Journal of Animal & Plant Sciences, 31(5): 2021, Page: 1477-1484
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2021.5.0349

COMPARISON OF ARFIMA, ARIMA AND ARTIFICIAL NEURAL MODELS TO FORECAST THE TOTAL FISHERIES PRODUCTION IN INDIA

Z. Chikr Elmezouar[1,2,3*], Ibrahim. M. Almanjahie[1,2], I. Ahmad[4], and A. Laksaci[1,2]

[1]Department of Mathematics, College of Science, King Khalid University, Abha, 62529, Saudi Arabia.
[2]Statisticsl Research and Studies Support Unit, King Khalid University, Abha, 62529, Saudi Arabia.
[3]Department of Mathematics, University Tahri Mohamed, Bechar 8000, Algeria.
[4] Department of Mathematics and Statistics, Faculty of Basic and Applied Sciences, International Islamic University, Islamabad, Pakistan.
[*]Corresponding Author's E-mail: zchikrelmezouar@kku.edu.sa

ABSTRACT

Autoregressive Integrated Moving Average (ARIMA) modeling is a statistical technique used for time series data in order to understand and forecast future trends in a better way. Recently, the ARIMA models have been employed in practice for modeling the data of total fisheries production in India. In this study, an important family of parametric time series modeling when the order of difference is fractional, called Autoregressive Fractional Integrated Moving Average (ARFIMA), has been proposed for modeling and forecasting the total fisheries production (metric tons) in India. For testing the fundamental assumption of stationarity, Augmented Dickey Fuller (ADF) test was used. We also used a nonparametric model such as Neural Network Autoregressive (NNAR) for investigating the behavior of the data. After the evaluation of different models and perform comparisons based on root mean square error (RMSE) and mean absolute percentage error (MAPE) values, the result indicated that ARFIMA (3, 0.48,0), ARIMA(1,2,1) and NNAR(3,1) were the best models. The current results reflected that ARFIMA model outperformed ARIMA and NNAR models in forecasting the total fisheries prediction. This could be suggested that the ARFIMA might be a remarkable selection for time series data modeling.

Key words: Time Series, Trends, ARIMA, ARFIMA, NNAR and Forecasting.

Published first online January 21, 2021 Published final August 07, 2021.

INTRODUCTION

Time-series observations are normally considered as a sequence taken serially over time. Understanding and describing the generating mechanism and forecasting future values based on past values are primary objectives of time series. To this end, statisticians have a well-known methodology based on regression analysis and linear time series models. Through literature, these statistical models with different techniques applied to certain phenomena to study and forecast the behavior of future values for the problem under study. Statistical methodologies such as regression analysis, univariate and multivariate time-series approaches were used by Stergiou *et al*. (1997) for modeling and forecasting monthly fisheries catches. These models also were used by Venugopalan and Srinath (1998) for modeling and forecasting the quarterly commercial landings of marine fishes.

Linear time-series models such as ARIMA was also used for analyzing and forecasting by many researchers. Sathianandan and Alagaraja (1998) used spectral time series models for studying analyzing marine fish species such as the Bombay duck, Mackerel and Oil sardine in India. Their results showed that the behavior of these species dipected a cyclical fluctuation and spectral decomposition of all India landings. Noble (1980) examined a ten-year cycle in the Mackerel fishery and investigated the forecasting of future values for the Mackerel fishery. ARIMA methodology was employed by Noble and Sathianandan (1991) for exploring the trend analysis in Mackerel catches of India. Sathianandan and Srinath (1995) conducted a study based on a time series analysis of India's marine fish landings.

