Original Paper

European Journal of Digital Economy Research
2022 Vol.3 Issue.2 pp.75-81

FORECASTING TÜRKİYE'S HOURLY ELECTRICITY PRODUCTION BY USING NONLINEAR AUTOREGRESSIVE MODELS

Mehmet ÖZSOY*

* Res. Assist., Süleyman Demirel University, Faculty of Economics and Administrative Sciences, Department of Business, mehmetozsoy@sdu.edu.tr, https://orcid.org/0000-0003-3204-7295.

ABSTRACT

Considering that industrialization is increasing rapidly, and technological enhancement is developing exponentially compared to the past, and considering that electrical energy is at the center of all these, it is one of the essentials of our age that this energy reaches its end users efficiently and at low cost. The aim of this study is to predict electricity generation with the help of the NARX model, an artificial neural network model, with the help of hourly data of Türkiye between 2016 and 2022 and to reveal the success of the model. As a result of the analysis, 99.83% coefficient of determination value was obtained with the NARX model.

Keywords: *Electricity Production, Forecast, NAR, NARX.*

1. INTRODUCTION

Energy resources are basically divided into two areas. These are primary and secondary energy sources. Primary energy sources are also divided into traditional and renewable energy sources. Oil, natural gas, and coal are traditional energy sources and wind, solar, and hydro powers are renewable energy sources. Electricity generated using these sources plays a central role in people's professional and daily work. As a result of the rapid development and changes in technology and industry, our need for electricity in almost every field is increasing day by day. With this increase in energy demand, it is possible to prevent uncertainties, supply-demand imbalances, and disruptions that may occur in the sector by using a reliable forecasting method. It can be said that the use of reliable forecasting methods in the energy sector reduces energy costs as well as provides uninterrupted energy (Soliman and Al-Kandari 2010).

In order to manage energy policies well, it is very important for countries to determine their future energy needs. There are basically three time horizons in electrical energy forecasting: short-term load forecasting, medium-term load forecasting and long-term load forecasting. These load forecasts are highly affected by periodic events and situations. There are significant differences between the electricity required in summer and winter, and these differences can be daily and weekly. By making short-term load forecasts, a significant cost reduction in electricity prices can be achieved. In line with an accurate forecast, a 1 percent reduction in the margin of error for a 10 GW service can result in a profit of approximately 1.6 million dollars (Dudek 2015). One-hour-ahead forecasting is an important element of the intraday electricity market and is also an important component and complement to the day-ahead market. Machine learning and deep learning techniques are frequently used to forecast electricity generation data. In addition, machine learning techniques produce much more successful results in situations where the amount of data is quite high, such as one hour ahead and a day-ahead, and therefore play a major role in stakeholder decision-making (Lantz 2019). When the studies using electric energy data are examined, it is seen that artificial neural networks are also used quite frequently in studies. Nonlinear Autoregressive (NAR) and Nonlinear Autoregressive model with exogenous variable (NARX) models, which are also used in this study, are also used for different purposes with similar data.

In this context, Türkiye, as a developing country, can have the opportunity to consolidate its position in the region and keep up with the progress of developed countries more easily by managing its energy policies correctly and increasing energy efficiency. According to the daily production amounts announced by the Energy Exchange Istanbul (EXIST), Türkiye's average hourly electricity demand is 28000 megawatt hours. This amount varies during the day. These data have been available since 2012. Approximately 20 percent of the hourly average electricity demand of 28000 MWh is met from coal, 19 percent from dams, 16 percent from wind and the rest from various other sources.

A brief review of the recent literature on forecasting the production, consumption and prices of electricity energy reveals that some authors forecast demand, some forecast price and some forecast load. In the studies on electricity load forecasting, authors have made short-term load forecasts using various models for only one region with a narrow scope (Şeker 2022; Shirzadi et al. 2021) and emphasized that the NARX model gives the most accurate results (Kim, Son, and Kim 2019). They have also forecasted long-term electricity load forecasts for the whole country using several models, including the NAR and NARX models (Essallah and Khedher 2022). In addition, some authors have also investigated price forecasting in addition to load forecasting and proposed a new NARX model (Mujeeb and Javaid 2019). It has been argued that by combining different models, they overcome each other's weaknesses and increase the success of the models in predicting electricity prices (Lehna, Scheller, and Herwartz 2022).

