International Journal of Scientific & Engineering Research Volume 9, Issue 8, Augsut-2018
ISSN 2229-5518

Program Development For Energy Demand Forecasting And Its Application For A Site

Abdurrahman BİÇER, Ahmet YONETKEN

Abstract— Demand forecasting is defined as a prediction of the amount of electrical energy, peak power demand and number of customers made by marketing companies in order to make energy planning for the future by setting different scenarios at certain time intervals [1].Demand forecasting is analyzed in three different topics; short, medium and long term as terms. In the application stage of this study, short term demand forecasting is emphasized, and two different forecasting models are created based on the total load and the region by using artificial neural network method. Forecasting process was performed with these two different models. The results were compared and the failure rate of the models was tried to be reduced. In creation of the models, Artificial Neural Network method was used. As a result, the average failure rate of the model based on the total load was 7.04% while the average failure rate of the model based on regional load was 7.66% according to the randomly selected forecasting values for a week.

Index Terms— Demand Forecasting, Artificial Neural Networks

1 Introduction

Electrical energy is an energy source generated in various power plants, which can be transmitted to residential areas with transmission lines and can be kept under control but cannot be stored.

Electric power systems are systems which contain components such as generator, transformer, transmission lines and distribution lines within itself; from the generation stage to the customers. Due to the high installation cost of the electric power systems, the maintenance and generation costs can be kept at the lowest level and the reliability of the system can be increased by the effective usage of the system. In addition, electricity pricing can also be performed in a planned manner. In order to provide all of these, the topic of forecasting the load demand is one of the most important parameters. It is obvious that the higher the accuracy of the forecasting is, the more successful the planning would be.

Demand forecasting is performed as three terms;

1. Short Term Demand Forecasting: Weekly Forecasts
2. Medium Term Demand Forecasting: Monthly Forecasts
3. Long Term Demand Forecasting: Annual Forecasts

The most important forecasting term for the systematical operation of the electric power system is the short term. In power systems, the energy demanded by the system during the day constantly changes nonlinearly. Correspondingly, the number of generators stepping in and stepping out according to the demand constantly changes. The planning of the generators which will step in and step out is also made according to the short term load demand forecasting. This planning provides an opportunity to intervene to the load changes of the system by being carried out in a healthy way and being monitored instantly. The short term forecasting contributes greatly to the load - generation balance of the power system being ensured. Thus, excessive energy would not be given to the system or the energy given would not be insufficient. The energy generated would be consumed as it is generated. A successful planning would be made and the power system would be used in a reliable manner in terms of efficiency, reliability of the system and the reduction of costs.

It is important to examine the load characteristics thoroughly in order to execute a successful forecasting operation. Load characteristics vary depending on time and weather conditions. The wise divided the load of the system into four separate components, at any given time [2].

$$Y = Yn + Yw + Ys + Yr$$

In this formula;

Y : represents the total load on the system,

Yn : represents the normal part of the load, i.e. the generalized load data set for each day, all year round,

Yw : represents the load values changing according to the weather depending on the season of the year,

Ys : represents the load value which leads to deviation apart from the generalized load type,

International Journal of Scientific & Engineering Research Volume 9, Issue 8, Augsut-2018
ISSN 2229-5518

Yr : represents a section which arises randomly and which has not been predicted before.

There are many studies in which Artificial Neural Networks are used for the energy demand forecasting. Var and Türkay had got a failure rate below 10% on average with their short term forecasting model they have created [3]. Kaysal, Hocaoğlu and Kaysal had created three different forecasting models; two with ANN and Regression analysis methods separately and one hybrid method with two of these methods. They had achieved a failure rate of 6,67% in the model that was using ANN, 7.49% with regression analysis and 5,26% with the hybrid model [4]. Islam, Baharudin, Raza and Nallgowonden had aimed to reduce the failure rate in the hybrid model that they have created with ANN and genetic algorithms [5]. Sahay and Tripathi had got 3,4% failure rate in the short term forecasting model they have created with ANN [6]. Liu, Xu, Shi, and Wei have used the hybrid model that they have created using fuzzy logic, ANN and expert systems methods in demand forecasting as well as in frequency control and the detection of transformer failures [7]. Topallı and Erkmen had achieved a failure rate of 1,6% with the short term forecasting model they have created with ANN[8].

In this study, the load data belonging to GEDAŞ and weather data belonging to MGM are taken as basis from the beginning of 2013 until the 6th month of 2017 for a district in Urla. With this data, two different short term forecasting models - region based and total load based - had been created by using ANN in a computer program called MATLAB and inferences had been obtained by comparing the results of these two models.

2 Materials and Methods

The Artificial Neural Networks method is based on the nervous system of the human brain which is developed in the computer environment. ANN has data processing parallel to the multiple order input data. The inductive outputs are obtained by making numerical calculations of ANNs having statistical algorithms.

ANNs ground on the items mentioned below on the basis of human nerve cell operation;

1- The data processing process consists of neurons.
2- Signals are transmitted by the links between neurons.
3- Each link between neurons has a weight value.
4- The net output of each neuron is obtained by transmitting the net input through an activation function[9].

