ISTANBUL TECHNICAL UNIVERSITY
FACULTY OF SCIENCE AND LETTERS
GRADUATION THESIS

SHORT AND LONG TERM FORECASTING OF ELECTRICITY DEMAND IN TURKEY

MURAT KOCAVELİOĞLU

Term : **2012-2013**

Department : **Mathematics Engineering**

Number : **090080054**

Thesis Supervisor
Dr. Eti MİZRAHİ

MAY 2013

İSTANBUL TEKNİK ÜNİVERSİTESİ

FEN EDEBİYAT FAKÜLTESİ BİTİRME

ÖDEVİ

TÜRKİYE ELEKTRİK TALEBİNİN KISA VE UZUN DÖNEM TAHMİNİ

MURAT

KOCAVELİOĞLU

Dönem : **2012-2013**

Bölüm : **Matematik Mühendisliği**

Numara : **090080054**

Bitirme Ödevini Yöneten Öğretim Üyesi

Dr. Eti MİZRAHİ

MAYIS 2013

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest gratitude to my supervisor Eti Mizrahi for her admired hardworking and planned personality, friendship, guidance, advice, motivation, encouragements, patience and insight throughout my term project process.

Special thanks to my close friends from IEEE family for their special help and friendship throughout my term project process and student life in ITU.

Lastly, I would like to express my endless gratitude to my dear big family and to all my friends for their endless support, love and encouragement.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

SUMMARY

In the developed markets, there are enough public and private market players for market liquidity so that open competition is exactly formed. Turkey's electricity market is at the beginning of this revolution (road to open competition). Turkish government is striving for matured electricity market with open competition by investments, incentives and privatization. When Turkey`s electricity market will be developed, players will have to make best strategy for trading in order to survive in the market. Correspondingly, in the electricity markets, supply equals how much demand is and also price rises where demand intersects with supply. So that price varies depending on demand directly. Due to this case, through making trading strategies needed short and long term demand forecast is subject of this work.

There are many simulations which are made by using mathematical model, neural network and fuzzy logic or etc. STLF works began at early 1960`s and one of the first published work was done by Heinemann *et al.* In 1966, it is first use for relationship between temperature and electric demand. In 1971, first electric demand forecasting system which was developed by Lijesen and Rosing used statistical approach. In 1987, Hagan and Behr used time series model. After 1990`s, more complicated systems were developed and there are new approaches. In 1991 Part *et al*. first ANN use in STLF. In 1995, last developed technology fuzzy logic and neural networks were used by Srinivasan. On the other hand in Turkey, the first STLF system was developed by Erkmen and Ozdogan in 1997 using ANN approach. In this work temperature data was not included as an input variable in ANN. Another example is in 2003 by Topalli and Erkmen for daily load forecasting using ANN in which hourly load and calendar were input variables only. However, In 2006 Topalli *et al*. developed a ANN model which included hourly load, calendar data and temperature data as input variable [10]. In addition, Ertugrul *et al.* used dynamic regression for long term load forecasting [2].

In this work, neural networks and mathematical models are applied comparatively. Mainly, the aim of this work is forecasting of electricity demand in Turkey with minimized error. The problem is what forecasting error is.

As a Short; in this work, electricity demand is forecasted for short and long term and their results are analyzed with theory and practice. In theory, either ARIMA or ANN can be developed differently for other try and also in practice, ANN have just applied in electricity sector but this work shows us ARIMA can be performed for electricity demand forecasting. Finally, although we cannot generalize this case, in this work it seems there is a case which is ARIMA`s load forecasting results can be better than ANN`s.

ÖZET

Gelişmiş elektrik piyasalarında özel ve kamu kuruluşu olan oyuncuların sayısı oldukça fazla ve serbest rekabet tam anlamıyla oluşmuş durumda. Türkiye elektrik piyasası ise bu evriminin henuz başlarında. Hükümet 2015'e kadar rekabetin yüksek olduğu ve piyasanın borsalaştığı bir durum için gerekli yatırımları, özelleştirmeleri ve teşvikleri yapıyor. Bu süreç içerisinde piyasada bundan önce olanın aksine hayatta kalabilmek için daha iyi strateji yapma gerekliliği doğuyor. Bunun sonucu olarak; uzun ve kısa dönem stratejilerin önemi günden güne artıyor. Bu stratejilerde alım satım fiyat odaklı olmasına rağmen; talep ve arzın kesiştiği noktada fiyatın oluşması ve talep edildiği kadar arz olma zarureti fiyatı doğrudan talep odaklı yapmıştır. Bu nedenledir ki; elektrik talep tahmini bu stratejilerde hayati önem taşımaktadır. Tahmin, modelleme ve fiyatlama gibi konuların olduğu yerde ise mühendislik başlar.

Elektrik talep tahmininde çeşitli matematiksel modeller, yapay zeka ve bulanık mantık modelleri, mevsimsel modeller, regresyon modelleri ve bir çok simulasyonlar kullanılmaktadır. 1966'da, sıcaklık ile elektrik talebi arasındaki ilişki ilk kez Heinmann tarafından kullanıldı. 1971'de ilk elektrik talep tahmin sistemi Lijesen ve Rosing tarafından yapıldı. 1987'de Hagan ve Behr zaman serileri modelini kullandı. 90`lardan sonra daha karmaşık modeller olan yapay sinir ağları ve bulanık mantık modelleri kullanılmaya başlandı. 1991'de Part ilk kez kısa dönemli elektrik talep tahminini yapay sinir ağları yöntemi kullanarak yaptı. Srinivasan 1995`de bulanık mantık yöntemiyle elektrik talebini tahmin etmeyi başardı. Bu yıllardan günümüze kullanılan yöntemler daha karmaşık hale gelmiştir ancak günümüzde hala hangi yöntemin daha iyi olduğu tartışılmaktadır.

Bu çalışmada kısa ve uzun dönem yük tahmini uygulaması yapıldı. Pratik olarak kısa dönemde yapay sinir ağları ve ARIMA modelleri karşılaştırmalı olarak uygulandı ve analiz edildi. Uzun dönemde ise regresyon analizi ve yapay sinir ağları modelleri kalibre edilerek kullanıldı ve sonuç olarak ARIMA elektrik talep tahmininde elektrik piyasalarında sık kullanılan yapay sinir ağları kadar gerçekçi ve uygulanabilir bir yöntem olduğunu gösterdi.

1. INTRODUCTION

Economy of “Emerging Turkey“ has been developed remarkably and also has great potential about new investments and trading opportunities for near future. In addition, all subjects in Turkey become more and more liberal as the government has privatized many corporation and dams with application of “sustainable growth strategy of economy”. In this context, it can be seen clearly electricity is the most affected industry by privatizations. In 1984, there was only one public corporation “TEK” (Turkish Electric Co.) which was in charge of electricity industry. In 2001 after the new laws, there are two main public corporation in industry; EPK and EPDK. Responsibility of EPDK is regulating the market and balancing fair competition (like BDDK in banking sector in Turkey.)[1] With development of the industry and rising of the competition, management of the companies has become harder. The strategic and financial plans and forecasts are had to be made. In the electricity industry, price and demand are the most critical variable. Indeed, the first aim is the estimation of electricity price. However, the price arise intersection of supply and demand. The demand is the forecasted variable and also the supply is shaped according to result of the demand forecast; in other word, supply is the decision of manager and also demand is action of the market. Because of this case, demand forecast has vital role in making long and short term strategy.

There are many methods for electric demand forecasting. Firstly, the forecasting have to be made for different time scales inasmuch as the fundamental rule of analysis is making more simply. Shortly, electric demand forecasting is divided into three groups; short term (hourly-daily), medium term (weekly) and long term (one or two years). Various methods can be used for different time scales and this work includes short and long term forecasting methods because the methods used for medium term is usually same with short term. Generally, artificial neural network, time series methods (Arima, Garch etc.) and support vector regression methods have been used for short term electric demand forecasting but it can be found uncommon methods such as Monte Carlo simulation. On the other hand, long term electricity demand has been forecasted by using artificial neural network, regression analysis and fuzzy logic generally.

In this work, fundamental methods of electric demand forecasting are examined. And also there is different analysis by profiling load data. By the way, this work consists of five main parts except for introduction; structure of electric demand, forecasting methods for electric demand, application of electric demand forecast, conclusion and references. In section 2, we explained electric demand structure and showed building blocks. In section 3, the methods which already used for load forecasting is explained in theoretical dimension. Lastly in section 4, there are applications for long term and short term load forecasting. For long term, ANN and regression analysis and for short term ANN and ARIMA are used. All results are compared. Lastly the conclusion is given in section 5.

2. STRUCTURE OF ELECTRICITY DEMAND

The term "*electricity demand*" basically refers to any kind of electricity used to satisfy human needs which can be device or anything consume electricity. However, it is not too basic to explain electricity demand perfectly because of multi dimensionality [9]. In addition, the term "*load*" can be signified "*electricity demand*" and also no longer load will be used.