The ARIMA models were reported to be quite the most popular used method (Zhang, 2003). However, the ARIMA modeling cannot capture many important features. Hence, a better option is needed. Fortunately, nowadays, time-series analysis has entered in a new field of a nonlinear domain. It is generally more suitable for better understanding and accurate description of time series dynamics during better multi-step-ahead forecasts, especially when the time series data is nonlinearly related to its past values (Fan and Yao, 2003). An elegant study of the all-India landings of oil sardine, Bombay duck and Mackerel using spectral decomposition was completed by Sathianandan and Alagaraja (1998). In another study, Nampoothiri and Balakrishna (2000) applied the nonparametric Threshold Autoregressive model for a time series data. Bishal *et al*. (2016) evaluated oil sardine landings for the period 1961-2008 in Kerala using nonlinear Exponential Smooth Transition Autoregressive (ESTAR) and Genetic Algorithm (GA) methods. Raymond *et al*. (1999) used artificial neural network (ANN), nonparametric approach, for time series data to model and predicted the fish yield in France based on a combination of some variables related to the environmental characteristics. Sun (2009) also used a new approach based on the ANN and produced a formula for forecasting fish stock recruitment. Mahalingaraya *et al*. (2018) compared the ARIMA model with ANN model and based on empirical results and found that the machine learning techniques outperformed the ARIMA model. The ANN is a general approach; however, with time series data, this model cannot mimic the actual process perfectly. An alternative nonparametric model that is capable of time series data and combines the approaches of ANN and Autoregressive process leading to a concise model is called Neural Network Autoregressive (NNAR). In this study, we will use this type of model and compare its performance to the candidate models.

One substantial family of parametric time series models, such as ARFIMA, has also been used to model and forecast the time series data. An important study for modeling and forecasting marine fish in Malaysia based on ARIMA and ARFIMA models was achieved by Shitan *et al*. (2008). Their results revealed the preference of the ARIMA. However, in this study we investigate the performance of ARIMA, ARFIMA and NNAR techniques for the total fisheries production in India and our study results reflected that ARFIMA model outperformed ARIMA and NNAR models in modelling and forecasting. This could be suggested that ARFIMA might be a remarkable selection for time series data modelling. Based on our knowledge, there is no study compared these candidate models and confirmed that the AFRIMA could be a better choice for modelling and forecasting the total fisheries production.

The rest of the paper is outlined as follows. Section "Materials and methods" deals with our methodology and details the used models. Section "Results and discussion" is devoted to the results and discussion. Finally, our conclusion is stated in "Conclusion" section.

MATERIALS AND METHODS

Economically speaking, fisheries production is more economical than livestock due to the lower costs involved. Developing reliable models for studying and analyzing the total fisheries production is usually of interest. In this study, we consider the total fisheries production time series data (metric tons) in India from the period 1960-2015. This time series data is taken from World bank website at https://data. worldbank. org/ indicator/ER.FSH.PROD.MT?locations=IN ARFIMA, ARIMA and NNAR models are considered for modelling and forecasting. Background regarding these models and methodology assessment are detailed below.

Autoregressive Integrated Moving Average (ARIMA) model: Consider that the random variable $\{Y_t\}$ can be modeled by ARIMA (p,d,q). Then, we define the $\{Y_t\}$ process by

$$\emptyset(B)(1-B)^d Y_t = \Theta(\mathrm{B})\epsilon_t,$$

where p represents the autoregressive terms, d indicates number of differences and carries positive integer values and q shows moving-average terms. The operator terms $\emptyset(B)$ and $\Theta(\mathrm{B})$ are given by

$$\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots \emptyset_p B^p$$

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots \theta_q B^q$$

which represent the AR and the MA operators of orders p and q, respectively.

Autoregressive Fractional Integrated Moving Average (ARFIMA) model: Processes with long memory are found in many real-life applications with fundamental importance in time series analysis. ARFIMA models were presented by Granger and Joyeux (1980) and also by Hosking (1981) to work with long memory series under discrete time domain. An $ARFIMA\,(p,d,q)$ process is given by

$$\emptyset(B)(1-B)^d Y_t = \Theta(\mathrm{B})\epsilon_t,$$

where $\{Y_t\}$ is the phenomenon of interest and $\epsilon_t \sim WN(0,\sigma^2)$; B is known as the backward shift operator; $\emptyset(B)$ and $\Theta(\mathrm{B})$ are recognized as polynomials of p and q degrees and represent the AR and the MA parts respectively. The operator $(1-B)^d$ is known as the fractional differencing operator defined by

$$(1-B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)B^k}{\Gamma(k+1)\Gamma(-d)}$$

with $\Gamma(.)$ is the gamma function. ARFIMA (p, d, q) is a stationary as well as invertible process if $|d| \leq 0.5$ with roots of the $\emptyset(B)$ and $\Theta(\mathrm{B})$ outside the unit circle.

