribution of the noise is a good fit then the parameter estimation is the best way of sampling noise but if it is hard to believe that the distribution is a good fit then the resampling residuals method is better and a safer method.

Instead of drawing the noise from a distribution the noise can be drawn from the observed residuals. The residuals are identical independent distributed and strong stationary, no matter where in time the variance will always be constant. The sampling must be drawn randomly from the set as the noise has to be independently and the residuals must be added at random time point in the model. The bootstrap is done with replacement as the noise is randomly picked and a bigger set is better. (Carpenter & Bithell, 2000)

If the variety in the residuals is big it is possible to use the block bootstrap. The block bootstrap is dividing the set of residuals into smaller blocks for specific periods where n samples are drawn from a block with N numbers of data point. N has to be much larger than n.

3. Modeling

With the fundamental theoretical background given in the previous chapter the model of this thesis will be explained. The path to the final model includes many tests of different models were on some trials and analyzes will be presented in this paragraph. The final model is built up to describe the variety during the days and the outcome of each hour. The fine resolution made it necessary to use deterministic variables but the correlation within the previous periods made it also necessary to use time invariant variables. The final model is an Autoregressive exogenous model, ARX which been introduced before within electricity demand modeling. Similar models have also been used within close related environments and where the same dynamic on fine resolution is tried to be modeled, e.g. (Mestekemper, 2011) and (Härdle & Trück, 2010).

3.1 Model structure

An appropriate model structure for the electricity consumption is a model that can describe the several trends that have different periods but also be general enough to explain different years depending on the input, in other words it is important that the model output is updated along with the update of the input data. Finding the right model structure is about finding the right method of describing the electricity consumption. The electricity demand could be described by fundamental variables or variables as autoregressive processes or moving average process or both.

A trade off was made when considering the final model as a complex academic model is not always suitable when it comes to actually using the model within operational companies. One of the objective was to implement and test the model operational. The ARX model is a suitable model due to its fundamental variables which makes its less abstract and more similar to existing models but also include the stochastic pattern.

In order to find a suitable model several, models were tried out to see the fit and to compare the results between. The models tested was first time series models which was found in papers where the electric demand was modeled. The issue with the stochastic processes of only time variant variables was the future description where their capacity, of describing the consumption properly, lasted maximum 10 days ahead when the resolution was on hourly basis.

3.1.1 Seasonal stochastic models

In the electricity consumption there are several different seasonal patterns which can be modeled by seasonal differentiation. The seasonal models is using a differentiating technique where the mean is being removed.

The most common models with seasonal description is SARIMA and SARMA. SARMA, Seasonal Autoregressive moving average process, removes the seasonal pattern from the data sequence before fitting the data to the ARMA model. SARIMA, Seasonal autoregressive integrated moving average model is using the same technique except that it is first differentiating seasonal and then differentiating the remaining data again to reach an ARMA process. The differentiation is done to achieve a stationary process which is necessary when time series analysis is applied.

It is not preferable to use the SARIMA model when it comes to long time horizon. This is because the model is constructed to model time series that are non - stationary. Non-stationary time series are often hard to forecast since they are not time invariant and can change shape over time. What SARIMA do is removing the non-stationarity of the data, modeling the stationary trends and then transforming it back to the non-stationary data.

3.1.2 Stochastic models

The ARMA model was tested to see if the seasonal pattern could be modeled without the seasonal differentiation. It was found that the best way of describing the consumption was by include the deterministic variables especially of the temperature which have a strong impact on the electricity consumption. Without fundamental variables it is hard to steer the simulation of future years if the model cannot find the temperature trend over the year. The ARMA model would have been a good estimated model if the time horizon was shorter than 5 years for example 10 days.

The next model that was tested was a linear regression model in order to describe the electricity consumption with only deterministic variables.

3.2 Deterministic model

The 5 years ahead prediction could become poor if it is only described by non-fundamental parameters as in a time series model. Due to the long time horizon will the trend hard to be predicted. In order to make a prediction it is better to use a normal year where the daily momentum is caught and add the specific predicted trends which often within this environment is on higher resolution and has to be interpolated down to hourly basis before added.

3.2.1 Deterministic parameters chosen

In order to find the vital variables to explain the electricity consumption, investigation was done for existing regression models of the electricity demand. There are some differences when modeling the electricity demand in the Nordic countries and the continental countries. Most of the variables fits in the Nordic countries when variables as cooling degree days is not significant enough to be contributive. The input variables of a regression model should be independent of the output variable but the output variable should be dependent of the input variables.

Starting from scratch adding one parameter at the time in order to see if there was an improvement of the model or if the parameter was insignificant. To ensure that the parameter was stable and significance was checked of the confidence interval. If a parameter was unstable it was seen very clear as the confidence interval was much larger relatively the parameter value.

The temperature was the first fundamental parameter experimented with. The temperature was shifted 1 hour since there is a delay between the changes in the temperature and the consumption. Another try out was to create a weight on the temperature as the model had a hard time reaching the tops of the consumption during winter time, The weight forced the model to stay at a lower temperature when the temperature was at the most extreme temperatures. The weight were though insignificant and was removed.

Another parameters were solar intensity which is explained by a cosines function where the electricity demand peaks in the beginning of the year and goes down beneath zero as the summer has a consumption less than base case. The base case is 24 base hours which are the base for every hour and from there the hours will be modified due to all the external variables.

Instead of seasonal differencing dummies for every day and hour was done. Some weekdays were highly correlated which made the correlation matrix close to singular. The correlation matrix is close to singular when a parameter is a linear combination of another parameter. To avoid this the weekdays were merged together. Monday and Friday were left for themselves as Monday has less consumption than the normal weekday because of the start-up phase after the weekend. Tuesday to Thursday were merged together since they are ordinary weekdays with not much difference. Friday has a trend of de-escalation before the weekend. Saturday and Sunday is merged together where the amount of consumed electricity during the weekend is less as ordinary jobs takes time off on weekends.

The hardest fundamental parameter to catch is the vacation. Vacation can be at all times during the year. I created a dummy variable for the vacation where the vacation parameter was most significant when the vacation was set during Christmas and until 6th of January, long weekend during Easter from Thursday to Monday. The summer is the hardest time as a lot of people is going away on holiday which makes the period more clearly but the vacation is very spread out during the summer period. The most significant was found when taking the whole month of July and half August. Though this period was clear enough to be discovered visually the parameter was not estimated good enough to have a significant effect on the simulation.

The public holiday is its own parameter as it differs from the vacation since everyone is free during this day.

The final regression model was the following,

$$Y_t^{regression} = \sum_{i=1}^{24} \beta_i + \beta_{25} \cos\left(2 * \pi * \frac{t}{365}\right) + \beta_{26} HDD_{t-1}$$

$$+ \sum_{i=27}^{51} \beta_i * (tue - thur)_t + \sum_{t=52}^{76} \beta_i * fri_t + \sum_{t=77}^{101} \beta_i * (sat - sun)_t$$

$$+\beta_{102} * Vac_t + \beta_{103} * Pub.\ Hol_t. + \varepsilon_t$$

- $\sum_{i=1}^{24} \beta_i$ constants that describes the base hour
- $cos\left(2 * \pi * \frac{t}{365}\right)$, *solar intensity, described by a cosine function*
- HDD_{t-1}, *Heating degree days with I hour delay in relation to the demand*
- $(tue - thur)_t, fri_t, (sat - sun)_t$, *24 hours of each weekday group because the daily trends differs between the weekday group.*
- Vac_t , *Vacation period*
- $Pub.\ Hol_t$, *Public holiday*

The model checking indicated of correlation within the residuals and in some specifically time delays. I decided to extend the model by adding stochastic autoregressive variables.

3.3 Extended stochastic model with deterministic variables – ARMAX and ARX

After having a model of fundamental parameters mentioned above it is seen in the plot of the ACF that there are still correlation between the residuals.

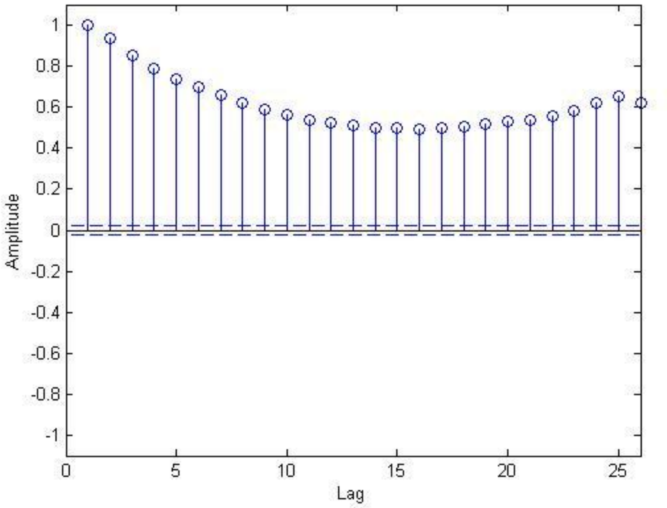


Figure 3.1 *The ACF plot of the residuals after the regression model. It has clearly correlation left especially at 24th hour.*

This means that not sufficient trends are caught with the deterministic model and since there are correlation between the data points between certain time points time variant parameters need to be added.

The ARMAX model used in (Härdle & Trück, 2010) is a dynamic system extended by deterministic variables. The model was tried but when analyzing the output, insignificance was found in the MA parameters. At last the ARX model was tested and chosen.

The addition of the autoregressive parameters to the deterministic model, turns the model into an ARX model, autoregressive exogenous model. The lags were found at the following hours back from time $t, t+1, t+2, t+23, t+24, t+25$.

$$A * Y_t = B * U_t + \varepsilon_t$$

3.4 Model order

The next step in the modeling was to identify the optimized model order. When using the mixed models with both AR variables and MA variables, e.g. SARIMA, SARMA, ARMA and ARMAX, I tried out the different orders of combinations and compared the Final prediction error, FPE, to find the best combination of orders.

The model order of the AR variables in the ARX model was possible to detect in an ACF plot where the lags were clearly seen. The lags were mostly seen in the nearby type of hours as the current hour t, meaning either in $t+1, t+2$ or $t+23, t+24, t+25$. I started include all lags in the model and by then I discovered in the model output which lags that were insignificant and if they could be excluded in order to highlight the significant parameters

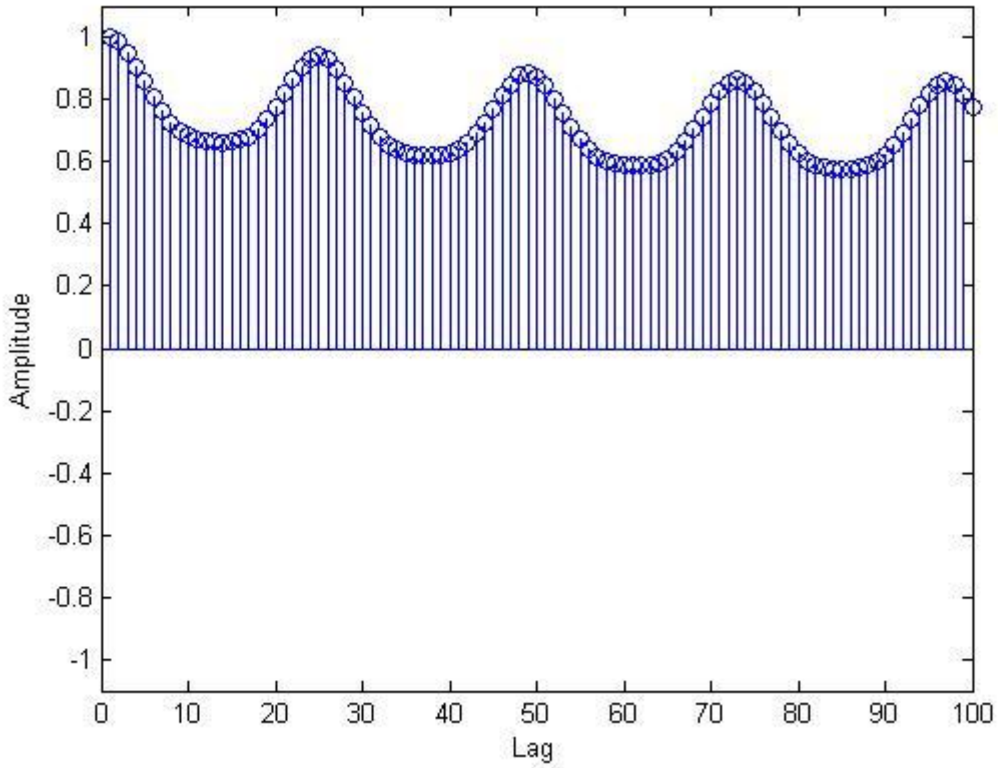


Figure 3.2 *The ACF plot of the observations modeled. The model order is hard to identify straight away but the correlations are strong every 24th hour. The model type is either an AR or an ARMA.*

3.5 Procedure of selection of deterministic parameters

When developing the deterministic part, the parameters was compared by their significance. For example were different delays in the temperature correlation to the demand tested, meaning how long time after the temperature has changed will it take until the effect is seen in the demand. The parameter of the temperature was most significant at 1 hour delay. The delays was tested from one hour delay to 24 hours delay. The existing model is using 3 hours delay The temperature was then transformed into heating day degree parameter, HDD. The HDD is supposed to catch the temperature when the correlation to the electricity consumption is high enough to make an impact on the model. The HDD variable refers to the temperature being measured when we are heating up our houses. To increase the significance of this parameter the summer temperatures will be excluded. The limits tested was between 13 and 20 where 17 degrees of Celsius gave the most significant parameter.

The number of variables from the regression model was determined by the loss function as described earlier. The parameters were also obtained in the model output were it was seen if the parameter was significant or not, if not it was excluded.

The number of parameters was counted to be 100 deterministic variables and autoregressive lags in 1,2,3,24,2. The final model structure with model order is written as,

$$\alpha_t y_t - \alpha_{t-1}y_{t-1} - \alpha_{t-2}y_{t-2} - \alpha_{t-3}y_{t-3} - \alpha_{t-24}y_{t-24} - \alpha_{t-25}y_{t-25} =$$

$$\sum_{i=1}^{24}\beta_i + \beta_{25}\cos\left(2*\pi*\frac{t}{365}\right) + \beta_{26}HDD_{t-1} + \sum_{i=27}^{51}\beta_i*(tue-thur)_t$$

$$+\sum_{t=52}^{76}\beta_i*fri_t +$$

$\sum_{t=77}^{101}\beta_i*(sat-sun)_t + \beta_{102}*Vac_t + \beta_{103}*Pub.\ Hol_t. + \varepsilon_t$

3.6 Parameter estimation

In order to optimize the parameters to fit the model to the data as good as possible the Prediction Error method was used. The prediction error method optimizes the parameters by minimizing the prediction error as described in theory section.

The prediction error for the model chosen is estimated by the following equation which is described in section 2.3.1

The autoregressive parameters are stationary if the roots are within the unit circle when looking at z transformed parameters which were explained in section 2. All parameters had roots within the unit circle. The sum of the parameters is between -1 and 1 then they are stabile since the simulation or prediction of the model can never increase more than the previous data.

When variables are correlated, or are too similar, the parameters tends to be insignificant or unstable as the confidence interval is way too large for the parameter value. This problem was seen within the dummy variables for each hour but was ignored since it was only a few

The parameters estimated for the fundamental variables in the ARX model was found very low. This means that most of the model was explained by the AR process and not by the external variables as the plan. The next hour was with 70% calculated from the previous hours and the rest was edited by the external variables.

When explaining a simulation for 5 years ahead it would feel more confident if the outcome was not all because of the hours before but due to the temperature for example. A simple test as splitting the model in two parts, one deterministic and one AR-process and see the difference in the parameters of the AR-process. If the AR- process would have lowered its parameter value it could be an idea to have the input data to the AR process as the difference between the observed data and a pre estimated regression model. and from there solve the parameter estimation non linearly. Another possibility could be to regularize the estimation, modify the parameter estimation to receive heavier weights on the external variables parameters.

The outcome after splitting the model was though almost the exact same value of the AR parameters as with the ARX model in with parameters estimated in 1 step. This means it would not help to regularize the estimation since the AR parameters will still be the same. And it would be the same as both methods are sub optimized, the estimation is forced to have higher external parameters and cannot optimize all parameters together.

The ARX model in 2 steps will be followed in the result part in order to compare the result.

3.7 Model check

As mentioned in the theory a common method to validate the model is by observing the residuals. In Figure 1 the residuals for the pure regression model still had correlation in the residuals especially in the hour close to the current hour. The residuals of the ARX model had some larger outliers but in general within an approved range see Figure 4.. Looking at the ACF plot for the residuals from the ARX model it does not have any significant lags and can then be seen as a good estimate of the observed data.

