ABSTRACT
The high growth rate of Turkish Economy in the last decade resulted in substantial increase of energy consumption. Both commercial and residential electricity consumption has increased significantly and due to this Turkey is planning to make new investments to meet this growing demand of electricity. The government is even considering building nuclear power plants which require high level of initial investments. To forecast the future demand of electricity becomes a key issue for giving this type of vital decision. In this research, we have forecasted the monthly electricity consumption of Turkey by using time series methods and artificial neural networks. The results of the methods are compared by using performance measures based on the difference between real and forecasted values. The best performing method has been proposed to make predictions for the future electricity consumption.

Keywords: Autoregressive Integrated Moving Average; Turkish electricity consumption; Time series analysis, Holt-Winter's Method, ARIMA, ANN.

1 INTRODUCTION
Modeling electric energy consumption is helpful in planning generation and distribution by power utilities [1]. Since the early 1970s, various estimation methods are used in various studies of energy demand. In most of these studies, the purpose is to measure the impact of economic activity and energy prices on energy demand [7].

In most cases, energy demand studies have adopted two different types of modeling; namely, “reduced form model” and “structural form model”. The former is a double-log linear demand model under which energy demand is assumed to be a direct linear function of energy price and real income. Kouris (1981), Drollas (1984) and Stewart (1991) developed this model in their studies [23, 24, 25]. The second model is a disaggregated demand model that separates energy demand into several number of demand equations and treats it as an indirect function of energy price and real income. Pindyck (1979) studied the second model [7].

There are many studies that concerns about the validity of the fixed coefficients assumption that uses constant elasticities for the entire sample period under study in the electricity demand equation employed by these methods. This characteristic of the model is questionable in view of the changes that could have taken place in the economy over such a long period of time affecting the demand for electricity [7].

Therefore, it is argued that if data is collected over a relatively long time period to estimate an electricity demand function, the possibility that the parameters in the regression may not be constant should be considered [7].

Previous methods make use of time series data to estimate energy demand but they do not analyze the data to set up its properties and thus they completely assume the data to be stationary. Engle et. al, (1989) [18] analyzed time series properties and estimating elasticity’s based on co integration and Error Correction Method (ECM), which enables full analysis of the properties of the relevant data before actual estimation and commonly used is the Augmented Dickey–Fuller (ADF) test. Subsequent improvements related to
this approach have been in the form of insertion of more specific energy-related variables in the model and the development of new methods to identify co integrating relationships, using the Autoregressive Distributed Lag Model (ARDL) and the Johansen Maximum Likelihood Model (JML) are especially popular [7].

2 FORECASTING ENERGY CONSUMPTION

There are many approaches include statistical models, neural networks, fuzzy logic and expert systems. A number of them focused on mid-term and long-term electricity demand and incorporated climate factors [3].


The studies on energy demand forecasting in Turkey dates back to 1960s. The tradition of energy forecasting by using simple regression techniques was initiated by the State Planning Organization (SPO). Similar studies later have been continued by the Ministry of Energy and Natural Resources of Turkey (MENR) and a number of academicians. These early forecasts consistently predicted much higher values than the consumptions that actually realized. Later, starting from 1984, several econometric modeling techniques have been employed. Among them, the Model for Analysis of Energy Demand (MAED), which was used in energy planning and policy making by MENR, has been the most commonly applied one [22].

Energy consumption in Turkey also shows a tendency to increase rapidly in parallel with the developments throughout the world. The changes that occurred in socioeconomic balances especially after 1980 have a significant contribution to this tendency [22].

3 METHODOLOGY

3.1 The Box-Jenkins (ARIMA) Methodology

Autoregressive integrated moving average modeling (ARIMA) was greatly improved by G.E.P. Box and G.M. Jenkins. They have contributed a great deal of effort to identify, fit and check the model validity [8]. Box-Jenkins methodology is different from the other forecasting methods as it does not assume any regular pattern in the history of the data series that will be forecasted. It heavily relies on its own past while describing the series and it’s an appropriate method for medium to long length time series data [4]. It iteratively improves the model until residuals are small, stochastically distributed and show no regular pattern [7][8].