There are studies in the literature where energy demand is met by using renewable energy sources and this demand is estimated using artificial neural networks, machine learning and other methods. Some of these studies used solar energy data, which is one of the renewable energy sources, and made forecasts with artificial neural networks and machine learning algorithms (Demolli et al. 2019; Gök, Yıldız, and Şekkeli 2019; Yousif and Kazem 2021). Krechowicz et al. (2022) examined the studies on machine learning forecasting using data from three different renewable energy sources and concluded that it has an important place in recent studies. In this study, forecasting is performed with the NARX model using hourly electricity data published by EXIST between 2012 and 2022.

2. MATERIAL AND METHOD

The data is collected from EXIST website for Türkiye with a time array of 2016 and 2022. The production and consumption values are normalized in order to get better results from the analysis. All the analyses are conducted using the MATLAB R2022a program. The normalization is done with MATLAB with the range between [0,1] and the code used for normalization was; "normalize(variable_name,'range')".

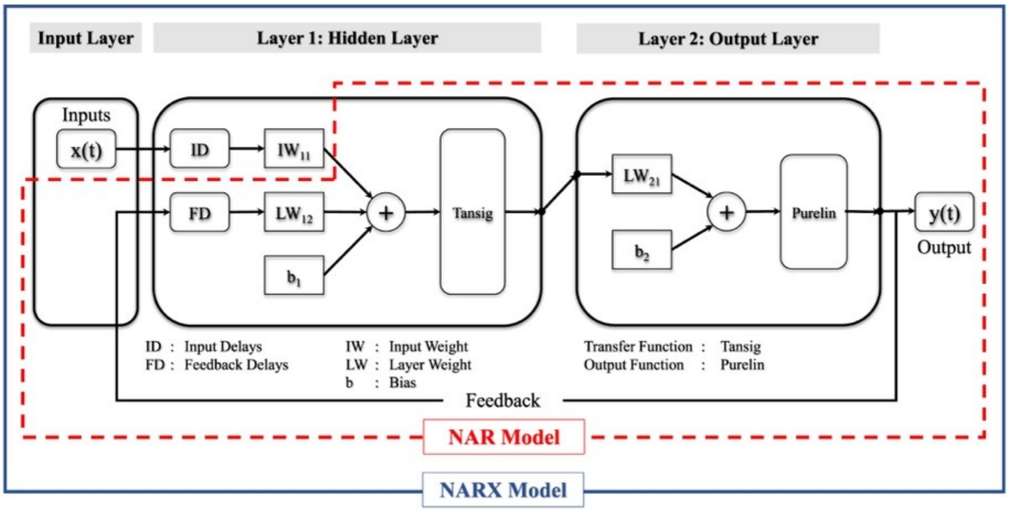
The Nonlinear Autoregressive (NAR) and with exogenous (NARX) neural network models used in the study are input-output based Recurrent Neural Networks (RNN) (Leontaritis and Billings 1985). NAR uses historical data to forecast the future of the same data. In addition to NAR, NARX uses exogenous input data together with the output data (Rogier and Mohamudally 2019). The NAR model can be interpreted mathematically in Equation 1 and the NARX model in Equation 2.

$$y(t) = f[y(t-1), y(t-2), \dots, y(t-dy)] \quad (1)$$

$$y(t) = f[y(t-1), y(t-2), \dots, y(t-dy), x(t-1), x(t-2), \dots, x(t-dx) \quad (2)$$

where f is the nonlinear function; $y(t-1)$, $y(t-2)$, ⋯, $y(t-dy)$ are the past values of the time series y; $x(t)$, $x(t-1)$, ⋯, $x(t-dx)$, are the current and past values of the exogenous time series x.