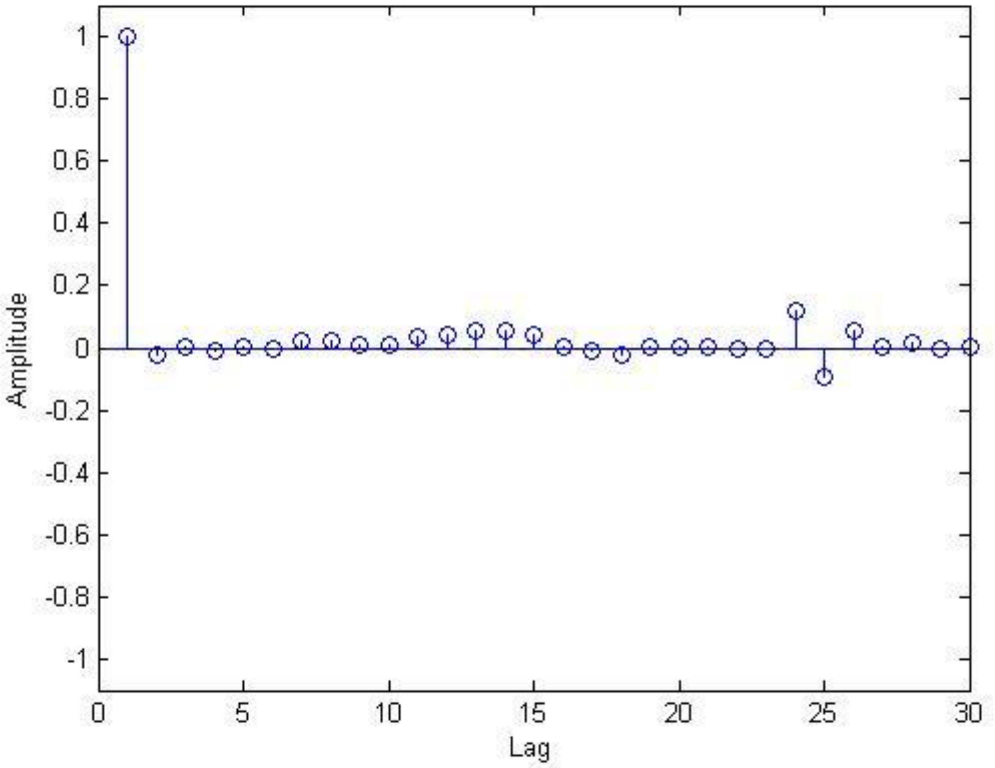


Figure 3.3 *The plot of ACF of the residuals between the model and the observed data. The correlation is very small in the lags of 24 but otherwise is it a good estimation which is approved to continue with.*

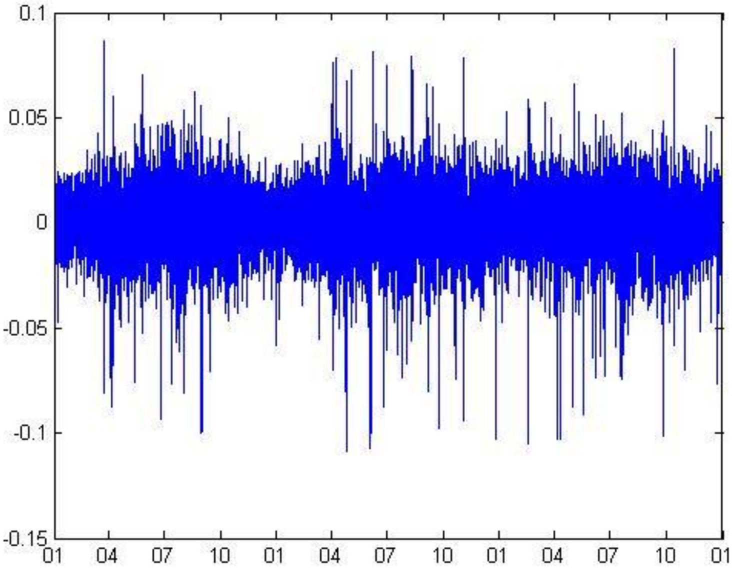


Figure 3.4 *The residuals after the ARX model, unit in percentage of the observed data.*

More detailed result will be presented in the Result in section 4.

3.8 Simulation

When simulating the first day, the AR process of the model needs input from the 25 hours before. The previous year's last 25 hours are used as input data and the sensitiveness of the initial values will be tested in section 4, Results. Otherwise the simulated data is used as input for the autoregressive parameters. In this section the method for simulating this model is going to be described.

3.8.1 Path of the simulation

The model is being simulated by one step prediction meaning, by calculating the model output one step at the time. The simulation will be a continuation of the first 25th observed data and therefor is the input of the autoregressive parameters. Certain deterministic parameters needs to be updated in each year, those who varies are days since the day of the week will be different, public holiday will also be different. In the price model the temperature input data are historical temperature data from the last 46 years. The output from the simulation will be 46 scenarios depending on the weather. The temperature used as input will be approximations in two ways, it will be distributed throughout the hours of the day based on a daily mean and approximated daily profile of temperature, will be written more of in part 3.6. For each step a noise is added and the time invariant variables are updated continuously as the next step is calculated.

3.9 Bootstrap of residuals, adding noise

After an ARX model is fitted to the electricity consumption data, the output of the ACF of the residuals do not contain any significant lags referring to correlation in the error data. The residuals has though outliers coming from misses of public holiday or vacation etc. When comparing to a normal distribution it is seen that the empirical distribution has too wide tails to be normal distributed. It fits better to a student t-distribution with 5 degrees of freedom.

Sampling from a distribution demand trust of the fit of the residuals to the distribution. If the distribution is not a good enough fit then the sampling could be misleading in the simulation. To simulate from a student t-distribution could imply a higher amount of outliers than before if the tails does not fit exactly. The t-distribution can be risky when it comes to the estimation of the variance. At low degrees of freedom the variance can tend to become very large as the variance has the following expression:

To avoid the issues of the outliers of the residuals another method can be applied called block bootstrap. The bootstrap method creates a distribution or a set of the data that is available. The set of data is then acting as a pool where new data set is being drawn. This method reduce the risk when simulating from a student t-distribution to either draw too many of an interval value that does not occur as often as it is drawn or draw outliers that is outside the interval of the real residuals.

The public holiday and the vacation days had residuals that were larger than the other days since the model did not succeed fully to describe these days. By also dividing the residuals obtained into blocks of the weekdays, Monday, Tuesday-Thursday, Friday, Saturday and Sunday the resampled residuals become even more accurate. The separation has to be balanced as the noise has to be generalized enough to be the same for other years than just that year, also called over fitted.

The residuals were first tried out to be divided into months and then hour type but the days had more similar output than hour types.

3.10 Simulation of 46 weather scenarios

Earlier in this thesis observed temperature was used but in the future the temperature will be unknown. The unknown temperature is weather scenarios from 46 years back which gives the simulation a sample space of the possible weather in each year.

One of the issues when it comes to the simulation was the updating of the input variables e.g. the temperature. The temperature has to be the temperature of the year simulated on hourly basis. The electricity price model for midterm forecasts at Vattenfall AB was using the temperature seen every year 46 years back. Each year would be used in the simulation and the outputs would be 46 according to the 46 weather scenarios.

The temperature data for 46 years back is on daily basis and needed to be interpolated to hourly basis. The interpolation was done by creating daily profiles for the temperature each month by polynomial fitting. The hourly temperature was found for three years which gave three months of data in order to fit the polynomial curve for each month. The polynomial curve for each month was then moved after the difference between the mean of the curve and the daily temperature the current day.

The 46 scenarios coming out of the simulation could then be used to interpret the different probabilities that year depending on the temperature.

3.10.1 Simulation forecast with economic growth modification

An existing model of the demand forecast the next five years on monthly basis is currently used. The monthly demand forecast is prioritizing the trend from a more out-zoomed perspective than describing the demand hour of the year. The monthly demand forecast is taking in consider the economic growth over the years. For example a decrease in economic growth can cause a reductions within the industrial consumption as some factories will shut down or reduce their production in order to save money. The monthly forecast is also taking under consideration the amount of heat pumps which is in use etc.

Since the model simulated is rather describing the dynamic during the year than profiles the years by forecasted events, is the information of the monthly demand forecast needed. As mentioned earlier the monthly forecast contains trends as economic growth over the years and the consequence in the electricity demand can be seen.

A fitted curve to the monthly forecast is compared to a fitted curve to the monthly demand simulated with normal weather. Normal weather is used in the simulation due to that the monthly forecast is independent of the temperature. The relation between the relatively difference in the curves will be the change of the factors that are not included in the ARX model. The relative difference will be multiplied by each hour after a simulation is done which becomes a parallel movement but different for each hour. The simulation will then also possess information and the future years gets a more distinct profile.

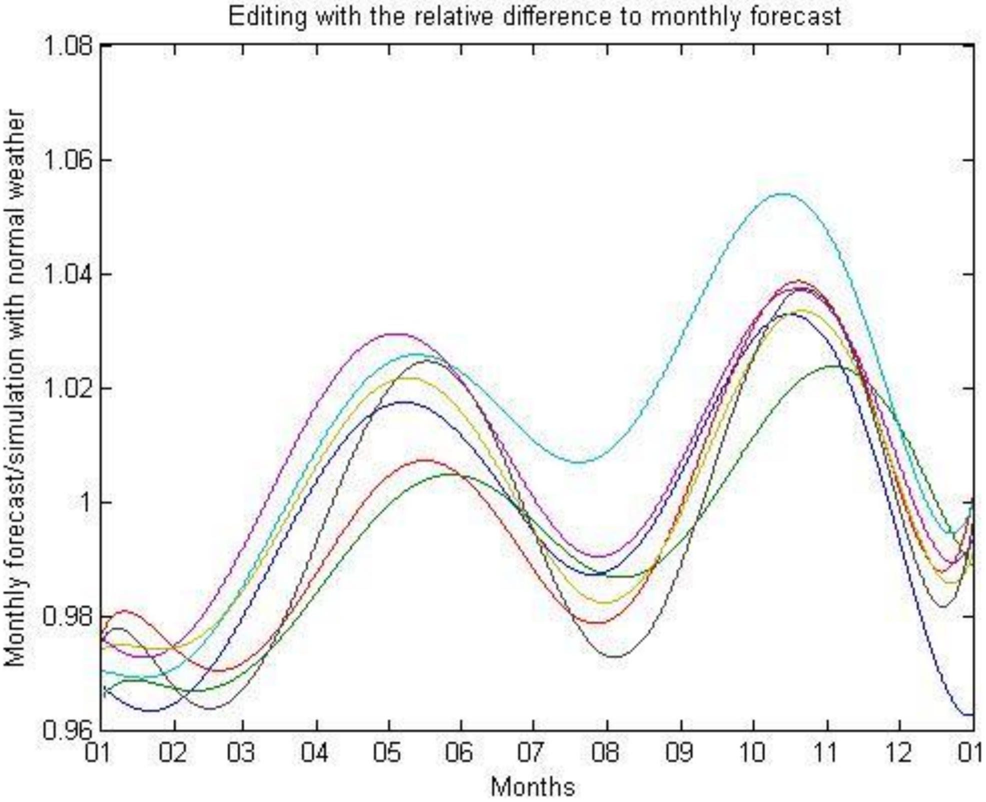


Figure 3.5 *Forecasting modification for seven years, 2010-2016.*

4 Results

To validate a model several aspects have to be considered. In this section a number of visual results will be presented and analyzed. As mentioned in the modeling part, I have tried to estimate the parameters of the ARX model in both a one-step and two-steps structure, see the modeling section. The results from using these two types of parameter estimation methods will be studied and compared. The interest in the comparison lies in the greater weight in the AR part of the two step structured ARX. In the first section of this chapter the results from the one-step structure ARX will be presented and analyzed. In the next section the results from the two-step structure ARX will be presented and compared with the results from one-step structure ARX.

4.1 The one-step structure ARX

Starting with the results from the ARX model in one-step structure means that all the parameters have been estimated simultaneously.

4.1.2 Simulating with known variables

The first way of testing the model is with all variables known; the observed initial values of the previous 24 hours electricity consumption and the observed temperature. The temperature observed is in fact also an approximation since it is a national temperature and the temperatures within Sweden are varying a lot from north to south.

The question is how good the model is given all correct variables? The model is simulated 100 times and the empirical 95% confidence interval of the simulated output is compared to the observed data for that simulated year. In order to see a variation with different variables two years are simulated, 2012 and 2013.

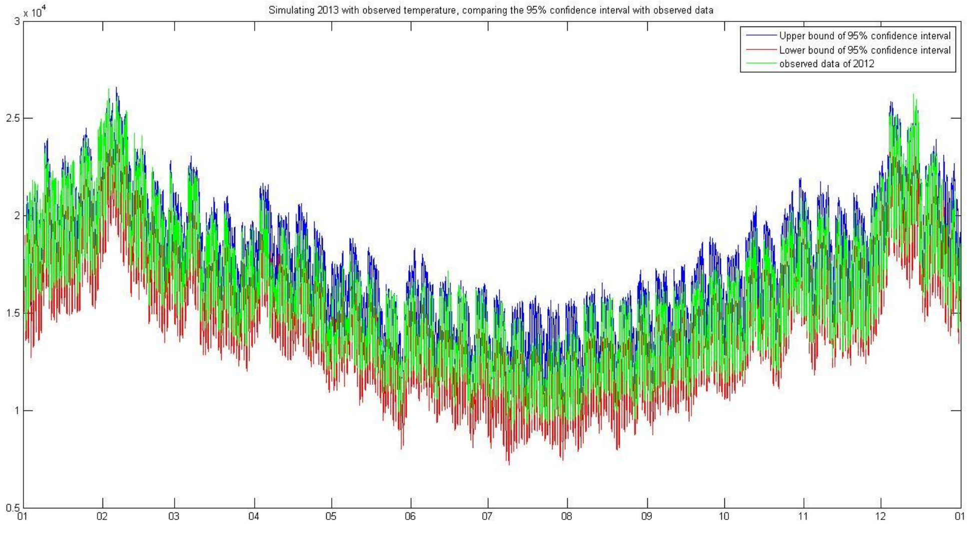


Figure 4.1 *Simulating 2012 with known external variables (observed temperature) and correct initial values. Comparing the output of the simulation with observed data of 2012.*

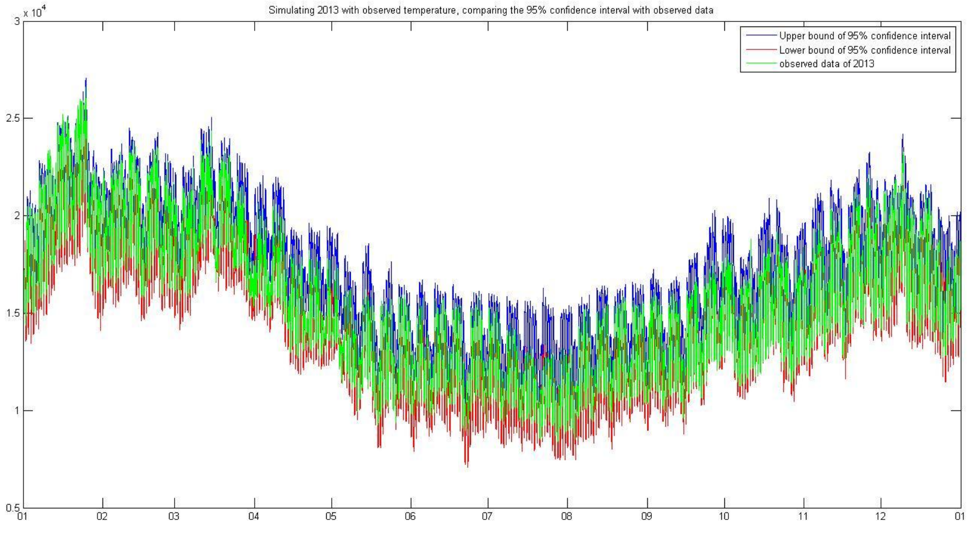


Figure 4.2 *Simulating 2013 with known external variables (observed temperature) and correct initial values. Comparing the output of the simulation with observed data of 2013.*

By studying figure 3.1 and 3.2 visually it is seen that the confidence interval (red and blue) follows the observed electricity consumption (green) when the temperature is changing relatively sharp. The confidence interval is closer to the observed consumption during periods of low temperature, e.g. December 2012 and January 2013, indicating the variable weight in the model, see figure 3.7 below. Compare this to summer when the temperature contribution is zero the confidence interval is more unchanging.

