Indian Journal of Geo Marine Sciences
Vol. 51 (09), September 2022, pp. 776-784
DOI: 10.56042/ijms.v51i09.2337

Comparative study of statistical and machine learning techniques for fish production forecasting in Andhra Pradesh under climate change scenario

S K Stephen[a], V K Yadav*[,a] & R R Kumar[b]

[a]ICAR-Central Institute of Fisheries Education (CIFE), Panch Marg, Off Yari Road, Versova, Andheri (W), Mumbai – 400 061, India
[b]ICAR-Indian Agricultural Statistical Research Institute (IASRI), Library Avenues, Pusa, New Delhi – 110 012, India

*[E-mail: vinodkumar@cife.edu.in]

Received 18 January 2022; revised 18 September 2022

The present study emphasizes the forecast of Andhra Pradesh's total marine fish production and the catch of commercially important fishes, *viz*., Indian Mackerel, Oil Sardine, Horse Mackerel, and Lesser Sardines for the next 5 years by different statistical and machine learning approaches under climate change scenario. Forecasting is done with and without the inclusion of climatic and environmental parameters in different models. Exogenous variables, *i.e*., climatic parameters such as Sea Surface Temperature (SST), wind speed, and environmental parameters such as Chlorophyll-*a*, diffusion attenuation coefficient, and Photo-synthetically Active Radiation (PAR), were used in the model. The following models like Non-linear Autoregressive (NAR) Artificial Neural Network (ANN) (NNAR-ANN), Auto-Regressive Integrated Moving Average (ARIMA), Empirical Mode Decomposition based Artificial Neural Network (EMD-ANN), are used to predict the fish catch data using time series quarterly catch data of commercially important fishes and total fish catch without the inclusion of climatic and environmental variables. Auto Regressive Integrated Moving Average method with inclusion of exogenous variables (ARIMAX) and Non-Linear Auto Regression with exogenous variables (NARX) models were used to forecast along with quarterly average data of environmental and climatic variables. The model developed predicts the total fish catch and also the catch of commercially important fish for the next 20 quarters. The developed model forecasts are compared based on the error measure, *i.e*., MAPE (Mean Absolute Percentage Error), and the results showed that the NARX model outperformed other models like ARIMAX, ARIMA, NNAR-ANN, and EMD-ANN. Implementation of management strategies considering the impact of climate change on fisheries will enhance sustainable fisheries and pave a pathway for the mitigation of climate change.

[**Keywords**: ARIMAX, EMD-ANN, Climate change, Marine fish production, NARX]

Introduction

A developing country like India faces a critical threat in terms of climate change, which is a major environmental problem across the globe. Climate change due to the environment continues to affect marine ecosystems. The change in temperature that drives dissolved oxygen, pH, salinity, and ocean currents and mix surface waters with deeper waters. The nutrient-rich waters affect the distribution and abundance of plankton, which is food for small fish and thus influence the total catch of fish.

As climate change has a very intense effect on the fisheries sector[1,2], certain management strategies and management decisions to mitigate the drastic changes in climate by policymakers help to achieve the goals of fisheries management through forecasted fisheries time-series data.

Active research works have been going on for years on time series modelling and forecasting as they have primordial importance in practical domains. Many models were studied in detail to understand the efficiency and accuracy of time series modeling and forecasting. Statistical models such as Auto-Regressive Integrated Moving Average (ARIMA) methods are used to predict the time series data. The assumption of linearity is the main disadvantage of this model. The time series has components that are non-linear which makes ARIMA models not appropriate for forecasting and modeling, so, the Artificial Neural Network (ANN) is the most commonly used machine learning technique to model and forecast the time series data with non-linear components. The empirical Mode Decomposition-based Artificial Neural Network (EMD-ANN) approach that hybridizes EMD and ANN is also an alternative to the models which does not capture the nonlinearity in the data. EMD decomposes a signal into a set of adaptive basis functions called intrinsic mode functions[3].

Many works have been done by researchers on short-term forecasting using renowned methods and models like Vector Auto-Regression (VAR), Auto-Regressive Integrated Moving Average (ARIMA), Neural Networks, etc.; but in the Indian context, forecasting marine fish catch under climate change scenario is limited. In India, many research works were done to forecast the marine fish catch with environmental variables by different models[4,5]. To model, the series with exogenous variables (environmental, climatic variables), Auto Regressive Integrated Moving Average with exogenous variables (ARIMAX), and Non-Linear Auto Regression with exogenous variables (NARX) model are used in this study. Machine learning technique, NARX model, captures the non-linear patterns in the data and has better accuracy in the prediction[6].