The ARIMA model is inherited from the ARIMA(p,q) model which has the form as,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t$$

where $Y_t$ is the response (dependent) variable at time $t$, $Y_{t-1}$, $Y_{t-2}$, ..., $Y_{t-p}$ are response variables at time lags $t-1$, $t-2$, ..., $t-p$, $\phi$’s are autoregressive parameters to be estimated, $\theta$’s are the moving average parameters to be estimated, $\epsilon$’s are random errors generally assumed to be independent, identically distributed variables sampled from a normal distribution with mean zero. Forecasts of the model rely on current and past values of the response $Y$ and current and past values of the errors $\epsilon$. The ARIMA (p,d,q) model is the combination of ARMA (p,d) and d times differencing applied to it. The notation of SARIMA model is represented as ARIMA(p,d,q) (P,D,Q)s where p is autoregressive order, q is moving average order and d is the times of differencing, P is the number of seasonal autoregressive (SAR) terms, D is the number of seasonal differences, Q is the number of seasonal moving average (SMA) terms and s is the periodicity of the season. Generally, the ARIMA(p,d,q)(P,D,Q)s equation is represented as follows [12].

$$\Phi_p(B) \Phi_d(B) \Phi_s(B^s) (1- B)^d Y_t = \delta + \Theta_q(B) \Theta_s(B^s) \epsilon_t$$

Where

$$\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$$

$$\Phi_d(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_d B^d)$$

$$\Phi_s(B^s) = (1 - \phi_1 B^s - \phi_2 B^{2s} - \ldots - \phi_s B^{ss})$$

$$\Theta_q(B) \Theta_s(B^s) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$$

are the nonseasonal and seasonal autoregressive and moving average parameters to be estimated, $\epsilon_t$ is the error term, $\delta$ is a constant term, $\phi$’s, $\theta$’s, $\Phi$’s, $\Theta$’s, $\epsilon$’s are random shocks that are assumed to be independent of each other. For abbreviation purpose in this model, a backshift operator ($B$) is used to represent the time series observation backward in time by k period and symbolized as $B^k$, meaning that $B^k Y_t = Y_{t-k}$. 

118
The term \( \alpha \) represents the d times nonseasonal difference while the term \( \gamma \) represents the D times seasonal difference.

The Box-Jenkins method is iteratively follows the identification, estimation and model checking steps until a satisfactory model is found and after that the forecasting is made. First, in order to identify the model correctly, the series data must be stationary, that is, the series must vary about a fixed level, and inconstant variance, trend and seasonality must be removed from the series data. In order to convert non stationary series to stationary series some modifications are applied to the series data. The inconstant variance can be removed from the data by applying methods such as the natural logarithm, square root or dividing the data to its moving standard deviation. Seasonality is corrected by taking the average of each seasonal/cyclic variables and subtracting it from the corresponding seasonal data. The trend in the non stationary data is often removed by differencing. The differencing can be done iteratively until data become stationary, but over differencing may introduce dependence while none exists.

After obtaining a stationary series data, the form of model must be identified. This is accomplished by matching the autocorrelation and partial autocorrelation functions of the series data with a theoretical pattern. Usually, if sample autocorrelations die out mildly and sample partial autocorrelations cut off, the model will require autoregressive terms. In case the sample autocorrelations cut off and the sample partial autocorrelations die out, the model will need moving average terms. If both sample autocorrelations and sample partial autocorrelations die out, autoregressive and moving average terms are needed. The number of significant terms in sample autocorrelations and sample partial autocorrelations determines the orders of the MA and AR parts [8]. The model identification phase ends up with determining the values of p, q, and d of ARIMA (p,d,q). If series data contains seasonality then seasonal values P, Q and D must be identified additionally. When the model is SARIMA, the seasonal components are determined according to spikes where the autocorrelation functions cut the confidence limits [6][8].