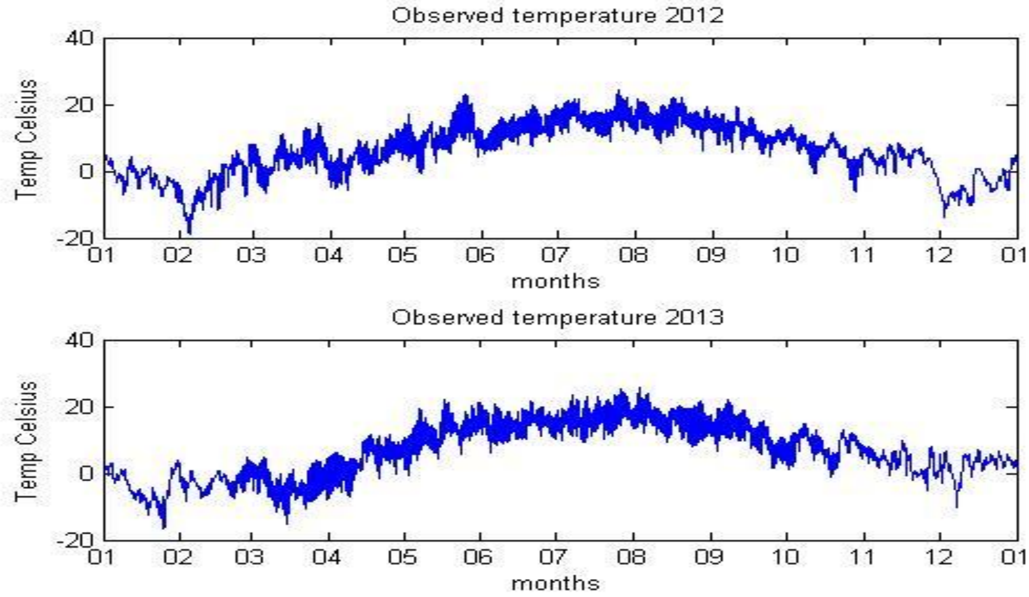


Figure 4.3.*Observed temperature 2012 and 2013. Note the colder temperatures in 2013 which is the cause of the higher electricity consumption in figure 3.1 and 3.2.*

A clear difference in the years of 2012 and 2013 is seen, meaning that the model is able to form the simulation depending on the variables of the year. The profiling is mostly due to the different temperature over the years.

Spring

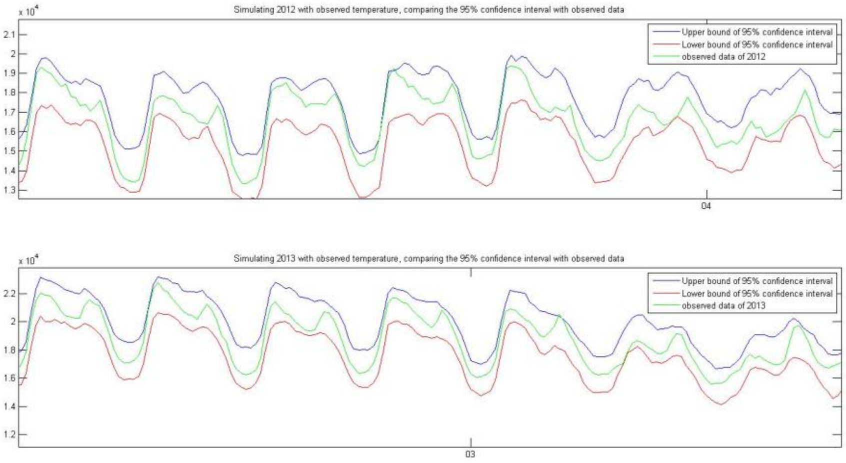


__Figure 4.4.__ Same content as the figure 3.1 (simulation of 2012) and 3.2 (simulation of 2013) but zoomed in at a week in March. Note the daily trend during this period of the year.

<u>Summer</u>

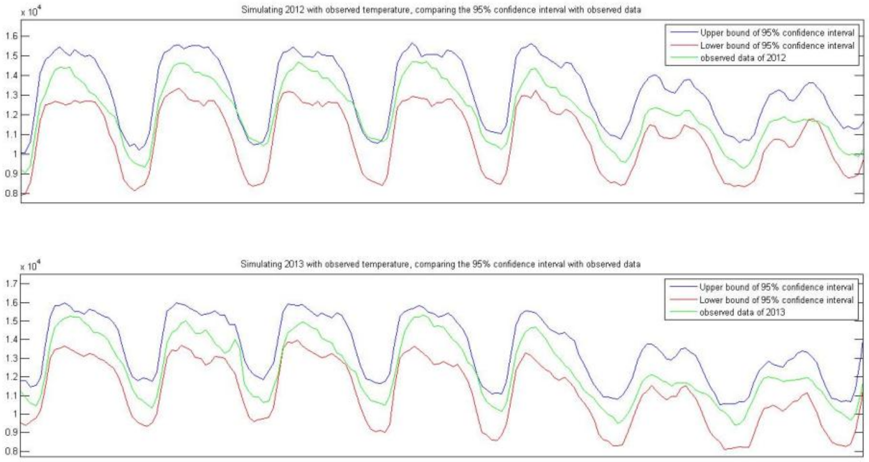


__Figure 4.5.__ Same content as the figure 3.1 (simulation of 2012) and 3.2 (simulation of 2013) but zoomed in at a week in June. Note the daily trend during the summer period.

Comparing the daily trend of the observed data during summer and the rest of the year, a clear difference is seen in the weekdays. During summer the weekdays only have a clear peak in the morning and during the rest of the year there is a peak both in the morning and in the evening, see figure 3.3-3.4. The simulation does not follow the daily trend in the summer as good as during colder periods of the year. It may be because the model does not include specific daily hour variables for the summer and so the estimated parameters for the daily hour variables are the same during the year. A consequence of this is that the simulation does not follow the tending decrease in consumption throughout the day. This is probably because more sun hours at night, reduction in industrial activity and vacations led to less electricity consuming activity during the night. Overall the simulated weakly pattern complies pretty well with the weakly pattern of the observed data with a reduced consumption during weekends. Daily trends during the weekend do not show any spectacular differences over the year.

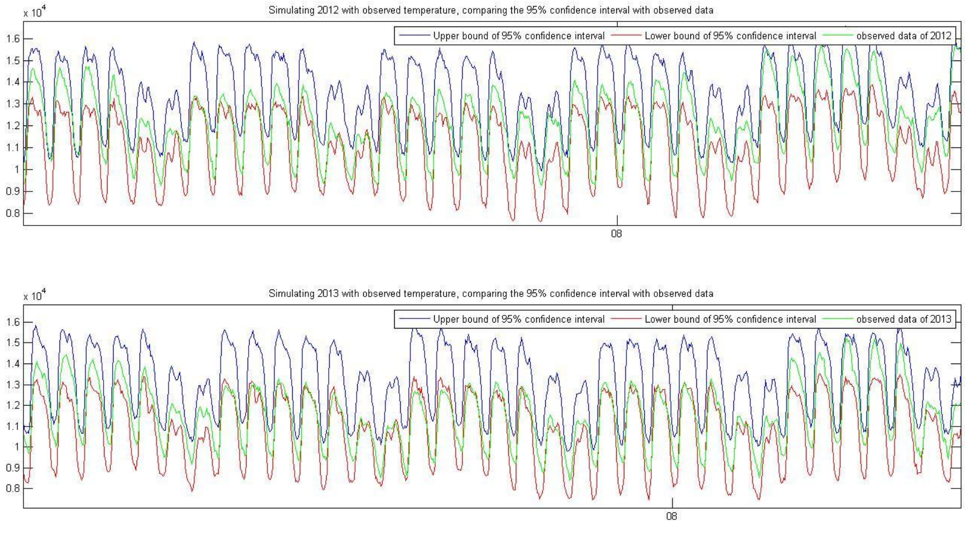


Figure 4.6. *Same content as the figure 3.1, simulation of 2012, and figure 3.2, simulation of 2013, but zoomed in at the vacation weeks in July/August.*

One of the weak points of the model is its disability to capture vacation periods. The confidence interval of the simulation is either too high or too low, meaning that the observed data is not centered in the interval, see Figure 3.5. In 2012 the vacation seem to be a little bit better fitted than in 2013 where the observed data is below the lower bound. The vacation period is very distinct as the curve suddenly decreases for approximately three weeks. The model has a dummy variable for the vacation but obviously the contribution is not enough to fit the observed data. One idea why this happens is that the vacation period data is quite small so the variable parameter has little impact, compared to other variables, in the least square parameter estimation.

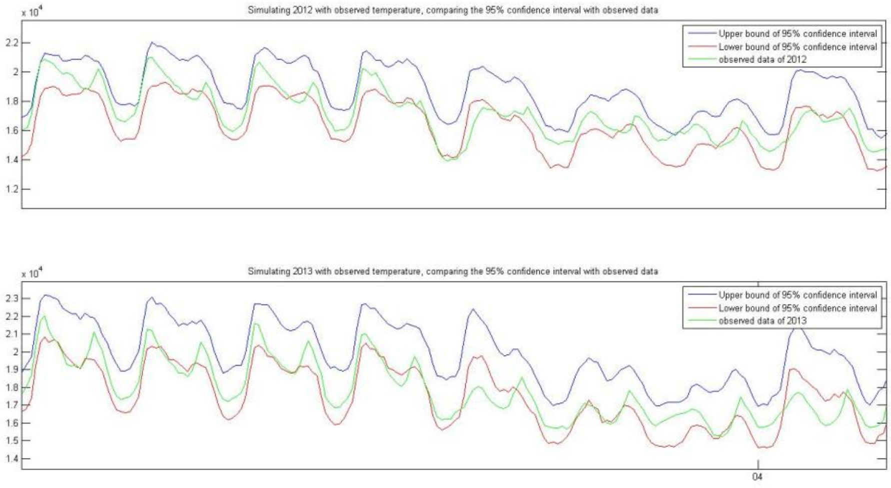


Figure 4.7. *Same content as Figure 3.1 (simulation of 2012) and Figure 3.2 (simulation of 2013) but zoomed in at the Easter week. Note the public holiday at Good Friday and Easter Monday.*

Another weakness of the model is the ability to model public holidays, with similar issues as for the vacation periods. The interval at these days sometimes misses the observed data as the day is modeled as a normal day and not a public holiday. The variable assigned for the public holiday is, as said about the vacation variable, not strong enough to fit an interval because of the minority of public holidays when estimating the parameters. In some cases, like the 6th of January, the interval fits the observed data very well. This may be since the 6th of January occur straight after the Christmas vacation and when simulating this day the consumption of the Christmas vacation days, that are the input data to the AR – part of the model, have decreased the level of the consumption. If the public holiday occurs randomly as the 1st of May it appears as a normal day in the simulation. The difficulty of using a general dummy variable for public holidays is that it gives a constant absolute contribution over the year. Most likely if a public holiday falls close to a weekend or during warm or cold periods of the year it should results in different consumption contribution. An example of when the simulation misses the public holidays is during Easter week where Friday and Monday are holidays. One can clearly see that the days are perceived as normal weekdays. The rest of the week is also poorly fitted as this period of the year can have very varied weather.

4.1.3 Relative spread size of the confidence interval

In figure 3.1 and 3.2 the 95% confidence interval seems to be surrounding the observed data quite well. But how big is the spread of the 95% confidence interval compared to the observed data? Figure 3.8 below is showing the relative difference between the bounds and the observed data for the years 2013 where the 95% confidence interval for 2012 is more or less the same. The upper bound is approximately 110% of the observed data, with some exceptions of outlying peaks, and the lower bound is approximately 90 %, also with some peaks excluded. Figure 3.8 gives a good picture of how the interval changes over the year. During the vacation weeks in summer time the bounds moves 10% up and are instead, 100%-120% of the observed data. Similar pattern can be seen at Christmas holiday. The reason for this, as mentioned previously, is that the model tends to overestimate the vacation consumption

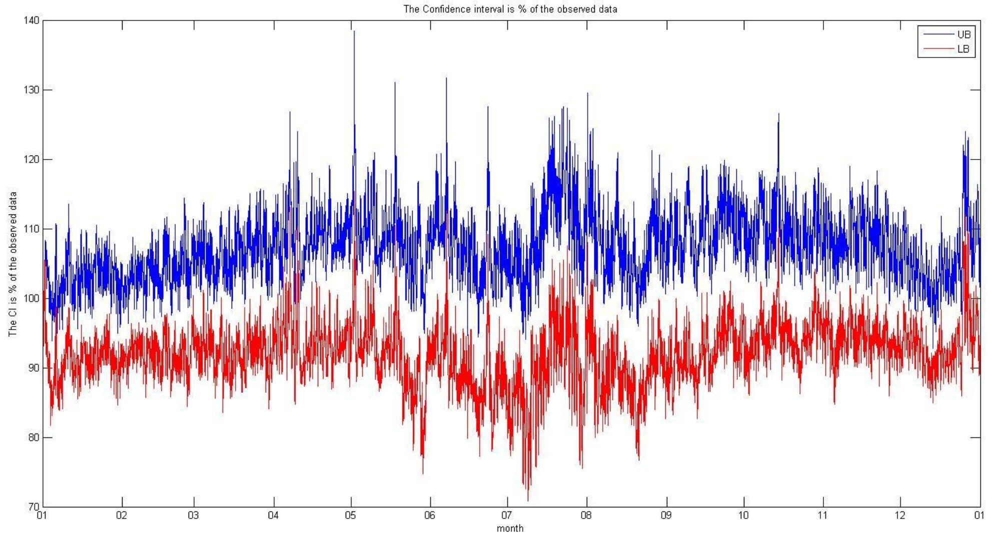


Figure 4.8. *The index ratio between the observed data and the 95% confidence interval in figure 3.1. Index = 100 when the observed data and the confidence interval are equivalent.*

4.1.4 Threshold

From a wide end perspective it did not seem to be many observed data points outside the confidence interval. But if we only look at the data points that lie outside the interval, how large are theses excesses and during which periods do they often happen? Figure 3.9 is showing the data points that were outside the 95% confidence interval for 2012 and 2013. The plot shows the relative level of the excess compared to the confidence interval, e.g. 0.05 means that the observed data point is 5% above the upper boundary.

Starting with 2012 the winter periods have intervals that are too low to capture the observed data. The confidence interval is 5% lower than the observed data. There are also some excesses during summer time were the interval appear to be too high for a short period as well as for the last days of the year.

If several time points after each other have data points outside the interval, they form together a smaller cluster. The clusters size mean in figure 3.9 is circa 3 hours. The biggest cluster size of 2012 is 25 hours and appears in the beginning of the year, otherwise there are only a few clusters with about 17 hours in size.

Looking at 2013 the interval is mostly too high in cases where the observed data lies outside the interval. The relative amplitude and the periods of the excesses resemble 2012. The mean of the cluster size is similar as for 2012, circa 4 hours, but there are some cluster during the year that are 21 hours.
What is more seen is that the model have periods during the year where the observed data are not randomly going outside the confidence interval. This is for example during winter when the extreme cold temperature is hard to capture or during summer as the model has a harder time to follow the daily trend.

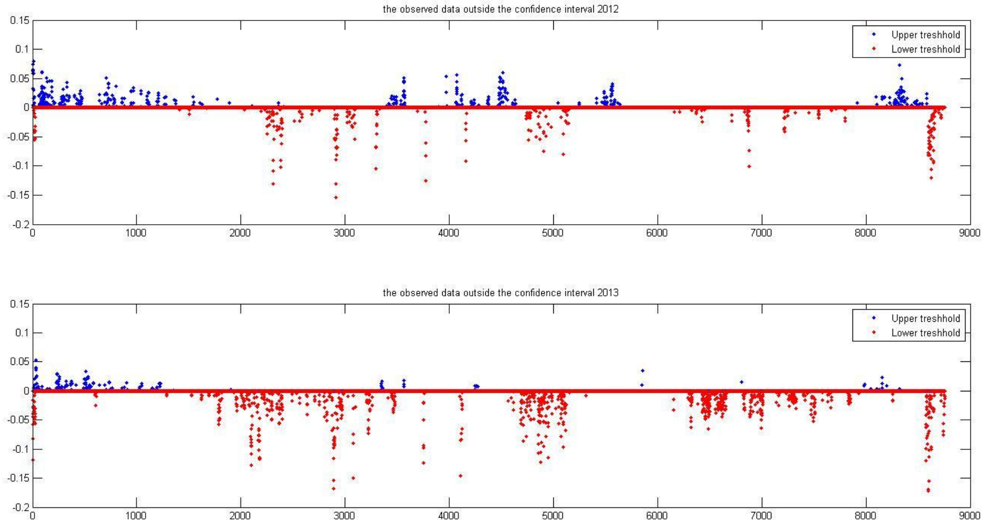


Figure 4.9. *The observed data points outside the 95% confidence interval. The blue dots are for data points exceeding the upper limit and the red dots are for data points that fall beneath the lower bound. Upper plot 2012 and lower 2013.*

Simulated year	**% outside the confidence interval**
2012	8,3%
2013	12,14%

Table 4.1 *The percentage amount hours when the observed electricity consumption was found outside the 95% confidence interval.*

As can be seen in table 3.1 the simulation of 2013 has more hours outside the confidence interval and particularly during summer time where the cluster size is often small but with a dozen having a cluster size of more than 12 hours. It is also seen that 2013 has 12% of the observed data outside the 95% confidence interval, which is perhaps a little much. Having smaller clusters are maybe inevitable since extreme consumption might be on a broader weekly or daily level and not on hourly level. The weak points of the model might be revealed with groups of clusters during the summer of 2013 and the same underestimation of some days of 2012 and 2013.