Biological Nervous System and Artificial Nerve Cell

Table 1. Comparison of Biological Nervous System and Artificial Neural Network[10]

Biological Nervous System	Artificial Neural Network
Neuron	Processor Element
Dendrite	Addition Function
Cell Body	Transfer Function
Axons	Artificial Neuron Outputs
Synapses	Weights

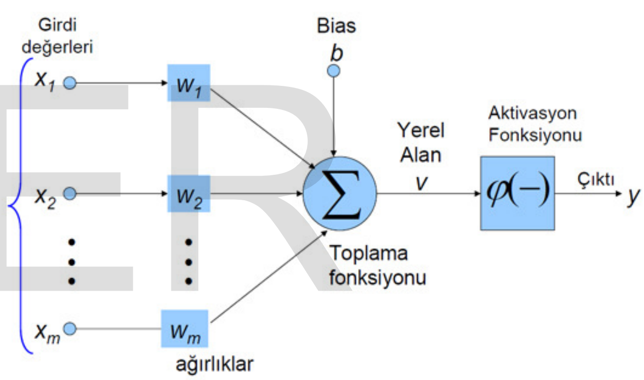

Fig..1 Artificial Nerve Cell

The x in the artificial nerve cell represents the inputs, w represents the weights, b represents the polarization value, v represents the activation potential and y represents the outputs. Inputs transmit data received from other nerves to the neural network. Weights are the coefficients that determine the neural effect of the data received through the inputs. The addition function is summed by multiplying the inputs by their own weights and the polarization value. At the end of the addition function, the data is sent to the activation function. The activation function is the last phase of the artificial nerve cell and provides a curvilinear match between the input and the output values. The most common activation functions are logarithmic sigmoid, tangential sigmoid and linear functions. Logarithmic sigmoid function is defined as a strictly increasing function that is linear and nonlinear. Tangential sigmoid function is a nonlinear and differentiable function. The linear function transmits the input value directly to the output.

International Journal of Scientific & Engineering Research, Volume 9, Issue 8, Augsut-2018
ISSN 2229-5518

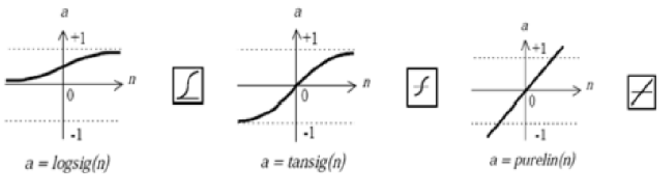

Fig. 2. Activation Functions [11]

2.1. Specifications of Artificial Neural Networks

ANN has specifications such as nonlinearity, learning, generalization, adaptability and fault tolerance. On account of learning, generalization and fault tolerance specifications, ANN can calculate and process data despite the input values that has not encountered during the learning phase.

2.2. Determination of the Network Structure

The input neuron, hidden layer, hidden neuron, and output neuron numbers should be determined for the network structure. There are no guaranteed methods in order to determine these numbers.

- The Number of Input Neurons: In the forecast problems based on cause and effect relationship, the independent variable number belonging to the problem will give the number of input neuron. However, it is not easy to determine the input number in forecasting problems based on time-series analysis. Tang and Fishwick had concluded that the input neuron number for univariate time-series analysis is equal to the Box-Jenkins AR (p) model's degree [12].
- The Number of Hidden Layers: Single hidden layer ANNs are not usually sufficient to solve the problem. Increase in the number of hidden layers can increase the accuracy of the forecast's truth value. On the contrary, the computation time increases and ANN might choose memorization instead of learning. This may cause the modelling not to be adapted to the system.
- The Number of Hidden Neurons: This number can be determined by starting with a low neuron number and increasing the hidden neuron number until the network performance tends to decrease.
- The Number of Output neurons: It is equal to the number of forecast term in the forecasting problems based on time-series analysis. If the forecast is based on cause and effect relationship, then it is equal to the number of dependent variable.

2.3. Determination of the Activation Function

It is the stage of determining which of the functions in Figure 2 is going to be used. There is no general method regarding which of these three functions is going to be chosen.

2.4. Normalization

Normalization is a very important step in terms of the speed of the network training. With this process, neuron outputs are placed on [0,1] or [-1,1] intervals.

2.5. Network Training

ANN training is a nonlinear minimization process. There are many optimization methods for training. Some of them are gradient steepest descent algorithm, Levenberg-Marquardt optimization, genetic algorithm and random search.

2.6. Arrangement of Data Sets

The available data set in the ANN forecasting process is divided into training-test or training-verification-test.Training set is used in the ANN training, the verification set is used to determine when to end the training process and the test data is used to compare ANN forecasts results.These groupings should not be randomly made in the data set. This proportioning directly affects the network performance.It is usually made as 70% - 15% - 15% or 80% - 10% - 10%.

2.7. Determination of the Performance

It is expected that the ANN model created for forecasting would be fast and accurate. The failure value in the forecast is the difference between the actual observed value and the forecasted value. Formulation as percentage is as follows;

$$MAPE = \frac{Actual\ Value - Forecasted\ Value}{Actual\ Value}\ 100 \quad (2.1)$$

3 Results and Discussion

In this study, short term energy demand forecasting models were created based on the load data belonging to GEDAŞ and weather data belonging to MGM (humidity and temperature) with the Artificial Neural Networks method by using Neural Network Toolbox in MATLAB program for a sample district of Urla county from the year 2013 to 2017. These models are two varieties - region based and total load based. Total load forecasting was endeavoured for the sample district with these two models.