After selecting a rough model, the parameters (phi’s and theta’s) for that model must be estimated. The parameters of the ARIMA models are estimated by minimizing the sum of squares of the fitting models. The significance of the parameters is tested with t-statistics and p-value. As the t-statistic is itself not informative, it is used for the calculation of p-value for hypothesis testing [8]. The null hypothesis (H0) states that all parameters of the model are zero against the alternative hypothesis (Ha) that some parameters are different than zero. If the p-value is less than \( \alpha \), representing the level of confidence as (1- \( \alpha \)) %, then null hypothesis is rejected, meaning that there are some parameters different than zero.

Before forecasting with model, the adequacy of the model must be checked by examining the information about the residuals, which should be random and normally distributed. This can be done very easily by examining the histogram, or the normal probability plot or a time sequence plot of residuals. The residual autocorrelations should be small and within the confidence intervals at the ACF and PACF plots. The overall adequacy of the model is provided by a chi-square (\( \chi^2 \)) test based on the Ljung-Box Q statistics. If the p-value regarding the Q statistic is very small (say, p-value < \( \alpha=0.05 \)), the built model is considered as inadequate. In another word, the built model can be considered as adequate, if the significance level of Ljung-Box Q statistics is greater than \( \alpha \) [4][8].

3.2 Holt-Winter's Method

Time series data is usually plotted as series on the vertical axis against the time usually on horizontal axis. The aim is to get a visual perception of the series data components such as trend, seasonality which can help an analyst to select the appropriate model with the potential to produce the best forecasts. Time series forecasting usually assumes that a time series is composed of a pattern and some random error. It is tried to separate the pattern from the random error by understanding the trend, its long-term fluctuations, and its seasonality, caused by seasonal use and demand, exist in the pattern. Time series often exhibit seasonal behaviors and this behavior has a tendency to repeat itself every L periods [9].

Holt-Winters method is also known as triple exponential smoothing and consists of three components, namely, level, trend and seasonality that change by time. If the fluctuation at a certain time of each seasonal term is nearly constant then the seasonality is considered as additive. On the other hand, if this seasonal fluctuation is proportional to a certain factor then, the seasonality is naturally considered as multiplicative. For this reason, there are two versions of Holt-Winter’s method which take into account the seasonality.

3.2.1 Holt-Winter's Multiplicative Method

The following equations are (taken from Russell Cheng web site [10]);

\[
L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1})
\]  
(3)

\[
b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}
\]  
(4)

\[
S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s}
\]  
(5)

\[
F_{t+m} = (L_t + b_t)mS_{t-s+m}
\]  
(6)

where \( s \) is the number of periods in one cycle of seasons e.g. number of months or quarters in a year. \( L_t \) and \( b_t \) are respectively (exponentially smoothed) estimates of the level and linear trend of the series at time \( t \), whilst \( F_{t+m} \) is the m-step-ahead forecast starting from time period \( t \). For initialization, one complete cycle of data, i.e. \( s \) values, is necessary. The parameters \( \alpha, \beta, \gamma \) should be between the interval (0, 1), and can be selected by minimizing mean absolute error (MAE) and square root of mean squared error values (RMSE).

Then set
For initialization of the starting trend, \(s + k\) time periods are needed. If the series is long enough then it is a good choice to use two complete cycles for this purpose.

\[
S_k = \frac{Y_k}{L_s} \text{ where } k = 1, 2, \ldots, s
\]  

(8)

Initial seasonal indices can be computed as

\[
b_s = \frac{1}{k} \left( \frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \cdots + \frac{Y_{s+k} - Y_k}{s} \right)
\]  

(9)

3.2.1. Holt-Winter’s Additive Method

The equations are (taken from Russell Cheng web site [10]);

\[
L_t = \alpha (Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1})
\]  

(10)

\[
b_t = \beta (L_t - L_{t-s}) + (1 - \beta)b_{t-1}
\]  

(11)

\[
S_t = \gamma (Y_t - L_t) + (1 - \gamma)S_{t-s}
\]  

(12)

\[
F_{t+m} = L_t + b_tm + S_{t-s+m}
\]  

(13)

where \(s\) is the number of periods in one cycle. The parameters \(\alpha, \beta, \gamma\) should be between the interval \((0, 1)\), and can again be selected by minimizing error values.