It seems load must be divided into four separate components;

L=Ln + Lw + Ls + Lr,

Where;

L: total load;

Ln: normal part of load that is standardized load profile for days and years. Indeed, normal part includes long term variables;

Lw: weather- oriented part of load, clearly seasonal;

Ls: special event part of load, unusual or special events which causes significant difference from usual;

Lr: random part of load is an unexplained component usually represented as zero mean white noise.

As it is known, where the demand intersects with the supply, the price is born so firstly the demand affects price [5]. However, in the competitive electricity market load may be influenced by price change and also electricity price has high variations as compared to other commodities. All of these affect demonstrates load function,

L=f(day, weather, special, random);

Where f is a highly nonlinear function that is difficult to define and hard to forecast with classical forecasting methods [5].

On the other hand, when we look at data of temperature which is main effect on load, it is too easy to see seasonality. In a year there are 4 different seasons and 2 shapes. In a season there are 3 different month and 3 shapes. And also months, weeks, days and hours are unique in local but all is seasonal. This case affects load data`s structure and make data seasonal. It can be seen in figure 2.1.

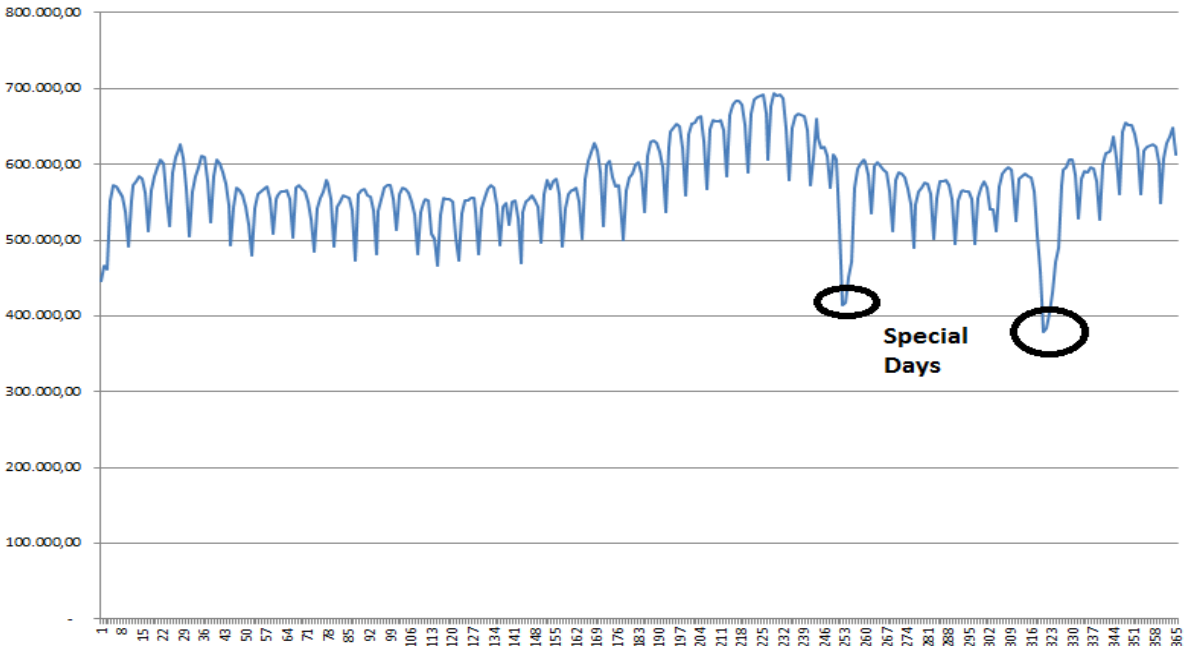

Figure 2.1 2011 Daily Load Data

For long term forecasting, structure is the same generally but Ln is main variable. Building of normal load is affected by GDP, population, HHI, historic load and date. Not only, all of them influence directly but also their effects appear with combination of them; so that, multiple linear regression is used for long term load forecasting.

3. FORECASTING METHODS USED FOR SHORT AND LONG TERM LOAD FORECASTING

The people always care about tomorrow from ancient times to the future. By this way, it is used many different techniques which shows what will happen. It begins with using of primitive methods and naturally bad results; however, by the time many different methods which has minimum error was born. The more science the more qualified forecast but in spite of everything there is no science or methods which can know what will happen tomorrow. If so the main aim in forecast is minimum error which means quality of forecast.

Firstly, forecasting have to be divided part by part for more effectiveness. For better forecasting, these several forecasting techniques are made three group; quantitative, qualitative and unpredictable. Shortly, if quantitative information is available it known as quantitative methods such as time series and explanatory. Similarly, if no quantitative information is available and also qualitative information is available it known as qualitative methods such as forecasting how a large increase in oil prices will affect the consumption of oil. The part 3 is unpredictable which means there is not enough information to forecast such as predicting the effects of interplanetary travel. Frankly, in this work it is used only quantitative methods and it continues only quantitative methods because it is available unbounded quantative data set for forecasting. Quantitative methods can be applied when three conditions exist:

1. Quantitative information about the past is available.
2. This information can be quantified in the form of numerical data.
3. It can be assumed that some aspects of the past pattern will continue into the future.

The last condition is known as the assumption of continuity; it is an underlying premise of all quantitative and many qualitative forecasting methods, no matter how complex they can be[8].

Load is multi-dimension variable so load forecasting methods includes deep mathematic and modeling. In last years, fuzzy logic and artificial intelligence are added the last generation methods for load forecasting. Shortly, it is examined in three general forecasting methods for long term and short term:

1. Time series methods for short and long term forecasting (ARIMA, ARMA, GARCH etc.).
2. Artifical intelligence and fuzzy logic for generally short term forecasting (ANN,Fuzzy Logic, Genetic Algorith etc.).
3. Regressional methods for generally long term and rarely short term(Multiple Linear Regression and Support Vector Regression)

In this groups, short term is hourly and also long term is 1-2 yearly. Finally, in this section only important load forecasting techniques which have minimum error, are examined

3.1 Artifical Neural Network

ANN is usually formed from many hundreds or thousands of simple processing units, connected in parallel and feeding forward in several layers. Because of the fast and inexpensive personal computers availability, the interest in ANN's has blossomed in the current digital world. The basic motive of the development of the ANN is to make the computers do what a human being cannot do

3.1.1 Benefits of ANN

- They are extremely powerful computational devices.
- Massive parallelism makes them very efficient
- They can learn and generalize from training data – so there is no need for enormous feats of
- programming.
- They are particularly fault tolerant – this is equivalent to the "graceful degradation" found in
- biological systems.
- They are very noise tolerant – so they can cope with situations where normal symbolic systems
- would have difficulty.

In principle, they can do anything a symbolic/logic system can do, and more

3.1.2 Modeling Approach

Neuron mentality is the main approach for ANN which is made from highly interconnected processing elements, called neurons as brain`s. Mathematical model is determined as

$$Qj = fj(\sum_{k=1}^{n} WjkXk) \qquad (j,k=1,2..s),$$

where

s: stands for the number of neurons;

Qj : is the output of a neuron;

fj : is a transfer function, which is differentiable and nondecreasing, usually represented using

a sigmoid function, such as a logistic sigmoid, tangent sigmoid, etc.;

Wjk: is an adjustable weight that represents the connection strength;

Xk: is the input of a neuron

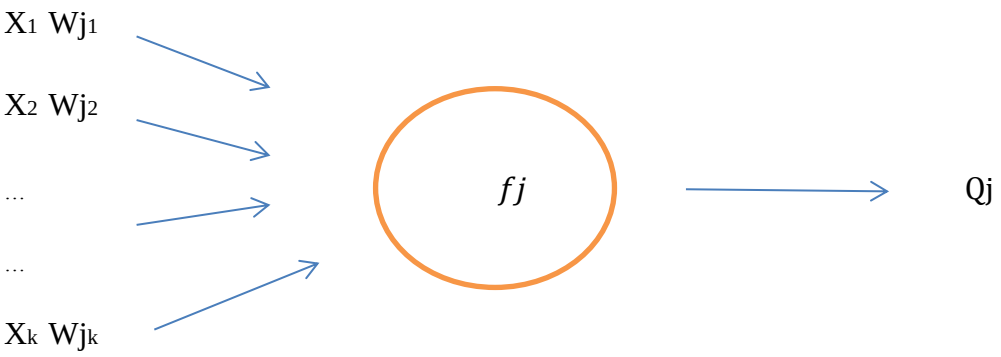

Figure 3.1 Neuron Structure

3.1.3 Network Structure

The three-layer fully connected feed-forward neural network which is generally used for load forecasting. It comprises of an input layer, one hidden layer and an output layer. Signal system is allowed only from the input layer to the hidden layer and from the hidden layer to the output layer. Input variables come from historical data, which are date, hour of the day, past system load, temperature and humidity, corresponding to the factors that affect the load. The outputs are the forecasting results, which in our case

are m = 24, i.e., one for each hour of the day. The number of inputs, the number of hidden nodes, transfer functions, scaling schemes, and training methods affect the forecasting performance and, hence, need to be chosen carefully.