3.1. Preparation of the Data Sets

The data set was prepared in Microsoft Excel. This data set was developed for the total load based forecasting model. Our data includes the periodfrom the beginning of 2013 until the

International Journal of Scientific & Engineering Research Volume 9, Issue 8, Augsut-2018
ISSN 2229-5518

end of June 2017.When all hourly values are considered to be available, 39384 data should exist. There are not 39329 data since there are values that could not be obtained due to the energy interruptions and analyser failures.35033 data were used in order to create the network and the rest were used as test data in order to calculate the forecasting's accuracy after the forecasting process.In other words, when the data set which is going to be used in network formation is visualised as a matrix; the year, month, day, hour, temperature, humidity, day specification and the total load value are found respectively on the columns of this matrix.Therefore, the number of columns will be 8. The total size of the matrix is 35033 x 8.

For the region-based forecasting model, a program is written in MATLAB in order to create each region's own data set. This is due to the fact thatdata is deficient or there are conditions in which the load value is 0. When the load is 0, it is excluded from the data set since it would affect the ANN training adversely and increase the failure rate.

In conclusion, 1 data set based on the total load and 2 data sets based on the region were created. The reason of having 2 data sets based on the region is that there are 2 TMs belonging to the sample district.

.

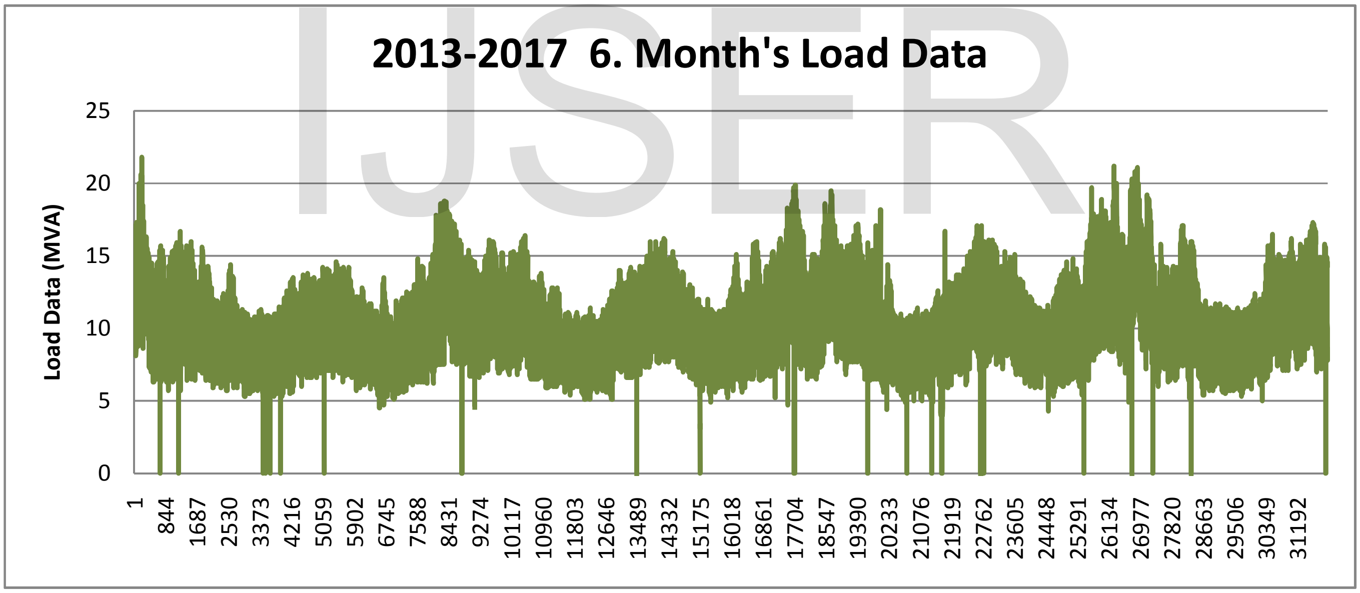

Fig. 3. Sample District's Load Data

International Journal of Scientific & Engineering Research Volume 9, Issue 8, Augsut-2018
ISSN 2229-5518