Figure 1. The Visualization of NAR and NARX Models



Source : (Chreng, Lee, and Tuy 2022)

The performance of the models was determined by the coefficient of determination (R^2) and mean squared errors (MSE). The mathematical interpretations of the values are given in Equations 3 and 4, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(P_{predicted}^{i} - P_{actual}^{i})^2}{\sum_{i=1}^{N}(P_{actual}^{i} - P_{predicted}^{i})^2} \quad (3)$$

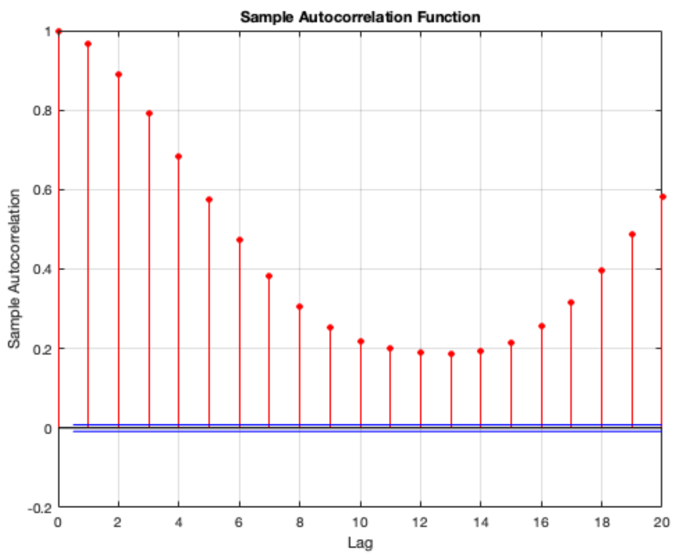
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(P_{predicted}^{i} - P_{actual}^{i})^2 \quad (4)$$

where P_{actual}^{i}, $P_{predicted}^{i}$ are the i_{th} actual and predicted values, N is the period. The coefficient of determination refers to how well the data fits, and the degree of the fit is evaluated by its distance to 1.

3. FINDINGS

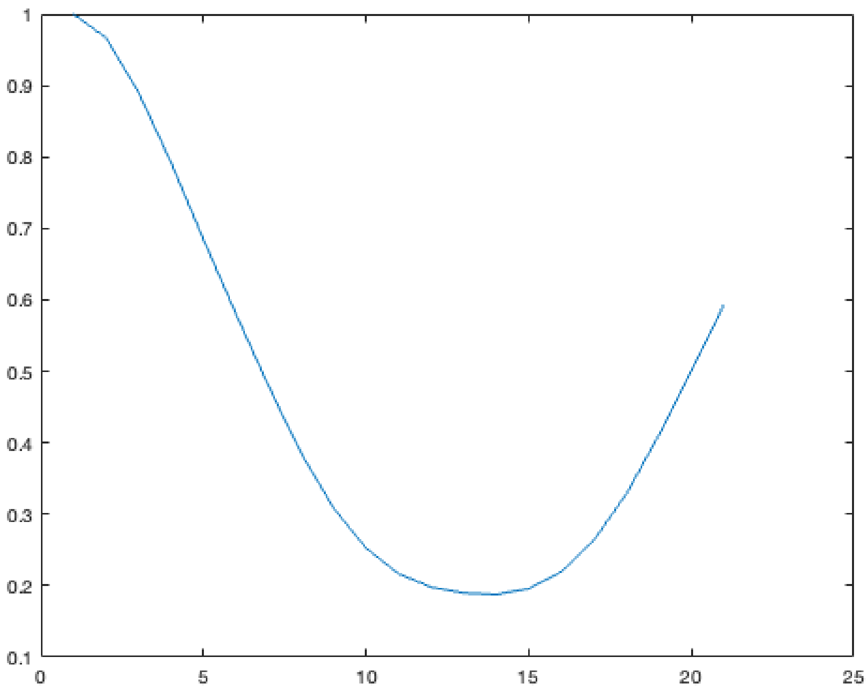
Since the NAR and NARX models used in the study contain autoregressive conditions, correlation test was first applied to both electricity generation and consumption data. As a result of the test, the autocorrelation of generation data is shown in Figure 2 and consumption data in Figure 3. As can be seen from the figure, the auto-correlation of both data sets is almost the same. The analysis shows that the correlation first decreases and then increases again as the lag increases.

