British Journal of Mathematics & Computer Science
6(5): 394-401, 2015, Article no.BJMCS.2015.086
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Forecasting the Number of Muslim Pilgrims Using NARX Neural Networks with a Comparison Study with Other Modern Methods

Esam A. Khan[1], Mahmoud A. Elgamal[1*] and Sameer M. Shaarawy[1]

[1]The Custodian of the Two Holy Mosques Institute for Hajj and Omra Research, Umm Al -Qura University, Makkah, Saudi Arabia.

Article Information

DOI: 10.9734/BJMCS/2015/14563
Editor(s):
(1) Paul Bracken, Department of Mathematics, The University of Texas-Pan American Edinburg, USA.
Reviewers:
(1) Anonymous, Egypt.
(2) Anonymous, Turkey.
(3) Tiago Andrade Mota, Mechatronics, Federal University of Bahia, Brazil.
(4) Anonymous, Egypt.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=735&id=6&aid=7681

Original Research Article

Received: 05 October 2014
Accepted: 15 December 2014
Published: 09 January 2015

Abstract

Pilgrimage (Hajj) of Muslims is considered the largest human gathering all over the world in which more than three millions move together through a very limited space in a short time period. The yearly number of pilgrims coming from outside Saudi Arabia, denoted by NPO for short, is more than two thirds of the total number of Pilgrims. Therefore forecasting the NPO is considered by Saudi Arabia as the most important indicator in determining the planning mechanism for future secure and comfortable hajj seasons. The main objective of this article is to employ the NARX neural networks to forecast the yearly series of NPO and to show that it gives better forecasts than Box–Jenkins and Bayesian Procedures. In order to achieve our objective, the NARX is used to forecast the future five observations and the results are compared with the results given in [1].

Keywords: Forecast, NARX, Bayesian analysis, box and Jenkins methodology, Mean Average Percentage Error (MAPE).

**Corresponding author: mahmoud_40@yahoo.com;*

1 Introduction

Pilgrimage (Hajj) of Muslims is considered the largest human gathering in which more than three millions Pilgrims move together through a very limited space in a short time period. This important event is repeated annually at the same time period and location, and the number of Pilgrims is increasing year after year. Moreover, Hajj is one of the main sources of gross national product in Saudi Arabia and very essential source of livelihood for many citizens and foreigners living in the Kingdom. In Hajj season, The Saudi government should provide many services to pilgrims such as security, food, housing, electricity, transportation and health care.

Of special interest for Saudi government is forecasting the yearly series of pilgrims coming from outside Saudi Arabia, denoted by NPO for short, since it is one of the most important indicators in determining the planning mechanism regarding these services for future Hajj seasons.

However, one may trace three important different procedures to forecast NPO. The first and well-known one is the methodology of Box and Jenkins. Their analysis has grown in popularity and is considered the prevailing methodology of time series analysis. They assume that a parsimonious stationary and invertible autoregressivemoving 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 (order determination), estimation, diagnostic checking, and forecasting). Their methodology have been widely used an explained by many others such as [2,3,4,5,6,7]. However, their identification technique is highly nonobjective and requires very good experience and skills.

The second approach of time series analysis is the Bayesian one. This approach is being developed and most of the Bayesian contributions have been occurred within the last three decades. For the theory and analysis of Bayesian procedure, the reader is referred to [8,9,10,5,11,12].

It is well known that ARIMA process are linear time series models; therefore they cannot represent the nonlinear behavior of many time series. The approach of nonlinear modeling the time series is suitable formost real-world problemsand the parametric models was developed to deal with them. To get an accurate forecast, the models must be known beforehand. Therefore, the model can not be used if the features of the data does not meet the assumptions of the model.

As an alternate to ARIMA models is the Artificial Neural Networks(ANN), as they have the ability to model the non linear problems. In addition to that, they are data-driven and do not require restriction on the data generation process.Therefore ANN are considered to be powerful methods when the mechanism of data generation is not known. Also ANN have the capability to globally approximate many kind of complex functional relationships [13].

The classical techniques for time series prediction require stationarity before performing the four phases of time series analysis, while most real time series are non-stationary. After ANN have been introduced, one can forecaste the original time series without the need to transform it to a stationary one.