Environmental and climatic variables were used to amplify the time series prediction performance. Models such as ARIMAX and NARX were used to predict the total fish catch landings and catch of commercially important fishes like Oil Sardine, Horse Mackerel, Indian Mackerel, and Lesser Sardines.

Materials and Methods

Study area and data

The present study was done in Andhra Pradesh, located on the east coast of India. The state comprises of 13 districts, in which 9 are coastal districts. The state has a long coastline of 974 km, with 2.97 lakh tonnes of annual marine fish production, and has been contributing significantly to the country's total fish production for the past few years in brackish water aquaculture, freshwater aquaculture, and marketing through an effective strategy. The study emphasizes secondary data of total marine fish catch, some of the important commercial fish catch data, along with data on climatic and environmental variables.

Quarterly average data of climatic variables [Sea Surface Temperature (SST), wind speed], environmental variables [diffuse attenuation coefficient (Kd_490 or Kd), chlorophyll-*a* (Chl-*a*), and Photosynthetically Active Radiation (PAR)] along Andhra Pradesh, was obtained from GIOVANNI for the period of 2002 – 2018. The quarterly landings of total fish catch and the catch of Indian Mackerel, Oil Sardine, Horse Mackerel, and Lesser Sardine fish estimate value were taken from ICAR-Central Marine Fisheries Research Institute (ICAR-CMFRI), Kochi for the same period to understand the impact of environmental and climatic variables on fish catch.

Ninety percent of the data, *i.e*., from the first quarter of 2002 to the fourth quarter of 2016 (60 data points from 1 to 60-time stamp as horizontal axis) were used as training data, and around 10 % of the data, *i.e*., from first quarter 2017 to fourth quarter 2018 (8 data points, *i.e*., from 61 to 68-time stamps as horizontal axis), were used to test and validate the data for all the models. Chl-*a*, SST, Wind speed, Kd, and PAR were expressed in mg/m^3, °C, m/s, m^{-1}, and Einstein/m^2/day, respectively, and fish landings were measured in ton.

Methodology

Auto-Regressive Integrated Moving Average (ARIMA)

ARIMA stands for Auto-Regressive Integrated Moving Average Method, in which auto regression is indicated by a pattern of growth or decline. Integration is indicated by a rate of change in the growth or decline in the data, and the noise between consecutive time points indicates the moving average.

An ARIMA model is given by:

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \quad \ldots (1)$$

Where,

$\phi(B) = 1-\phi_1 B-\phi_2 B^2 -\cdots-\phi_p B^p$ (Autoregressive parameter); $\theta(B) = 1-\theta_1 B-\theta_2 B^2 -\cdots-\theta_p B^q$ (Moving average parameter); ε_t = white noise or error term; d = differencing term; B = Backshift operator *i.e*. $B^a Y_t = Y_{t-a}$.

ARIMAX

The generalization of the ARIMA model is the ARIMAX model, which includes an exogenous input variable. The model ARIMA error and the presence of exogenous variables is ARIMAX (p,d,q,k) model. In statistical form, ARIMAX model can be written as:

$$Y_t = \theta_0 + \sum_{i=1}^{p}\delta_i Y_{t-i} + \sum_{j=1}^{q}\theta_j e_{t-j} + \sum_{l=1}^{k} b_l F_{t-l-1} + e_t \quad \ldots (ii)$$

Where, F is an exogenous variable of order k

Using back operator, ARIMAX model[8] may be written as:

$$y_t = \varphi(B^s)\phi(B)\nabla^d\nabla_s^D y(t) + \Theta(B^s)\theta(B)\varepsilon(t) + \beta x_t \quad \ldots (iii)$$

Artificial Neural Network (ANN)

Forecasting of non-linear data can be done in various ways; amongst them, the most efficient model is Artificial Neural Network[5,9]. Neural Networks are widely used for the forecasting of fish landings as the data of fish landings is highly non-linear. Neural networks have an input layer, a hidden layer, and an output layer with simple non-linear computing units that mimic human neural systems, and the hidden layers are commonly referred to as neurons (Fig. 1).

To amplify the efficiency of the learning results of ANN[10], the input data has to be pre-processed into a numeric range. Non-Linear Auto Regressive Artificial Neural networks (NNAR-ANN) approaches in the prediction of time series data are widely used.

NARX

Non-Linear Auto Regressive Artificial Neural networks (NNAR-ANN) approaches in the prediction of time series data are widely used.