The Neural Network Autoregressive (NNAR) model: The NNAR is constructed by using the lagged values as inputs to a neural network. For simplicity, we abbreviate the NNAR model by using $NNAR(p,k)$ which considers the p lagged inputs as $(y_{t-1} \dots, y_{t-p})$ and k refers to the number of nodes; for more details see Hyndman and Athanasopoulos (2019). Figure 1 shows the nonlinear version of the NNAR. The relationship between the output $(\hat{y}_t)$ and the inputs $(y_{t-1}.y_{t-2}.\dots.y_{t-p})$ has the following mathematical representation:

$$\hat{y}_t = \upsilon_0\left\{W_{00} + \sum_h W_{h0}\vartheta_h\left(W_{ch} + \sum_i W_{ih}y_{t-i}\right)\right\} + \epsilon_t \qquad (3.7)$$

where $\{W_{ch}\}$ denotes the weights for the connections between the constant input and the hidden neurons and $\{W_{00}\}$ denotes the weight of the direct connection between the constant input and the output. The weights $\{W_{ih}\}$ denote the weights for the connections between the inputs data and the hidden neurons. The weight $\{W_{h0}\}$ denotes the weights for the connections between the hidden neurons and the output. The functions ϑ_h and υ_0 denote the activation functions used at the hidden layer and at the output, respectively.

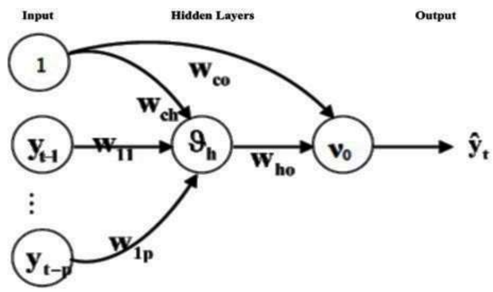

Figure 1: NNAR for time series forecasting with p inputs one hidden layer of one neuron.

Statistical test for stationarity: Before applying any of the above models, the assumption stationarity for time series should be checked. We used the Augmented Dickey Fuller Test (ADF) test since it is a widely accepted statistical test being used to check whether or not a given time series is stationary. The null hypothesis for this test is that the data are non-stationary. In case the time series is stationary, the next step is modeling the time series data by applying the above models but the number of terms that are needed for achieving a valid model is questionable. In this article, we used and examined the ACF (Autocorrelation) and PACF (Partial Autocorrelation) plots technique to determine AR and MA terms needed.

Model Evaluation criteria: In the present study, evaluation criteria based on root mean square error (RMSE) and mean absolute percentage errors (MAPE) are considered to check and determine the performance of the applied models. The expression of RMSE is represented by

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}}$$

where $e_t = Y_t - \hat{Y}_t$. In particular, the terms of Y_t and $\hat{Y}_t$ depicts the actual observation and fitted value at time t, respectively. The expression of MAPE is represented by

$$MAPE = \left(\left(\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right|\right)/n\right)100\%$$

Also, we used the well-known R^2 and adjusted R^2 for more checking and justification.

RESULTS AND DISCUSSION

An attempt was made in the current study to evaluate the predictive performances of three modeling techniques, i.e., ARFIMA, ARIMA and NNAR using the total fisheries production (metric tons) data in India. Figure 2 indicates the time series plot of total fisheries production (metric tons) in India.

The test of stationarity: ARFIMA, ARIMA, NNAR models provide the assumption of stationarity. To investigate the stationarity, we need to assure through the unit root test. There are no certain principles to identify the exact approach need to be adopted in a particular scenario. Therefore, in the present study for the total fisheries production, the unit root based on augmented Dickey and Fuller (ADF) procedure was implemented. In this test, the alternative hypothesis is that the process is stationary. Table 1 shows the results of the executed test. With a lag order of 3, the time series becomes stationary since the ADF test indicates statistically significant ($p-$ value < 0.05) when the integrated order equals 2, i.e., I(2).