The initial values of \(L_t\) and \(b_t\) can be computed as in the multiplicative case. The initial seasonal indices can be taken as

\[
S_k = Y_k - L_s \text{ where } k = 1, 2, \ldots, s
\]  

(14)

3.3 Artificial Neural Network

Since the emergence of artificial intelligence concept, the most controversial issue is whether machines are intelligent or not. There are still many ongoing arguments on this subject. Even though a lot of people still believe that machines cannot be intelligent, many ongoing studies on automation systems are being conducted to gain ability to make decisions by using artificial intelligence. Today, control mechanisms have features that almost don’t require any human intervention which, eventually, ends up with more efficient and faster production. Consequently, the ultimate goal is to make machines more intelligent by programming.

There are many approaches within the concept of artificial intelligence such as genetic algorithms, fuzzy logic, expert systems, artificial neural networks and intelligent agents which are programmed to solve different problems effectively. These approaches/methods can be adapted to solve each technical problem according to the characteristics of the problem. Among these approaches, artificial neural networks go to the fore by predicting the information by learning from experience. As learning methods are very vital in artificial intelligence, many studies are still being conducted to produce best learning rules. Some of these rules are Hebb’s rule, Hopfield’s rule, Delta’s rule and Kohonen’s rule among which the Hebb’s rule is the oldest and most used one.

In a neural network system model, it is supposed that the input is processed in neurons and it is transformed between neurons by signals. Each unit has its own value and these values are multiplied with signals. The sum of these values is called as net input. The output is calculated with activation function. The input of the activation function is our net input. Here is an example of a neural network, [13] as seen at Figure 1.

![Fig. 1. An example of neural network](image)

As you see in the Figure 1, the neural network consists of inputs, weights, summation function, activation function and outputs.

![Fig. 2. Multilayer perceptron [17]](image)

As learning mechanism, mostly the backward chaining is used in multilayer perception. After backward chaining is invented, neural networks are more popular as a research area [16]. This algorithm helps decreasing the errors starting with outputs and ending with inputs. That is why it is called backward chaining.
In literature, neural network was used especially for energy consuming forecasting. As an example of time series study, the interested person can investigate the study of [15]. Zhang (2003) [14] used ARIMA and artificial neural network together in his study.

4 ENERGY CONSUMPTION FORECAST IN TURKEY: 1970-2010

4.1 The ARIMA Modeling of the Energy Consumption

The statistical software package of IBM SPSS Statistics 19 and Minitab 16 are employed to develop and use the ARIMA models. This section contains how the four stages, namely identification, estimation, diagnostic checking and forecasting, of ARIMA model applied to the electric consumption data.

4.1.1 Identification

As a first step to model identification, monthly electric consumption time series data is plotted for the years 1970 to 2010. When the time series plot is examined Figure 3, the main features of the data are: (1) the overall level rises almost linearly until 1985 and after that there are exponential fluctuations. (2) the variance is not constant meaning that transformation is necessary (3) there is also seasonal pattern of monthly length 12 which can be more easily seen when the yearly data superimposed on the same graph as seen in the Figure 4.

After the application of natural logarithm transformation to the data (see Figure 5), the variance became constant but, trend and seasonality can be seen.

Fig. 3. Time Series Plot of Electric Consumption between 1970 and 2010.

Fig. 4. Seasonal Pattern of the Electric Consumption Data.

Fig. 5. Natural Logarithm Transformation of the Electric Consumption Data.

Fig. 6. Natural Logarithm and Differencing (1) Transformation of the Electric Consumption Data.