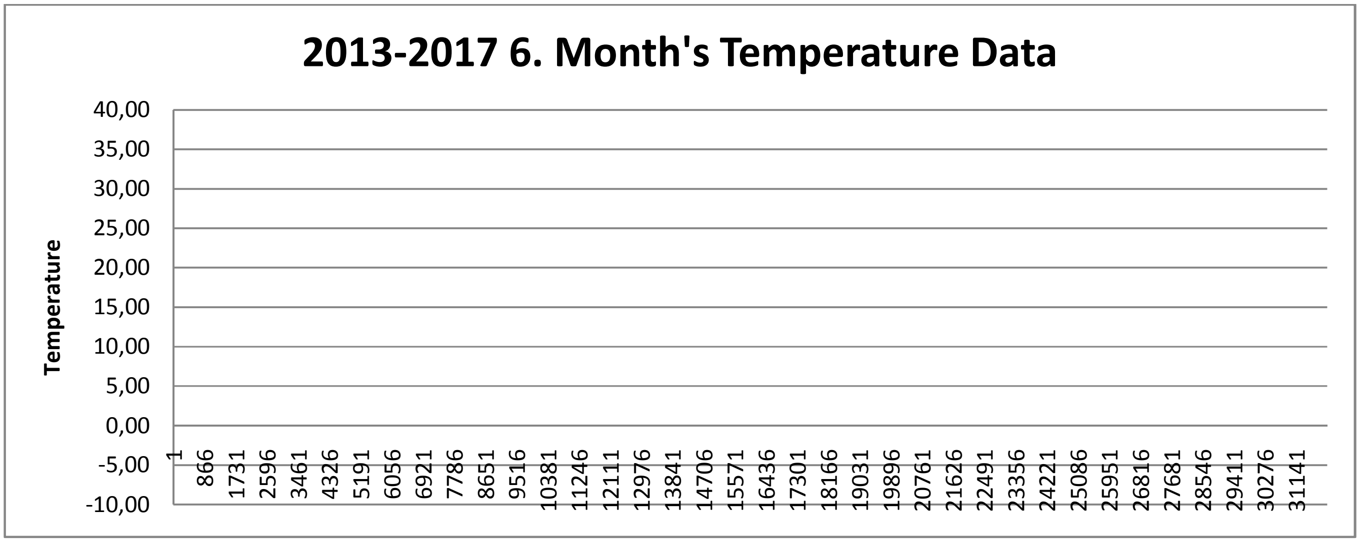

Fig.4. Sample District's Temperature Data

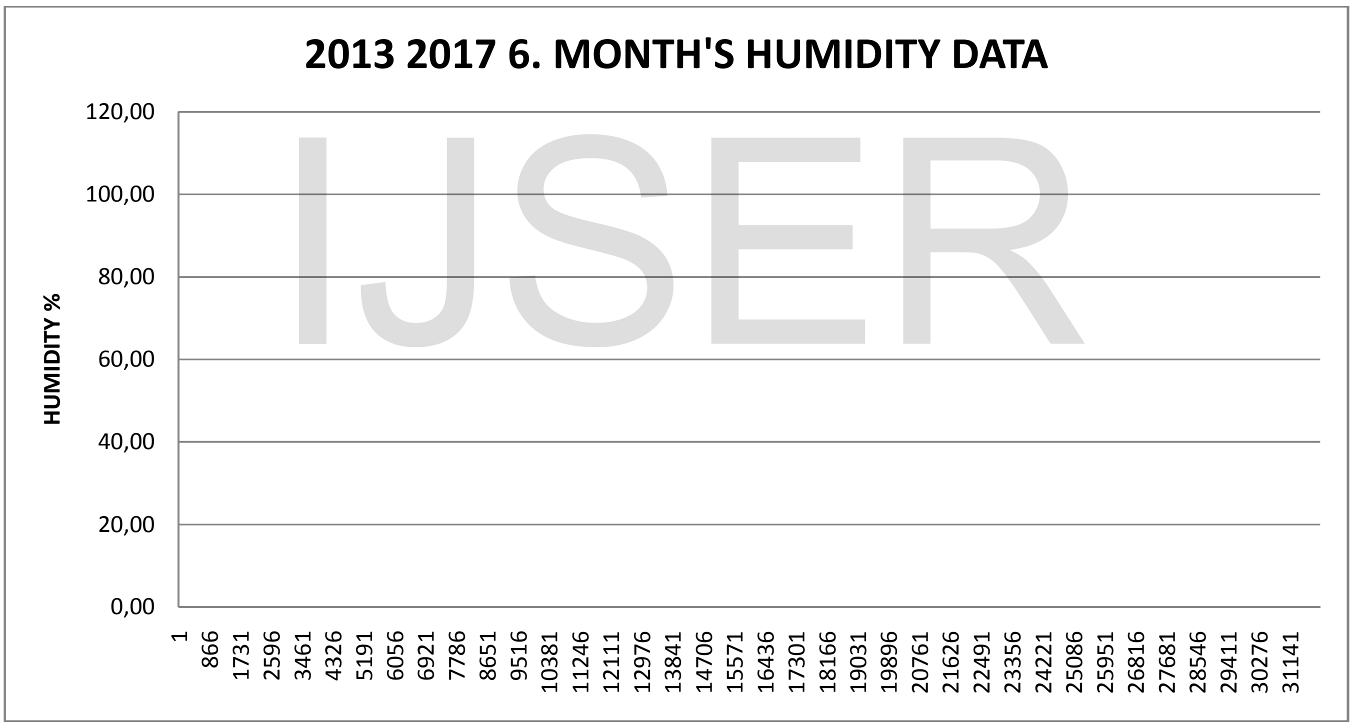

Fig. 5. Sample District's Humidity Data

HAVA SICAKLIĞI (°C)	BAĞIL NEM (%)																		
	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95
50	45	48	53	58	66	69	76	83	91	99									
49	44	47	51	55	61	66	72	79	86	94									
48	43	46	49	53	58	63	68	75	81	88	96								
47	42	45	48	51	55	60	65	70	76	83	90	98							
46	41	43	46	49	53	57	62	67	72	78	85	91	99						
45	41	43	45	48	52	56	62	65	70	76	82	88	96						
44	40	42	44	46	49	52	57	61	66	71	77	83	89	96					
43	39	40	42	44	47	50	54	58	62	67	72	77	83	90	97				
42	38	39	41	43	45	48	51	54	58	62	67	72	78	83	90	96			
41	37	38	39	41	43	45	48	51	55	59	63	67	72	78	83	89	96		
40	36	37	38	39	41	43	46	48	51	55	59	63	67	72	77	83	88	95	
39	35	36	37	38	39	41	43	46	48	51	55	58	62	67	71	76	81	87	93
38	35	35	36	37	38	40	42	44	47	50	53	56	60	64	68	73	78	83	89
37	34	34	35	36	37	38	40	42	44	46	49	52	56	59	63	67	72	76	81
36	33	33	34	34	35	36	38	39	41	43	46	48	51	55	58	62	66	70	74
35	32	32	33	33	34	35	36	37	39	41	43	45	48	50	53	57	60	64	68
34	31	31	32	32	32	33	34	35	37	38	40	42	44	46	49	52	55	58	61
33	31	31	31	31	32	32	33	34	36	37	39	40	42	45	47	49	52	55	58
32	30	30	30	30	31	31	32	33	34	35	36	38	39	41	43	45	47	50	53
31	29	29	29	29	29	30	30	31	32	33	34	35	36	38	40	41	43	45	47
30	28	28	28	28	28	29	29	30	30	31	32	33	34	35	36	38	39	41	42
29	27	27	27	27	28	28	28	28	29	30	30	31	32	32	33	34	36	37	38
28	26	26	26	27	27	27	27	27	28	28	29	29	30	30	31	32	32	33	34
27	26	26	26	26	26	27	27	27	27	28	28	28	29	29	30	30	31	31	32