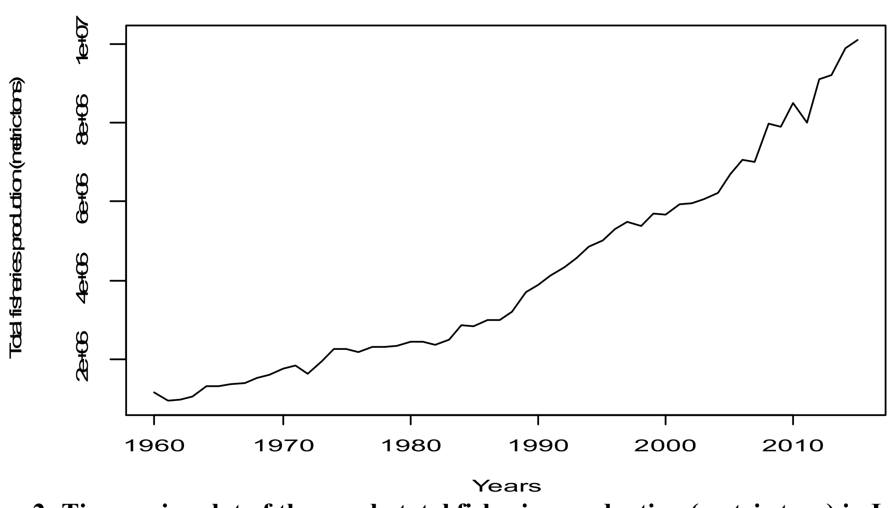

Figure 2: Time series plot of the yearly total fisheries production (metric tons) in India.

Table 1: Augmented Dickey-Fuller Test.

Integrated	Dickey-Fuller	Lag order	p-value
$d = 0$	-2.3944	3	0.4157
$d = 1$	-2.6486	3	0.3135
$d = 2$	-5.1952	3	0.010

Inspections Autocorrelation Function (ACF)and Partial Autocorrelation Function (PACF): Inspecting ACF and PACF are essential to identify the order of the MA and AR processes. Figures 3(a) and 3(b) shows the ACF and PACF graphs of the second difference of the basic series. These figures depict the behavior of ACF and PACF of the total fisheries production (metric tons) in India.