Fig. 7. ACF of Natural Logarithm and Differencing (1) Transformation of the Electric Consumption Data.
Taking the additional nonseasonal differences, it still is not meaningful about ACF and PACF plots that seem stationary. So it is applied the seasonal differencing to the data and get the Figure 9, Figure 10 and Figure 11. ACF and PACF now seem more stationary and have only a few significant coefficients at the seasonal and nonseasonal levels either they cut off or die down fairly quickly. Therefore, the final transformed series data can be considered as stationary and ARIMA modeling procedures can be implemented now. Both ACF and PACF have some significant spikes at nonseasonal, seasonal and near-seasonal lags. Even though additional seasonal differencing may help reduce these extra spikes, it is not common to use seasonal second differencing, nor do more than two total differencing (nonseasonal and seasonal). That is why; it is decided to ignore these spikes at near-seasonal and near-half-seasonal lags [1]. These significant lags also may be the reflections of the strong seasonality and often disappear at the residuals of ACF (Figure 10) and PACF (Figure 11) whose model don’t introduce the coefficients of those lags. It is also known that excessive differencing may also introduce artificial patterns that result in forecast errors with large variance [11].

The ACF in Figure 10 cuts off to zero suddenly after lag 1 at the nonseasonal level while the PACF in Figure 11 almost dies down meaning MA (1), nonseasonal moving average. Both ACF and PACF dies down at the seasonal or near seasonal lags that suggests to add SAR (1) and SMA(1) to the model. Thus the final model is the combination of them and stated as ARIMA (0,1,1)(1,1,1)12. The PACF in Figure 11 also reminds of the AR(1) pattern, that will result in ARIMA (1,1,1)(1,1,1)12 model as a whole but, the former model is simple and has slightly better residual errors than the later. Also principle of parsimony advises us to prefer simple models over complex ones.
4.1.2 Estimation

The parameters of the model ARIMA (0,1,1)(1,1,1)_{12} are estimated by using the IBM SPSS 19 package. The model fit statistics are Stationary R-squared 0.411, R-squared 0.997, RMSE 241.743, MAPE 2.719, MAE 146.693, MaxAPE 13.382, MaxAE 1331.077. 

The formulation of the ARIMA(0,1,1)(1,1,1)_{12} model corresponds to the \( \phi(B) \Phi(B^2)(1-B)^{12} (1-B)^{12} Y_t = \theta(B) \Theta(B^2) \varepsilon_t \) equation which can be expanded as in Appendix 1 and represented from to the \( \Phi(B^2)(1-B)(1-B) Y_t = \Theta(B) \Theta(B^2) \varepsilon_t \)

\[
Y_t = \epsilon_t - \theta(1) \hat{\epsilon}_{t-1} + \theta(2) \hat{\epsilon}_{t-2} + \cdots + \theta(p) \hat{\epsilon}_{t-p} + \Phi(1) Y_{t-1} - Y_{t-12} - \Phi(2) Y_{t-2} - \cdots - \Phi(q) Y_{t-q} - Y_{t-12q}
\]

In this work, it is used the natural logarithmic transformation for the electric consumption data before application of the SARIMA, so the above formulation is based on this data.

\[
Y_t = \epsilon_t - \theta(1) \hat{\epsilon}_{t-1} + \theta(2) \hat{\epsilon}_{t-2} + \cdots + \theta(p) \hat{\epsilon}_{t-p} + \Phi(1) Y_{t-1} - Y_{t-12} - \Phi(2) Y_{t-2} - \cdots - \Phi(q) Y_{t-q} - Y_{t-12q}
\]

And

\[
\hat{\epsilon}_t = Y_t - \hat{Y}_t
\]

The estimated ARIMA Model Parameters and their corresponding values are 0.540 for the MA(1), \( \theta_1 \), 0.270 for the SAR(1), \( \Phi_1 \), and 0.816 for the SMA(1), \( \Theta_1 \), while addition of 1 set of nonseasonal and seasonal differencing is applied.