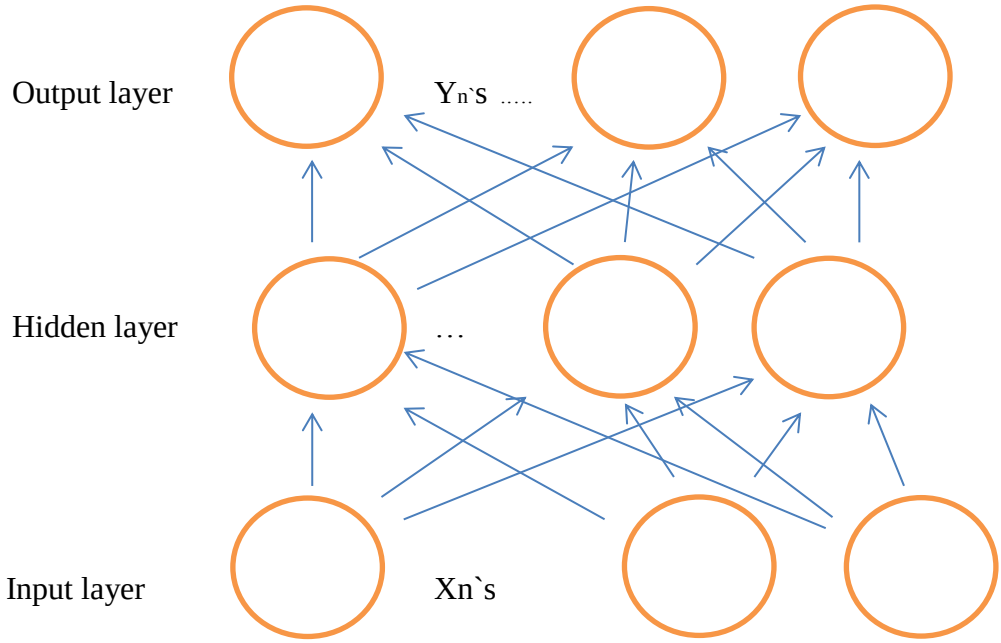

Figure 3.2 Structure of ANN

3.1.4 Training

ANN training basically consists on determining the network parameters such as weights and others, that allow achieving the desired objective based on the available training sets. Supervised manner is generally used in multi-layer feed-forward neural networks' training. As a training method, back-propagation is employed here Back-propagation is an iterative procedure that has three steps during each iteration:

- *Forward:* It is the process of output calculation for given inputs.
- *Backward:* The errors at the output layer are propagated backwards toward the input layer, with the partial derivatives of the performance with respect to the weights and biases calculated in each layer.
- *Weight adjustment:* A multivariate nonlinear numeric optimization algorithm finds the weights that minimize the error based on the gradient.

Training stops when the performance has been minimized to the goal, the performance gradient falls below a minimum gradient, the maximum number of epoch is reached, or the maximum amount of time has been exceeded. The error function, which is necessary in the back-propagation training, is the sum-squared error (SSE) and described as follows

$$E = 1/2 \sum_{p=1}^{7} \sum_{j=1}^{n} (tpj - Opj)^2$$

where tpj and Opj are the target output and the actual output j for input pattern p, respectively.

The learning function used in the training process is a gradient descent with momentum weight / bias function, which enables to compute the weight change for a given neuron. It can be expressed as:

$$dW = m.dWprev + (1 - m).lr.gW$$

where $dWprev$ is the previous weight change, gW is the weight gradient with respect to the performance, lr is the learning rate, and m is the momentum. Different learning rates and momentum have a great influence on the convergence properties.

3.2 Box Jenkins Method

It is known as Box Jenkins methods which is combined application of Autoregressive and Moving Avarage methods. Box Jenkins Method has a two parts, Seasonal and unseasonal and also unseasonal is generally known as ARIMA(p,d,q) models. In this equation, p is degree of autoregression model, d is differantial and q is degree of moving avarage model. Seasonal one is knowns as ARIMA(p,d,q)(P,D,Q) where P is degree of seasonal autoregression model, D is seasonal differrantial and Q is degree of seasonal moving avarage model [7].

P,D,Q degrees

$$P(B\hat{}s\ )\Delta_s^d\ y = Q\ (B\hat{}s\ )e$$

are determined. Δ_s is seasonal differancial processor for monthly S=12.The aim of using Δ_s^d is making stable equation.

Generally in Box Jenkins Models, forecasting rely on average value of historic data, real value of variables and random shocks. This model is using an approach for determine which equation is suitable for data which is used for forecasting. Determining, forecasting of parameters, test of suitibility and forecasting for present. If the determined equation is not enough, process continue with using a auxilary model to develop until find a solution that is satisfier. For good understanding of ARIMA, AR and MA are must be known so that it is determined in follow shortly [3].

3.2.1 Autoregressive Model AR(p)

Many observed time series exhibit serial autocorrelation; that is, linear association between lagged observations. This suggests past observations might predict current observations. The autoregressive (AR) process models the conditional mean of y_t as a function of past observations, $y_{t-1}, y_{t-2} \ldots y_{t-p}$. An AR process that depends on p past observations is called an AR model of degree p, denoted by AR(p).

The form of the AR(p) model in Econometric;

$y_t = c + \psi_1 y_{t-1} + \ldots + \psi_p y_{t-p} + \beta_t,$

where β_t is an uncorrelated innovation process with mean zero.

In lag operator polynomial notation, $L^i y_t = y_{t-i}$. Define the degree p AR lag operator polynomial $\psi(L) = (1 - \psi_1 L - \ldots \beta_p L^p)$. You can write the AR(p) model as

$\psi(L)\, y_t = c + \beta_t.$

3.2.2 Moving Average Model MA(q)

The moving average (MA) model captures serial autocorrelation in a time series y_t by expressing the conditional mean of y_t as a function of past innovations, $\beta_{t-1}, \beta_{t-2} \ldots \beta_{t-q}$. An MA model that depends on q past innovations is called an MA model of degree q, denoted by MA(q).

The form of the MA(q) model in Econometric;

$y_t = c + \psi_{t+} \beta_t \psi_1 + \ldots + \beta_t \psi_{t-q},$

where ψ_t is an uncorrelated innovation process with mean zero. For an MA process, the unconditional mean of y_t is $\mu = c$

In lag operator polynomial notation, , $L^i y_t = y_{t-i.}$ Define the degree q MA lag operator polynomial $\beta(L)=(1+ \beta_1 L+\ldots+\beta_q L^q)$.

$y_t = \mu + \beta(L)\ \psi_{t.}$

3.3 Multiple Linear Regression

Regression analysis is a modelling tecxhnique for analysing the relationship between a continuous (real-valued) dependent variable *y* and one or more independent variables $x_1, x_2, \ldots, x_k$. The goal in regression analysis is to identify a function that describes, as closely as possible, the relationship between these variables so that the value of the dependent variables can be predicted using a range of independent variables values[6].

In the multiple linear regression method, the load is found in terms of explanatory (independent) variable such as weather and other variables which influence the electrical load. The load model using this method is expressed in the form as

$$y=\beta_o + x_1\beta_1 +.. + x_\kappa\beta_\kappa +\varepsilon$$

Where *y* is the load, *x* is the affecting factors, b is regression parameters with respect to *x* and e is an error term. The error term ε has a mean value equal to
zero and constant variance. Since parameters β_κ are unknown, they should be estimated from observations of y and x_κ. Let

$$y'=b_0+b_1.x_1+\ldots+b_k x_k$$

The difference between the actual load value of y and the predicted value y' would, on average, tend toward 0, for this reason it can be assumed that the error term in equation has an average, or expected, value of 0 if the probability distributions for the dependent variable y at the various level of the independent variable are normally distributed (bell shaped). the error term can therefore be omited in calculating parameters.
Then, the least square estimates method is used to minimize the sum of squared residuals(SSE) to obtain the parameters bi

$$B=[b_0\ b_1\ ..\ b_k]^T = (X^T X)^{-1}.\ X^T.Y$$

After parameters are calculated, this model can be used for prediction. Assuming that all the independent variables have been correctly identified and therefor the standard error will be small. The standard error is obtained by the equation below:

$$SSE= \sum (y_k - y'_k)^2$$

R^2=1- SSE / SSE*

In SSE*, y'_k is not used and also y''_k is used is average value of y'_k.

Experience about the load to be modelled helps an initial identification of the candidate of explanatory (independent) variables. To determine the significance of regression parameters, the F statistical test is performed and to determine the significance of each of these coefficients; is performed by calculating t ratios. A goodness of fit measurement is represented by the R2 statistic which ranges from 0 to 1 and indicates the proportion of the total variation in the dependent variable Y around its average that is counted for by the independent variable in the estimated regression function. The closer the R2 statistic to the value 1, the better the estimated regression function fits the data.

4 APPLICATION

In this part, application of forecasting is demonstrated and also there are three main parts: data analysis, short term load forecasting (STLF) and long term load forecasting (LTLF). The first part is about structure of Turkey`s load data and results of analysis. In the second part, results of STLF with ARIMA and ANN are showed and in the last one, LTLF with using regression and ANN is demonstrated.

