Forecasting of Muslim Pilgrims UsingNonlinear Auto-regressive Neural NetworkModels with a Comparison of other Modern Methods

Mahmoud Elgamal
Faculty of Computer and Information Sciences, Ain Shams uni., Egypt

Abstract
Yearly Muslims pilgrimage is considered one of the largest human collection all over the world, as nearly 3 million move together through a very limited space in a short time period. The number of pilgrims coming from outside Saudi Arabia (NPO) account for two thirds of pilgrims, therefore forecasting the NPO is considered by Saudi Arabia as a vital indicator in determining the planning mechanism for future secure and easy hajj seasons. The aim of this paper is to exploit the Nonlinear Autoregressive models with exogenous inputs (NARX) neural networks to forecast the yearly series of NPO and to show that it gives better forecasts than Box and Jenkins techniques. First the last five observation held out and conduct NARX methodology to predict them and the result was promising with error of about 4.4% which is less than half of Box and Jenkins result (9.6%). Afterwards NARX employed to forecast the future five years of NPO.

Keywords: forecasting, Muslim Pilgrims, NARX, ARIMA, forecast errors.

1 Introduction

Muslims pilgrimage (Hajj) is counted the largest human gathering, where more than three million pilgrims move together through a very limited space in a short time period. This important event is repeated annually at the same period of time and location, and the number of pilgrims is increasing yearly. In addition to that, Hajj considered one of the main resources of gross national product in Saudi Arabia and very essential source of sustenance for many people living in Saudi Arabia. During Hajj season, Saudi authorities should provide many services to pilgrims such as security, food, housing, electricity, transportation and health care. Annual forecasting the number of NPO, is of special concern to Saudi government, as it is the most crucial indicator in planning mechanisms regarding theservices for future Hajj seasons.

The available NPO data represented by a time series consists of 50 annual values, which will be used to predict the future values.

A time series is an ordered sequence of data samples which are recorded over a time interval. Time series data includes a variety of features, for example, few of data series may possess seasonality, few reveal trends, i.e., exponential or linear and some are trendless. Time series analysis and forecasting models are implemented in order to extract meaningful statistics, other characteristics of the time series data and to predict future values based on previously observed values. The early and well-known one is the methodology to forecast time series is of Box and Jenkins, which have grown in popularity and is considered the prevailing methodology of time series analysis. They assume that a parsimonious stationary and invertible autoregressive moving average (ARMA) process could present the time series at hand (or a transformation of the series) such that one can perform the four phases of time series analysis (Identification, estimation, diagnostic checking, and forecasting). Their methodology has been widely used and explained by many others such as [1, 2, 4, 11, 12, 13, 14]. However, their identification technique is highly nonobjective and requires very good experience and skills. The nonlinear approach modeling of the time series is suitable for most real-world problems and the parametric models was developed to deal with them. In order to get an accurate forecast, the models must be known beforehand, therefore, the model cannot be used if the features of the data do not meet the assumptions of the model.

The formulation and preparation of a nonlinear model to a specific type of data set is a very challenging task as there are too many unknown nonlinear patterns and a specified nonlinear model may not be sufficient to acquire all the significant representations or features.

Artificial neural networks, are actually nonlinear data-driven approaches as contradicted to the statistical model-based nonlinear methods and are capable of performing nonlinear modeling without a priori knowledge about the relationships between input and output variables.

Artificial Neural Network, denoted by ANN for short, model takes input and produces one or more output, in-between input and output variables, the ANN does not require any presumption on logical or analytical forms [8]. A neural network gains the knowledge over the system dynamics by examining the patterns between input data and corresponding outputs, and becomes able to use this knowledge to predict a system’s output [9].

The purpose of this research is to establish and train a network that can well predict and forecast the NPO with optimization of neural network parameters. In order to achieve this, several NARX models were trained with different parameter settings. The NARX model trained to predict a future five years of NPO.

The paper is organized as follows: section (2) introduces and explains the proposed NARX networks structure, section (3),section (4), and finally the conclusion.

2 NARX Networks

The NARX[5, 10, 3] uses the past values of the actual time series to be predicted and past values of other inputs to make predictions about the future value of the target series. This is a powerful class of models which has been demonstrated that they are well suited for modeling nonlinear systems and specially time series. Some important qualities about NARX networks with gradient-descending learning gradient algorithmhave been reported: (1) learning is more effective in NARX networks than in other neural network(the gradient descent is better in NARX) and (2) these networks converge much faster and generalize better than other networks [10, 3].
NARX is an important class of discrete-time nonlinear systems that can be mathematically represented as
$y(n+1) = f[y(n),\cdots,y(n-d_y+1);u(n),u(n-1),\cdots,u(n-d_u+1)]$, (1)

where $u(n) \in R$ and $y(n) \in R$ denote, respectively, the input and output of the model at discrete time step n, while du,and $d_y \geq 1$ and $du \leq d_y$, are the input memory and output-memory orders, respectively. Equation (1) may be written in vector form as

$$y(n+1) = f[y(n);u(n)] \qquad (2)$$

where the vectors y(n) and u(n) denote the output and input regressors, respectively. The nonlinear mapping $f(\cdot)$ is generally unknown and can be approximated, for instance, by a standard multilayer perceptron network. The resulting connectionist architecture is then called a NARX network, a powerful class of dynamical models which has been shown to be computationally equivalent to Turing machines [5]. Figure (2.1) shows the topology of a three-hidden-layer NARX network, where the hat symbol (^) is used to denote approximated values or functions.