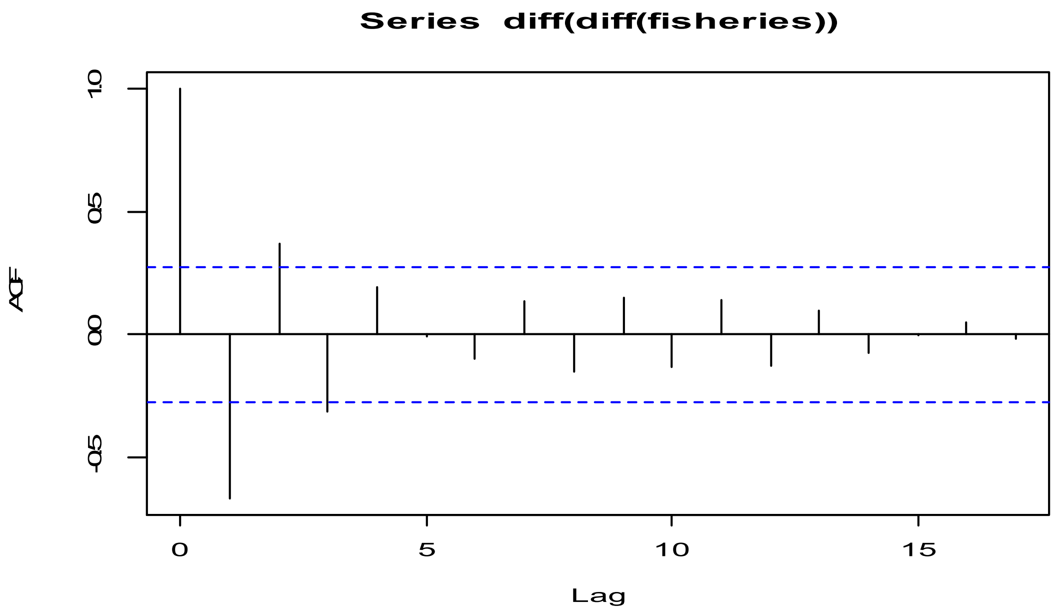

Figure 3(a) : ACF of the second difference of the total fisheries production

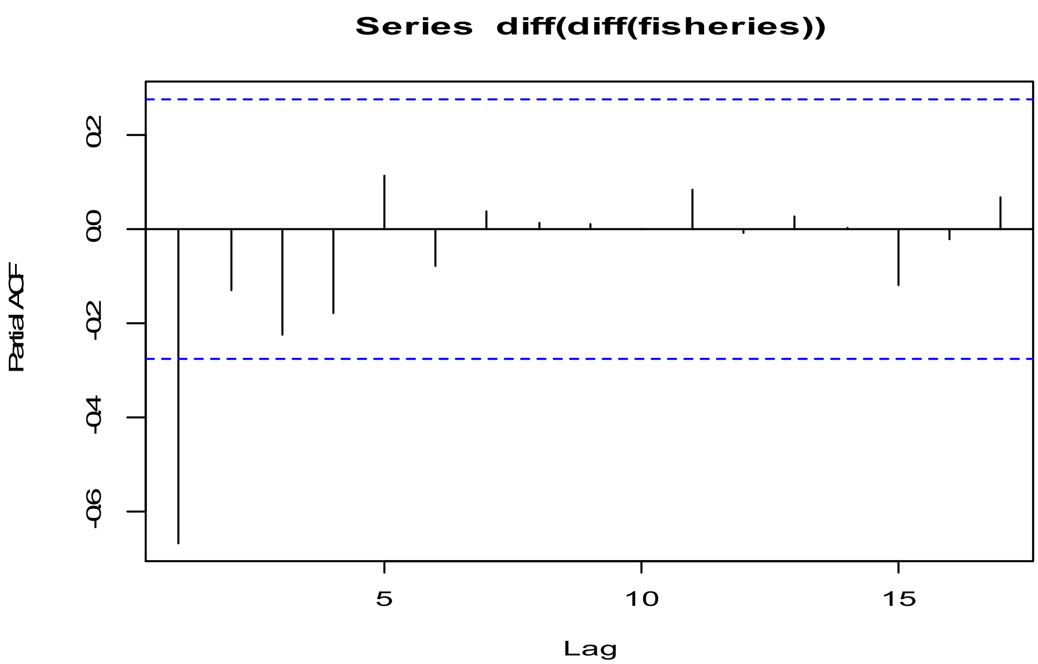

Figure 3(b) : PACF of the second difference of the total fisheries production"

Based on the result displayed in Figure 3(a), the ACF pattern depicts a single negative spike at lag 1, giving an indication of $MA(1)$, while in Figure 2(b) the pattern of PACF with a single negative spike at lag 1 is evidence of $AR(1)$.

Fitting of ARIMA, ARFIMA and NN models

Fitting ARIMA Model: Here, it was aimed to select the best one among candidate ARIMA models for the total fisheries production in India. From Figure 3, it was understood that $ARIMA$ (1,2,1) model might be the best one among candidate ARIMA models. However, we used the previously mentioned criteria to test the candidate ARIMA models. The comparison results are presented in Table 2.

It was determined that the model $ARIMA\ (1, 2, 1)$ was the best model for modeling the total fisheries production with the smallest values produced by RMSE and MAPE (Table 2). Also, the results given by the R^2 and adjusted R^2 indicated that the predictor(s), based on $ARIMA\ (1, 2, 1)$, accounted for 75% explained variability towards the total fisheries production. After the selection of the best ARIMA model, the next step is the estimation of the parameters of the selected model. The estimation results of the best-selected model are shown in Table 3.

The residual plots of ACF and PACF, for $ARIMA$ (1,2,1) model, are presented in Figure 4. The plots show the correlation among residuals of the series and this result fulfil the purpose of the designated criterion.

Table 2: Comparison of candidate ARIMA models.

Model	RMSE	R^2	Adjusted R^2	MAPE
ARIMA (1,2,1)	221156	0.75	0.74	0.0176
ARIMA (1,2,0)	251694	0.71	0.71	0.0593
ARIMA (0,2,1)	262221	0.69	0.68	0.0488

Table 3: Model estimation from 1960 to 2012 for ARIMA (1,2,1).

Coefficients the parameters	Coefficients	standard error	p-value
AR1	-0.67	0.13	< 0.001
MA1	-0.80	0.11	< 0.001