The main objective of this article is to use the NARX approach to forecast the NPO and show that it gives better forecasts than Bayesian and Box-Jenkins approaches.

The remainder of this paper is structured as follows. Section 2 introduces and explains the proposed NARX network structure. Section 3 is devoted to forecast the NPO usingthe proposed NARX network. Section 4is dedicated to evaluate the forecast performance of the proposed NARX network and compare the achieved numerical results with Bayesian and the traditional Box and Jenkins approaches. Finally, the paper is concluded in Section 5.

2 NARX Networks

NARX model stands for"Nonlinear Autoregressive models with eXogenous input", and therefore it is called NARX recurrent neural networks [14,15,16]. The NARX network 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. NARX has been shown to be a powerful tool that suits non-linear systems modelling. NARX with gradient descent learning algorithm learns more efficiently than other NNT. In addition,NARX converges much faster and generalize better than other NNT [14,16].

2.1 NARX Archticture

The NARX model can be implemented by using feed forward neural network with embedded memory (first tapped delay line 'TDL') plus a delayed connection from the output of the second layer to input (second taped delay line) see Fig. 2.1.

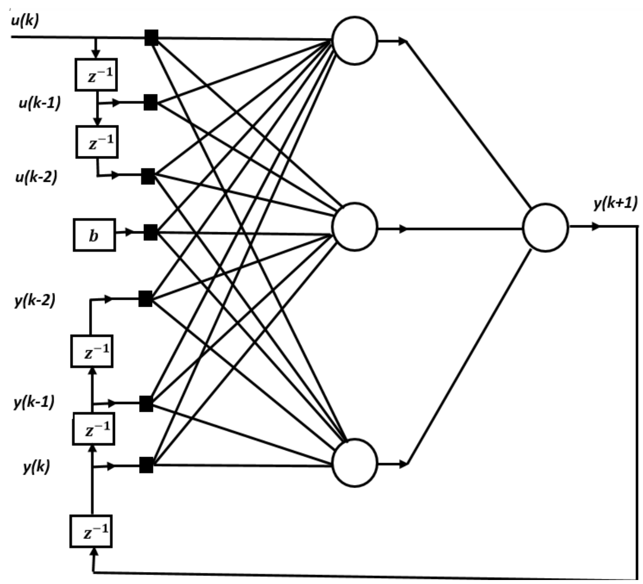


Fig. 2.1. NARX model with tapped delay line at input.

For the model shown in figure (2.1), let RNN(d_u,d_y;N) denote the neural net with d_u ≡ input delays, d_y ≡ output delays and N ≡ the number of neurons in the input layer. The general prediction equations for computing the next value of time series (output) y(k+1) using the model in figure (2.1) is the following:

$$y(k+1) = \Phi_0\Big[w_{b0} + \sum_{h=1}^{N} w_{ho}.\Phi_h\big(w_{ho} + \sum_{i=0}^{d_u} w_{ih}u(k-i) + \sum_{j=0}^{d_y} w_{jh}.y(k-j)\big)\Big]$$

Where u(k),u(k− 1),⋯ ,u(k − d_u) are the past observations, and y(k),y(k − 1),⋯ ,y(k − d_y) are the past outputs and are used as inputs in Fig. 2.1 [17].

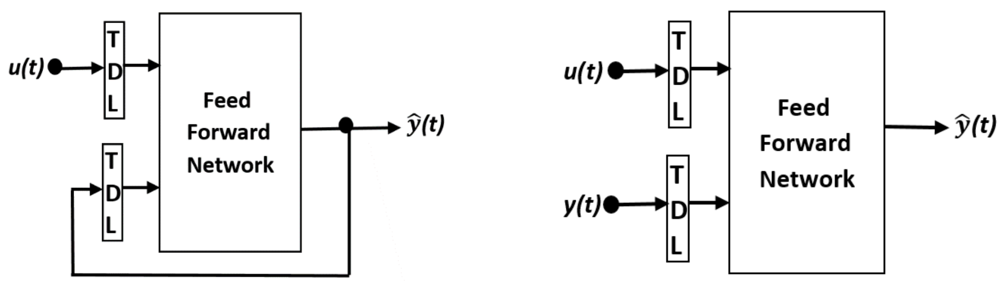
2.2 Learning Algorithm

To compute the gradient, it is necessary to use dynamic back-propagation algorithm; which is time consuming and more likely to trap at local minima [18].