4.1.5 Distribution comparison

To compare the hourly distribution an empirical distribution is observed for each month. The observed data consists of data from 2012 and 2013 and the simulated data for the same years.

Looking at the mean hour of each distribution it is seen that for the winter months the distributions of the simulations is matching the observed mean hour very well for both years. Though both years overestimate and underestimate the same monthly hours. During summer the mean hour are underestimated a little in 2012, which also can be seen in figure 3.9, and in 2013 it overestimates the summer hours a little. The autumn months is not following the trend of the observed mean but is both above and below and ends up in December with a good match.

As expected, by viewing the first plots in this chapter, the summertime mean deviates more from the observed mean than other moths. The lower mean hour springs from the misses at the peaks, the lowest temperatures, and will cause a lower monthly consumption.

The standard deviation should be higher for the simulated values as it is taken from a simulation with 100 simulations with noise added. The spread of the noise is creating the higher standard deviation of course than an observed output. The standard deviation is though lower many months but during summer are the standard deviation higher than the observed outcome for both years.

Month	**Mean 2012 (10^4)**	**Observed mean 2012 (10^4)**	**Mean 2013 (10^4)**	**Observed mean 2013 (10^4)**	**Std 2012 (10^3)**	**Observed std 2012 (10^3)**	**Std 2013 (10^3)**	**Observed std 2013 (10^3)**
Jan	1.9128	1.9711	2.0055	2.0453	2.2915	2.3814	2.5788	2.8022
Feb	2.0271	2.0730	1.9680	1.9844	2.5655	2.7063	2.0088	2.0984
Mar	1.7234	1.7350	1.9551	1.9205	2.0119	2.0727	1.9737	1.9610
Apr	1.6652	1.6394	1.6821	1.6165	1.9635	1.8388	2.0083	1.8170
May	1.4116	1.4080	1.3732	1.3293	1.9218	1.7473	1.9674	1.7839
Jun	1.3377	1.3626	1.2447	1.2512	1.8649	1.7467	1.8313	1.7422
Jul	1.2060	1.2107	1.2091	1.1631	1.7978	1.5273	1.8467	1.5330
Aug	1.2661	1.2916	1.2581	1.2526	1.8497	1.8376	1.8430	1.8400
Sep	1.4076	1.3931	1.4067	1.3650	1.9334	1.8437	2.0957	1.9725
Oct	1.6208	1.5829	1.5743	1.4991	2.2144	2.2376	2.0262	1.9835
Nov	1.7196	1.6989	1.7426	1.6901	2.1270	2.2446	2.2368	2.3099
Dec	2.0119	2.0258	1.7645	1.7464	2.4007	2.7688	2.2130	2.4006

***Table 4.2** The mean hour of each month. A comparison between observed data and simulated data with known temperature.*

The cumulative distribution is compared between the simulation of 2012 and the observed data of 2012. It is seen that in spring are the distributions following each other quite well.

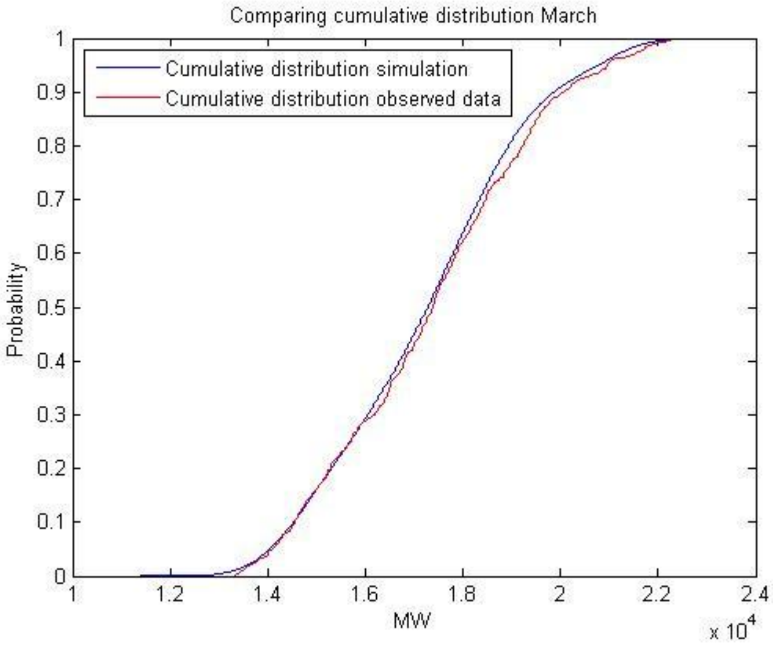


Figure 4.10. *The empirical cumulative distribution of the hours in March.*

The weak periods of the model, as has been told before, the summer time when the temperature does not have a significant impact the cumulative distribution of the simulation is increasing a little bit faster than the observed data but the most important is that the simulation is surrounding the observed cumulative distribution in order to catch the sample space.

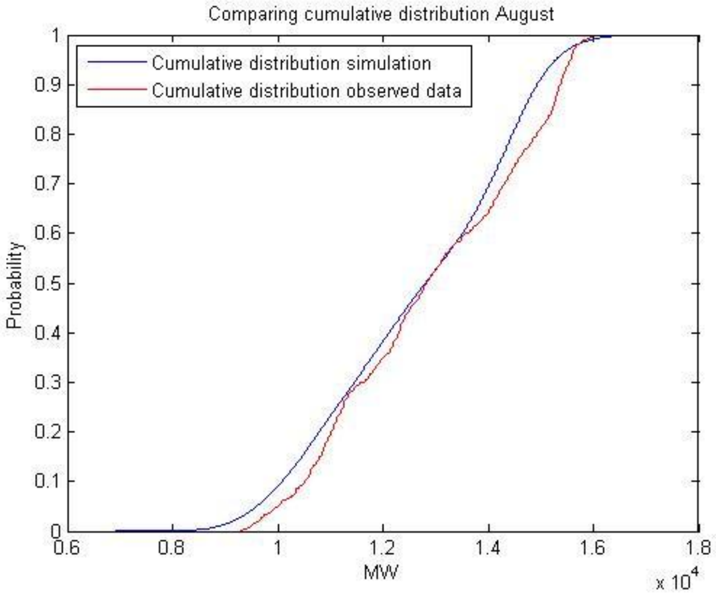


Figure 4.11. *The Cumulative distribution of the hours in August.*

4.1.6 How sensitive is the model given external data but unknown initial value?

We have seen the result when all variables where known but let see what happens if the initial values are changed. The simulation of the one-step estimated ARX model is 90% output from the AR process and 10% from the fundamental variables. The conclusion should be that the model is very sensitive to the initial value as the simulation is mostly built up by the previous hours. Two tests are done, one where the initial value is changed by 10% and another test where the initial value is changed with 300%.

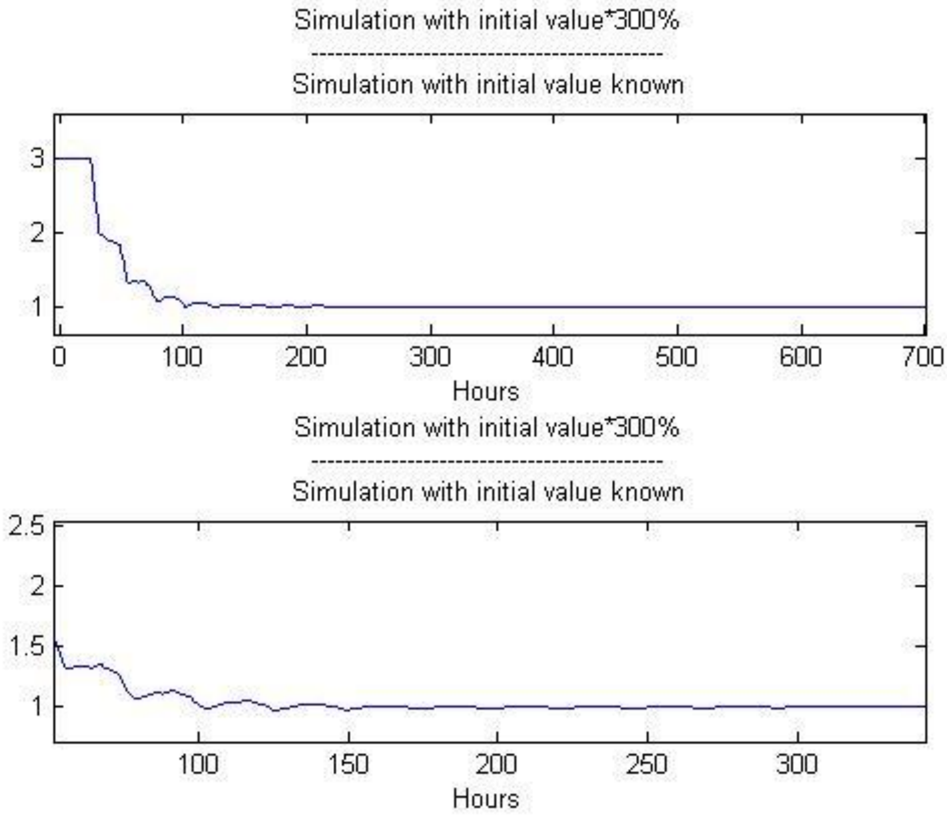


Figure 4.12. *The robustness of the simulation is tested by triple the initial value and see when it is totally restored, not affected by the initial value anymore. The lower figure is zoomed in.*

Figure 3.10 is showing the relative difference between the simulation with the modified initial values and the first simulation with observed initial values. Surprisingly the figure shows that the output from the simulation is stabilizing very fast. The impact of the initial value disappeared after 200 hours, which is approximately 8 days. This means that if the last day of a certain year were completely miscalculated it would only affect the first couple of days when simulating the following year.

Since the one step structured ARX model is explaining the electricity consumption with mostly autoregressive variables this test can be interpreted differently. The model is a good simulation up to 8 days as the model is built up by the last observed value.

4.1.7 Simulating with different weather scenarios

We now change the other variable in the model. Since the future temperature cannot be known the model will be simulated with 46 different historical weather scenarios instead. The aim of using the 46 years is that the simulated outcome space is a result of the historical temperature. The 95% confidence interval should become wider, than in the case of simulating 2013 with known temperature, since warmer and colder year are included in the simulation.

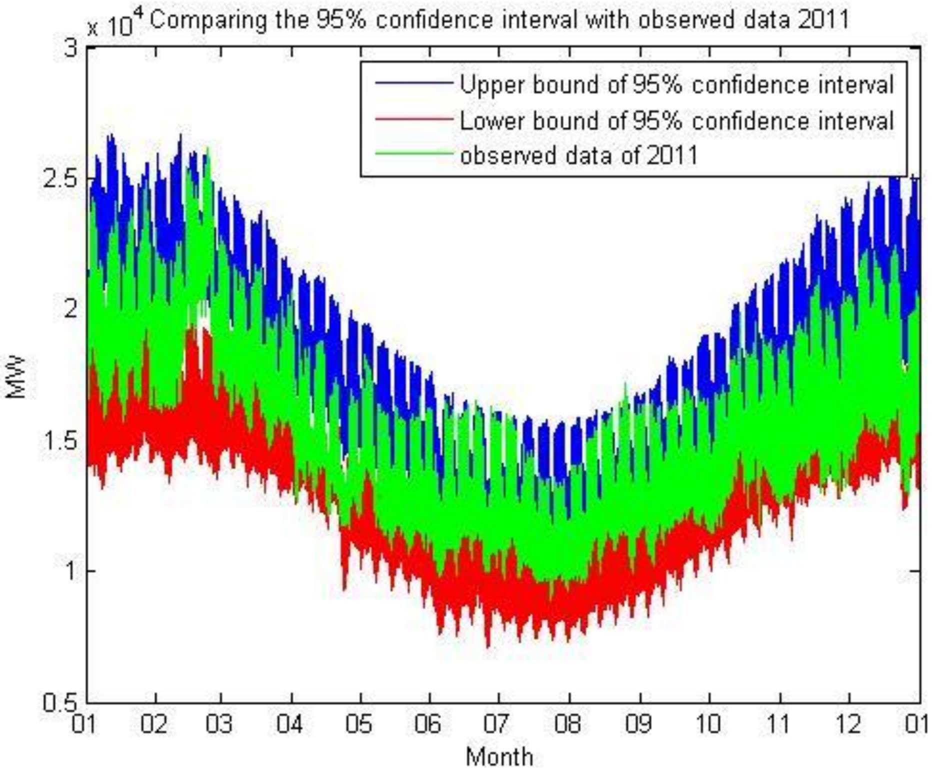


Figure 4.13 *Simulating 2011 with unknown temperature variable, 46 weather scenarios. Comparing the output of the simulation with observed data of 2011.*

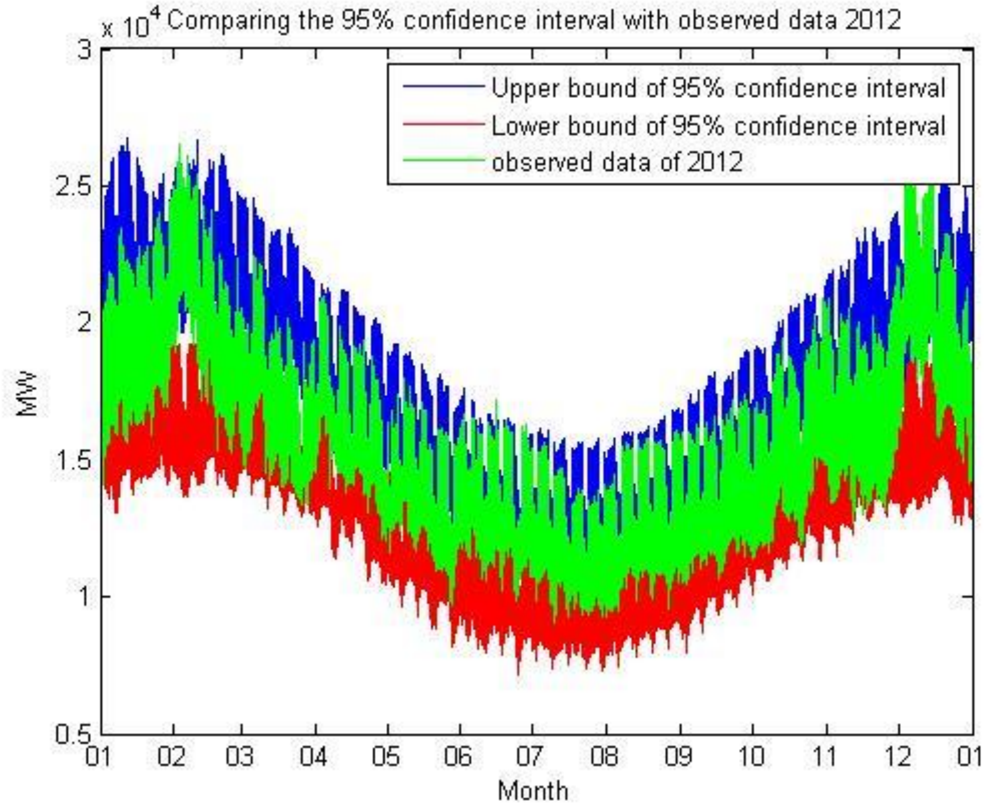


Figure 4.14 *Simulating 2012 with unknown temperature variable, 46 weather scenarios. Comparing the output of the simulation with observed data of 2012.*