NARX network has two modes in training:

- parallel mode in which the output is fed back to the input of the feedforward neural network as part of the standard NARX architecture, as shown in figure (2.2-a) below, in which the TDL means the tapped delay line.
- series-parallel mode in which the true output is used instead of feeding back the estimated output, as shown in figure (2.2-b).

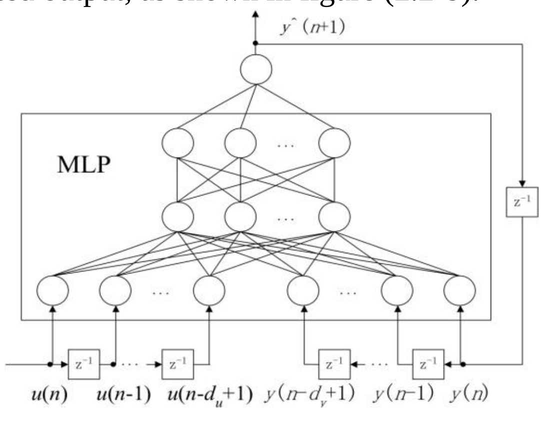
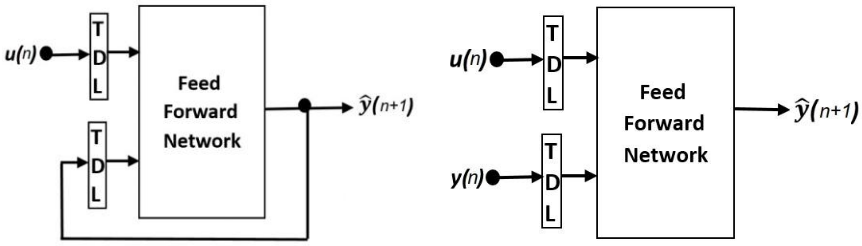


Figure 2.1:NARX network with, delayed inputs(d_u) and delayed outputs(d_y), and unit time delay(z^{-1}).



(a) NARX parallel architecture (b) NARX series-sarallel architecture

Figure 2.2: NARX network architecture.

In this paper series-parallel is adopted because it has a purely feedforward architecture, and static backpropagation can be used for training. Firstly, in the training phase NARX series-parallel mode used and then converted to the parallel mode to perform the prediction.

3 Experimental results

The time series data of number of pilgrims coming from outside the kingdom of Saudi Arabia (NPO) consists of 50 observations based on Hijri calendar2 as shown in table (1).

Table 1: NPO from 1390 ~ 1439.

Year	NPO	Year	NPO	Year	NPO
1390	431270	1407	960386	1424	1419706
1391	479339	1408	762755	1425	1534759
1392	645182	1409	774560	1426	1557447
1393	607755	1410	828993	1427	1654407
1394	918777	1411	720102	1428	1707814
1395	894573	1412	1012917	1429	1729841
1396	719040	1413	992813	1430	1613965
1397	739319	1414	995611	1431	1799601
1398	830236	1415	1043274	1432	1828195
1399	862520	1416	1080465	1433	1752932
1400	812892	1417	1168591	1434	1379531
1401	879368	1418	1132344	1435	1389053
1402	853555	1419	1056730	1436	1384941
1403	1003911	1420	1267555	1437	1325372
1404	919671	1421	1363992	1438	1752014
1405	851761	1422	1354184	1439	1758722
1406	856718	1423	1431012		

[2]The years are written using Lunar Calendar (AH) from 1390(1971) 1439(2018); one lunar year is shorter than Gregorian year by about 11 days (see http://en.wikipedia.org/wiki/Islamic Calenda).

In all the experiments performed, a one-step-ahead prediction is considered; that is, the actual observed values of all lagged samples are used as inputs. (If multistep-ahead predictions are required then, it is possible to proceed by adding the first one-step-ahead prediction to time series, and then the new time series is used to predict the second step-ahead, and so on).

In all experiments, the following NARX parameters are set as follow: Maximum number of epochs to train = 1000, Performance goal = 0, Minimum performance gradient = 1E-10, input delay = 1:2, feedback delay = 1:2, and Number of hidden neurons is 10,15,20,30 respectively. But found the best 10 neurons in the hidden layer.