A Non-linear Auto Regressive model with the inclusion of exogenous inputs (NARX) predicts the time series data with past values of series, external series, which can be either multidimensional or single. The equation that models the NARX network behavior for time series prediction is mentioned below.

$$Y(t)=f(x(t-1),\ x(t-2)\ldots\ldots,\ x(t-a),y(t-1),\ y(t-2),\ \ldots\ldots\ldots y(t-a))+ e(t) \quad \ldots \text{(iv)}$$

Empirical Mode Decomposition-based Artificial Neural Network (EMD-ANN)

Empirical Mode Decomposition (EMD) was first proposed by Huang[12]. EMD-ANN is commonly used for data that is non-stationary and non-linear. In this method, original time series data are decomposed into "mono component functions" called Intrinsic Mode Functions (IMFs), which are obtained by the superposition of the different frequency and amplitude waves and by the elimination of asymmetric signals with respect to the zero level. In the Empirical Mode Decomposition based Artificial Neural Network (EMD-ANN) model, first of all, original data series are decomposed into IMFs, and ANN is applied to all-individual IMFs.

EMD-ANN is a repetitive process that converts a particular signal into different IMFs with different amplitudes and frequencies. To get the IMFs, the conditions are:

a) The number of zero-crossing and extrema must be equal or differ by one in the entire dataset.
b) At any point, the mean value of local maxima and local minima at any point should be zero.

Results and Discussion

Forecasting the fish production with and without climatic and environmental parameters

Forecasting climatic and environmental time series data are integral for climate management because it allows policymakers to develop necessary strategies and implement management decisions to achieve goals during uncontrollable events. In this study, the fish production with and without the climate change scenario's is forecasted with the help of various statistical and machine learning models.

Seasonal forecasting of fish production with the SARIMA model

The best SARIMA model for forecasting is selected based on the Bayesian information criterion and Akaike information criterion (not shown here)

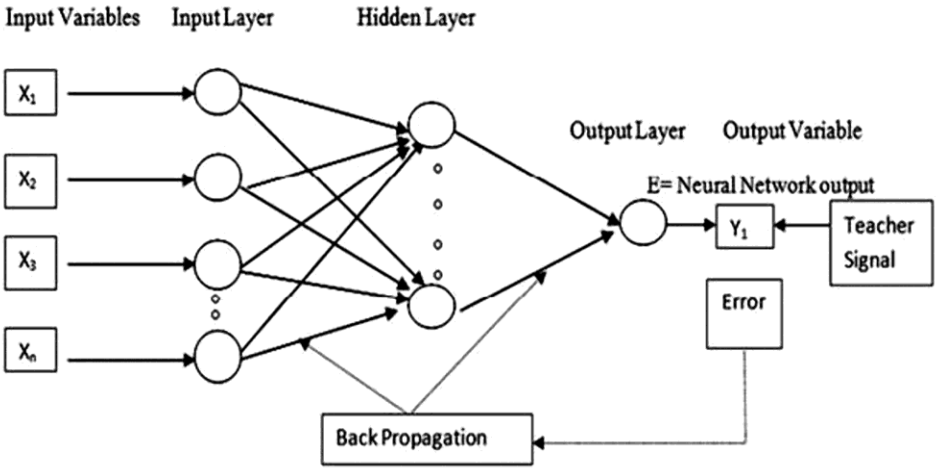

Fig. 1 — Structure of artificial neural network (Yadav *et al*.[11])

(Table 1) and forecasted for the next 20 quarters using SPSS software (graph is not shown here). Indian Mackeral and Horse Mackerel show good forecasts with the SARIMA model with better accuracy and minimal errors (MAPE) comparatively with Oil Sardine catch and Lesser Sardine catch that did not show a good forecast where the error percentages are also high in these fish catch forecasts. The total fish production data of Andhra Pradesh shows a better forecast where the error percentages (MAPE) are low.

ARIMAX model forecasting of fish production data

ARIMAX, in a similar way to ARIMA, is a statistical model used to predict time series data. ARIMAX forecasts data with the inclusion of exogenous variables. In the present study, fish catch data under the influence of climate change is forecasted by taking the required parameters such as SST, PAR, Kd-490, Chl-*a*, and Wind speed in SPSS software. Best models were developed for the different fish catches based on the explained methodology, and those developed models are used to forecast the fish catches. The ARIMAX model with MAPE in the forecast of the catch of selected fish species and the total fish catch is shown in Table 2. Forecast efficiencies are measured with error measures such as MAPE, and the errors are relatively lower than the SARIMA models, except for Oil Sardines (Table 2).

Table 3 shows the MAPE of fish catches with the best model developed compared to all other fishes, and total fish catch showed a better forecast with minimal error. As the data of Oil Sardine is highly non-linear, it showed very poor accuracy of the forecast and very high errors.