4.1 Data Analysis

Load data has clearly seasonality in different time intervals. The following graphs demonstrate three different seasonality; hours in a day, days in a month and days in a year.

First one, twenty four hours in a day have unique shape. This shape is relevant with working hours of companies. As it is seen, weekends especially Sundays are lower then

others. And also, the Fridays are different from each other because of Islamic effect (%98 of Turkish people choose Islam).

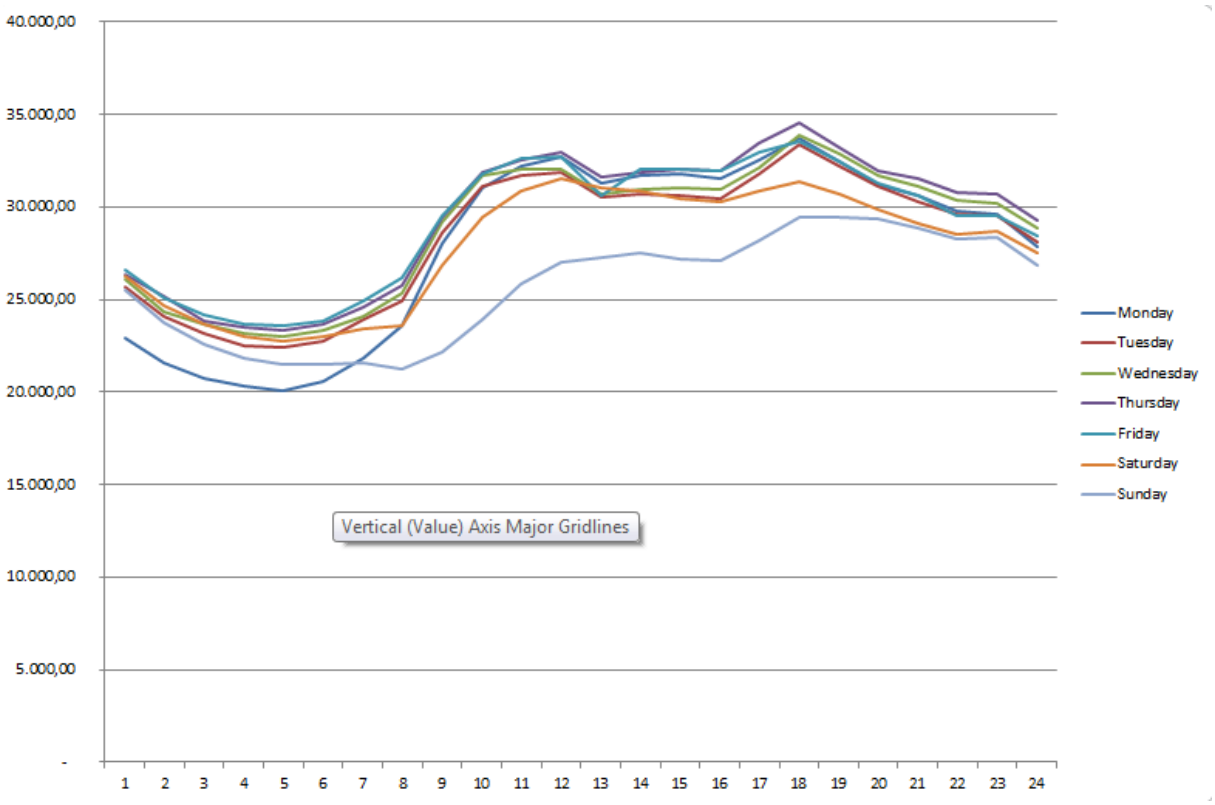

Figure 4.1 Hourly Load Graph in Seven Days of a Week in August`12

On the other hand, there are mid and long term seasonality for load data in Turkey. Following graphs show that the weeks are same shape if there is not special day. And also let look at the data generally from 2000 to 2012. Graph goes like sinx and peak points are summer and winter and also data have generally a upward trend which is effect of GDP. It can be seen clearly in 2008-2009 range there is a level effect that born from economic crisis.

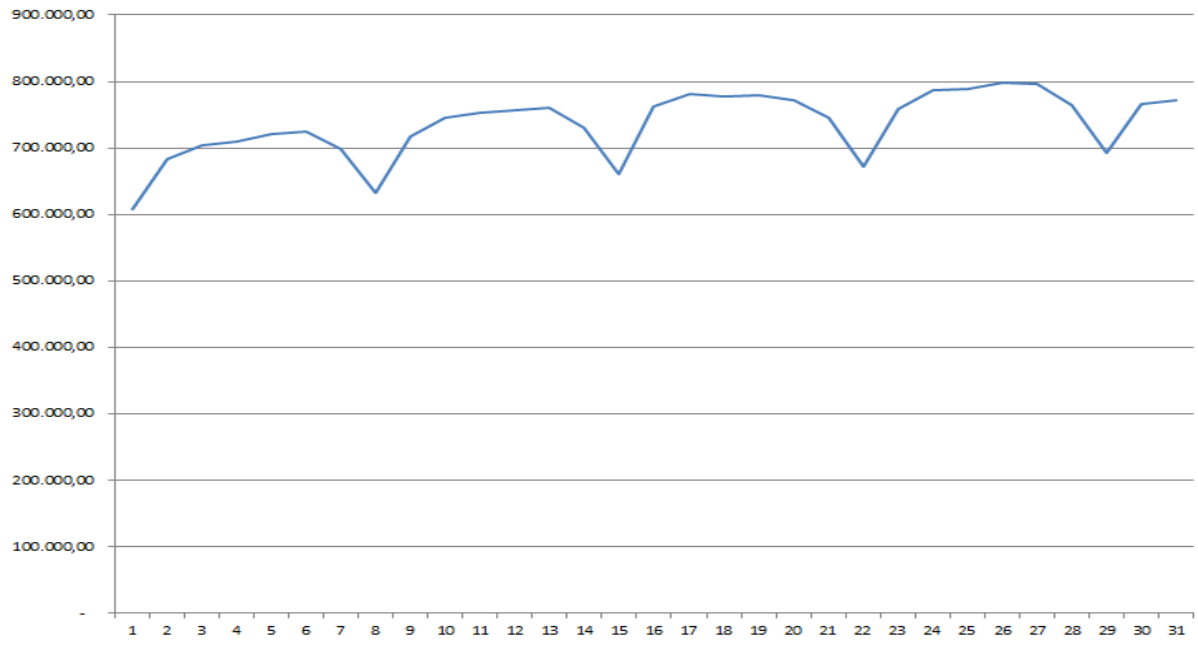

Figure 4.2 Daily Load Graph in Seven Days of a Week

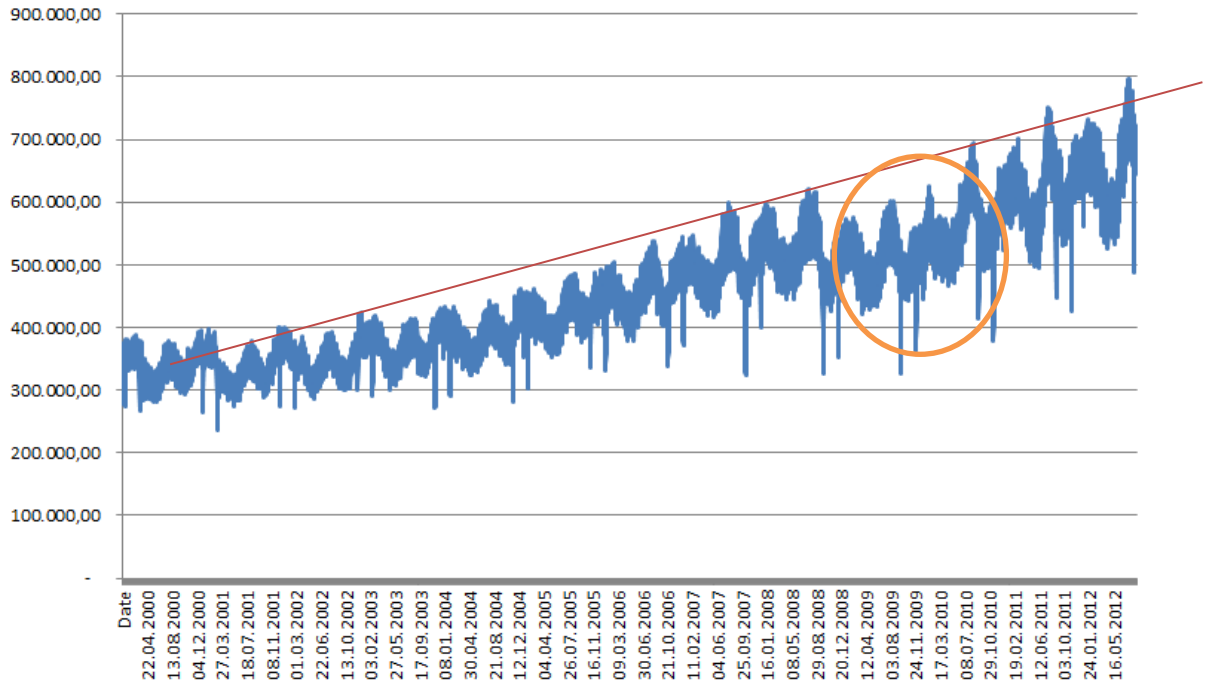

Figure 4.3 Daily Load Graph 2000 - 2012

4.2 Short Term Load Forecasting (STLF)

In this part there are ANN and ARIMA applications` results and analysis and also processes are showed and explained. In short term load forecasting process 29-30-31.08.2012`s load data is forecasted. And also ARIMA, ANN and real data are compared.