The aim of this research is to perform an accurate prediction of NPO, in order to do this, our strategy is based on the consideration of several numbers of trained models, the best accuracy model nominated as a forecaster for further deployments. The performance of each model is based on the percentages of the absolute errors (PAE) which can be computed as expressed in equation (3)

$$\text{PAE} = \left|\frac{T-P}{T}\right| \times 100, \tag{3}$$

where T, and P are the true and predicted values respectively.

Before forecasting the future observations, the NARX-model has been used to check its ability to forecast the future observations. In order to evaluate the forecast performance of our proposed approach, a small portion of the NPO data at the end of the data are reserved only for forecasting evaluation, these data are referred to as hold-out sample, or post-sample, and in principle are not used in model or forecasting when evaluating forecast performance. The last 5 observations (about 10% of the whole data) are reserved as the hold-out sample (post-sample). The first 45 observations were used to forecast the next five observations using NARX approach; then the five forecasts were compared with the five real observations and the MAPE errors were calculated. The results are reported in table (2) and shown in figure (3.3).

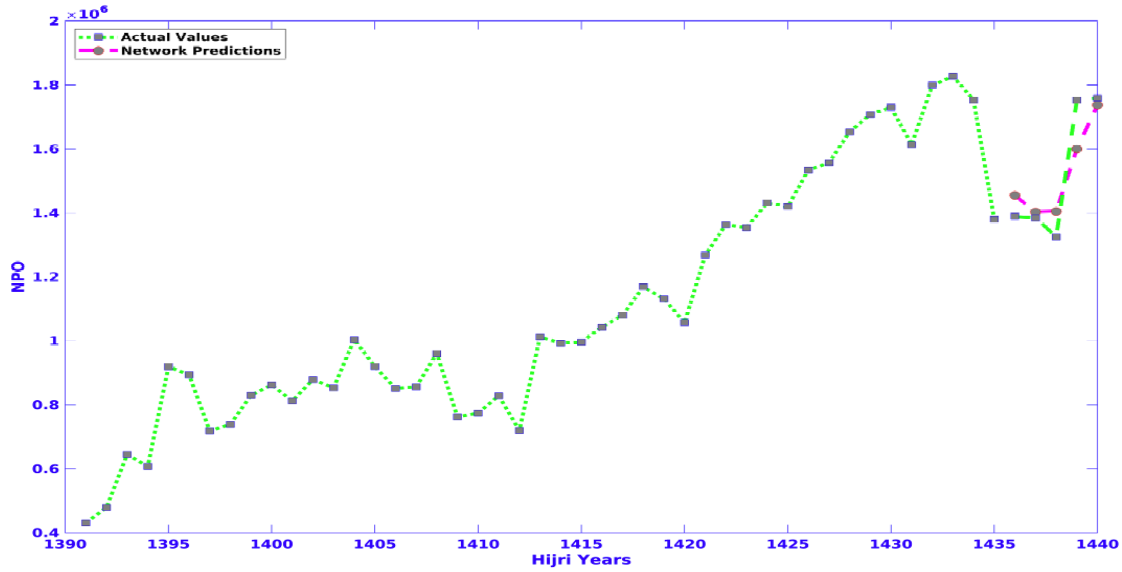


Figure 3.3: The true NPO values and predicted values of the last five values.

Table 2: Last 5 observations and their predicted values.

Year	True Value	Predicted Value	PAE error (%)
1435	1389053	1455098	4.755
1436	1384941	1402917	1.298
1437	1325372	1404705	5.986
1438	1752014	1600112	8.670
1439	1758722	1737514	1.206
Mean			4.383

Finally, the NARX- model has been used to forecast the next five future years of NPO. The point forecasts for these observations are given by table (3) and shown in figure (3.4).

Table 3: Future 5 forecasts.

Year	1440	1441	1442	1443	1444
Point forecast	1702178	1781900	1802913	1863536	1877564

4 Comparative Study

In this section, we compare our result with *Automatic Time Series Forecasting package* [6], which is used to automatically determine the appropriate time series model, andestimate the parameters. Moreover, the algorithms of that package are applicable to both seasonal and non-seasonal data.

For the sake of comparison, the following three criteria of errors [7] are calculated:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{true\ value - forcasted\ value}{true\ value}\right| \times 100$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|true\ value - forcasted\ value\right|$$

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(true\ value - forcasted\ value)^2\right]^{\frac{1}{2}}$$

Where n is the total number of observations in the hold-out sample.
Table (4) shows the results after conducting the simulation on NPO data using Box-Jenkins and NARX.