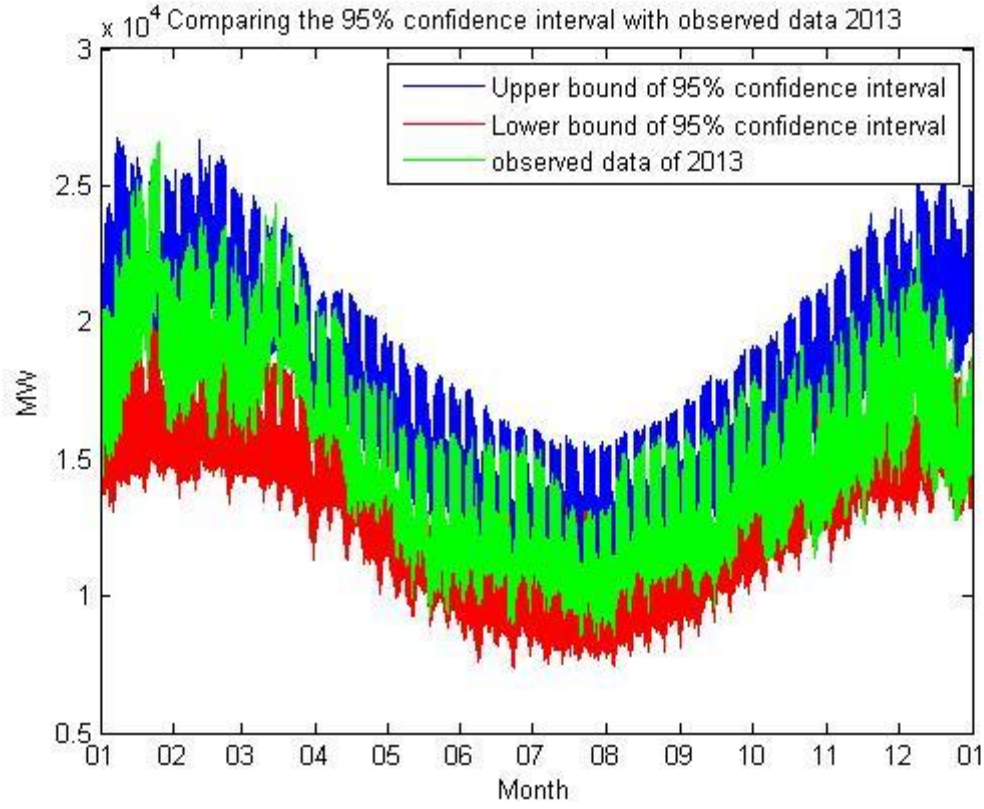


Figure 4.15 *Simulating 2013 with unknown temperature variable, 46 weather scenarios. Comparing the output of the simulation with observed data of 2013.*

In Figure 3.14 we see the relative difference between the simulated confidence interval with all weathers scenarios and the observed data of 2013 .The 95% confidence bands are around 80%-120% and is shifting quite a lot during summer time. One unexpected observation is that the interval is not covering the cold months in the beginning of the year 2013. A closer look is telling us that the coldest days in February and January are seen in a 99% confidence interval. Though the 99% confidence interval is not following the observed data in its curves and then is not describing the possible outcome very good.

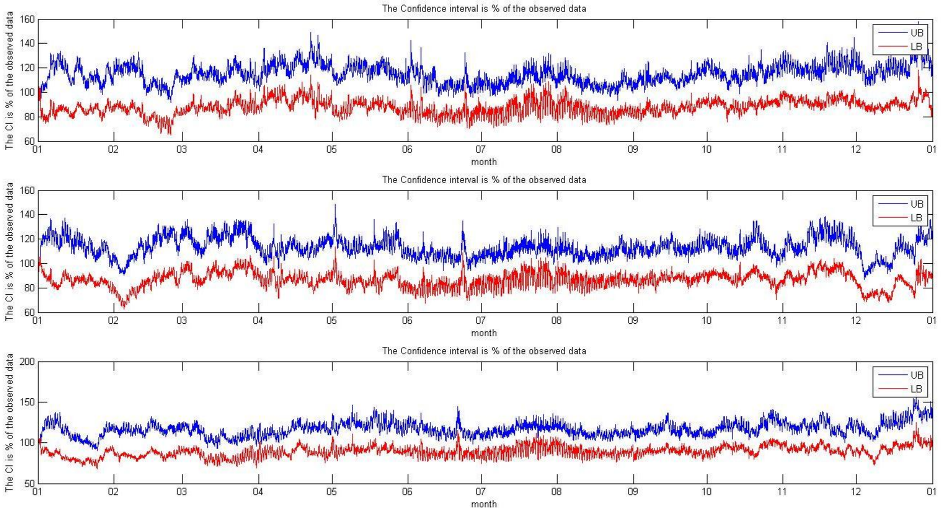


Figure 4.16 *The 95% confidence interval that was seen in figure 3.11-3.13 this relative relation with the observed data for respective year. If the output is 1 then the observed data and the confidence interval are equivalent.*

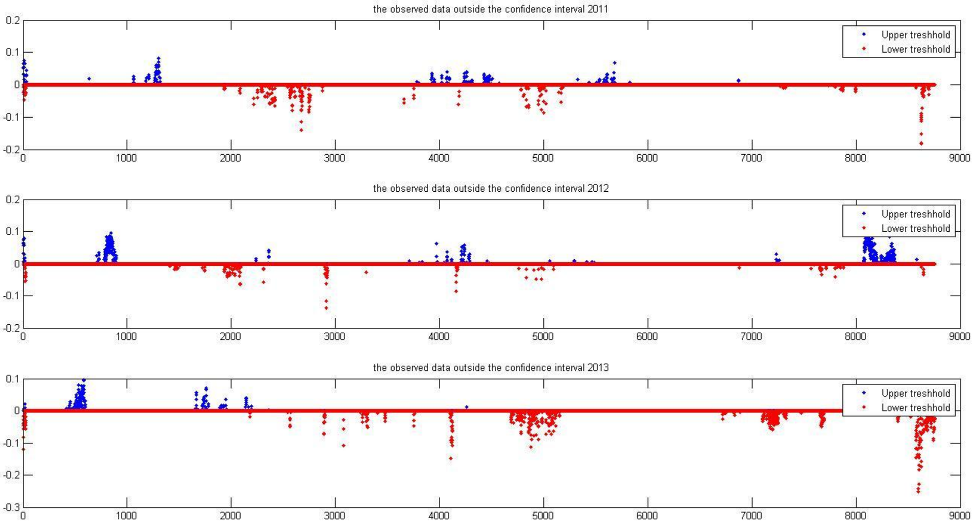


Figure 4.17. *The observed data that was left outside the 95% confidence interval for simulations with weather scenarios 2011-2013. The blue dots are for data point*

exceeding the upper limit and the red dots is for those data points that falls beneath the lower bound. Upper plot 2011 then 2012 and lowest 2013.

Simulated year	% outside the confidence interval
2011	5%
2012	6,4%
2013	8,2%

Table 4.3 *The percentage amount hours when the observed electricity consumption was found outside the 95% confidence interval.*

4.1.8 Temperature correlation comparison

The temperature is an important factor when calculating the electricity demand. A way of verifying that the model has the right relation to the temperature is by comparing the observed correlation, between the temperature and the electricity consumption and the correlation in the simulation, with the approximated temperature used.

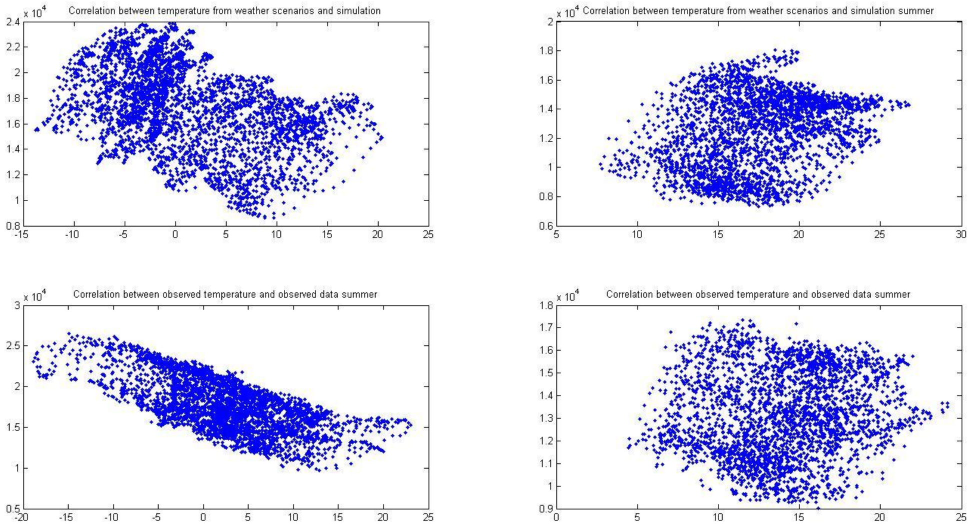


Figure 4.18. *The pairwise observation between the temperature and the electricity consumption. The upper row, first seen with the temperature of the spring with temperature form 2007 and in figure, first row second column, is during summer period. Second row is the correlation between the observed temperature and the observed electricity consumption from 2012.*

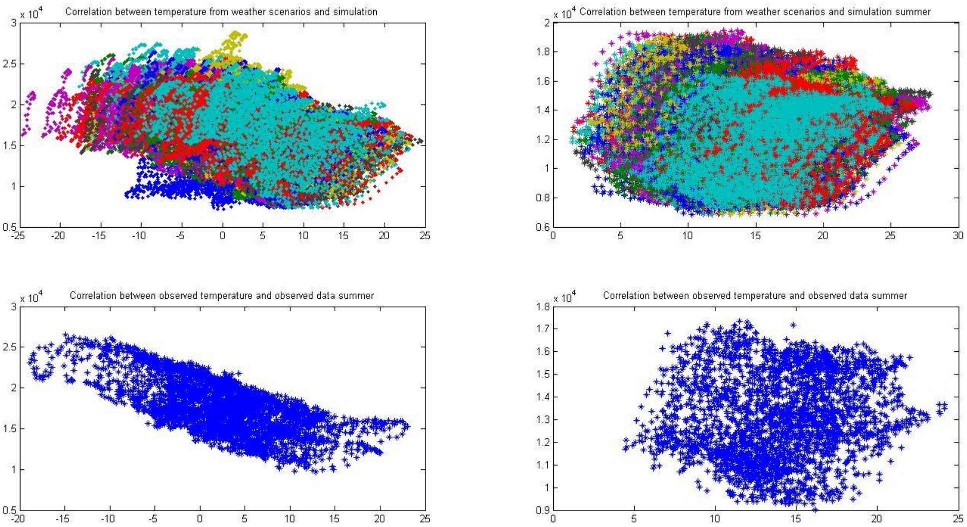


Figure 4.19. *The pairwise observation between the temperature and the electricity consumption. The upper row, first seen with the temperature of the spring and in figure, first row second column is during summer period. Second row is the correlation between the observed temperature and the observed electricity consumption.*

Looking at the figures that are describing the pairwise observation during wintertime it is seen that both figures flatten out when the temperature is around -12 degrees. The pairwise observation to the temperature in the simulated case is noticeable less dependent to the simulated demand due to the big weight on the AR process in the model. The observed pairwise observation has a strong dependency. During summer time the simulated pairwise observation is more similar to the observed pairwise observation and is nearly independent. This because the temperature variable do not affect the demand when the heating consumption is less.

This part will be compared to the 2 step ARX model where the pairwise observation were much more significant.

4.2 The two-step structure ARX

The results of the two-step structure ARX will here be presented and analyzed. They will mostly be compared with the results of the one-step structure ARX to avoid repetition of the comments.

4.2.1 Simulating with known variables

The confidence intervals are following the temperature good, for example it captures the sharp turn in January 2013 caused by quickly decreasing temperature very well, see figure 3.17. In a broad view there seems to be no clear differences between the one-step ARX model and the two-step ARX model.

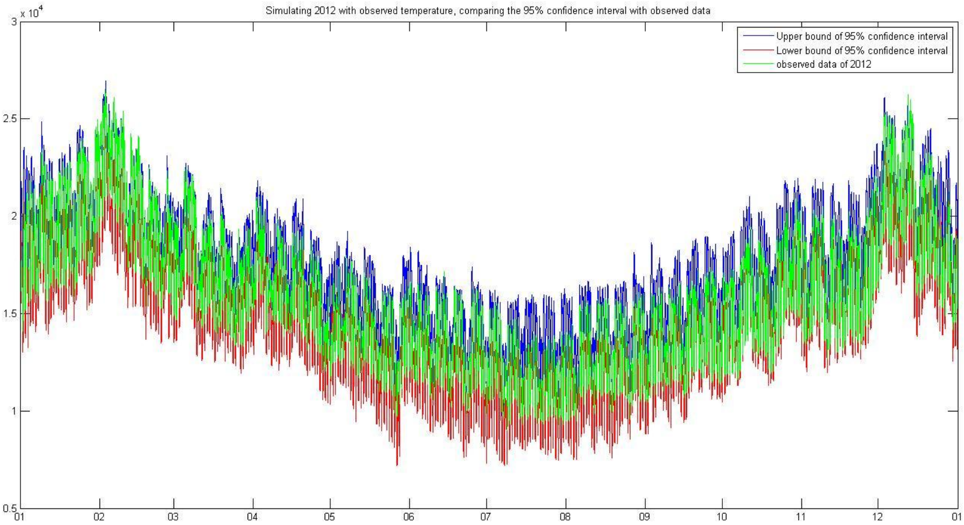


Figure 4.20 *Simulating 2012 with known variables. Comparing the output of the simulation with observed data of 2012.*

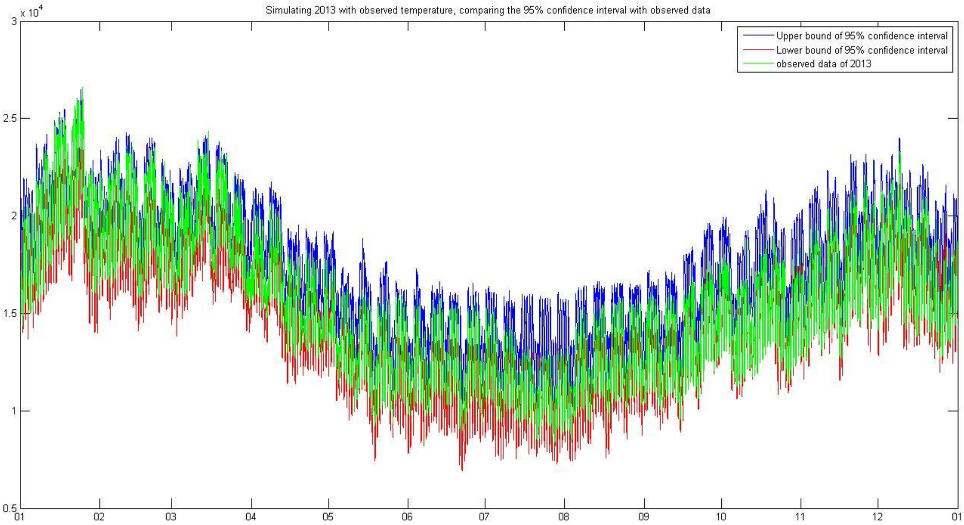


***Figure 4.21** Simulating 2013 with known variables, observed temperature. Comparing the output of the simulation with observed data of 2013.*

Comparing the years 2012 and 2013 in figure 3.18 the simulation of 2012 succeeds a little bit better with the 95% -confidence interval but the simulation of 2013 has a better daily profile during the vacation.

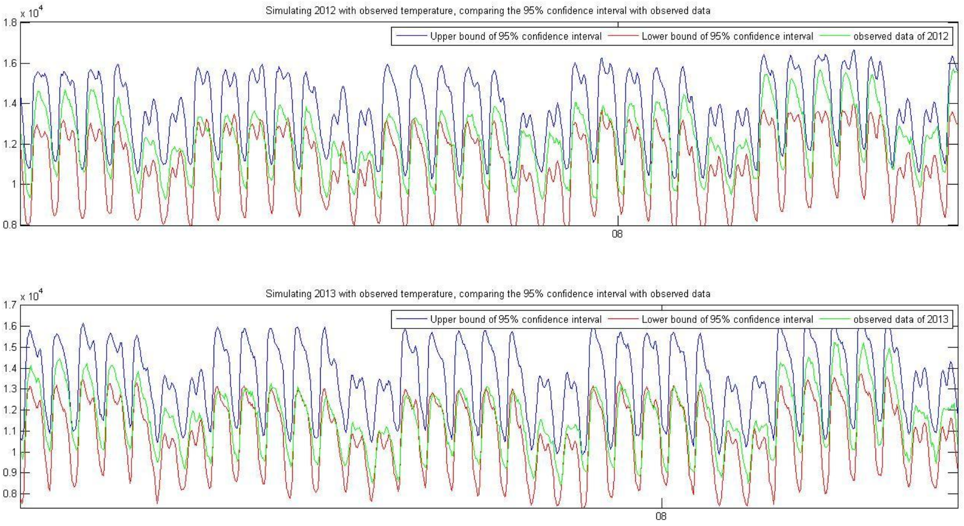


***Figure 4.22.** Same content as figure 3.16 (simulation of 2012) and figure 3.17 (simulation of 2013) but zoomed in at the vacation weeks in July/August.*

Zooming in at the summer period, see figure 3.19, the observed data for the weekdays in June have a more distinct summer profile than the output from the simulations, although some days the simulation fits better than others.