The training method adopted in this paper uses the advantage of availability at the training time of the true output set.

"Output of NARX network can be considered as an estimate of the output of some nonlinear or dynamic system that we are trying to model. The output is fed back to the input of the feedforward neural network as part of the standard NARX architecture, as shown in Fig. 2.2-left. Because the true output is available during the training of the network, one could create a series-parallel architecture [19], in which the true output is used instead of feeding back the estimated output, as shown in Fig. 2.2-right. This has two advantages:

a. The first is that the input to the feedforward network is more accurate.
b. The second is that the resulting network has a purely feedforward architecture, and static backpropagation can be used for training." [20].



Parallel architecture **Series-parallel architecture**

Fig. 2.2. Architecture of NARX network.

All of the training is done in series-parallel mode, including the validation and testing steps. The typical workflow is to fully create the network in series-parallel mode, and only when it has been trained (which includes validation and testing steps) it is transformed to parallel mode for multistep-ahead prediction.

One of the problems that occur during training is over fitting. Due to this, error in early stage is very small, but when new data is presented to the network, the error is large.

The solution to this problem is Bayesian regularization [21]. The weight and bias values updated according to Levenberg-Marquardt [21] optimization. It minimizes a grouping of squared errors and weights, and generates a network that generalizes well.

3 NARX Analysis of NPO Data

The time series of number of pilgrims coming from outside the kingdom of Saudi Arabia (NPO) consists of 44 observations (from year 1390AH[1] to year 1433 AH).

In this experiment, NARX-net of *Matlab Neural Network Toolbox*was deployed to develop the prediction program for the future number of pilgrims, with 4-delayed input and 9-hidden layers and that configuration was chosen for best performance.

[1] *The years ate written using Lunar Calendar (AH) from 1390(1971) ~1433(2012); one lunar year is shorter than Gregorian year by about 11 days (see http://en.wikipedia.org/wiki/Islamic_Calenda).*

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 solely for forecasting evaluation. In statistical literature, 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 12% of the whole data) are reserved as the *hold-out sample (post-sample).* The first 39 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 percentages of the absolute errors were calculated. The results are reported in Tables 3.1.

Table 3.1. The last five observations and their forecasts

Year	Real observations	Forecasts	Percentage of the absolute error (P.A.E) (%)
1429	1729841	1695484	1.986
1430	1613965	1648932	2.167
1431	1799601	1706119	5.195
1432	1828195	1767791	3.304
1433	1752932	1725265	1.578
Mean			2.846

Inspecting the results given Table 3.1, one may conclude that NARX- model gives very good forecasts since the mean of the percentages of the absolute errors (P.A.E) is 2.846%. Fig. 3.1 shows the NPO forecasts of the last five observations, and Fig. 3.2 shows the NPO forecasts of the five future observations. It might be important to mention that the percentages of the absolute errors are calculated using the formula

$$P.A.E = \left| \frac{real\ \ observatio\ \ n - forecast}{real\ \ observatio\ \ n} \right| .100$$