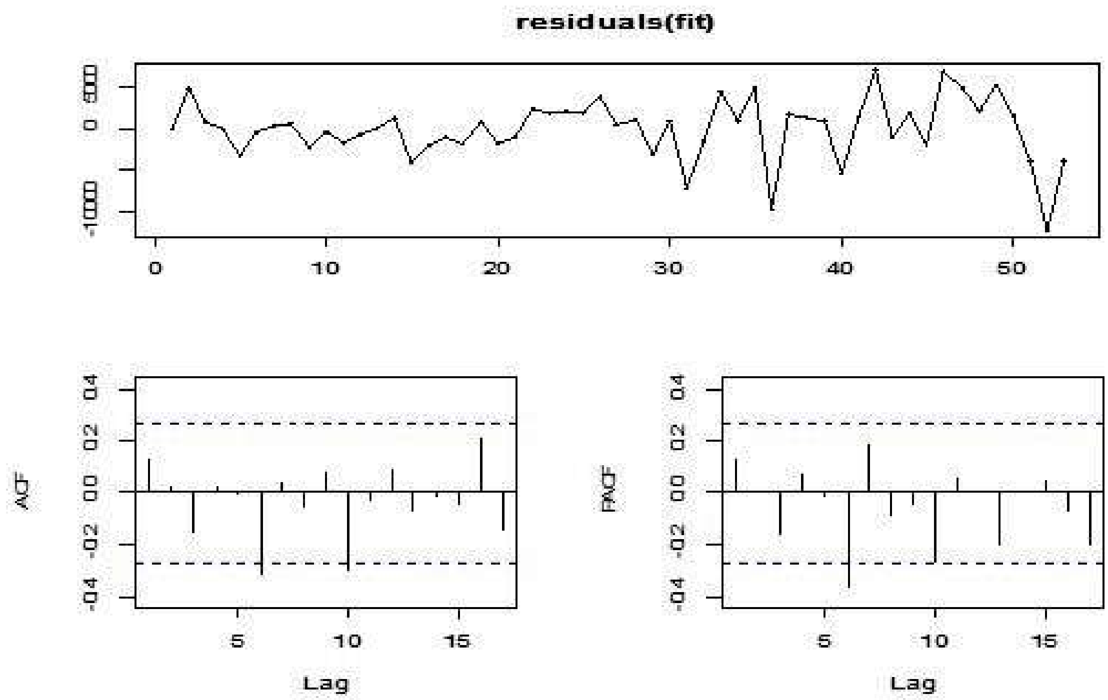

Figure 4: Residual ACF and PACF of ARIMA (1,2,1)

Fitting ARFIMA Model: In this study, the ARFIMA model was also evaluated for total fisheries production data and by using the criteria mentioned early with careful checking, we found $ARFIMA(3,0.48,0)$ as the most appropriate model for this data among the class of other ARFIMA models. The estimates of the parameters are reported in Table 4.

Table 4: Results of Estimating the parameters.

Coefficient	Estimate	P-value
D	0.48	0.002
AR1	0.63	< 0.001
AR2	0.78	< 0.001
AR3	-0.43	< 0.001

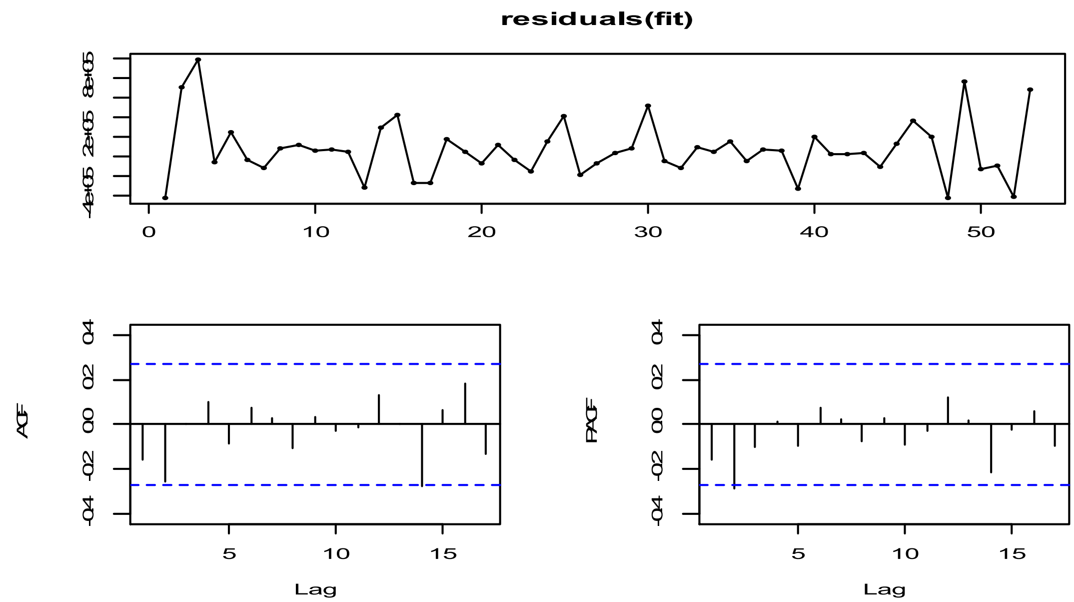

Figure 5: ACF and PACF of ARFIMA (3,0.48,0).