Summer

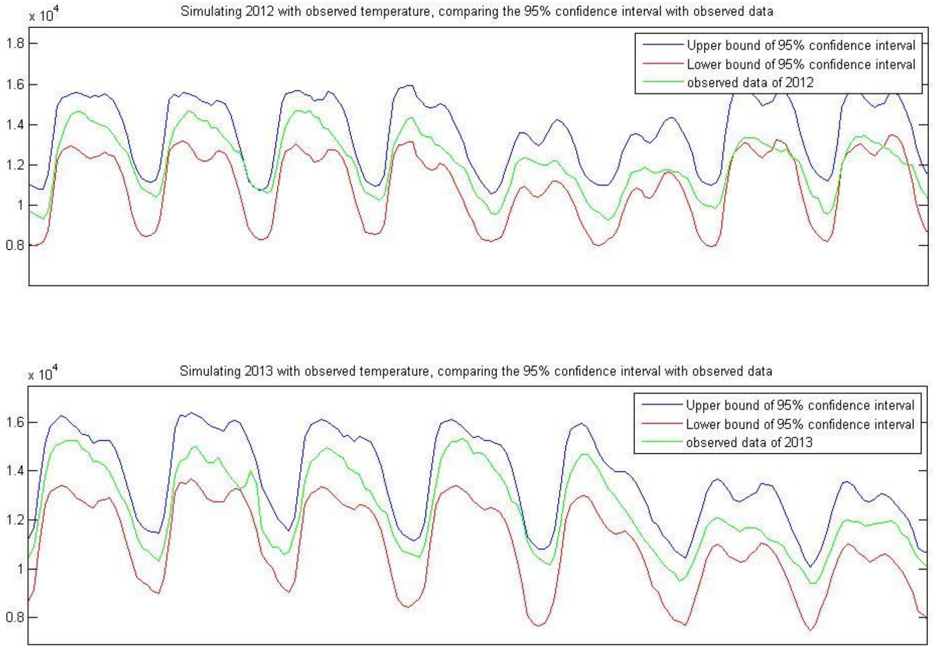
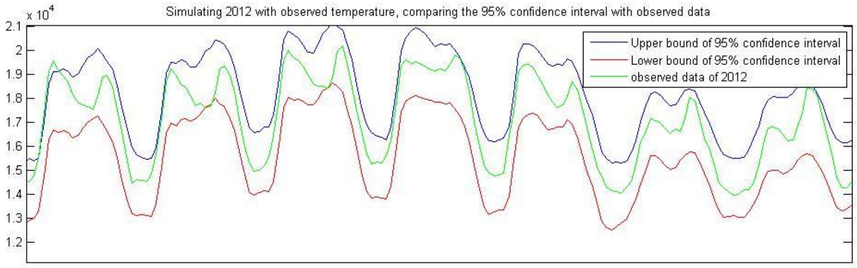


Figure 4.23. *Same content as figure 3.16 (simulation of 2012) and figure 3.17 (simulation of 2013) but zoomed in in a week in June. Note the daily trend during the summer period.*

In figure 3.20 it is seen that the two-step ARX model has a little more significant daily trend than the one-step ARX model. This is probably due to the temperature and hourly variables, which have more significant parameters. The two-step ARX model follows the daily trend better in the simulation of 2013 than 2012.

Spring



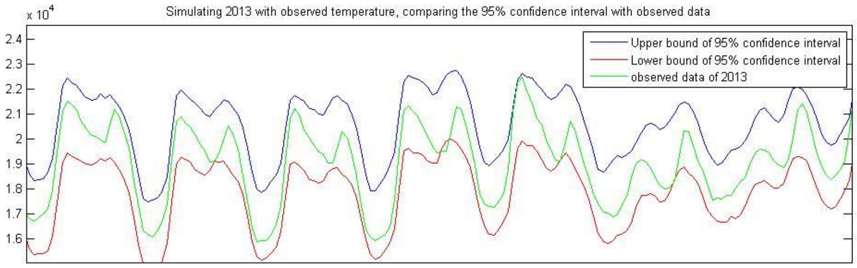


Figure 4.24. *Same content as figure 3.16 (simulation of 2012) and figure 3.17 (simulation of 2013) but zoomed in at a week in March. Note the daily trend with one morning peak and one evening peak.*

The idea of this model was as described above to be able to catch rapid changes due to external factors, better than in the one step ARX model, e.g. the public holidays, since the external part has more weight. In 3.21 the simulation struggles to find the public holidays during the Easter week and they tend to look like normal weekdays.

This result was not very surprisingly as the parameters for these periods are hard to estimate due to the leak of data. Vacation is three weeks during summer and the data used for estimating the model parameters were three years. Obviously it is not enough since the vacation period and the public holidays are not distinct enough. The interval seems slightly better than with the one-step ARX model since the lower interval follows the public holiday trend fairly good. The differences are however not very distinct.

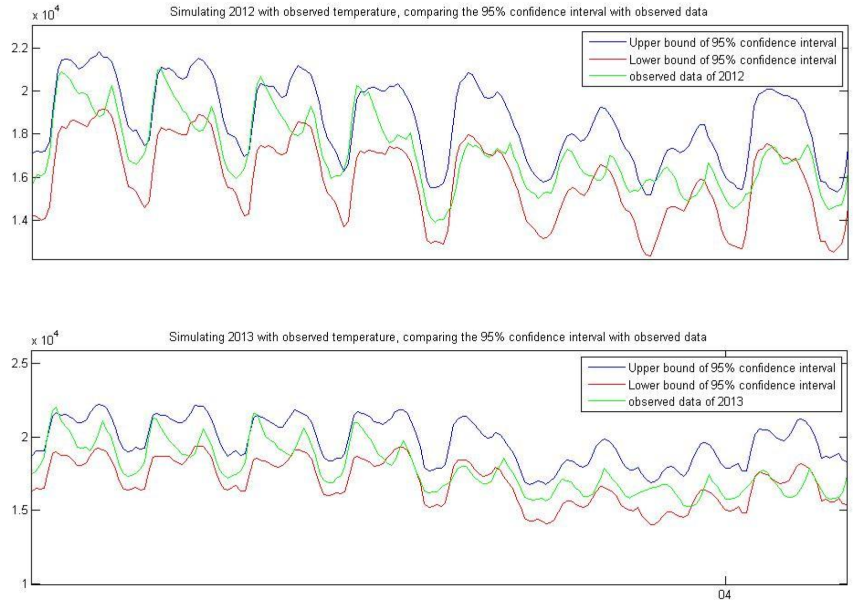


Figure 4.25. *Same content as the figure 3.16 (simulation of 2012) and figure 3.17 (simulation of 2013) but zoomed in at the Easter week. Note the public holiday at Good Friday and Easter Monday.*

4.2.2 Relative spread size of the confidence interval

The confidence interval is approximately between 90% and 110% of the observed data as can be seen in the following figure 3.22. More or less the same spread as with the one-step structure ARX model.

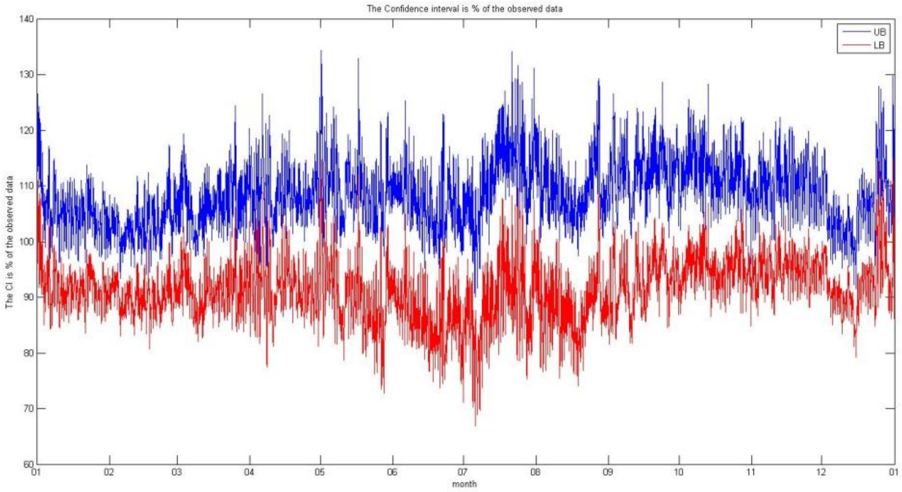


Figure 4.26. *The index ratio between the observed data and the 95% confidence interval in figure 3.17. Index = 100 when the observed data and the confidence interval are equivalent.*

4.2.3 Threshold

The simulation underestimated the effect of the cold temperature a little as grouped clusters are seen during wintertime. During summertime 2012 the excess is also increasing. This could be because of the none-existing temperature relationship during summertime. When the temperature is not affecting the model the AR process plays a bigger role. If the AR process poorly simulates one hour it will follow into the next day and this explains the grouped clusters. In the plot of 2013 the wintertime been better simulated. The bad indication could be that the threshold excess is not very random between the years but that the grouped clusters occur at same periods. Either the two years where similarly difficult to model during these periods or the model are weaker in these areas of the year.

The threshold excesses are similar to the one-step structure ARX model and no distinct differences are seen.

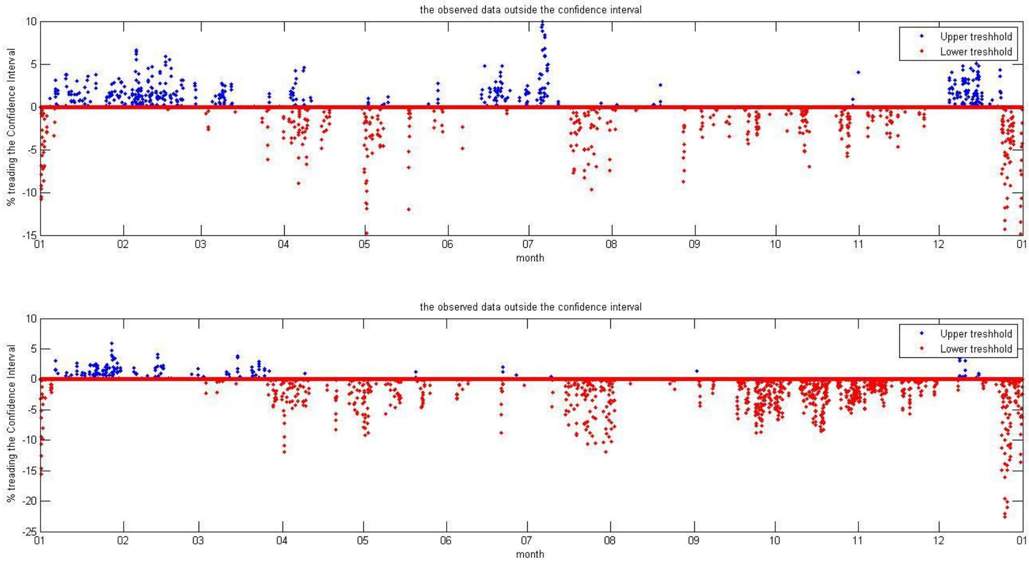


Figure 4.27. *The observed data that was left outside the 95% confidence interval. The blue dots is for data point exceeding the upper limit and the red dots is for those data points that falls beneath the lower bound. Upper plot 2012 and lower 2013.*

Simulated year	**% outside the confidence interval**
2012	3%
2013	3,4%

Table 4.4 *The percentage amount hours when the observed electricity consumption was found outside the 95% confidence interval.*

4.2.4 Distribution comparison

There are no big differences seen in the output of the mean hour for each month compared to the one-step structure ARX model. They are more or less overestimating and underestimating the same periods.

Month	**Mean 2012** (10^4)	**Observed mean 2012** (10^4)	**Mean 2013** (10^4)	**Observed mean 2013** (10^4)	**Std 2012** (10^3)	**Observed std 2012** (10^3)	**Std 2013** (10^3)	**Observed std 2013** (10^3)
Jan	1.9128	1.9711	2.0055	2.0453	2.2915	2.3814	2.8022	2.8022
Feb	2.0271	2.0730	1.9680	1.9844	2.5655	2.7063	2.0984	2.0984
Mar	1.7234	1.7350	1.9551	1.9205	2.0119	2.0727	1.9610	1.9610
Apr	1.6652	1.6394	1.6821	1.6165	1.9635	1.8388	1.8170	1.8170
May	1.4116	1.4080	1.3732	1.3293	1.9218	1.7473	1.7839	1.7839
Jun	1.3377	1.3626	1.2447	1.2512	1.8649	1.7467	1.7422	1.7422
Jul	1.2060	1.2107	1.2091	1.1631	1.7978	1.5273	1.5330	1.5330
Aug	1.2661	1.2916	1.2581	1.2526	1.8497	1.8376	1.8400	1.8400
Sep	1.4076	1.3931	1.4067	1.3650	1.9334	1.8437	1.9725	1.9725
Oct	1.6208	1.5829	1.5743	1.4991	2.2144	2.2376	1.9835	1.9835
Nov	1.7196	1.6989	1.7426	1.6901	2.1270	2.2446	2.3099	2.3099
Dec	2.0119	2.0258	1.7645	1.7464	2.4007	2.7688	2.4006	2.4006

Table 4.5 *The mean hour for each month of the year and the standard deviation for each month of the year. A comparison to between the observed data and the simulation with known temperature.*

The simulated empirical cumulative distribution for the month of March is following the output of the observed data as well it did with the one-step structure ARX model. The simulated model gives slightly fewer hours with higher electricity demand than the observed data but the difference between the distributions is very small.

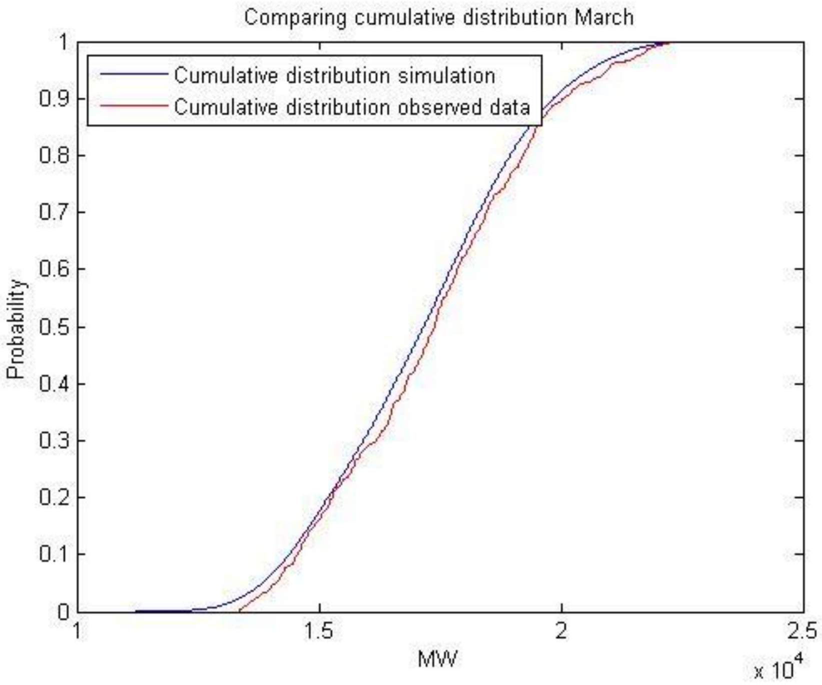


Figure 4.28. *The Cumulative distribution of the hours in March 2012.*

The tails in the cumulative distribution of August are larger for the simulated data than the observed data. Otherwise they are following each other quite well. Looking at figure 3.16 the confidence interval is wider than the observed data. During summer the temperature parameter is mostly zero so the simulation is less restricted.