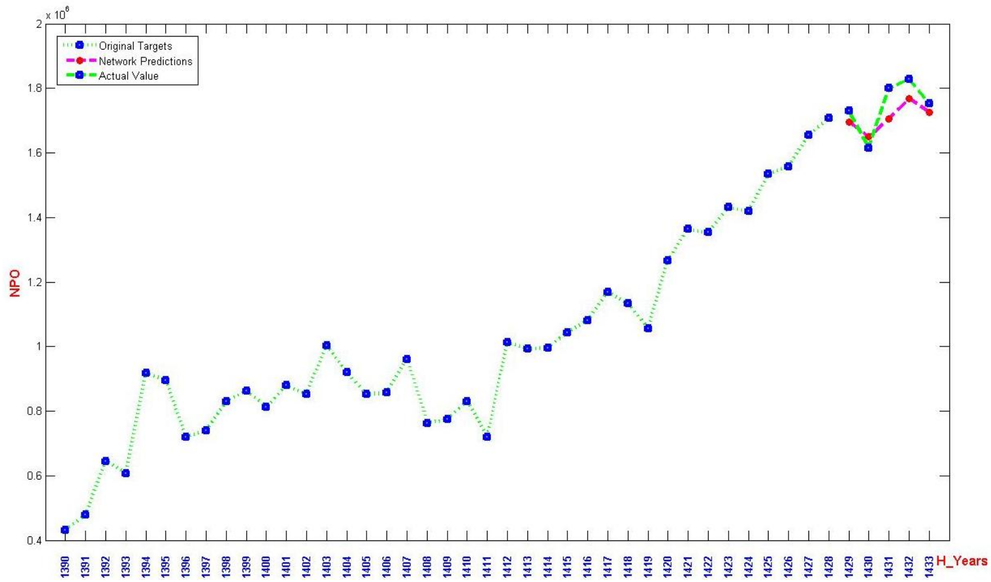


Fig. 3.1. The true NPO and forecasts of the last five observations

Finally, the NARX- model has been used to forecast the next five future observations. The point forecasts for these observations are given by Table 3.2.

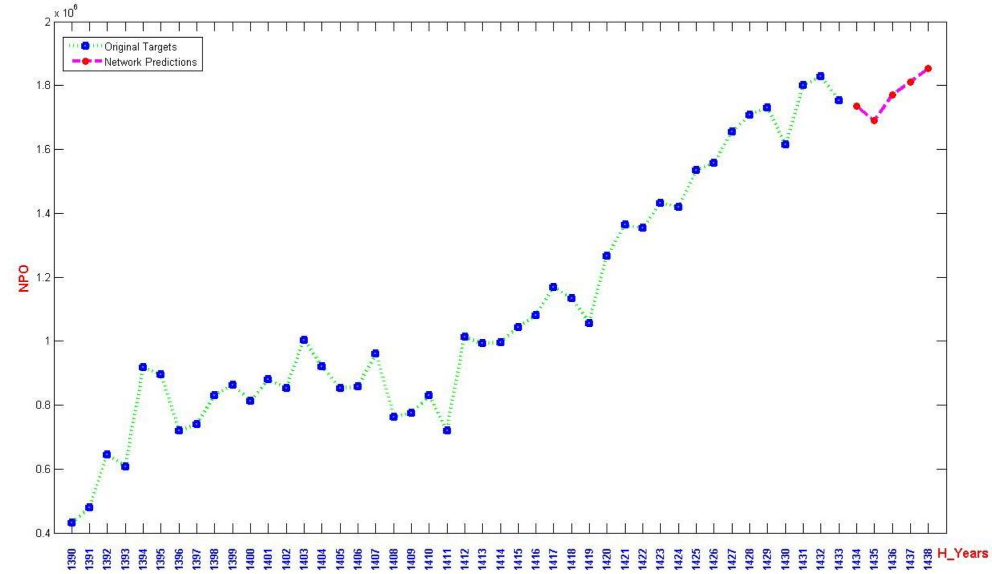


Fig. 3.2. The true NPO and future five observations

Table 3.2. The future five forecasts

Year	1434	1435	1436	1437	1438
Point forecast	1852816	1811862	1769373	1689825	1734789

4 A Comparative Study

As we have mentioned before, we analyzed the same NPO data using Bayesian and the traditional Box-Jenkins approaches [1]. In the previous section, we used their proposed evaluation technique to evaluate our proposed NARX approach in forecasting the hold-out sample (the last five observations).The main objective of this section is to compare the numerical results achieved by the proposed NARX approach, given in the previous section, with the numerical results achieved in [1]. In order to achieve the main objective, the following three criteria are calculated for the NARX, Bayesian and Box-Jenkins approaches and the numerical results achieved are reported in Tables 4.1.

$$MAPE = \frac{1}{m}\sum_{t=1}^{m}\left|\frac{true\ value - forecast}{true\ value}\right|.100$$

$$MAD = \frac{1}{m}\sum_{t=1}^{m}\left|true\ \ value{-}forecast\right|$$

$$RMSE = \left[\frac{1}{m}\sum_{t=1}^{m}(true\ \ value{-}forecast)^2\right]^{1/2}$$

Where, m is the total number of observations in the hold-out sample (post-sample).