The plots in Figure 5 show the absence of serial correlation in the residual's series. Therefore, $ARFIMA(3,0.48,0)$ is considered an adequate choice.

Fitting NNAR Model: We fitted NNAR with 3 inputs and 1 neuron and 1 output (i.e. 3-1-1 network with 6 weights). We find that the $RMSE = 200299.8$ and $MAPE = 0.04283$ for training.

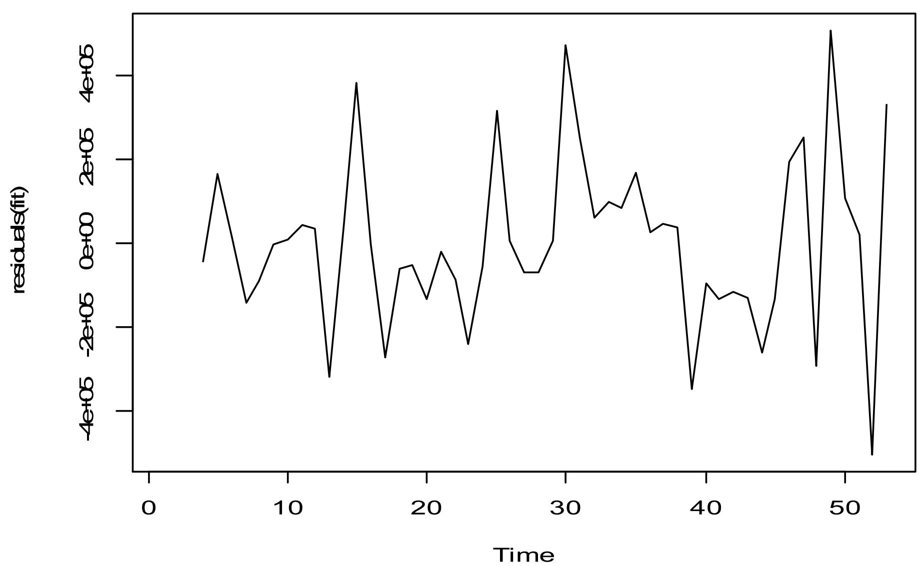

Figure 6: Residuals of NNAR fit

Figure 6. Show that the errors are independent. In other words, there is no absolute threshold seems in the fitted forecasted model.

Forecast of ARIMA, ARFIMA and NNAR Models: After the selection of ARIMA(1,2,1), ARFIMA(3,0.48,0) and NNAR(3,1), we move further towards the forecasting step. Forecasting results of ARIMA, ARFIMA and NNAR models for the next three years (2013-2015) are determined and shown in Table 5.

Table 5: Forecasting performance of ARFIMA and ARIMA models.

Forecast	ARIMA	ARFIMA	NNAR	Original
2013	8833440	**8837405**	8763791	9222391
2014	9471599	**9663359**	9480414	9884999
2015	**9500459**	9468081	9353054	10100057

From Table 5, the predicted values of ARFIMA (3, 0.48, 0)) are in closer agreement to the observed values as compared to the predicted values of ARIMA(1, 2, 1) and NNAR(3,1), especially in the first two years. To identify the parsimonious model between these types of model, we finally compare the forecasted values of ARFIMA (3, 0.48, 0), ARIMA (1, 2, 1) and NNAR(3,1) with the observed values by computing the RMSE and MAPE.

Table 6: Comparison between the two models in forecasting.

Model forecast	RMSE	MAPE
ARFIMA (3,0.49,0)	**445994**	**0.0422**
ARIMA (1,2,1)	476690	0.0477
NNAR (3,1)	557379	0.0548

MAPE can be considered as a statistical tool to evaluate how accurate the forecast results. Generally, from Table 6, the forecast results illustrate significant concluding remarks as the MAPE is less than 5% for most of the selected models. This result is considered excellent according to the rough rule of thumb to predict values. Moreover, the values of RMSE and MAPE are less for $ARFIMA\,(3, 0.48, 0)$ than those for $ARIMA\,(1, 2, 1)$ and $NNAR(3,1)$. We, therefore, conclude that ARFIMA model is a better choice than the ARIMA model and NNAR.