First of all, ARIMA`s using process provide quality analysis because it is used differently from all of mathematical modeler. At the beginning, in-day profile is created with all profile data equals average value of 24 hours' historic data (high dependence of near past) and calculate percentage distribution. After that, it is known every days have unique structure so that time schedule historic data is not used instead the days have same characteristic whose historic data is used for profiling (ex: if we force for Sunday forecasting only-Sunday`s historic data has to be created).

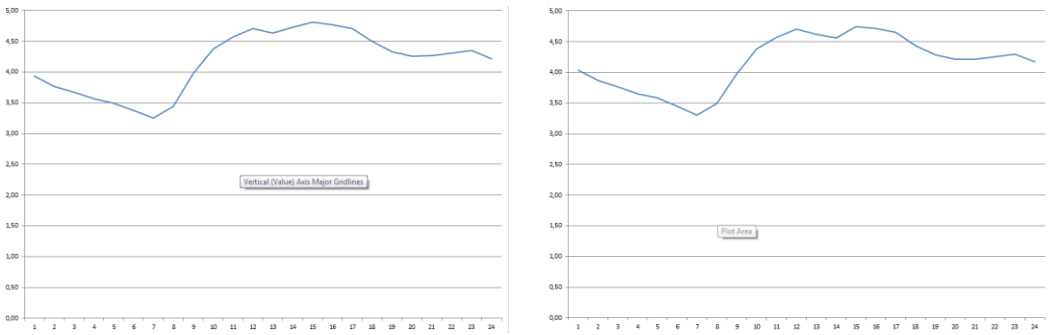

Figure 4.4 Hourly Load Profile of Wednesday-Thursday and Friday

And then ARIMA structure can be created. In this work (P,D,Q)=(1,1,0) and (SP,SD,SQ)=(2,2,2). In line with this, from graph of partial autocorrelation function (PACF) and autocorrelation function (ACF) weekly seasonality can be seen clearly.

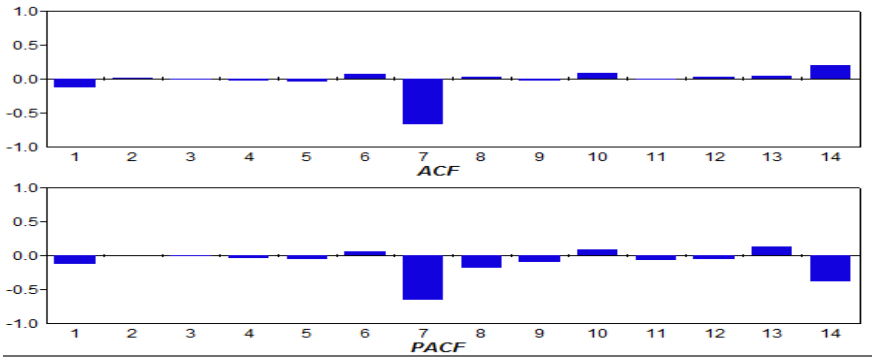

Figure 4.5 First Difference Graph of ARIMA model

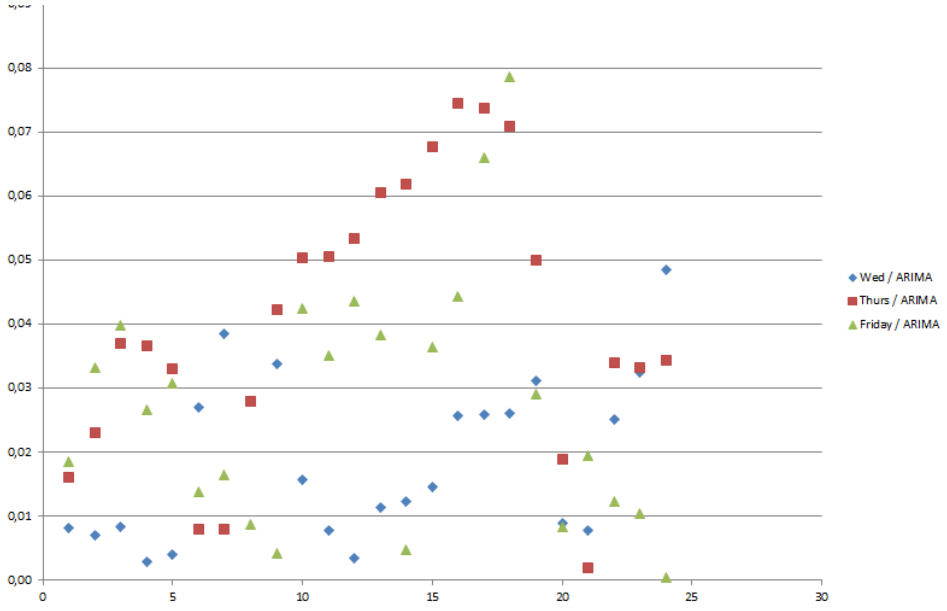

Figure 4.6 Hourly Error Graph of ARIMA Model

As a result, average daily ARIMA forecasting errors are %2 for Wednesday (29.08.12), %4 for Thursday (30.08.12) and %3 for Friday (31.08.12). And hour by hour errors are showed the graph above.

On the other hand, traditionally ANN is used for STLF and better than the others relatively but in this work main aim is making alternative solution for load forecasting. In addition, some of modelers indicate the problem is the more complex the more error. Because, the problem obtains too much dimensions to minimize error with the more complexity.

Secondly, ANN is implemented for STLF and also there are many software or toolboxes for STLF with ANN (Matlab Toolbox and other different software). Because of that only structure of process is indicated. A broad spectrum of factors affect the system`s load level such as trend effects, cyclic-time effects, and weather effects, random effects like human activities, load management and thunderstorms. Thus the load profile is dynamic in nature with temporal, seasonal and annual variations. As inputs it is taken the past 24 load and the day of the week. It is chosen for Turkey region and used the daily temperature, humidity and wind speed as input parameters.

The inputs given are:

- Hourly load demand for the full day.
- Day of the week.
- Min/Max/ Average daily temperature.
- Min/Max daily Humidity.

And the output obtained was the predicted hourly load demand for the next day. The flow chart is shown below.

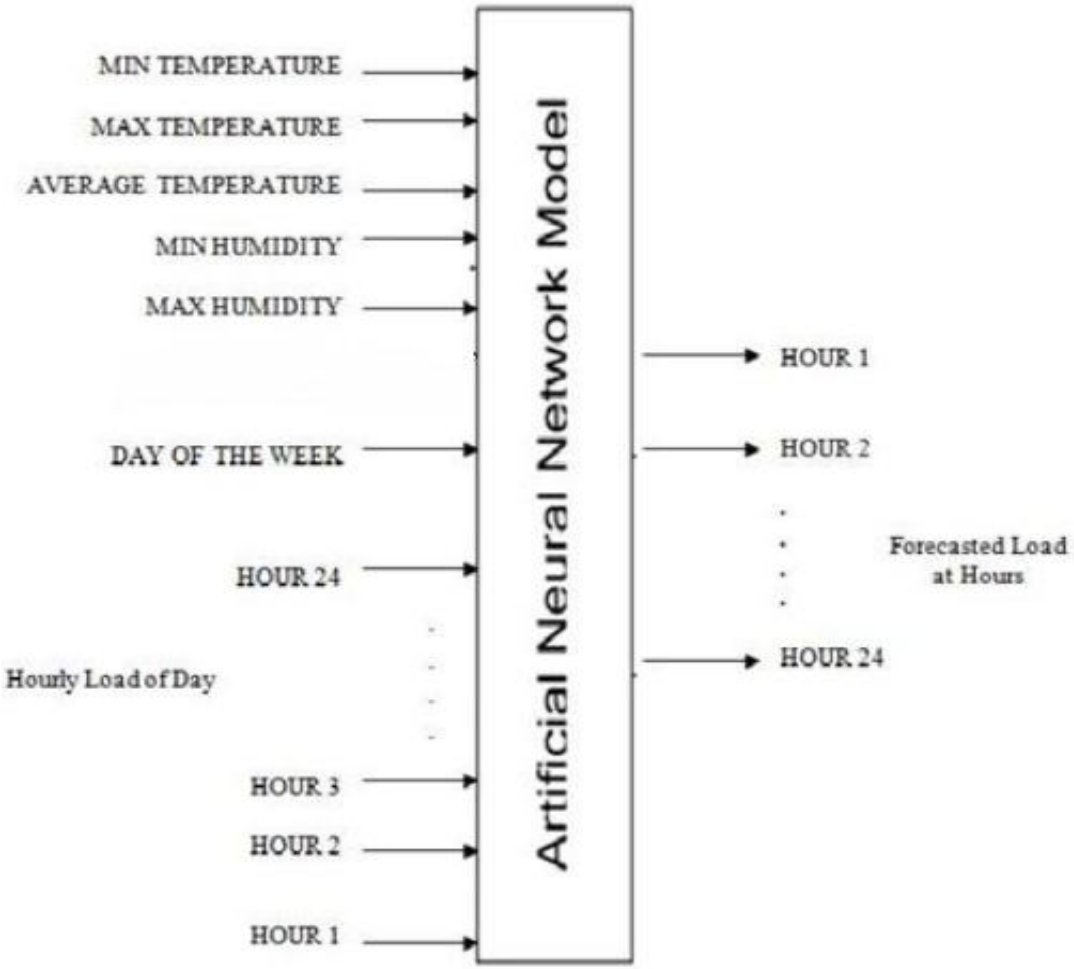

Table 4.7 ANN Model Structure

ANN STLF model is developed using MetrixND software. The hourly load, weather, and calendar data was imported to the project files developed. The hourly