NARX

In the present study, the NARX model is used to forecast the fish catch data under the climate change scenario in R software 3.5.3. The typical NARX model is shown as NNAR (p, size), *p* represents the number of seasonal lags that are used as inputs, and size refers to the neurons in the hidden layer. To find the best model in the NNAR, the values of *p* and size were fitted and checked for the tentative model's accuracy with MAPE, and the best model was NNAR (4,7) based on accuracy. The fish landings for the next five years (20 quarters) are forecasted in this study with the best model. Figures 2 to 6 show the forecasts for total fish production, catch of Indian Mackerel, Oil Sardine, Horse Mackerel, and Lesser Sardines, respectively, for the next five years, *i.e*., 2019-2023 (blue line graph).

EMD-ANN

There are crucial steps in the EMD-ANN model, including data decomposition by EMD and forecasting by the ANN. The EMD decomposition acts as a dyadic filter bank, and the obtained IMFs are in a range from high to low, that indicates the local characteristic time scale by itself. Therefore, these components of data have a periodic pattern, as seen in Figures 7 – 11.

Table 2 — The ARIMAX model and MAPE in the forecast of the catches of selected fish species and total fish catch

	Model	MAPE
TOTAL	ARIMAX(2,1,2)(0,0,1)	15.89
Horse Mackerel	ARIMAX(1,1,2)(0,0,1)	60.44
Indian Mackerel	ARIMAX(2,1,2)(0,0,1)	43.48
Oil Sardines	ARIMAX(1,1,1)(0,0,0)	1521.33
Lesser Sardines	ARIMAX(2,1,3)(0,0,0)	54.63

Table 1 — SARIMA models of fish production

	P	d	q	P	D	Q	Best model (p,d,q) (P,D,Q)	MAPE
Total	3	1	3	0	0	2	(3,1,2) (0,0,2)	18.469
Indian Mackerel	1	1	2	0	0	1	(1,1,1) (0,0,1)	57.844
Horse Mackerel	3	1	3	0	0	1	(3,1,2) (0,0,1)	58.865
Oil Sardines	1	1	2	0	0	0	(1,1,0) (0,0,0)	896.407
Lesser Sardines	1	0	3	0	0	0	(1,0,3) (0,0,0)	88.302

Table 3 — Model forecast comparison

	MAPE (Mean Absolute Percentage Error)				
	SARIMA	NNAR-ANN	ARIMAX	NARX	EMD-ANN
TOTAL	18.469	12.12	15.89	0.79	1.86
Horse Mackerel	57.844	43.41	60.44	2.90	5.51
Indian Mackerel	58.865	44.17	43.48	1.22	2.02
Oil Sardines	896.407	1346.86	1521.33	4.37	4.87
Lesser Sardines	88.302	38.53	54.63	1.89	2.63