Conclusion: In this paper, we carried out a time series analysis and forecasting of total fisheries production (metric tons) in India. The ACF of the total fisheries production depicted the persistence feature of a long memory process. This leads to the $ARFIMA(3,0.48,0)$ model as an appropriate model for fitting such data. Similarly, we also employed the most commonly used ARIMA. An opposite $ARIMA(1, 2, 1)$ model was recognized and fitted. The data is also fitted using the $NNAR(3,1)$ model in order to compare its performance with the other proposed models. Even though both models fit the data very well, forecasts obtained using the $ARFIMA(3,0.48,0)$ model are in close agreement to the actual values as compared to the forecasts obtained from $ARIMA(1,2,1)$ and $NNAR(3,1)$ models. It is also worth to note that forecast evaluation using RMSE and MAPE revealed the superiority of the ARFIMA model than ARIMA and NNAR models. We, therefore, conclude that the total fisheries production data could be modeled in a better way by using the ARFIMA model than ARIMA and NNAR models.

Acknowledgments: The authors are very grateful to the Deanship of Scientific Research at King Khalid University, Kingdom of Saudi Arabia, for funding this work through research groups program under Project no. R.G.P.1/189/41.

REFERENCES

Bishal Gurung, K. S. N., P. Singh Nilakanta and A. Grover (2016). Analysis of cyclical fish landings through ESTAR nonlinear time-series approach. *Indian J. Fish.,* 63(2): 110-113.

Fan, J. and Q. X. Yao (2003). *Nonlinear time series: Nonparametric and parametric methods*. Springer, New York.

Granger, C. W. J. and R. Joyeux (1980). An Introduction to Long-Memory Time Series Models and Fractional Differencing. J. Time Series Analysis, 1(1): 15-29.

Hyndman, R. J. and G. Athanasopoulos (2019). *Forecasting: Principles and Practice*, 3nd edition, OTexts: Melbourne, Australia.

Hosking, J. R. M. (1981). Fractional Differencing. *Biometrika*, 68(1): 165-176.

Mahalingaraya, S. Rathod, K. Sinha, R.S. Shekhawat and S. Chavan (2018). Statistical modeling and forecasting of total fish production of India: a time series perspective. *Int. J. Curr. Microbiol. App. Sci*., 7(3): 1698-1707

Nampoothiri, C. K. and N. Balakrishna (2000). Threshold autoregressive model for a time-series data. *J. Indian Soc. Agr. Stat*., 53: 151-160.

Noble, A. (1980). Is there a ten-year cycle in the mackerel fishery? *Seafood Exp. J.*, 12(4): 9-13.

Noble, A. and T. V. Sathianandan (1991). Trend analysis in all India mackerel catches using ARIMA models. *Indian J. Fish.,* 38(2): 119-122.

Raymond, L., B. Sovan and M. Jacques (1999). Predicting fish yield of African lakes using neural networks. *Ecological Modelling*, 120: 325-335.

Sathianandan, T. V. and K. Alagaraja, (1998). Spectral decomposition of the all India landings of oil sardine, mackerel and Bombay duck. *Indian J. Fish.,* 45(1): 13-20.

Sathianandan, T. V. and K. Srinath (1995). Time series analysis of marine fish landings in India. *J. Mar. Biol. Ass. India*, 37(2): 171-178.

Shitan, M. P. M. J. Wee, L. Y. Chin and L. Y. Siew (2008). ARIMA and integrated ARFIMA models for Forecasting Annual Demersal and Pelagic Marine Fish Production in Malaysia. Malaysian J. Mathematical Sciences, 2(2): 41-54.

Stergiou, K. I., E. D. Christou and G. Petrakis (1997). Modelling and forecasting monthly fisheries catch: Comparison of regression, univariate and multivariate time series methods. *Fish. Res.,* 29: 55-95.

Sun, L. (2009). Forecasting Fish Stock Recruitment and Planning Optimal harvesting strategies by Using Neural Network. J. Computers, 4(11): 1075-1082.

Venugopalan, R. and M. Srinath (1998). Modelling and forecasting fish catches: comparison of regression, univariate and multivariate time series methods. *Indian J. Fish.,* 45(3): 227-237.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing,* 50: 159–175.