load between 1/1/2004 and 12/31/2011 is divided randomly into two datasets: 75% of it for training dataset and remaining 25% for testing datasets. ANN models were developed for each hour of the day. The ANN architecture used in this study contains an input layer, an output layer and a hidden layer. Hidden layer has 3 neurons. The first neuron of the hidden layer was fed with calendar data like holidays, week of day, weekend, month, etc. The variables that represented under or over estimated months or days, such as the last months of 2008 and first 3 months of 2009 with low consumption due to the economic crisis, were also sent to this hidden layer neuron. The second and third hidden layer neurons were fed with temperature data like daily average, maximum, and minimum temperatures. Similarly, ANN models for daily energy (sum of all hourly loads) and daily peak (maximum daily load) were developed. When required, the estimated daily energy or peak was used as an input variable and fed to the second and the third neurons of the hidden layer. The continuous input data variables were scaled between 0.1 and 0.9, the remaining binary input variables were used as they are. Linear activation function was used for the first neuron of the hidden layer, and sigmoid activation function was used for the second and third neurons of the hidden layer. The activation function of the output layer was chosen as linear. The learning algorithm used for training the networks was the standard BP learning algorithm. The training was performed with 100 iterations repeated five times. The trial with the highest R2 value was selected and then the testing dataset was run by the software to assess the model prediction performance.

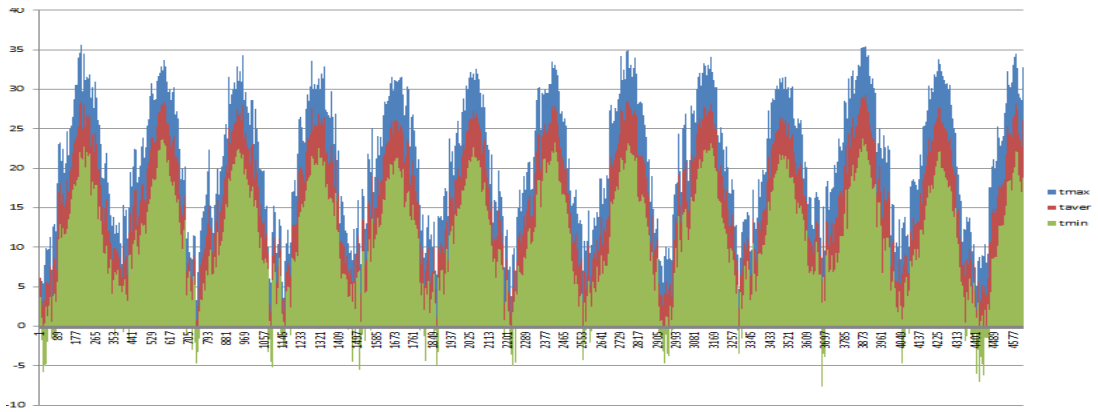

Figure 4.8 Average-Low-High Temperature Data 2000-2012

As a result, comperative analysis results indicated that not only ANN is useful but also with a good pre-analysis ARIMA can make nearly with different approach. Average daily ANN forecasting errors are %5 for Wednesday (29.08.12), %5 for Thursday

(30.08.12) and %4 for Friday (31.08.12). And hour by hour errors are showed the graph below.

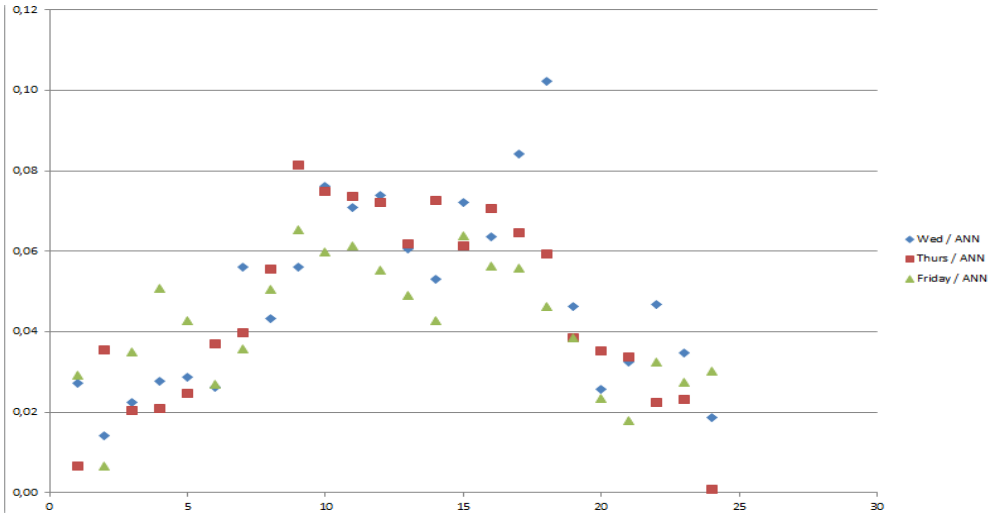

Figure 4.9 Hourly Error Graph of Forecast with ANN

Comperative graph can be show forecasting result with details.

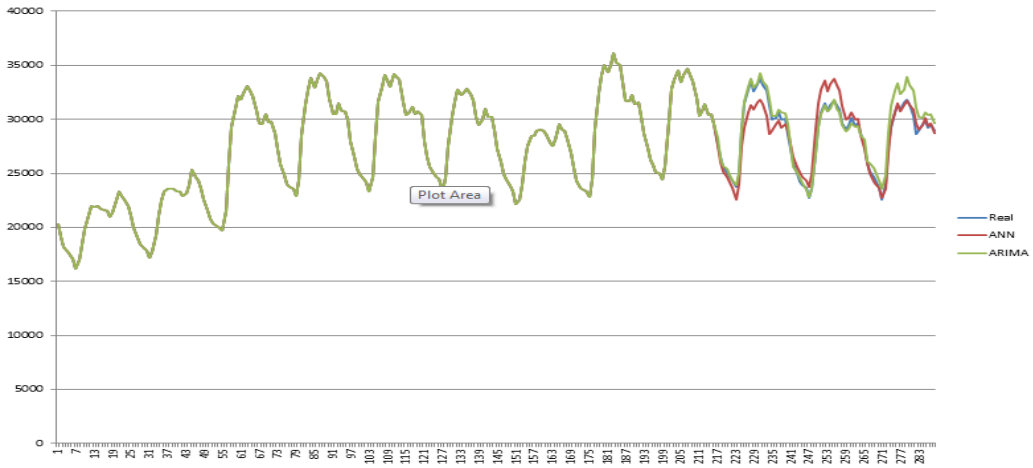

Figure 4.10 Real Load Data for 08.2012 & Forecast with ANN & Forecast with ARIMA

Consequently, ANN and ARIMA methods were applied and error analysis were showed above. This work is a example of ARIMA may be more realistic if it is applied with detailed analysis.

4.3 Long Term Load Forecasting (LTLF)

In long term load forecasting, regression analysis is the most important factor. There is an two part ın process; firstly making a profile or shape for short term (with using ARIMA or ANN) and secondly getting closed to real and adapt for long term with regression model[4].

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,919007							
R Square	0,844573							
Adjusted R Square	0,84305							
Standard Error	1,61E+18							
Observations	104							
ANOVA								
	df	*SS*	*MS*	*F*	*gnificance F*			
Regression	1	1,44E+39	1,44E+39	554,2584	5,02E-43			
Residual	102	2,65E+38	2,6E+36					
Total	103	1,71E+39						
	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95,0%*	*pper 95,0%*
Intercept	-1,2E+19	9,58E+17	-13,0003	2,25E-23	-1,4E+19	-1,1E+19	-1,4E+19	-1,1E+19
X Variable 1	1,4E+12	5,95E+10	23,54269	5,02E-43	1,28E+12	1,52E+12	1,28E+12	1,52E+12

Figure 4.11 Regression Analysis Results of Data

In the regression analysis, elasticity of Gross domestic product, household income, population and number of household are calculated and the most qualified result is this equation [12]:

Load=GDP^EL.GDP+HHI^EL.HHI+POP^EL.POP+NHH^EL.NHH+C

EL.X=elasticity of data x,

GDP= Gross domestic product,

HHI= household income,

NHH= number of household,

POP= Population.