26	25	25	25	26	26	26	26	26	26	27	27	27	27	27	28	28	28	28	29
25	25	25	25	25	25	26	26	26	26	26	26	26	27	27	27	27	27	27	27

Fig. 6. Influence of the Humidity Rate on Air Temperature [13]

3.2. Creation of the Interface

The user interface will be developed by means of MATLAB GUI in order to provide ease of use for the program users. Users can easily make forecast demand with this interface,

First, the "Database" interface was created to control the developed ANN and the database that is going to be used in the forecasting process. With this interface, new data can be added to the database and the ANN update can be provided. The Database Interface created is shown in Figure 7.

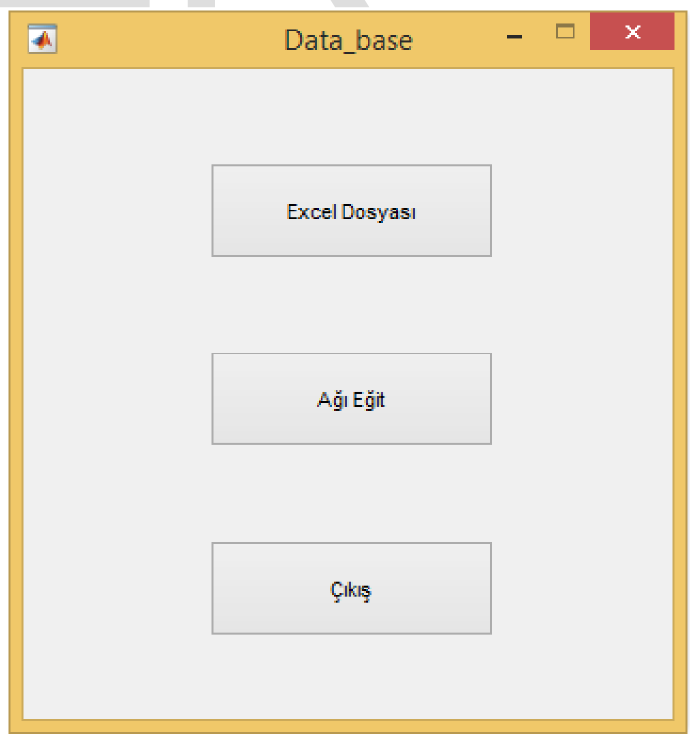

Fig.1. Data Base User Interface

International Journal of Scientific & Engineering Research, Volume ….., 9, Issue.8, Augsut-2018
ISSN 2229-5518

The Excel file that we created the data set is opened with the "Excel File" button in the Database Interface . New data can be added to the opened file. After the data is changed, the "Train Network" button is used to train the network with the new data.

After this interface, in other words, after the network training has been completed, the Load_Forecasting interface has been created in order to perform the forecasting process. Figure 8 shows the interface of the Forecasting Program. The "Enter Data" button opens the Excel file in which the temperature and the humidity values for the forecasted day will be entered."Data base" button is used to switch to the Database interface.