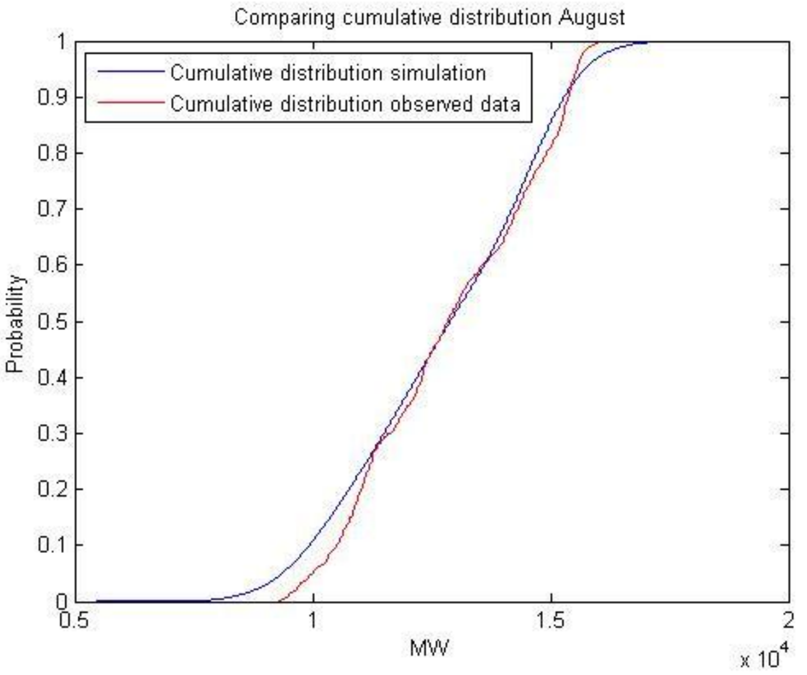


__Figure 4.29.__ The Cumulative distribution of the hours in August.

4.2.5 How sensitive is the model given external data but unknown initial value?

The test is done by having the initial value tripled in amplitude. There is hardly any difference in the results of the one-step structure ARX model. The simulation also recovers very well, and after approximately 200 hours (8 days) the outcome of the simulation has restored as if the initial value was known.

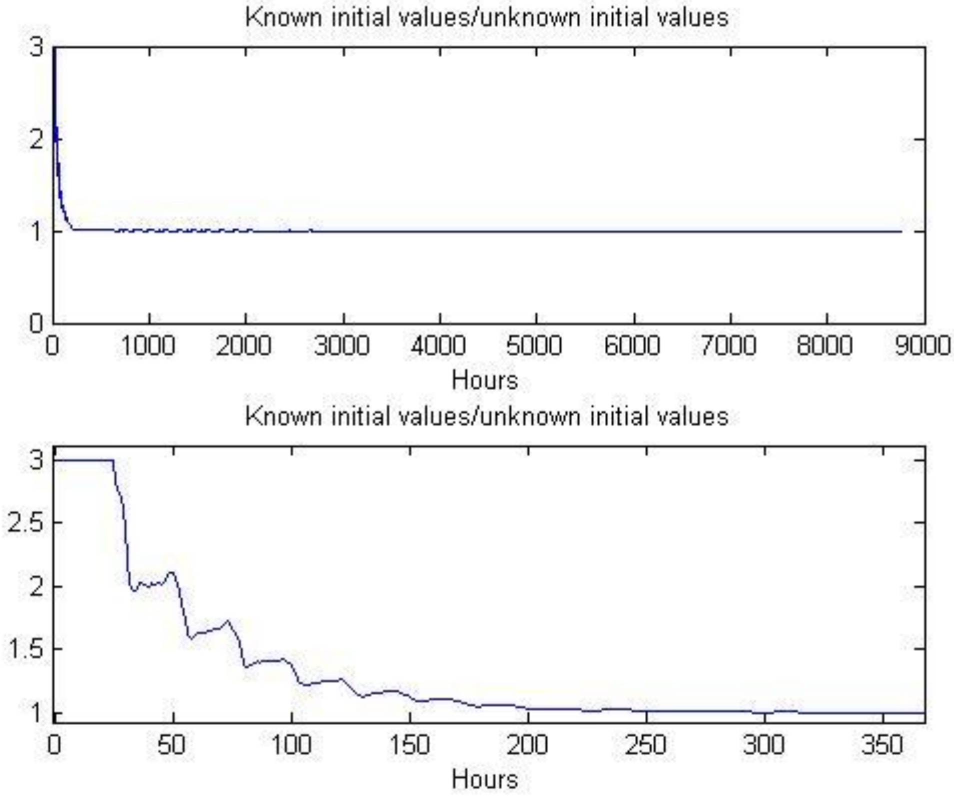


__Figure 4.30.__ The robustness of the simulation is tested by triple the initial value and see when it is totally restored, not affected by the initial value anymore.

4.2.6 Simulating with different weather scenarios

When simulating with all weathers scenarios the output of the coldest temperatures, which include February 2013, does not appears in the 97% confidence interval and is first seen in the 99% confidence interval. The tradeoff is that the wider the confidence interval the more the yearly trend (that resembles a sinus curve shape) diminishes. Figures 3.25-3.27 resemble the output from the one structure ARX; the confidence interval becomes more smooth and wider when simulating with a mixture of weather scenarios.

Increasing the 95% confidence interval of the weather scenarios is not raising the upper bounds a lot but rather decrease the lower bounds the more. This is probably due to the spread of cold winter comparing to the spread of warmer winters. The smaller spread in the upper bound causes of the underestimation at extreme cold temperature of the model.

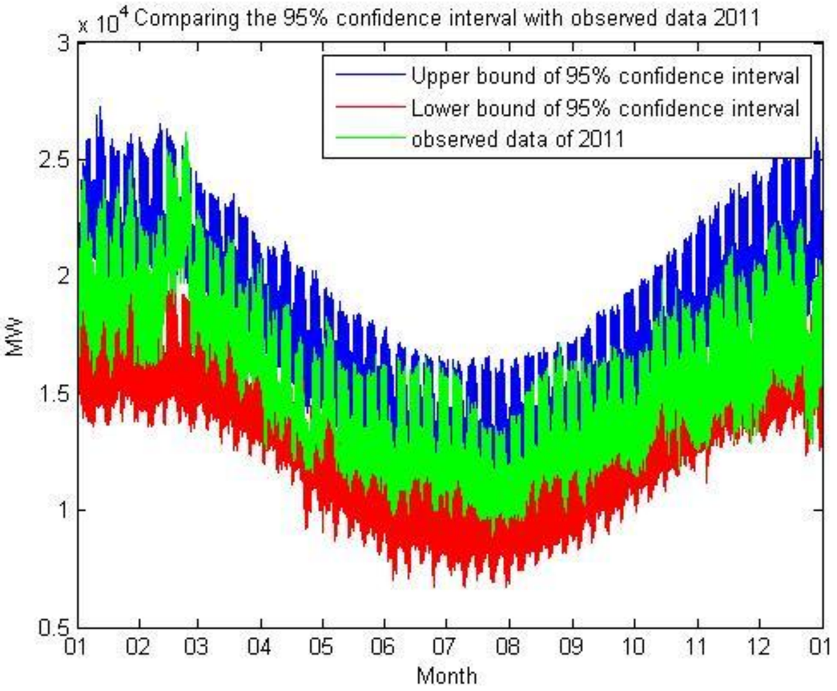


Figure 4.31 *Simulating 2011 with unknown temperature variable, 46 weather scenarios. Comparing the output of the simulation with observed data of 2011.*

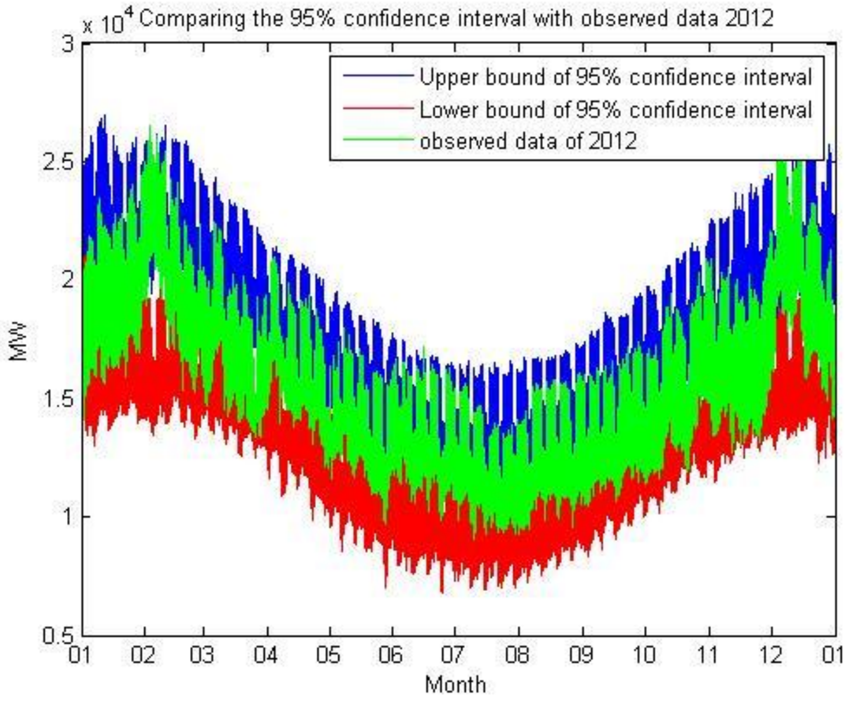


Figure 4.32 *Simulating 2012 with unknown temperature variable, 46 weather scenarios. Comparing the output of the simulation with observed data of 2012.*

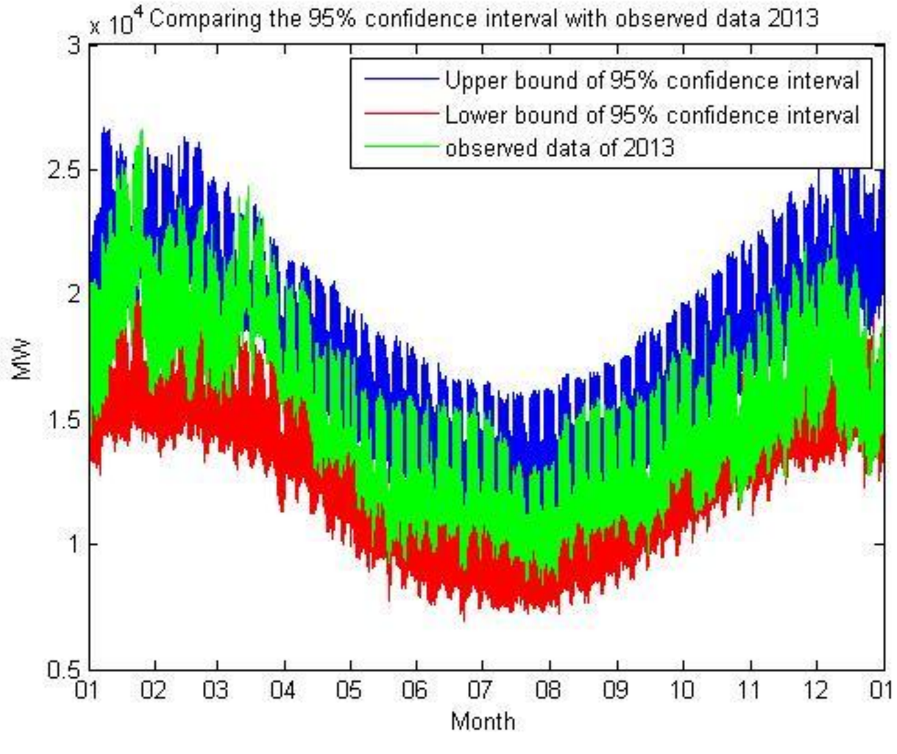


Figure 4.33 *Simulating 2013 with unknown temperature variable, 46 weather scenarios. Comparing the output of the simulation with observed data of 2013.*

In order to investigate if the spread of the confidence interval is in order and functional or if it is a badly description of possible outcomes the amount of demand hours is counted and the percentual amount is analyzed. Though it is seen that the variation during summer is very low but in the confidence interval for the simulations it has a wider spread than other periods. The low bounds during summer period are due to the absence of temperature variable. Since heating day degrees is used is the temperature variable not included in the model when the temperature is higher than 17 degrees Celsius.

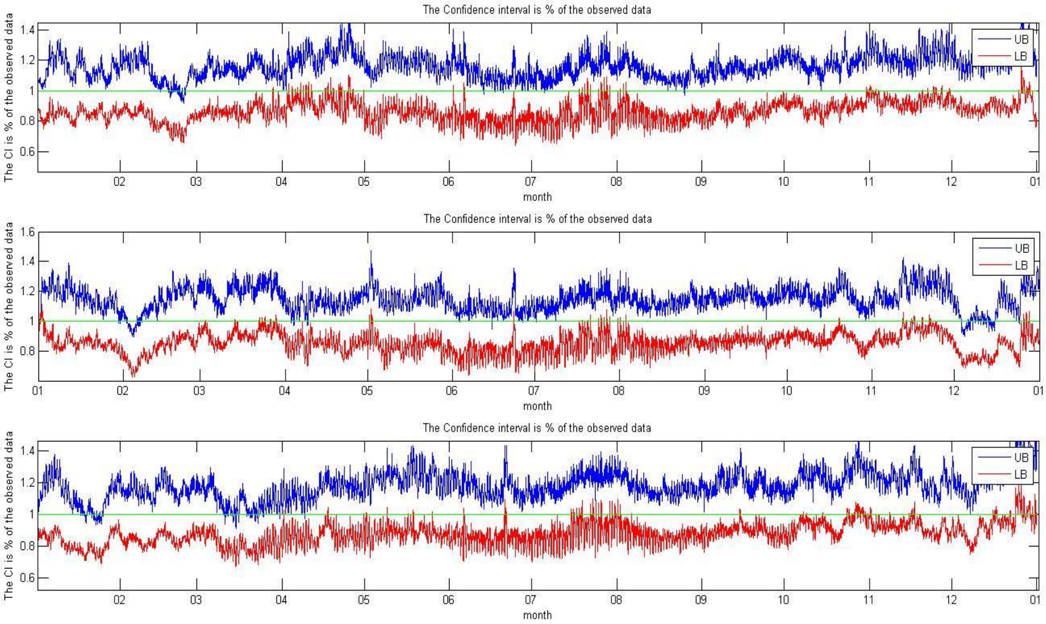


Figure 4.34 *The 95% confidence interval that was seen in figure 3.25-3.27 this relative relation with the observed data for respective year. If the output is 1 then the observed data and the confidence interval are equivalent.*

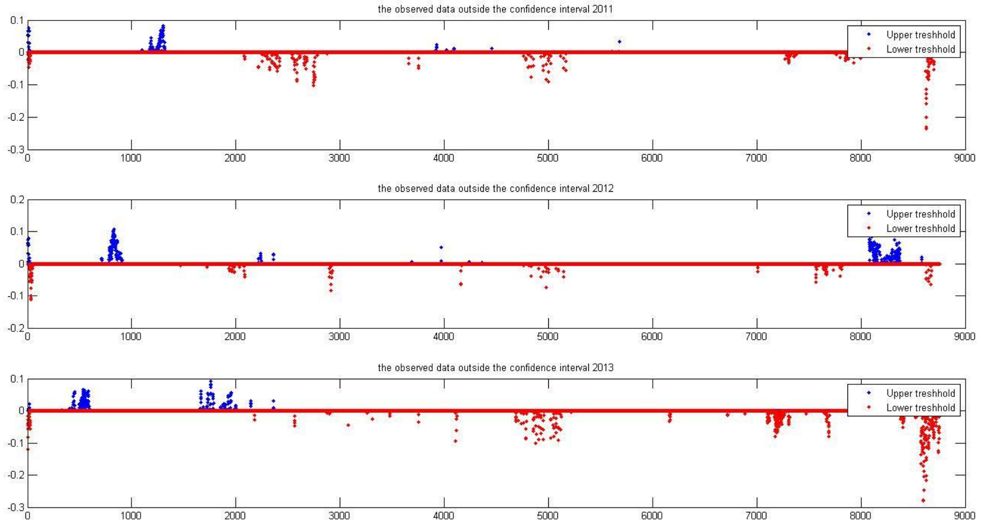


Figure 4.35. *The observed data that was left outside the 95% confidence interval for simulations with weather scenarios 2011-2013. The blue dots is for data point exceeding the upper limit and the red dots is for those data points that falls beneath the lower bound. Upper plot 2011 then 2012 and lowest 2013*

Simulated year	% outside the confidence interval
2011	4%
2012	5%
2013	6%

Table 4.6 *The percentage amount of hours when the observed electricity consumption was found outside the 95% confidence interval of the simulation with weather scenarios.*

4.2.7 Temperature correlation comparison

The simulation of the two-step ARX model should have a more significant correlation seen with the temperature, as there is more weight on the parameter than in 1 step ARX model. Looking at the correlation of the non-summer month in figure 3.32, first row first column, this is also seen. The correlation output is more packed into a line than in the case with the one step ARX model simulation. Figure 3.32 is showing the correlation when using temperatures from 2007 when figure 3.33 has for all weather years at the same time.