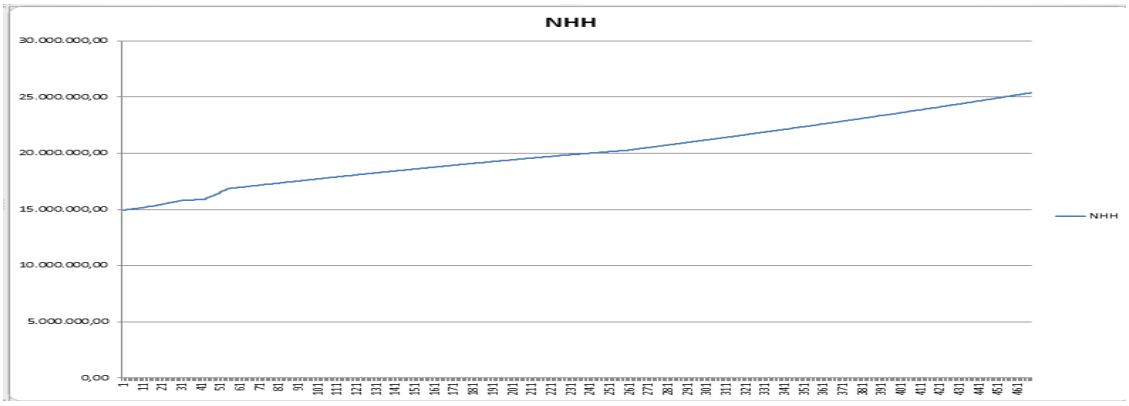

Figure 4.12 Historic with World Bank Forecasting (until 2040) of Turkey`s NHH

Figure 4.13 Historic with World Bank Forecasting (until 2040) of Turkey`s GPD

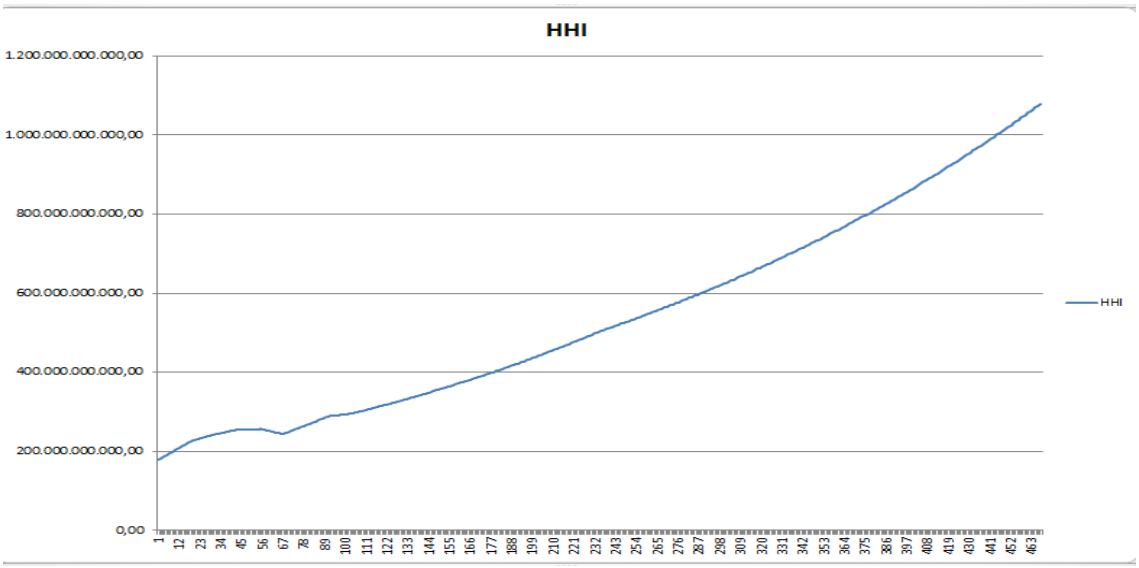

Figure 4.14 Historic with World Bank Forecasting (until 2040) of Turkey`s HHI

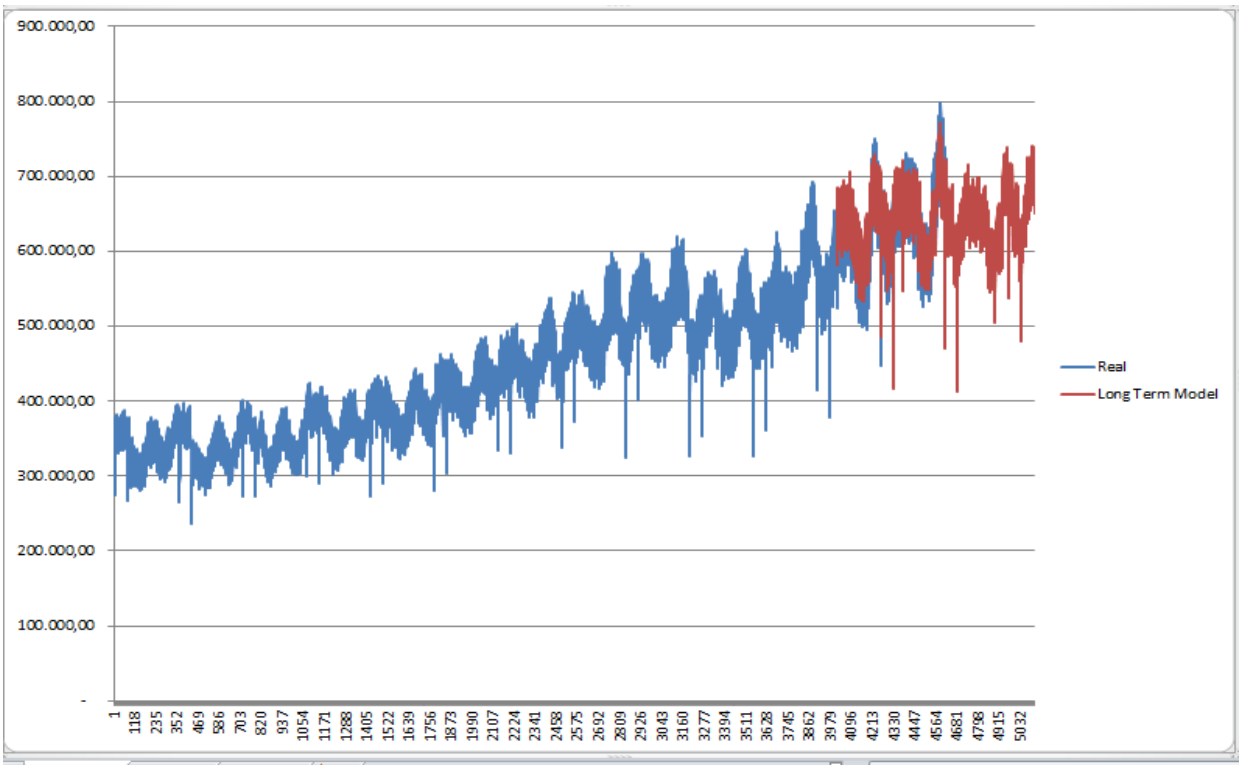

Figure 4.15 Real and Forecast Data of Turkey Load for Long Term

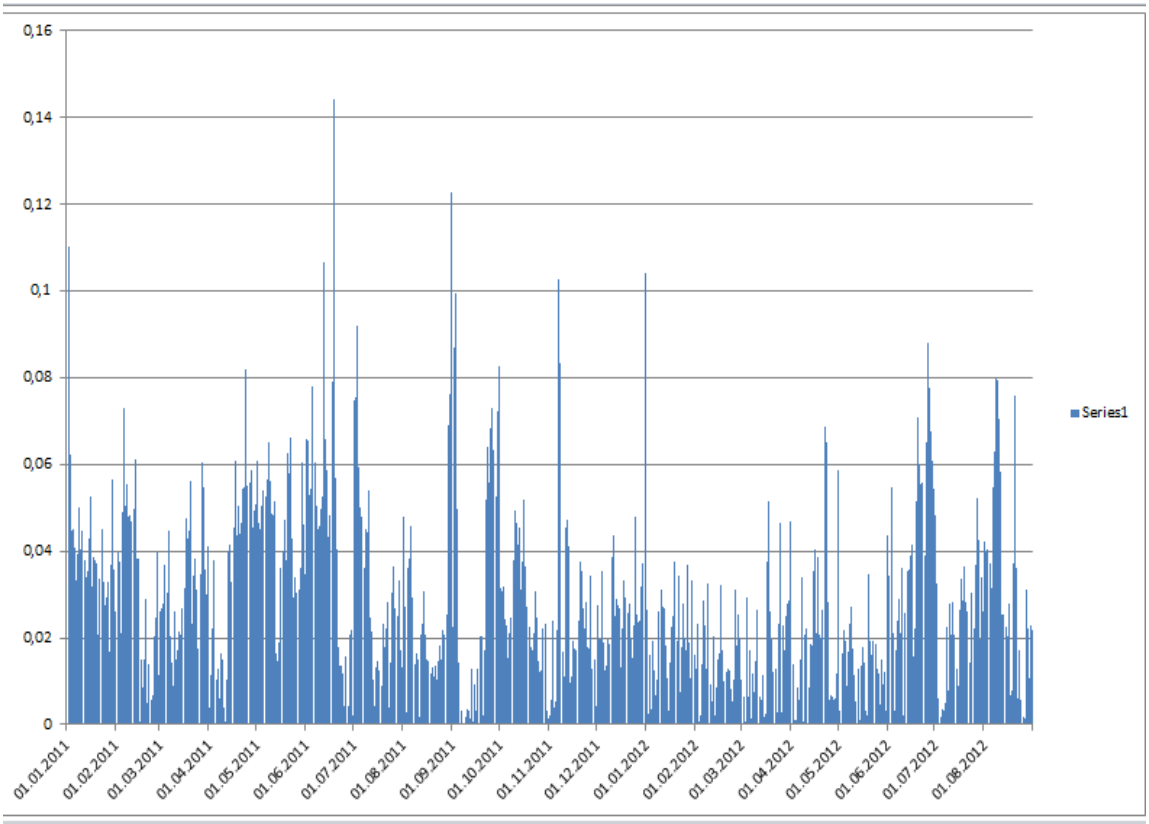

Figure 4.16 Percentage Error of LTLF

In addition after making load shape and regression model, the need is making combine of this two methods` results with calibration. Calibration was made with ND Matrix software but mantalitiy is clearly certain. Mainly, load shape means characteristic of load data without seasonality with ANN or ARIMA; however, load data continues on same way without regression. Regression analysis affects forecasted load shape and makes long term load forecasting structure. Lastly, daily error graph is above and average error is %2,9. As known, it is a really good and optimistic result for long term forecasting. Because, it is expected that error of STLF is generally less then LTLF. In addition, LTLF is vital role for electricity trading contracts as strategies are made according to mid and long term forecasting. And also ANN or ARIMA with regression model is ideal method for LTLF for “now”.

5 CONCLUSION

The players in developped energy markets in the world need quality forecasts and reports for making best strategy; correspondingly, new forecasting methods can rise from day to day with developing software technology which brings coding for applied mathematics and other different simulations. It means that forecasting is getting more and more easy with coding so that ARIMA and ANN can be easily applied by softwares but vital point in this process is thinking quality and modeling capability. In this work ARIMA and ANN applied comperatively for STLF and also ANN with regression was applied for LTLF and it was demonstrated that not only ANN is one of the best answers for STLF but also ARIMA is too. Shortly, LTLF and STLF were applied and examined theorically and practically. Load data can be affected many variable such as GDP, temperature and etc.; correspondingly, according to this variables ANN was applied. On the other hand, Load data is a time series data and this variables` effect is seasonally stable and forecastable and this case make load seasonal and forecastable with ARIMA so that ARIMA was applied. For LTLF, combination of two methods were applied and all result analysis of these applications was made. As a result, Turkey Load which is getting more and more important day by day, can be forcastable with good result (good result= error less then %5).

REFERENCES

1. Ahmet Toker, O. K. (2010). (Abdel, 2010)*Türkiye Kısa Süreli Elektrik Talebinin Saatlik Olarak Tahmin Edilmesi.*

2. Ertuğrul, M. (2011). *Türkiye`de Elektrik Tüketimi Büyüme İlişkisi: Dinamik Analiz.*

3. J.V.Ringwood, D. F. (2000). *Forecasting Electricity Demand on Short, Medium and Long Time Scales Using Neural Networks.*

4. Jennifer Hinman, E. H. (2009). *Modeling And Forecasting Short-Term Electricity Load Using Regression Analysis.*

5. Karaçayır, A. (2012). *Short Term Electricity Load Forecasting With Multiple Linear Regression And Artifical Neural Network.*

6. Marin Matijas, M. V. (2011). *Supplier Short Term Load Forecasting Using Support Vector Regression And Exogenous Input.* Journal Of Electrical Engineering.

7. Spyor Makridakis, S. W. (1998). *Forecasting Methods And Applictions.* Wiley.

8. Abdel S. (2010). *Electrical Load Forecasting*

9. Sheikh K.S.(2012). *Short Term Load Forecasting with ANN.*

10. Bilgic M.(2009). *Forecasting Turkey`s Short Term Hourly Load with ANN*

Murat Kocavelioglu / Grad`13 ITU Engineering Math

Past Work: Junior Trader at Emerging Market Intrinsic Co.

05547523739 / Seyrantepe Quarter Demirtas St. 5/11

Kagithane / Istanbul – Turkey

mkocavelioglu@gmail.com