Table 4.1. The forecasting performance of the proposed approaches

Criterion	Box-Jenkins	Bayesian	NARX
Mean Absolute Percentage Error(MAPE)	9.6	5.7	2.6
Mean Absolute Deviation(MAD)	175628.2	122409.6	64026.6
Root Mean Squared Error(RMSE)	196237	135948	19514

Inspecting the results given by Tables 4.1 one may conclude that our proposed NARX approach is much more accurate than Bayesian and Box-Jenkins approaches in forecasting the NPO.

5 Summary and Conclusion

The authors have proposed to use the NARX approach to forecast the series of number of Pilgrims coming from outside the Kingdom of Saudi Arabia from year 1390AH to year 1433AH. Point forecasts for the next five future years are provided by the authors using the proposed NARX approach. The forecasts of the last five observations achieved by our proposed NARX approach are compared with the results of Bayesian and Box-Jenkins approaches given in [1]. It has been shown that the proposed NARX approach gives much better forecasts thanthose achievedby Bayesianand the traditional Box-Jenkins approaches.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Shaarawy S, Khan E, Elgamal M. Modeling and forecasting the number of pilgrims coming from outside the Kingdom of Saudi Arabia using Bayesian and Box-Jenkins Approaches. International Journal of Advanced Computer Science and Applications (IJACSA). 2014;5(4):199-207.

[2] Box G, Jenkins G. Time series analysis, forecasting and control. Holden-Day, San Francisco; 1970.

[3] Harvey A. Time series models, 2nd edition. The MIT Press; 1993.

[4] Liu L. Time series analysis and forecasting. 2nd edition. Scientific Computing Association Corp, USA; 2009.

[5] Newbold P. Bayesian estimation of Box and Jenkins transfer function model for noise models. Journal of the Royal Statistical Society, Series B. 1973;35(2):323-336.

[6] Priestley M. Spectral analysis of time series. Academic Press, London; 1981.

[7] Wei WWS. Time series analysis: univariate and multivariate methods. Addison Wesley, Reading, MA; 2005.

[8] Broemling L, Shaarawy S. Time series analysis: a Bayesian analysis in the time domain. Bayesian Analysis of Time Series and Dynamic Model, edited by Spall J; 1988.

[9] Chow G. Multi-period prediction from stochastic difference equations by Bayesian methods. Chapter 8 of Studies in Bayesian Econometrics and Statistics, edited by S.E. Fienberg and A. Zellner. North-Holland, Amsterdam; 1975.

[10] Monahan J. Fully Bayesian analysis of ARIMA time series models. Journal of Econometrics. 1983;21:307-331.

[11] Shaarawy S, Ali S. Bayesian identification of seasonal autoregressive models. Communications in Statistics-Theory and Methods. 2003;32(5):1067-1084.

[12] Shaarawy S, Ali S. Bayesian identification of multivariate autoregressive processes. Communications in Statistics-Theory and Methods. 2008;37(5):791-802.

[13] Zhang GP. Neural networks for time-series forecasting, in Handbook of natural computing. G. Rozenberg, J. N. Kok, and T. Back, Eds., Springer, Berlin, Germany, 2012;461-477.

[14] Gao Y, Er MJ. NARMAX time series model prediction: feed forward and recurrent fuzzy neural network approaches. Fuzzy Sets and Systems. 2005;150(2):331-350.

[15] Haykin S. Neural networks, Second Edition, Pearson Education; 1999.

[16] Lin T, Bill GH, Tino P, Giles C. Learning long-term dependencies in NARX recurrent neural networks, IEEE Transactions on Neural Networks. 1996;7(6):1329-1351.

[17] DIACONESCU E. The use of NARX Neural Networks to predict Chaotic time series, 9th WSEAS International Conference on AUTOMATION and INFORMATION (ICAI'08), Bucharest, Romania; 2008.

[18] Mandic DP, Chambers JA. Recurrent neural networks for prediction. John Wiley & Sons;. 2001

[19] Narendra, Kumpati S, Kannan Parthasarathy. Learning automata approach to hierarchical multiobjective analysis. IEEE Transactions on Systems, Man and Cybernetics. 1991;20(1):263-272.

[20] Mathworks; 2014. Available:http://www.mathworks.com/help/nnet/ug/design-time-series-narx-feedback-neural-networks.html.

[21] Kaur H, Salaria DS. bayesian regularization based neural network tool for software effort estimation. Global Journal of Computer Science and Technology Neural & Artificial Intelligence; 2013.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=735&id=6&aid=7681