Figure 2. Autocorrelation of Electricity Production



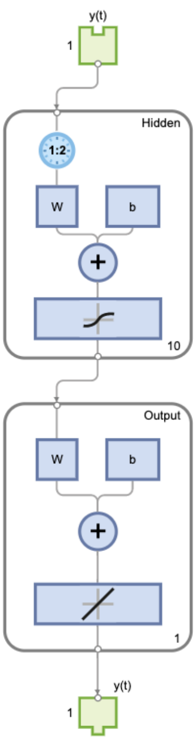
A correlation value of 50% was determined for the lag to be used in both data sets and this percentage corresponds to a lag of 5 units in the data sets. Therefore, while constructing the NARX model, a lag of 5 lines was applied to both production and consumption data and the model was made ready for analysis.

Figure 3. Autocorrelation of Electricity Consumption



The NAR model using lagged data is shown in Figure 4. In the model, electricity generation data is used both as input and output. 70% of the data is allocated for training, 15% for validation, and the remaining 15% for testing. Levenberg-Marquardt algorithm was used for training the network. The success of the algorithm in solving non-linear real-life problems and its ability to produce fast solutions played a major role in the selection of this algorithm. Since there is no definite rule for determining the number of hidden layers and the number of layers is determined by trial and error in the literature, the trial-and-error method was used in this study.

Figure 4. NAR Model with 10 Hidden Layers for Electricity Production



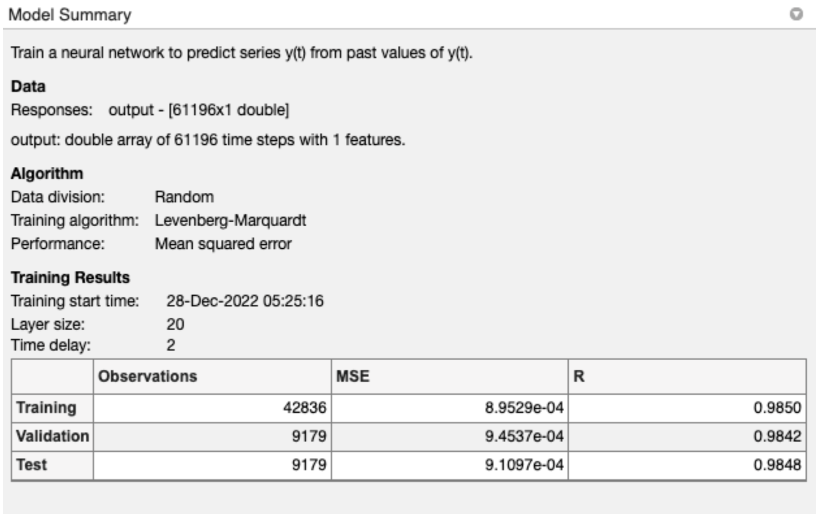
In the NAR model, models with 2, 5, 10, 15, 20, and 40 hidden layers were created and analyzed respectively. When the R^2 and MSE values of the models were examined, no significant difference was observed and it was concluded that increasing the number of layers would only increase the analysis time and the performance of the NAR models with only 10, 15, and 20 hidden layers are given in Table 1 below.

Table 1. Performance Values of NAR Models with Different Hidden Layers

Number of Hidden Layers	R^2	MSE
10	0.9846	0.0009156
15	0.9848	0.0009063
20	0.9850	0.0008952

When the results are analyzed, it is seen that the 20-layer model performs better with the lowest error and the highest level of coefficient of determination. The performance values of the 20 hidden layer NAR model for training, validation, and test data are also given in Figure 5.