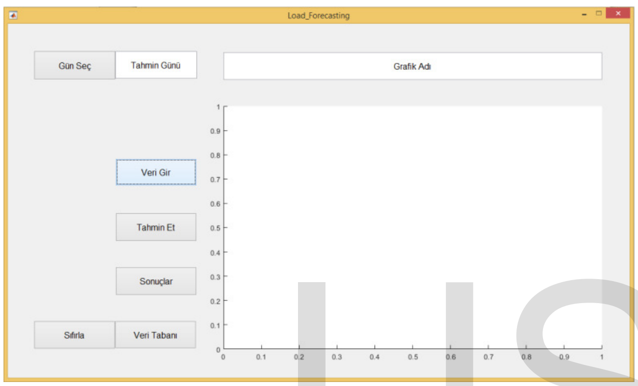

Fig. 8. User Interface for Forecasting Process

The day to be forecasted is selected from the "Select Day" button. When this button is clicked, the calendar shown in Figure 9 opens and the day which is going to be forecasted is selected and then "ok" button is clicked. After the day to be forecasted is determined, forecasting is performed by clicking the "Forecast" button. After the forecasting is performed, on the chart area in the Load_Forecasting interface, the charts of the forecast based on the two different models created and the chart of the actual values are generated in order to make comparison with the forecasted value. The "Results" button should be clicked to see the forecast results numerically. If forecasting process is going to be performed for another day, the interface is reset by clicking the "Reset" button, and the forecasting process is performed by selecting the day again.

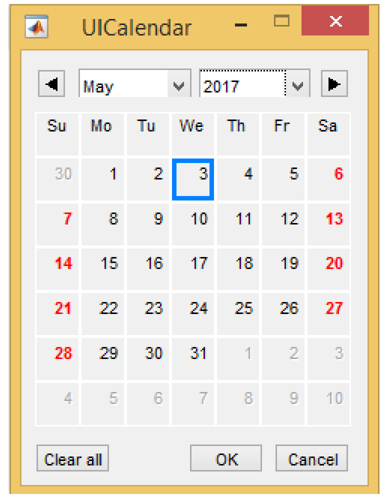

Fig. 9. The Calendar to Select the Forecast Day

3.3. Exporting Data to MATLAB

Input and output values in the database should be transferred to MATLAB for the creation and training of ANN. Firstly, all data is transferred to MATLAB as a 35033x8 matrix. Then, input and output matrices need to be separated for ANN training. The input matrix is assigned to a size of 35033x7, and the output matrix is assigned to a size of 35033x1. Since ANNs operate with columns, these matrices are transposed to be in accordance with ANN.

3.4. Normalization Process

Some operations should be performed in order to train the network and increase its performance. These operations transform input values into appropriate forms for network training. These operations are called "Normalization". The normalization process creates a suitable-formed database for the network training by being applied to the data set.

With normalization process, all the input and the output values are assigned to values within a certain range, making it easier for the network to read the values. This increases the performance of the network training by reducing its duration.

In this study, "mapminmax" function is used in MATLAB for normalization process. With this function, all of the input values are assigned to the values between -1 and 1. The "minmax" function in MATLAB is used for the normalization process of the output values. With this function, all of the output values are assigned to the values between 0,1 to 0,9.

International Journal of Scientific & Engineering Research, Volume 9, Issue 8, Augsut-2018
ISSN 2229-5518

After the network is created, the output values are converted to their actual form by applying the "Denormalization" process. The reason for the denormalization process is to compare the actual values with the network output values. Denormalization process is the reverse of the function created for normalization.

$[P_n, \sim]$ = mapminmax(P); Normalization of Input Values

mm = minmax(T_n);

min = mm(1);

max = mm(2);

$T_n = 0,8\frac{T-T_{min}}{T_{max}-T_{min}} + 0,1$ Normalization of Output Values

P : Represents the input values,

P_n : Represents the normalized input values,

T : Represents the output values,

T_n : Represents the normalized output values,

T_{min} : Represents the smallest value in the output values set,

T_{max} : Represents the largest value in the output values set.

3.5. Creation and Training of ANN

Various ANNs have been created for feed-forward as ANN feed type. Among these ANNs created, it was tried to attain the network with the highest accuracy.

The network with the highest accuracy rate was selected by experimenting with different number of neurons in the layers, various activation functions, optimization algorithms for training, and number of iterations.

The "trainlm" training function which is developed with the Leveberg-Marquardt optimization algorithm within MATLAB is used for the feed-forward network training that was created. The structure of the created network is shown in Figure 10.

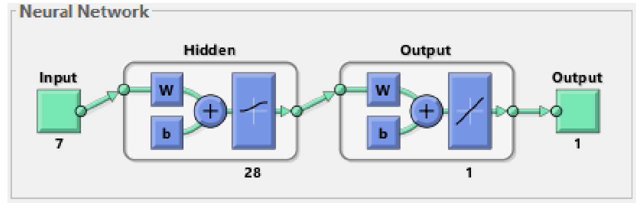

Fig. 10. The Structure of the ANN Created

This generated ANN has 2 layers; the input and the output layer. The number of neurons in the input layer was determined as 28 as the end of trials. In the output layer, only the "logsig" activation function was used.