4.1.3 Diagnostic checking

From plotting and looking the residual errors, they look random, normally distributed, no outliers and their mean is approximately zero. Although there are several slight amounts of autocorrelations at some lags, the overall appearance of the plots is good (see Figure 12, Figure 13 and Figure 14).

The Ljung-Box statistic is a test of the relationship between the residuals and a large value shows that the residuals are related. When the observations are completely random (i.e. form a white noise process), it is usually expected that there would be no significant autocorrelations among the observed values of the series and there will be no need to fit a model. The Ljung-Box Q Statistics null hypothesis states that the series is white noise, meaning uncorrelated. For our model, the test statistics rejects the no-autocorrelation hypothesis at a high level of significance (see Table 1, p-value=0.001). This means that the residuals are not white noise, and the proposed model is fully adequate for this series with significant autocorrelations.

Table 1. Model Statistics

<table>
<thead>
<tr>
<th>Model Fit statistics</th>
<th>Ljung-Box Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationary R-squared</td>
<td>0.411</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.997</td>
</tr>
<tr>
<td>RMSE</td>
<td>241.743</td>
</tr>
<tr>
<td>MAPE</td>
<td>2.719</td>
</tr>
<tr>
<td>MAE</td>
<td>146.693</td>
</tr>
<tr>
<td>Statistics</td>
<td>39.715</td>
</tr>
<tr>
<td>DF</td>
<td>15</td>
</tr>
<tr>
<td>Sig.</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 2: SARIMA Model Parameters

<table>
<thead>
<tr>
<th>Consumption Log</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA(1)</td>
<td>0.540</td>
<td>0.039</td>
<td>13.943</td>
<td>0.000</td>
</tr>
<tr>
<td>AR, Seasonal(1)</td>
<td>0.270</td>
<td>0.066</td>
<td>4.391</td>
<td>0.000</td>
</tr>
<tr>
<td>Seasonal Difference</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA, Seasonal(1)</td>
<td>0.816</td>
<td>0.044</td>
<td>18.665</td>
<td>0.000</td>
</tr>
</tbody>
</table>

As it can be seen at Table 2 which deals with the final estimate of parameters for ARIMA (0, 1, 1)(1,1,1)_{12}, it is observed that the p-values of MA (1), SAR (1) and SMA (1) parameters are less than 0.05 and are therefore significant.

Fig. 12. ACF of Residuals

Fig. 13. PACF of Residuals

Fig. 14. Plots of Residuals
4.1.4 Forecasting

After the diagnostic checking of the model, the forecasting is made for the next 12 months by using the constructed model and presented at the Results and Discussions part.

4.2 Holt-Winter’s Method

The optimum value of parameters (\( \alpha \), \( \beta \) and \( \gamma \)) of Holt-Winters method can be found by using the maximum likelihood estimation and minimization of the sum of the squared errors. In this study, the parameters of both the additive and multiplicative Holt-Winter’s Method are estimated by using Stata/SE 11.0 software package.

4.2.1 Additive Method

The estimated optimal weights of model parameters are 0.5804 for alpha, 0.0020 for beta and 0.2909 for gamma with accuracy measures of MAE 166.3 and RMSE 283.2378. The residuals plots of the data seem to satisfy the randomness (see Figure 15, Figure 16 and Figure 17).

4.2.2 Multiplicative Method

The estimated optimal weights of model parameters are 0.5039 for alpha, 0.0063 for beta and 0.2343 for gamma with accuracy measures of the MAE 162.26 and RMSE 273.4088. The residuals plots of the data seem to satisfy the randomness (see Figure 18, Figure 19 and Figure 20).
In this work, an electricity consumption estimation model is developed by using artificial neural network. The MLP training process is based on supervised learning algorithm. To generate ANN network, Java based open source Neuroph software is used (http://neuroph.sourceforge.net/index.html, January 2012). Neuroph is simple and functional software with Java libraries for developing ANN models. Training and test data are separately given to the software with the other necessary selected model parameters. With training data set, the ANN network learns the structure of the data, and then with test data set, the quality of the learning of the network is tested.