Figure 5. The Performance Values of the NAR Model with 20 Hidden Layers



Model Summary

Train a neural network to predict series y(t) from past values of y(t).

Data
Responses: output - [61196x1 double]
output: double array of 61196 time steps with 1 features.

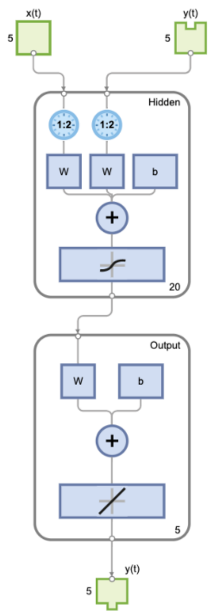
Algorithm
Data division: Random
Training algorithm: Levenberg-Marquardt
Performance: Mean squared error

Training Results
Training start time: 28-Dec-2022 05:25:16
Layer size: 20
Time delay: 2

	Observations	MSE	R
Training	42836	8.9529e-04	0.9850
Validation	9179	9.4537e-04	0.9842
Test	9179	9.1097e-04	0.9848

After the analysis results of the NAR model are obtained, production and consumption data with 5 hours lags are used in the NARX model as explained earlier. As in the NAR model, 75% of the data is reserved as training data and the Levenberg-Marquardt algorithm is again chosen as the training algorithm. The NARX model with 20 hidden layers and 2 input and 1 output variable is shown in Figure 6.

Figure 6. NARX Model with 20 Hidden Layers to Forecast Electricity Production by Using Electricity Consumption



In the NARX model, as in the NAR model, models with 2, 5, 10, 15 and 20 hidden layers were created respectively. Although there is almost no performance difference between these models, the best performance, albeit very small, is obtained from the model with 20 hidden layers. The R^2 and MSE values of the models are given in Table 2.

Table 2. Performance Values of NARX with Different Hidden Layers

Number of Hidden Layer	R^2	MSE
10	0.9982	0.00010596
15	0.9982	0.00011014
20	0.9983	0.00010269

As a result of the analysis of NARX models, the highest R^2 value of 0.9983 was obtained. This high value indicates the success of the prediction performance of the model. In addition, compared to the NAR model, an R^2 increase of more than 1% and a lower error rate were obtained.

4. CONCLUSION

Regardless of the source, the uninterrupted, high quality, and low cost of the electrical energy produced affect both the daily and professional lives of people. Considering that industrialization is increasing rapidly, and technological enhancements are developing exponentially compared to the past, and considering that electrical energy is at the center of these, it is one of the essentials of our age that this energy reaches its end users uninterrupted, efficiently, and at low cost.

In this study, Türkiye's hourly electricity generation planning is forecasted by including both its own data and hourly electricity consumption data. With this forecasting, the performances of NAR and NARX models from RNN models were compared, and according to the results of the analysis, the NARX model, that is, the use of past and delayed data of electricity consumption as well as past and delayed data of electricity generation, increased the performance in the forecasting of hourly generation and allowed to obtain a result that will contribute to the planning for the uninterrupted supply of electricity with less margin of error. At the end of the study, a prediction accuracy of 99.83% was obtained with the NARX model, which shows the success of the model. Likewise, the error rate measured by the MSE value of the model was also found to be quite low.

When the analysis results are analyzed, it is seen that increasing the number of hidden layers does not affect the model performance to a great extent. This is thought to be due to the high autocorrelation with 61,196 lines of hourly data between 2016 and 2022. It is also thought that the study can be improved by adding other variables affecting electricity generation to the model. In addition, it is thought that more meaningful and useful results can be obtained from electrical energy data by using different machine learning and deep learning methods if sufficient data is available.

