IJARIIT
INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 11, Issue 3 - V11I3-1404)
Available online at: https://www.ijariit.com

Time Series Forecasting through Hybrid ARIMA-ANN Modelling for Rice in Odisha

Madhu Chhanda Kishan
madhuchhandakishan@gmail.com
Odisha University of Agriculture and Technology, Bhubaneswar, Odisha

ABSTRACT

Rice, being the staple food grain of Odisha, holds a crucial place in the state's economy and food security. Rice holds around 69% of the total cultivable area in Odisha, making it crucial to have an accurate forecast of its status for stakeholders in agriculture. Modelling and forecasting of time series dataset of yield and production of rice from 1970-71 to 2019-20 is carried out in this study, using Auto Regressive integrated Moving Average (ARIMA), Artificial Neural Network (ANN) and Hybrid ARIMA-ANN methodologies. ARIMA is a linear modelling approach where whereas ANN is more of a non-linear modelling technique. The hybrid ARIMA-ANN methodology integrates the strengths of both models to effectively capture both linear and non-linear patterns within the dataset under study. It was found that ARIMA(1,1,1) with constant and under the developed ANN models, the Neural Network Autoregression(NNAR) of order NNAR(3,2) came out to be the best fitted model for both of the variables under study. ARIMA(1,1,1)-NNAR(1,1) is found to be suitable for both yield and production of rice in Odisha. All three models are compared using accuracy measures like RMSE and MAPE, and the hybrid methodology is found to be superior to others.

Keywords: *ARIMA, ANN, ARIMA-ANN, Rice, Forecasting*

INTRODUCTION

Odisha is the fourth largest contributor of Paddy pool of Food Corporation of India contributing around 9% of the total rice production in the country. Rice covers about 69% of the cultivated area which translates to 63% of the total area under food grains in Odisha.(Agricultural Statistics,Odisha,2020). Rice is cultivated on an area of 4.45 million hectares, which can be classified into seven different ecosystems: irrigated kharif(27.4%), rainfed upland (19.1%), medium land(12.4%),shallow lowland(22.5%),semi-deep (7.9%), deep (3.4%) and irrigated rabi (7.4%). Farmers have their own system of classification of rice environment such as uplands,medium lands and lowlands, primarily on the basis of land topography and water regime (Das, 2014). A significant fraction of the agricultural community depends of it for their livelihood, and greater percentage of the population depends on it for basic sustenance.

Rice being the staple food, plays a crucial role in the state's economy and food security. Critical analysis of production and productivity is pre-requisite for proper knowledge base on ecology and appropriate research/development efforts for harvesting maximum possible potential (Tripathy et al., 2014). As production and marketable surplus are inversely varying with farmer's income, it is essential to have proper forecast to stabilize the price and ensured profit through appropriate surplus and deficit management. However, the state's rice production has faced challenges due to various factors, including climate change, water scarcity, and inefficient agricultural practices.(Pradhan et al., 2020). So accurate forecasting is essential for effective planning and decision-making in agricultural sector.

In this paper, the proposed models are Auto-regressive integrate moving average, Artificial neural network and hybrid ARIMA-ANN methodology to forecast the yield and production of rice in Odisha. ARIMA model (Box et al.,1994) have been performing exceptionally well for forecasting agricultural time series data. Pradhan & Dash (2024) used ARIMA model to forecast kharif sweet potato production In Odisha. Kakti et. al.(2022) have used ANN technique to forecast yield of Rapeseed and Mustard in the Bramhaputra Valley of Assam. Rathod et al., (2018) have used artificial intelligence techniques TDNN and NLSVR to model and forecast oilseed production in India. Mishra & Singh (2013) have done a study on the price forecasting of groundnut oil in Delhi by using ARIMA and ANN. Agricultural data are usually combination of linear and non-linear patterns. Lately some research indicates that combining different models enhances the accuracy of forecasting as compared to individual model.

One of the most popular hybrid technique combines ARIMA and ANN models given by Zhang(2003) captures both the linear and non-linear aspect of agricultural data. Bhardwaj et al. (2022) have proposed hybrid ARIMA-ANN model to predict sugarcane production. The suggested hybrid approach processes the original dataset using ARIMA to capture the fundamental patterns and the residuals from ARIMA model are then passed to the ANN for additional refinement.

METHODOLOGY

The data

The study utilises data on the yield and production Of rice in Odisha for the period from 1976-77 to 2019-20 sourced from "Five Decades of Odisha Agricultural Statistics 2020" by the Directorate Agriculture and Food production, Government of Odisha, 2020. The area, yield and production are expressed in '000 ha, kg/ha and '000 MT and respectively. The data from 1976-2016 is used for model building i.e. training data set and 2017-2020 is used for model validation i.e, testing data set.