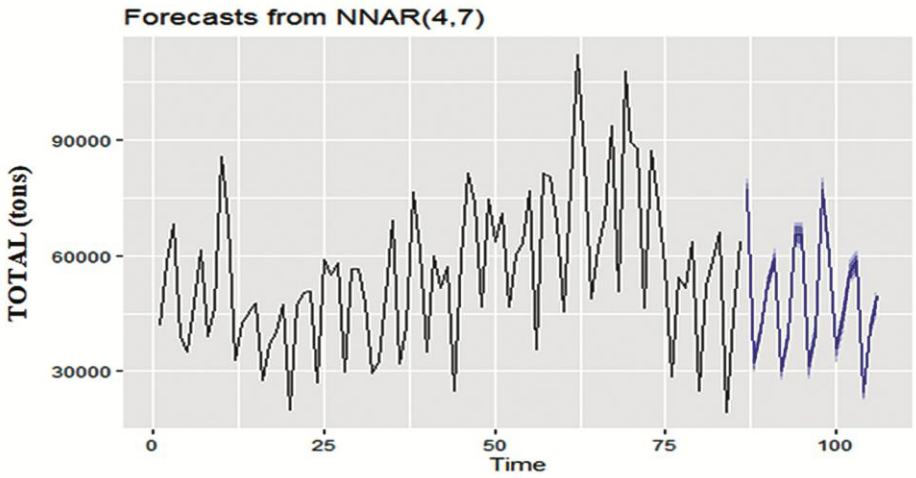

Fig. 2 — NARX model forecast of total marine fish production of Andhra Pradesh

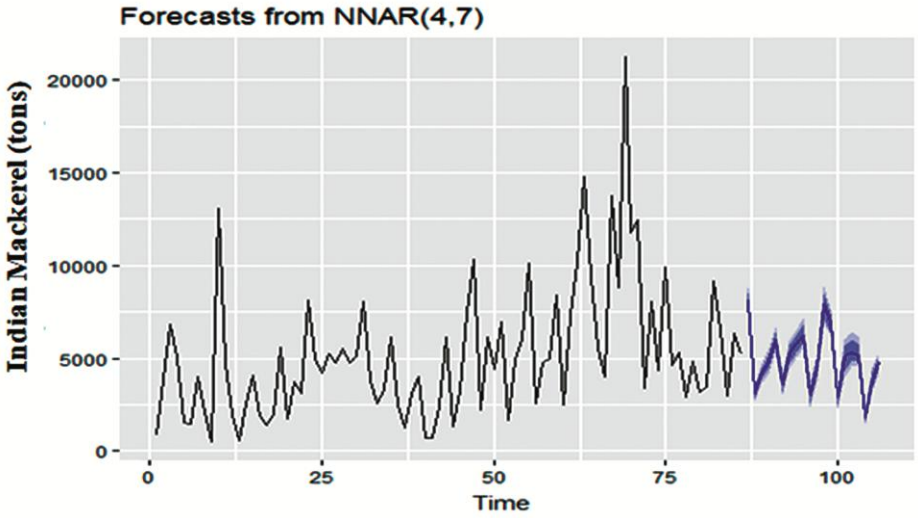

Fig. 3 — NARX model forecast of Indian Mackerel

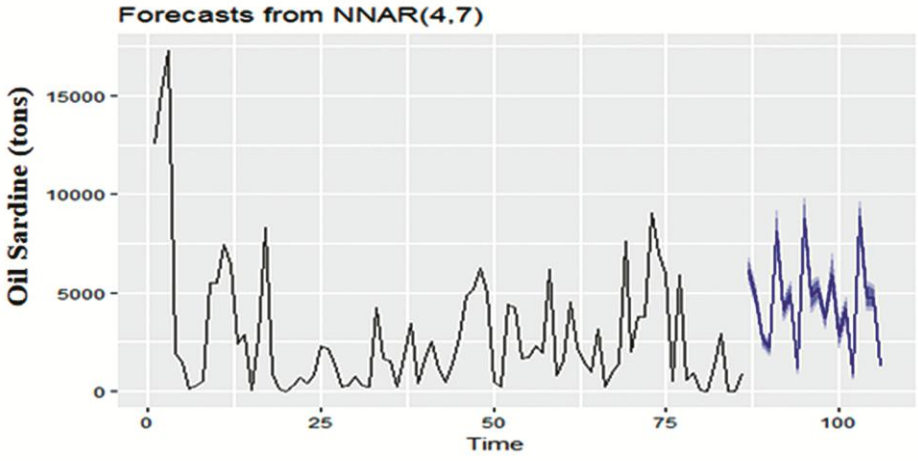

Fig. 4 — NARX model forecast of Oil Sardine

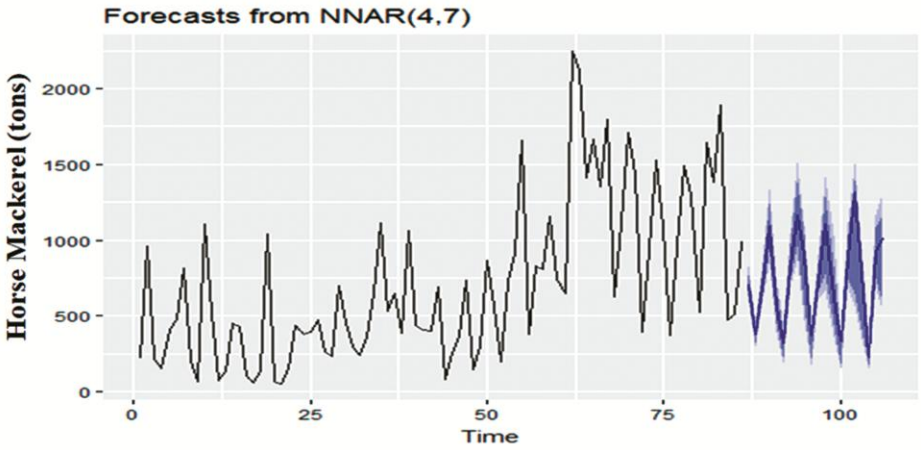

Fig. 5 — NARX model forecast of Horse Mackerel

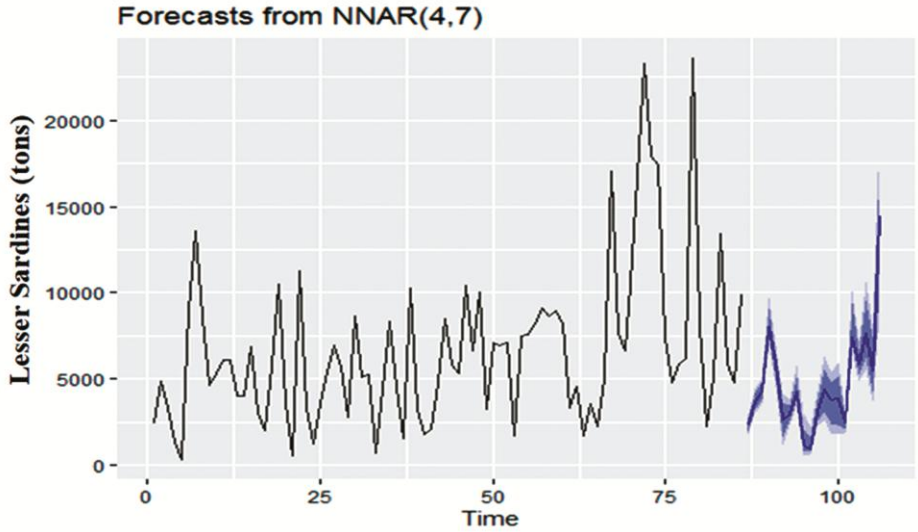

Fig. 6 — NARX model forecast of Lesser Sardines

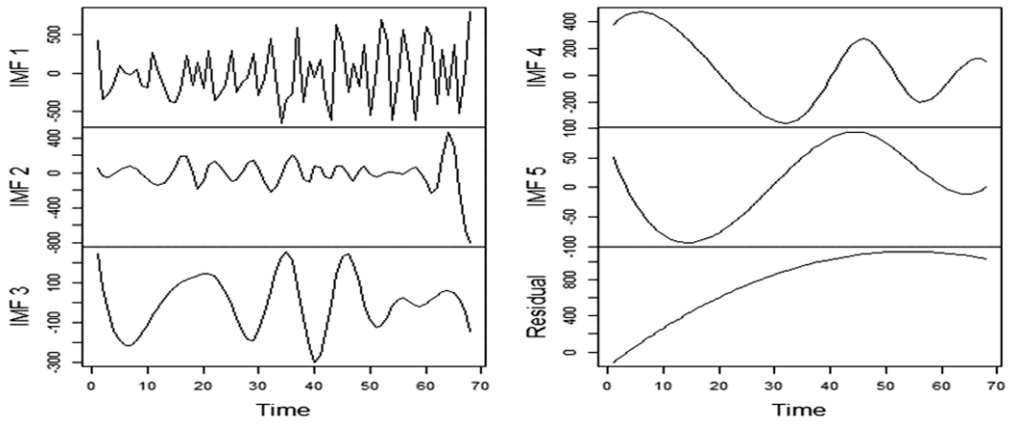

Fig. 7 — IMFs for Horse Mackeral

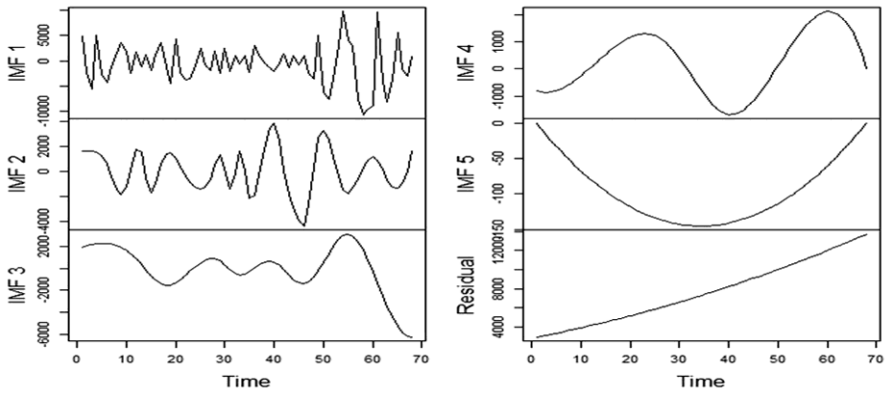

Fig. 8 — IMFs for Lesser Sardines

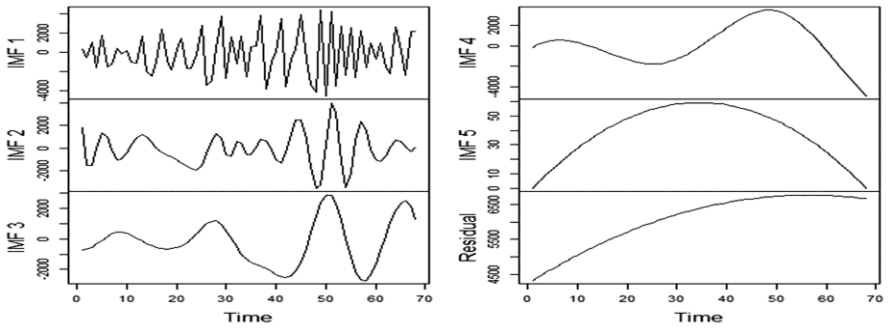

Fig. 9 — IMFs for Indian Mackerel

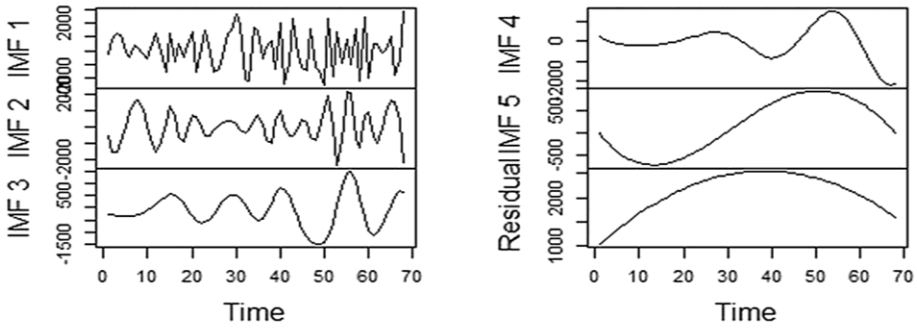

Fig. 10 — IMFs for Oil Sardine

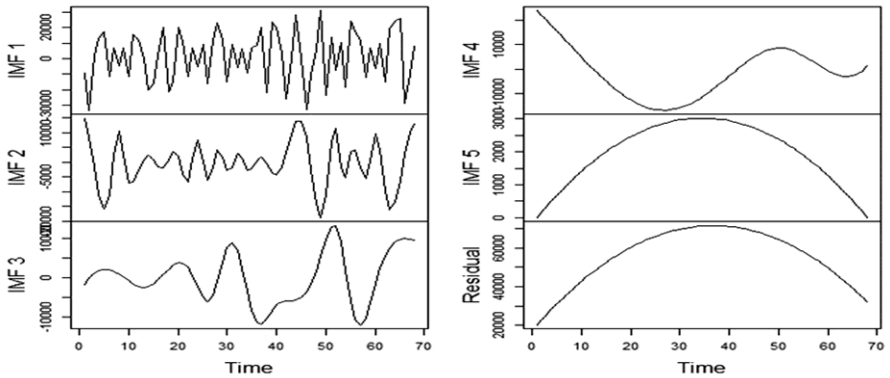

Fig. 11 — IMFs for total fish production

A three-layer feed-forward ANN with a logistic activation function was used in the second stage. The size of input data is a significant factor for computational requirements in ANN, and the neurons in the hidden layer of the network are iterated to minimize the error between the actual and desired output during testing and learning rates of the network.