REFERENCES

1. CHRENG, K., LEE, H. S. & TUY, S. (2022), "Electricity Demand Prediction for Sustainable Development in Cambodia Using Recurrent Neural Networks with ERA5 Reanalysis Climate Variables", Energy Reports 8, 76–81, DOI: 10.1016/j.egyr.2022.01.025.
2. DEMOLLİ, H., ECEMİŞ, A., DOKUZ, A. & GOKCEK, M. (2019), "Makine Öğrenmesi Algoritmalarıyla Güneş Enerjisi Tahmini: Niğde İli Örneği", International Turkic World Congress on Science and Engineering, (pp. 775-783).
3. DUDEK, G. (2015), "Pattern Similarity-Based Methods for Short-Term Load Forecasting–Part 1: Principles", Applied Soft Computing 37, 277–287.
4. ESSALLAH, S. & KHEDHER, A. (2022), "Long-term Load Forecasting in the Smart Grid Framework Using Time Series and

Econometric Approaches", In Renewable Energy Production and Distribution, (pp. 389-413), Academic Press.

5. GÖK, A. O., YILDIZ, C. & ŞEKKELİ, M. (2019), "Yapay Sinir Ağları Kullanarak Kısa Dönem Güneş Enerjisi Santrali Üretim Tahmini: Kahramanmaraş Örnek Çalışması", Uluslararası Doğu Anadolu Fen Mühendislik ve Tasarım Dergisi, 1(2), 186–195.
6. KIM, Y., SON, H. & KIM, S. (2019), "Short Term Electricity Load Forecasting for Institutional Buildings", Energy Reports 5, 1270–80, DOI: 10.1016/j.egyr.2019.08.086.
7. KRECHOWICZ, A., KRECHOWICZ, M. & POCZETA, K. (2022), "Machine Learning Approaches to Predict Electricity Production from Renewable Energy Sources", Energies, 15(23), 9146.
8. LANTZ, B. (2019), "Machine Learning with R: Expert Techniques for Predictive Modeling, 3rd Edition", Packt Publishing.
9. LEHNA, M., SCHELLER, F. & HERWARTZ, H. (2022), "Forecasting Day-Ahead Electricity Prices: A Comparison of Time Series and Neural Network Models Taking External Regressors into Account", Energy Economics 98, 105742, DOI: 10.1016/j.eneco.2021.105742.
10. LEONTARITIS, I. J. & BILLINGS, S. A. (1985), "Input-Output Parametric Models for Non-Linear Systems Part I: Deterministic Non-Linear Systems", International Journal of Control, 41(2), 303–328.
11. MUJEEB, S. & JAVAID, N. (2019), "ESAENARX and DE-RELM: Novel Schemes for Big Data Predictive Analytics of Electricity Load and Price", Sustainable Cities and Society, 51, 101642, DOI: 10.1016/j.scs.2019.101642.
12. ROGIER, J. K. & MOHAMUDALLY, N. (2019), "Forecasting Photovoltaic Power Generation via an IoT Network Using Nonlinear Autoregressive Neural Network", Procedia Computer Science, 151, 643–650, DOI: 10.1016/j.procs.2019.04.086.
13. SHIRZADI, N., NIZAMI, A., KHAZEN, M. & NİK-BAKHT, M. (2021), "Medium-Term Regional Electricity Load Forecasting through Machine Learning and Deep Learning", Designs, 5(2), 27, DOI: 10.3390/designs5020027.
14. SOLIMAN, S. A. & AL-KANDARI, A. M. (2010), "Electrical Load Forecasting: Modeling and Model Construction", Elsevier Science.
15. ŞEKER, M. (2022), "Long Term Electricity Load Forecasting Based on Regional Load Model Using Optimization Techniques: A Case Study", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 44(1), 21–43, DOI: 10.1080/15567036.2021.1945170.
16. YOUSIF, J. H. & KAZEM, H. A. (2021), "Prediction and Evaluation of Photovoltaic-Thermal Energy Systems Production Using Artificial Neural Network and Experimental Dataset", Case Studies in Thermal Engineering, 27, 101297, DOI: 10.1016/j.csite.2021.101297.