Organizations

ITU IEEE

Board of Directors / Treasurer (July 2012 to July 2013) Administrative Committee / Sponsorship Team Leader (Sept 2011 to July 2012)

IEEE GOLD

Member

May 2013 to Present

Experience

Junior Trader at Emerging Market Intrinsic Group
June 2013 - …

Intern at IS Investment and Sec.
April 2013 - May 2013 (1 month)

Part Timer Energy Trading at Dogus Energy Trading Co.
September 2012 - March 2013 (7 months)

Working with the team making an effort for trading and forecasting.
Application of Energy Demand Forecasting, Hydroptimization and Pricing and also searching for energy demand forecasting methods

Intern / Enterprise Risk Management at Dogus Holding
August 2012 - November 2012 (4 months)

Prepared sectorel (one by one; automotive, construction, real estate, tourism, energy) external risk report and assisted enterprise risk speacialist.

Market Risk Management / Part Timer at Burgan Bank
February 2012 - July 2012 (6 months)

Preparing daily VaR report with analyzing Profit&Lost and reporting backtesting to the manager. Analyzing portfolio risk of the bank and preparing the bank limit raport and reporting risk/limit/excess values.

Internship And Junior Risk Analyst at Riskturk
June 2011 - November 2011 (6 months)

Intern at Financial Risk;

Researching about Quantative Finance, Financial Eng. and Financial Maths. Practices: Options pricing, VaR Calculation models, Merket Risk, Backtesting

Technological Using: SQL(intermedate), Excel(intermedate), VBA(New), Java(Only observation)

Part Timer & Junior Risk Analyst at Market Risk Management Module (FX, Interest Rate, Value At Risk, ES, PV, VaR , Derivatives etc.) From Sep. 2011 to Nov. 2011

Web-Based Software Developer at Anatolia system
June 2011 - June 2011 (1 month)

Internship at web developing by using PHP,CSS,HTML and rarely Jscript.

According to my dialy, my working are creation contact form of a site and sticky footer, studying about opencart and wordpress, translation of amaount of code from ASP to PHP

Projects

Time Series Analysis and Data Forecasting
December 2012 to Present
Members:Murat Kocavelioglu, Semih Yildirim, Bugrahan Ayanoglu

Deterministic data modelling with using Matlab and eviews6(AR, MA, ARMA, our own models). And forecasting with ARCH model.

Short and Long Term Electricity Demand Forecasting (Grad Project)
January 2013 to May 2013
Member: Murat Kocavelioglu with Doc. Eti Mizrahi

Searching and application on electric load forecasting with ARIMA, Regression Model and ANN.

Skills & Expertise

PHP
C
Market Risk Derivatives
Financial Risk
Time Management
Microsoft Excel Microsoft SQL Server
Matlab
Eviews
Data Modeling
Forecasting

Courses

Bachelor of Science (BSc), Applied Mathematics in Istanbul Technical University
Financial Mathematics
Optimization
Time Series Analysis Operational Research Optimal Control Systems Economics
Economics and Society
International Relationship and Glob.

Education

Istanbul Technical University
Bachelor of Science (BSc), Applied Mathematics, 2008 - 2013
Activities and Societies: IEEE R8 Member, ITU IEEE and Mathematics/Computer Student Branches, Football
Player at My Own Faculty Team,

Fatih Samiha Ayverdi Anatolian High School
78, Science, 2004 - 2008
Activities and Societies: Licensed Basketball Player, Science Club member

---

Languages

English	(Full professional proficiency)
German	(Limited working proficiency)
Turkish	(Native or bilingual proficiency)

Summary

My goal is looking at my first five years and becoming determined something about my future and then. Shortly my aim is making decision about which sector i will have prefered.For my best choice, i will try working at few industry so i will taste a few working environment. After the six year in my career way, i will have had to be determined for my way and i am working for the best, always.

Specialties

HTML, CSS, PHP, MYSQL, C, Ms. Office Softwares.

Interests

Sports (Basketball, football, Tennis..) Chess, novels, computers and career researcing

---