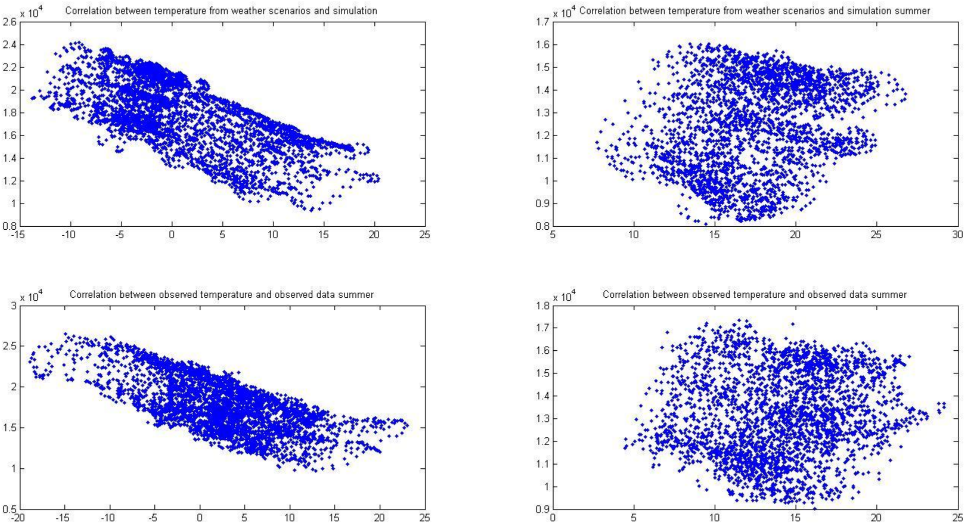


Figure 4.36. *The correlation between the temperature and the electricity consumption for 2007. The upper row, the correlation between the weather scenario of 2007 and the simulation done with this temperature. Upper row first column is during spring and second column is during summer time. Second row is the correlation between the observed temperature and the observed electricity consumption 2012.*

Continuing by analyzing the correlation between all-weather scenarios and their respectively simulation the result seems good. The non-summer months is more similar to the observed correlation than before. The summer period with all-weather scenarios looks a little bit different which is because of the cold summers which created correlation to the demand as heating were used also during summer time.

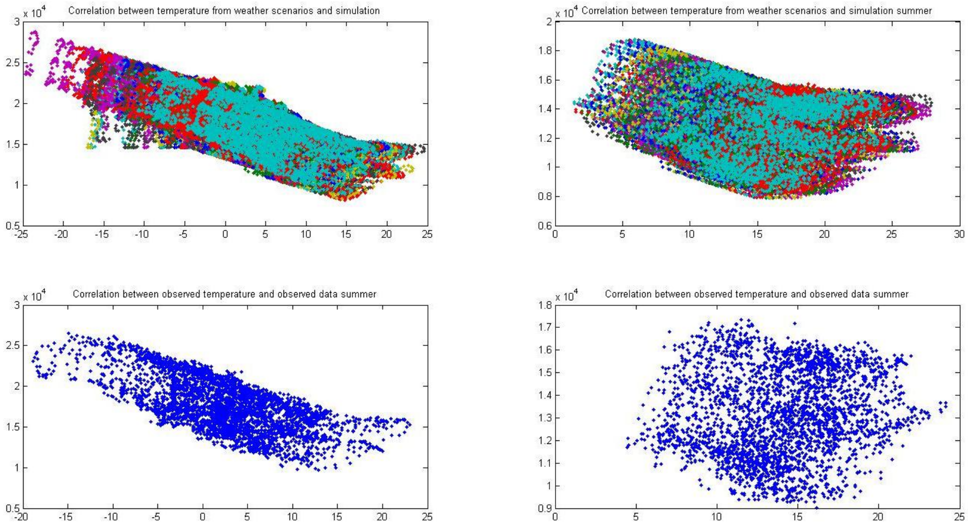


Figure 4.37. *The correlation between the temperature and the electricity consumption for 2007. The upper row, the correlation between the weather scenario*

of 2007 and the simulation done with this temperature. Upper row first column is during spring and second column is during summer time. Second row is the correlation between the observed temperature and the observed electricity consumption 2012.

4. 3 Result of the EMPS model

The main objective of this project was to create a better electricity price forecast by transform from weekly outcome to hourly outcome of the price prognosis. In order to receive an improvement by the transformation, more information had to be received by the input data which also had to be on hourly basis. The demand input data was one of the factors. The validation of the demand model is hard to do by observing the outcome from the price model since the impact of the demand is not the most significant variable in the EMPS model.

4.3.1 Changing input demand data from annually to hourly

The previous input data was an annually consumption which was distributed by annually-, weekly- and daily profiles. The demand data simulated by the model has 46 weather scenarios which are included in the model. In the old demand data only a normal temperature is used and the weather scenarios is built in as input to the price model instead.

If we disregard the noise added to the model, just simulate the model 46 times with the 46 different weather scenarios without noise and use these simulations as input data to the price prognosis model. The outcome of the price mode using the old demand data and the demand data without noise form the model can be compared by looking at the weekly trend and the daily trend.

During summer time the prognosis, with the new demand data, had much more distinct variation between day and night. The rest of the year did not show any distinct different.

4.3.2 Including noise to the demand data

The demand data simulated by the model has noise added which includes the error estimation of the model. The model is simulated with the 46 weather scenarios 20 times each to include the spread of the noise. If the demand has enough impact on the price prognosis the spread of the price prognosis will be wider.

A comparison was done between the spread of the price prognosis, when simulated with the old demand data and the 46 weather scenarios, and when simulated with the new demand data, 46 weather scenarios and 20 simulation on each weather scenarios to include the noise spread.

The analysis of the comparison was that the spread was almost equal. This means that the impact of the demand on the price model is not big enough to be able to notice the difference between the outcomes with the noise.

5. Discussions and conclusions

This section will analyze the result from a wider perspective as the section of the result was more detailed. I will also go through the main objectives of the project , what goals were accomplished and how a tradeoff between the objectives had to be considered.

5.1 Conclusion of the main objectives in the project

The objectives were to create a model that explained the hourly electricity consumption in the Nordic countries, one for each country. The model should be buildup by annually, weekly and daily trends but also be temperature sensitive. The objective of the model was also to include economic trends and be able to identify changes in the consumption connected to the economic development. Quantifications of the factors should be estimated and possibilities of analyzing the variation of the factors.

The model is describing the annually, weekly and daily trend and also follows changes along with the temperature. According to the main objectives and looking at the simulated outcome, has the model accomplished the main tasks and the outcome is a realistic demand year saying the model is good.

5.2 Discussion and conclusion of Results

One of the weakness of the model is when the temperature is unusually low. It does not captures the most extreme peaks as the heating consumption increases with the extremely cold temperatures. One of the explanations is the approximation of the temperature that is done. The input data of the temperature is on daily basis and is transformed to hourly basis by observing the average daily profile of the temperature each month. Otherwise the simulation of the model is changing fairly good along with the temperature.

Another weakness is when public holidays occurs and during vacation. The parameters of the variables are estimated with use of four years of data. Each year have approximately 10 public holidays and the conclusion is that the leak of data results in weakly estimated parameters. The parameter of the public holiday should be more significant than it is if more data was available and handled which would contribute to a more distinct difference between a normal day and a public holiday.

The same issue is seen during the vacation period and the variable of the vacation. The consumption is not decreasing sufficient, either during the summer or the winter. This is probably also due to the leak of data.

One strength of the model is the spread of the confidence interval is relatively tight. If the confidence interval is too tight it would imply more risk and if it is too wide it would give less information about the possible output. Looking at the threshold plots the 95 % confidence interval is a little too tight as the amount of observed demand hours that is outside the interval is slightly too high. But when all the weather scenarios are included in the simulation the 95% confidence interval is broadened a bit. This suits the variation of the observed demand better as the threshold plot shows less demand hours outside the interval. Comparing to how much the consumption changes from one year to another it is relatively good size of the spread of the interval from the simulation with the weather scenarios.

The annually trend is followed very good from summer to winter and so is also seen for the weekly trend. The daily trend is more distinct when the temperature variable is more distinct, though not extreme, but comparing the summer period with the rest of the year as the heating day degree is only effective below 17 degrees of Celsius. The observed daily trend is different during summer and the rest of the year. In the summer time the weekdays often only have one peak of the day unlike the rest of the year where two peaks is seen, one in the morning and one in the afternoon.

The hour variables are the same for the whole year, meaning there are no specific weekday hours for the summer and for the winter. Since the daily trend differs depending on the period the parameter estimation might be bit of a compromise as the summer only have one peak and the winter two. The afternoon peak in the winter time is not distinct as in the observed data but the afternoon peak also appears, not as distinct as in the winter, in the summer time as well.

Though would the including of almost twice as many variables not work since they have many hours which are very similar and the parameters would probably be insignificant.

In the chapter of Result the two different structures of ARX was compared. The aim when trying the two step structured ARX was to increase the significance of the fundamental parameters and by then increase the impact of them in the simulation. This was only seen in the plot of correlation between temperature and demand. The 2 step ARX model had way more temperature correlation during winter/spring than the 1 step ARX model. Otherwise the results were very similar.

When first looking at the residuals after the simulation I did not think of the definition of the inversed autocorrelation and suspected that the simulation was not good. This was also one of the reason why I thought the 2 step ARX model would be better. With the definition of the inverse ACF it is seen that the residuals after simulating with an AR process will in the inverse ACF correspond to a MA process. The residuals were then correct simulated with both structures of the models. The definition of the inverse ACF of an AR process is seen in section 2.1.1.1. (Madsen, Time Series Analysis, 2008)

The last analysis of the result and the main application of the model is the EMPS model. The EMPS model did expand the variation during the summer period while the rest of the year did was varying and did not show any distinct trend that differed from the old data. The analysis of the spread when include the noise in the new simulated demand data did not indicate of a high impact of from the noise. The simulations of the demand with the different noise did come out as more or less the same price.

5.4 Future work

As mentioned before there are possibilities to improve the model when it comes to days that differs from the normal day, e.g. vacation, public holiday and days with more extreme temperature.

In order to capture the cold days better a better transform of the temperature could be done as estimating model with several variables that affects the way of how the temperature affect the demand. Example of variables would be wind, humid and as well where in the country the temperature is measured. The north of Sweden for example does not react in the same way as south of Sweden to different weather scenarios.

One of the hopes of the model was to include an economic indicator as a variable. This was hard to fulfill since the variables that could possible explain economic develop was at minimum on quarterly basis. A future work could be to find the industries that has hourly balancing which it could be possible to find the trend of economic development. But still with that information it is hard to see such trend as the industries are normally cutting down the production in larger scales instead of a bit at a time. Another possible solution would be to use the MIDAS model, see appendix B, which uses a technique that makes it possible to have input variables with different sample frequencies. The method would make it possible to keep the low resolution and not approximate the quarterly based data to hours. The advantage of the method is that it can keep the more information of the raw data but it struggles with the amount of parameters which can quickly be too many due to the way of including lags. (Ghysels, Santa-Clara, & Valkanov, 2004)

Another objective that was too time consuming to include was to find the trend of moving the consumption from day to night due to lower electricity prices. To be confident in such a result it requires further investigation of how consumers perceive around this and create a representation of a possible future change.

Final entry is to create an interface between the output of the simulation of the model to EFI, the EMPS model. Today there is no connection to the EFI system and a future interface could either implement the simulation in EFI or create a link to the MATLAB files so the data will be sent directly as input data to the system.

7. References

Akaike, H. (1969). Fitting autoregressive model for prediction. *Ann. Inst. Statist. Math., Vol. 21*, 243-247.

Alfares, H. K., & Nazeeruddin, M. (2002). Electric load forecasting: literature survey and classification of methods. *nternational Journal of Systems Science, vol. 33, number 1,* , 23±34.

BOFELLI, J. V., & MURRAY, F. T. (2001). Forecasting Electricity Demand on Short, Medium and Long Time Scales Using Neural Networks. *Journal of intelligence and Robotic Systems Vol 31*, 129-147.

Brown, B. G., Katz, R. W., & Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of climate and applied meterology Vol. 23*, 1184-1195.

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *STATISTICS IN MEDICINE vol. 19*, 1141-1164.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The MIDAS Touch: Mixed Data Sampling Regression Models. *Working paper*.

Härdle, W. K., & Trück, S. (2010). *The dynamics of hourly electricity prices, SFB 649 Discussion paper 2010-013, , ISSN 1860-5664.* Retrieved from Humboldt-Universität: http://sfb649.wiwi.hu-berlin.de

Jakobsson, A. (n.d.). Time series analysis and signal modelling. Lund university.

Madsen, H. (2008). Time Series Analysis. Chapman & Hall/CRC.

Mestekemper, D.-W. M. (2011). *Energy Demand Forecasts and Dynamic Water Temperature Management, Dissertation for the degree of doctor.* Universität Bielefeld, (referee: Kauermann, Prof. Dr. G.; Kneib, Prof. Dr. T.).

Singh, A. K., Ibraheem, Khatoon, S., & Muazzam, M. (2013). An Overview of Electricity Demand Forecasting Techniques. *Network and Complex Systems, Vol.3, No.3, ISSN 2224-610X (Paper)*, 38-48.

SINTEF, EMPS. (n.d.). Retrieved from http://www.sintef.no/home/SINTEF-Energy-Research/Project-work/Hydro-thermal-operation-and-expansion-planning/EMPS/)

Söderström, T., & Stoica, P. (1989). Chapter 7. In *System Identification.* Prentice Hall.

University of Baltimore. (n.d.). Retrieved from http://home.ubalt.edu/ntsbarsh/stat-data/GraphForecast.gif

Appendix A

Dynamic Factor Model

A model that estimates the co-movement between several time series. The dimension of the original data set, the number of input variables, is reduced as the variables are projected onto one and another. The information from the original data set is still there but compressed into a lower dimensional data set. The dynamic factor model finds the common factors by using normalization.

A modified way of using dynamic factor models is by letting the factors be functions of observable variables, this is called semi parametric factor model. To identify the common factors the DSFM uses a simulating technique which shows that for any set of estimated factors there exists a set of transformed factors with the same covariance structure as the original set. This means that it is possible to interpretation can be done on any probable set of factors.

In an
orthogonal L-factor model an observable J-dimensional random vector

$$Y_{t,j} = m_{0,j} + Z_{t,1}m_{1,j} + \cdots + Z_{t,L}m_{l,j} + \varepsilon_{t,j}$$

Where, $Y_{t,j}$, is the demand at time t and dimension j and can be considered as a multi-dimensional time series. $Z_{t,l}$ are the common factors, $\varepsilon_{t,j}$ are the error, $m_{l,j}$ are factor loadings.
The advantage of the dynamic factor model is if sufficiently of the variation in Y_t can be explained by the L common factors, $Z_{t,l}$, the feasibility will increase along with the reduction of dimension.
(Mestekemper, 2011)

Appendix B

MIDAS

Time series regression model with time series at different sampling frequencies. A base frequency is used where the frequencies of the other variables are described by the base frequency, e.g. if the base frequency is one year and *m*=4, the variable has a frequency of a quarter.

The formula is,

$$Y_t = \beta_0 + B\left(L^{1/m}\right)X_t^{(m)} + \varepsilon_t^{(m)}$$

Where

$$B\left(L^{1/m}\right) = \sum_{j=0}^{j_{max}} B(j)L^{j/m}$$

And Y_t is the demand, $B\left(L^{j/m}\right)$ is a suitable polynomial where the lags *m* of $X_{t-j/m}^{(m)}$ is included as parameters.

The MIDAS regression model includes much more information as the variables with higher resolution does not need to be aggregated and also the variables with lower resolution does not need to be approximated to higher resolution. The model also become more flexible along with the different sample frequencies.

The cost of the increased set of information and the flexibility is the distribution of the parameters. Imagine a model with many variables which has a suitable polynomial $B\left(L^{j/m}\right)$ with many lags of $X_{t-j/m}^{(m)}$ data. The number of parameters to estimate would be many as well. The best possible would be to capture as much as possible of the information from the MIDAS regression but keep down the number of parameters. The number of parameters are reduced by methods within distributed lag models. (Ghysels, Santa-Clara, & Valkanov, 2004)