Model forecast comparison

In the present study, time-series data from 2002 – 2018 was used to forecast total fish catch, Indian Mackerel catch, Oil Sardine catch, Horse Mackerel catch, and Lesser Sardine catch of Andhra Pradesh for the next five years, *i.e*., 2019 – 2023, by different models with and without the influence of climatic and environmental variables. Table 3 shows the forecasts of different models with error measures like Mean Absolute Percentage Error (MAPE)

MAPE = (|Actual-Forecast|/Actual)*100

NNAR-ANN showed better accuracy (MAPE = 12.12) than the SARIMA model (MAPE = 18.469) except for Oil Sardines, in which there is no intervention of the exogenous variable. ANN even outperformed the ARIMAX model forecast of all the fish catch data except the catch data of Indian Mackerels. The performance of the EMD-ANN model outperformed the SARIMA and NNAR-ANN models. Compared to all the model forecast accuracies, the NARX model showed far better forecast accuracy with very minimal error. Similar results were reported by Yadav *et al.*[5] and Paul *et al.*[13].

Summary and Conclusion

Forecasting and modeling of time series have much importance in various practical domains, and to improve the efficiency of time series modeling and forecasting, many models have been proposed in the literature. Forecasting fish production time series data under climate change scenarios is integral for climate management as it allows for the development of strategies and management decisions by policymakers to goals during uncontrollable events. The present study on forecasting of fish production data with and without the climate change scenario's with the help of statistical and machine learning models reveals that the NNAR-ANN exhibit better accuracy (MAPE = 12.12) than the SARIMA model (MAPE = 18.469), in which there is no intervention of the exogenous variable. ANN even outperformed the ARIMAX model forecast of all the fish catch data except the catch data of Indian Mackerels.