The consumption data from 1970 to 2009 are considered as training data and 2010 consumption data as test data. Years and months of data are selected as input while the consumption is considered as output. As it is selected the sigmoid function for activation function, normalization of the data fields between (0,1) is necessary. The normalization of the input – output data is maintained by formulation. For normalization of consumption data, DataMin value is set as 350 and DataMax value is set as 20,000. The years are normalized between 1970 and 2025 while the months are normalized dividing each month to 12. The model is constructed with; an input layer with two nodes, two hidden layers with six nodes each and an output layer with one node. Additionally a bias node is included in each layer except the output layer. The model structure created with Neuroph software can be seen in the Figure 21 below.

Backpropagation with momentum learning method is selected while creating the model. The parameters used in this study are: momentum rate 0.8, learning rate 0.5 and error rate 0.0001. The software run stops when the error rate decreased to the value that was set for training. According to the defined parameters, Neuroph software was executed and the training process was performed. The training error rates were demonstrated on the Table 4 below. In the next step, the model was verified with the test data and the learning level was measured as in Table 4.

### Table 3. Training and Test Measures

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>272.44</td>
<td>389.59</td>
</tr>
<tr>
<td>Test</td>
<td>1136.36</td>
<td>1329.27</td>
</tr>
</tbody>
</table>

During the training phase, the trend was usually easy part for ANN model to capture but, the major challenge was to overlap the fitted values with the actual values to meet the seasonality pattern needs. When the ANN model is trained normally, the fitted values showed a very similar behavior to simple linear regression and a seasonal pattern cannot be seen. However, when the model is trained up to a significant level in order to minimize the errors, the ANN model produced fitted values that match the seasonal patterns somehow. But this overtraining did not end up with good predictions as it almost removed the trend from the ANN weights. This can be easily seen at Table 3 by comparing the training measures with the test measures. There is great deal of difference between training and test measures. Even though the training seems significant, the ANN failed to produce good estimates for testing and it can be seen at the Figure 22.
5 RESULTS AND DISCUSSIONS

The real consumption values of 2011 and the forecasted values produced by each method are given in Table 4. Their mean absolute mean error values and root mean squared error values are also given. In forecasting 2011 monthly consumption values Winter’s additive method performs the best; both MAE and RMSE values are lower than other methods. SARIMA is the second best performer. ANN’s performance is quite poorer than others. This is also obvious in figure 23 the ANN forecasts are constantly lower than the real values.

We recommend to use Winter’s additive method to forecast monthly electricity consumption of Turkey. Our results have shown that this method is really successful for predicting the real consumption values. The forecasted values can be used confidently for making short term energy planning such as purchasing and selling electricity decisions.