Not all the data in the data set was used for network training.70% of the data was used for network training, 15% for verification and 15% for program testing.

The created network ensures the training by predicated on the performance of the failure rate squares (MSE) method. The determination of the generated network's parameters is shown in Figure 11.

```
hiddenSizes = [28];
trainFcn ='trainlm';
net = feedforwardnet(hiddenSizes,trainFcn);
net.layers{1}.transferFcn = 'logsig';
net.trainParam.show = 1000;
net.trainParam.epochs = 5000;
net.trainParam.goal = 0.000001;
net.trainParam.max_fail = 100;

net.PerformFcn = 'mse';
net = init(net);
net.divideParam.trainRatio = 0.7;
net.divideParam.valRatio = 0.15;
net.divideParam.testRatio = 0.15;
```

Fig. 11. The Parameters that the Network Generates

International Journal of Scientific & Engineering Research, Volume ….9, Issue 8, Augsut-2018
ISSN 2229-5518

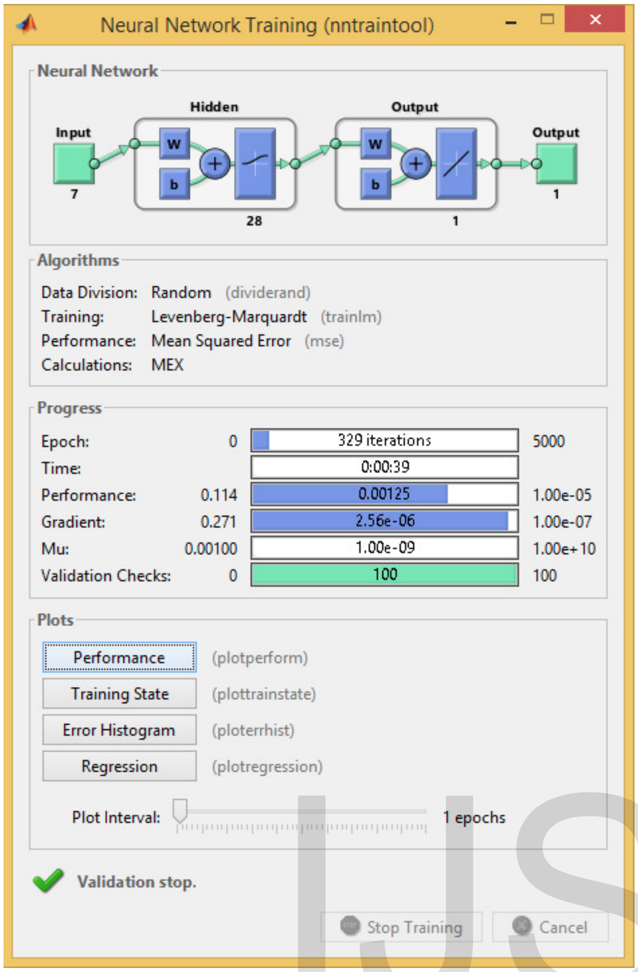

Fig. 12. Training Phase of the Created ANN

Figure 12 shows the network training phase. From this tab, line of regressions can be attained in order to determine the structure of the network that was created, the algorithm used, and the method of error performance, the performance chart of the training and the compliance of the system.

The chart of the created network's performance is shown in Figure 13. When the error target of the network is set to 10-5, the created ANN approximately approaches to the value of 10-3.

In Figure 14, the network regression lines are shown. The system's compatibility is determined by the regression lines. The parameter that generates the regression line is calculated by the ratio of the forecasted values to the actual values[56]. The regression value takes a value between 0 and 1. A value of 0 indicates that the system is incompatible; a value of 1 indicates that the system is compatible. Therefore, it is desired that the regression coefficient is close to 1.The regression value should be between 0,9 and 1 so that the compability can be defined as adequate. In our application, this value was calculated as 0,95 on average.

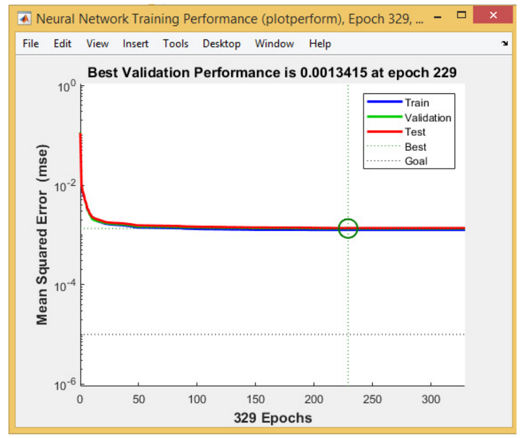

Fig. 13. Training Performance of the Created Network

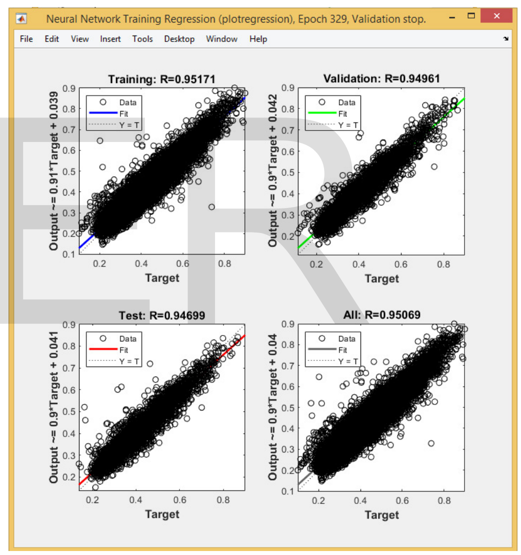

Fig. 14. Regression Lines of the Created Network

4 Conclusion

We had mentioned that two different models were created in practice.