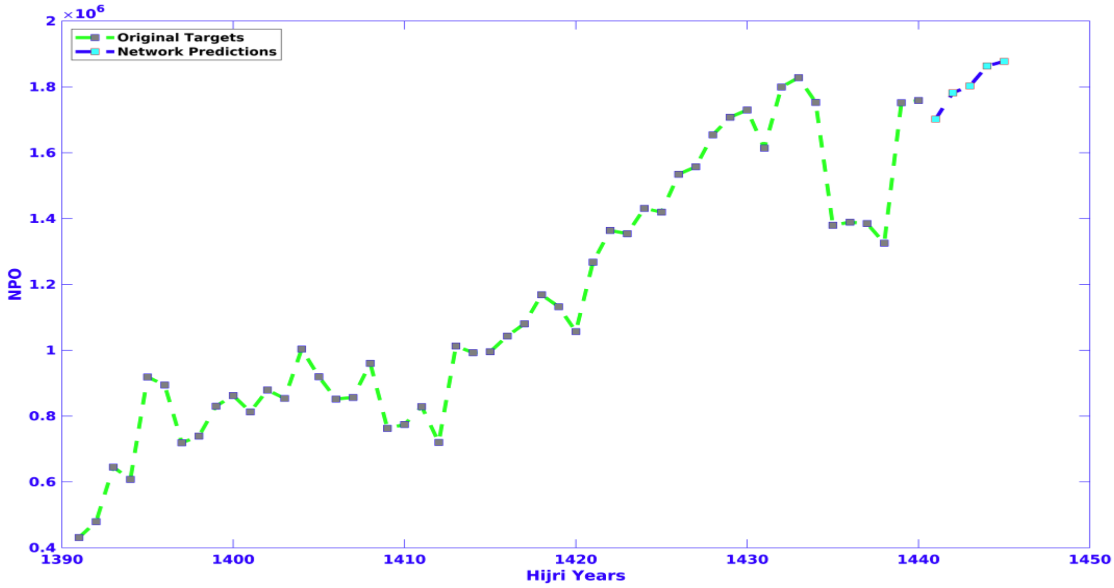


Figure 3.4: The forecasted NPO values of the next five years.

Table 4: Comparison of Box-Jenkins and NARX results of the five hold-out samples.

Error	Box-Jenkins	NARX
Mean Absolute Percentage Error (MAPE)	9.597	4.383
Mean Absolute Deviation (MAD)	16.4153e+04	6.7293e+04
Root Mean Squared Error (RMSE)	23.899e+04	8.3070e+04

5 Conclusion

In this paper, the powerful NARX technique engaged to forecast five future years of NPO, at beginning out-sampled predictions done by eliminating five values from the NPO time series and run the technique to get result of 4.4% accuracy despite the high fluctuation at the end of the time series compared with 9.6% of Box and Jenkins methodology.
Motivated by NARX superiority over Box and Jenkins, a forecast of five years were predicted using NARX.

References

[1] Box, G.E.P., Jenkins, G.M, Reinsel, G.C. and Ljung, G.M. (2016). Time Series Analysis: Forecasting and Control, 5th Edition, John Wiley & Sons.

[2] Box G. and Jenkins G.(1970). Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco.

[3] Gao Y., Er M. J.*(2005*). NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches, Fuzzy Sets and Systems, *Vol. 150, No. 2, pp.331-350.*

[4] Harvey A.(1993). Time Series Models, 2nd edition, The MIT Press.

[5] Haykin S. (1999). Neural Networks, Second Edition, Pearson Education.

[6] Hyndman R.J. and Khandakar Y. (2008).Automatic time series forecasting: the forecast package for R, Journal of Statistical Software 27(3).

[7] Hyndman R.J. and Athanasopoulos G. (2018).Forecasting: principles and practice. (1st edition). OTexts.

[8] Koskivaara, Eija. (2003).Artificial neural networks in auditing: state of the art, Turku Centre for Computer Science.

[9] Leondes, Cornelius T. (2002).Intelligent Systems: Technology and Applications, Six Volume Set. Vol. 1. CRC Press, chapter 1, neural network techniques and their engineering applications.

[10] Lin T., Bill G. Horne, Peter Tino, C. Lee Giles. (1996).Learning long-term dependencies in NARX recurrent neural networks, IEEE Transactions on Neural Networks, Vol. 7, No. 6, pp. 1329-1351 .

[11] Liu L. (2009). Time Series Analysis and Forecasting.(2nd edition). Scientific Computing Association Corp, USA.

[12] Newbold P. (1973). Bayesian Estimation of Box and Jenkins Transfer Function Model for Noise Models, Journal of the Royal Statistical Society, Series B, vol. 35. No. 2, pp. 323-336.

[13] Priestley M. (1981).Spectral Analysis of Time Series. (1st Edition). Academic Press, London.

[14] Wei W.W.S.(2005). Time Series Analysis: Univariate and Multivariate Methods. (2nd Edition). Addison Wesley, Reading, MA.