Table 4. The Comparison of 2011 Forecasts

<table>
<thead>
<tr>
<th>Method</th>
<th>Real</th>
<th>ANN</th>
<th>Winter’s Multip.</th>
<th>Winter’s Add.</th>
<th>SARIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>16926.19</td>
<td>14704.33</td>
<td>16785.68</td>
<td>16774.95</td>
<td>16483.81</td>
<td></td>
</tr>
<tr>
<td>15392.17</td>
<td>14299.54</td>
<td>15243.35</td>
<td>15315.34</td>
<td>15052.93</td>
<td></td>
</tr>
<tr>
<td>16596.39</td>
<td>15993.00</td>
<td>16057.08</td>
<td>16174.70</td>
<td>15995.09</td>
<td></td>
</tr>
<tr>
<td>15279.88</td>
<td>13776.75</td>
<td>14851.74</td>
<td>15109.59</td>
<td>15034.08</td>
<td></td>
</tr>
<tr>
<td>15213.85</td>
<td>13684.49</td>
<td>15066.03</td>
<td>15427.42</td>
<td>15429.74</td>
<td></td>
</tr>
<tr>
<td>15636.83</td>
<td>13928.15</td>
<td>15383.88</td>
<td>15873.98</td>
<td>15870.10</td>
<td></td>
</tr>
<tr>
<td>18414.47</td>
<td>15545.35</td>
<td>17262.36</td>
<td>17832.08</td>
<td>17960.77</td>
<td></td>
</tr>
<tr>
<td>17871.84</td>
<td>16478.72</td>
<td>17535.16</td>
<td>18235.59</td>
<td>18566.25</td>
<td></td>
</tr>
<tr>
<td>16577.64</td>
<td>14964.36</td>
<td>15495.92</td>
<td>16206.44</td>
<td>16206.07</td>
<td></td>
</tr>
<tr>
<td>16378.35</td>
<td>13945.84</td>
<td>15583.17</td>
<td>16404.40</td>
<td>15910.24</td>
<td></td>
</tr>
<tr>
<td>16586.5</td>
<td>13936.61</td>
<td>16257.19</td>
<td>17192.66</td>
<td>15848.78</td>
<td></td>
</tr>
<tr>
<td>18355.08</td>
<td>13998.89</td>
<td>16081.95</td>
<td>19160.40</td>
<td>17746.06</td>
<td></td>
</tr>
<tr>
<td>MAE</td>
<td>2222.80</td>
<td>481.72</td>
<td>326.94</td>
<td>459.28</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>2375.28</td>
<td>591.29</td>
<td>397.80</td>
<td>487.23</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 23. The Comparison Plot of 2011 Predictions with 2011 Real Values

REFERENCES


Mucuk, M., Uysal, D., (2009), Turkey’s Energy Demand, Current Research Journal of Social Sciences 1(3), 123-128


Russell Cheng’s Home Page (online) http://www.personal.soton.ac.uk/che/Teaching/MATH6011/Forecasting-Chap3.htm (Access date: January 02, 2012).


APPENDIX 1

The general SARIMA model is:

\[ \phi_p(B)\Phi_P(B^p)(1-B^p)^Q(1-B)^dY_t = \delta + \theta_q(B)\Theta_Q(B^q)\epsilon_t \]

The formulation of the ARIMA(0,1,1)(1,1,1) model expressed as:

\[ \phi_1(B)\Phi_1(B^{12})(1-B^{12})^1(1-B)^1Y_t = \theta_1(B)\Theta_1(B^{12})\epsilon_t \]
\[
=> \Phi_1(B^{12})(1 - B^{12})(1 - B)Y_t = \Theta_1(B)\Theta_1(B^{12})\varepsilon_t
\]

\[
=> (1 - \Phi_1B^{12})(1 - B^{12})(1 - B)Y_t = (1 - \Theta_1B)(1 - \Theta_1B^{12})\varepsilon_t
\]

\[
=> (1 - \Phi_1B^{12})(1 - B^{12})(Y_t - Y_{t-1}) = \varepsilon_t (1 - \Theta_1B)(1 - \Theta_1B^{12})
\]

\[
=> (1 - \Phi_1B^{12})(Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}) = (\varepsilon_t - \Theta_1\varepsilon_{t-1})(1 - \Theta_1B^{12})
\]

\[
=> Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} - \Phi_1Y_{t-12} + \Phi_1Y_{t-13} + \Phi_1Y_{t-24} - \Phi_1Y_{t-25} = \varepsilon_t - \Theta_1\varepsilon_{t-13} - \Theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-13}
\]

\[
=> Y_t - (Y_{t-1} + Y_{t-12} - Y_{t-13}) + \Phi_1(Y_{t-13} - Y_{t-12} + Y_{t-24} - Y_{t-25}) = \varepsilon_t - \Theta_1\varepsilon_{t-13} - \Theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-13}
\]

\[
=> Y_t = \varepsilon_t - \Theta_1\varepsilon_{t-13} - \Theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-13} + (Y_{t-1} + Y_{t-12} - Y_{t-13}) - \Phi_1(Y_{t-13} - Y_{t-12} + Y_{t-24} - Y_{t-25})
\]