1. Model; total load based ANN model

2. Model; regional load based ANN model.

According to these models, weekly forecast values for each day of the week are given in Table 2.

International Journal of Scientific & Engineering Research,Volume 9, Issue 8, Augsut-2018
ISSN 2229-5518

Table 2.Daily MAPE Values between 24 April to 30 April

Day	Model 1 MAPE	Model 2 MAPE
1 (24 April 2017)	7,93	9,03
2 (25 April 2017)	6,88	7,53
3 (26 April 2017)	4,47	5,14
4 (27 April 2017)	9,71	11,08
5 (28 April 2017)	6,74	5,39
6 (29 April 2017)	7,48	9,12
7 (30 April 2017)	6,07	6,36
AVERAGE	**7,04**	**7,66**

MSE had been used when the network performance is calculated. MAPE had been used for the performance value of the forecast, that is, the error value.

The MAPE calculation is as follows;

MAPE = [(Actual Value – Forecasted Value) / Forecasted Value] x 100

MAPE : Average failure rate

Actual Value : Measured Value

Forecasted Value : Indicates the value forecasted by ANN.

Lewis had made the evaluations of the models having MAPE value below 10% were "very good", between 10% to 20% were "good", between 20% to 50% were "acceptable" and value above 50% were "faulty" in 1982. On the basis of Lewis' evaluation, our study had achieved its objective by keeping the MAPE value below 10% in two different models..

Although both models had achieved their goals, it had been concluded that Model 1 (region based) was more successful.

There were some fluctuations in the forecast due to the fact that the region selected for the application is a touristic place and the population changes accordingly; data used in the network being deficient as the result of the power cuts due to some reasons such as electric poles being tilted over and line breakages since it is a windy region.

REFERENCES

[1] EPDK, Elektrik Piyasası Talep Tahminleri Yönetmeliği, Ankara, Mayıs 2016.

[2] Bilge, B., "Kısa Dönem Yük Tahmini", TEİAŞ Milli Yük Tevzi İşletme Müdürlüğü, Gölbaşı, ANKARA.

[3] Var, H., Türkay, E.B., "Yapay Sinir Ağları Kullanılarak Kısa Dönem Elektrik Yükü Tahmini" İstanbul Teknik Üniversitesi, İstanbul, 2014.

[4] Kaysal, K., Hocaoğlu, F.O., Kaysal, A., "Yük Tahmini için Hibrit (YSA ve Regresyon) Model" Afyon Kocatepe Üniversitesi, Afyonkarahisar, 2014.

[5] Islam, B., Baharudin, Z., Raza, M.Q., Nallagownden, P.,"Optimization of Neural Network Architecture Using Genetic Algorithm for Load Forecasting", Department of Electrical & Electronics Engineering, Universiti Teknologi PETRONAS, Tronoh, Preak, Malaysia, 2014.

[6] Sahay, K.B., Tripathi, M.M., "Day Ahead Hourly Load Forecast of PJM Electricity Market and ISO New England Market by Using Artificial Neural Network", Department of Electrical Engineering, Delhi Technological University, New Delhi, India, 2013.

[7] Liu, Q., Xu, H., Shi, C., Wei, S., "The Application of Intelligent Control Technology in Power Systems", International Conference on Mechatronics and Automation, pp. 2034-2038,August 2014.

[8] Topallı, A.K., Erkmen, I. and Topalli, I.," Intelligent short-term load forecasting in Turkey, Electrical Power & Energy Systems", 28, 437-447, 2006.

[9] Hamzaçebi C., "Yapay Sinir Ağları: Tahmin Amaçlı Kullanımı Matlab ve Neurosolution Uygulamaları" Ekin Basım Dağıtım, Bursa, 2011, ss. 11.

[10] T. Bakırtaş, S. Karbuz ve M. Bildirici, "An Econometric Analysis of Electricity Demand in TurkeyMETU Studies in Development", 27, (1-2), 2000, s.23-34.

[11] Doğan, G., Çıngı, H., "Yapay Sinir Ağları Kullanılarak Türkiye'de ki Özel Bir Sigorta Şirketinde Portföy Değerlendirmesi", Yüksek Lisans Tezi, Hacettepe Üniversitesi, 2010.

[12] Tang, Z., Fishwick, P.A., "Feedforward Neural Nets as Models for Time Series Forecasting", ORSA Journal on Computing, 1993.

[13]URL, MGM Sıcaklık-Nem Arası Bağıntı https://www.mgm.gov.tr/genel/sss.aspx?s =hissedilensicaklik

[14] Colin D. LEWIS, Industrial and Business Forecasting Methods, Butterworths Publishing: London, 1982, s. 40