The accurate forecasts are obtained by the inclusion of exogenous variables, *i.e*., environmental and climatic variables, in the model. Compared to all the model forecast accuracies, the NARX model showed far better forecast accuracy with very minimal error. NARX method was found better for predicting fish production compared to other methods such as ARIMAX and EMD-ANN. Fishes are poikilothermic, and hence, with the change in temperature, changes in distribution and assemblage do occur. Climate change is a very complicated scenario as the change in SST, PAR, and Chlorophyll-*a* will affect marine fish production. Hence the prediction of catch based on these parameters will provide a general trend. However, the effect of other factors influencing climate change needs to be considered while modeling the effect on fish abundance. The prediction of the influence of climate change on fish catch will give the roadmap to facilitate the management strategies for enhancing sustainable fisheries and adaptation of strategies during climate change.

Acknowledgments

This paper forms part of the M.F.Sc dissertation of the first author. The authors sincerely thank the director ICAR-CIFE for constant encouragement and for providing the necessary facilities for the study.

Conflict of Interest

The authors would like to declare that there are no conflicts of interest to publish this research papers in the journal.

Ethical Statement

This material is the authors' own original work, which has not been previously published elsewhere. The paper is not currently being considered for publication elsewhere. The paper properly credits the meaningful contributions of co-authors. All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

Author Contributions

The authors like to certify that, the first author (SKS) of this paper had contributed towards the preparation of the paper such as conceptualization, data collection, and drafting of the manuscript; the second author (VKY) contributed in data analysis,

drafting and editing of manuscript; and the third author (RRK) contributed in some part of EMD-ANN analysis and overall supervision.

References

1 Brander K, Impacts of climate change on fisheries, *J Mar Syst*, 79 (3-4) (2010) 389-402.

2 Das M K, Sharma A P, Sahu S K, Srivastava P K & Rej A, Impacts and vulnerability of inland fisheries to climate change in the Ganga River system in India, *Aquat Ecosyst Health Manag*, 16 (4) (2013) 415-424.

3 Liu H, Chen C, Tian H Q & Li Y F, A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks, *Renew Energy*, 48 (2012) 545-556.

4 Madhavan N, Thirumalai V D, Ajith J K & Sravani K, Prediction of Mackerel Landings Using MODIS Chlorophyll-*a*, Pathfinder SST, and SeaWiFS PAR, *Indian J Nat Sci*, 5 (29) (2015) 4858-4871.

5 Yadav V K, Jahageerdar S, Ramasubramanian V, Bharti V S & Adinarayana J, Use of different approaches to model catch per unit effort (CPUE) abundance of fish, *Indian J Geo-Mar Sci*, 45 (12) (2016) 1677-1687.

6 Yadav V K, Jahageerdar S & Adinarayana J, Modeling Framework to Study the Influence of Environmental Variables for Forecasting the Quarterly Landing of Total Fish Catch and Catch of Small Major Pelagic Fish of North-West Maharashtra Coast of India reference to selected pelagic fishes of Gujarat and Maharashtra coast of India, *Nat Acad Sci Lett*, 43 (6) (2020) 515–518.

7 Naskar M, Chandra G, Sahu S K & Raman R K, A Modeling Framework to Quantify the Influence of Hydrology on the Abundance of a Migratory Indian Shad, the Hilsa *Tenualosa ilisha*, *N Am J Fish Manag*, 37 (6) (2017) 1208-1219.

8 Raman R K, Mohanty S K, Bhatta K S, Karna S K, Sahoo A K, *et al*., Time series forecasting model for fisheries in Chilika lagoon (a Ramsar site, 1981), Odisha, India: a case study, *Wetl Ecol Manag*, 26 (4) (2018) 677-687.

9 Sun L, Xiao H, Li S & Yang D, Forecasting Fish Stock Recruitment and Planning Optimal Harvesting Strategies by Using Neural Network, *JCP*, 4 (11) (2009) 1075-1082.

10 Kim K J & Lee W B, Stock market prediction using artificial neural networks with optimal feature transformation, *Neural Comput Appl*, 13 (3) (2004) 255-260.

11 Yadav V K, Krishnan M, Biradar R S, Kumar N R & Bharti V S, A comparative study of neural-network & fuzzy time series forecasting techniques-Case study, Marine fish production forecasting, *Indian J Geo-Mar Sci*, 42 (6) (2013) 707-716

12 Huang N E, Shen Z, Long S R, Wu M C, Shih H H, *et al*., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, 454, (1998) 903-995.

13 Paul R K & Sinha K, Forecasting crop yield: ARIMAX and NARX model, *RASHI*, 1 (1) (2016) 77-85.