ARIMA

Box and Jenkins developed an outstanding and simple technique to time series forecasting in the year 1976, and it became one of the most used approach until the late 90's. The future forecasted values are assumed to be a linear combination of past time series and past errors in this method of forecasting.

ARIMA stands for Autoregressive Integrated Moving Average models. It is one of the univariate timeseries forecasting model that projects the future values of a series, based on its own inertia. An ARIMA model is usually stated as ARIMA(p,d,q), where p shows the order of the autoregressive components, d shows the degree of differencing and q is the order of the moving average. The general form of ARIMA is :

$$X_t = \theta_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots \ldots\ldots + \varphi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \ldots . - \theta_q e_{t-q} \qquad (1)$$

Where, X_t= the actual time series data at time t
e_t= random error at time t
$\theta_j (j = 1,2,3 \ldots . q)$ *and* $\varphi_i (i = 1,2,3 \ldots\ldots p)$ are the model parameters, p shows the autoagressive and q shows the moving average terms are in polynomial. (Singh,2021).

ARIMA modelling is done under four stages. These are model identification, parameters estimation, diagnostic checking and forecasting. The data is first divided into two segments; training and testing segments. To implement ARIMA, it is necessary that the data is stationary. Augmented Dickey Fuller test is done to check the stationarity of the data. Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots are examined to determine the order of the Moving Average and Auto-Regressive (AR) component, respectively. Based on the identified parameters, best fitted ARIMA model is identified with low AIC, RMSE, MAPE values. Then the residuals are checked for normality and auto correlation in diagnostic checking. Lastly, the best fitted model is cross validated with the testing data and then it is used for forecasting.

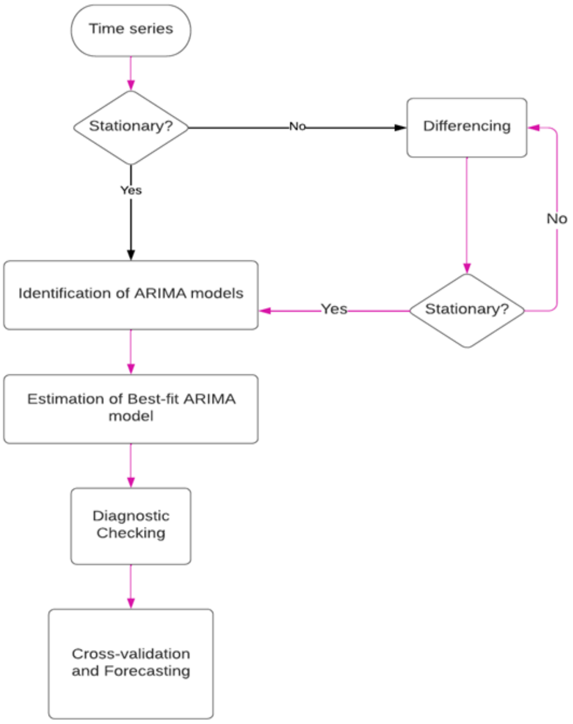

Time series
Stationary?
No
Differencing
Yes
No
Identification of ARIMA models
Yes
Stationary?
Estimation of Best-fit ARIMA model
Diagnostic Checking
Cross-validation and Forecasting

Fig 2: Flowchart of ARIMA model

Neural Networks

Neural Networks are simulated networks with interconnected simple processing neurons which aim to mimic the function of the brain central nervous system (McCulloch and Pitts, 1943). Artificial neural network (ANN) modelling is used to model the non-linear data to solve the problems for which conventional statistical methods are not suitable. The main neural network topology consists of an input layers, an output layer and usually one or more hidden layers. Each layer consists of nodes connected to nodes at the adjacent layers.Each connection link has as an associated weight which multiplies the signal being transmitted. Then each neurons applies an activation function to set it's net input to determine it's output signal.

The relationship between the output y_t and the inputs (y_{t-1},$y_{t-2,\dots\dots}$ y_{t-p}) can be mathematically represented as

$$\mathbf{y_t = f\left(\sum_{j=0}^{q} w_j g\left(\sum_{i=0}^{p} w_{ij}\, y_{t-i}\right)\right)} \qquad \textbf{(2)}$$

where, w_j and w_{ij} are model parameters, called as connection weights. (i=0,1….p and j=0,1,…q)p= no of input nodes,q= no of output nodes, g= activation fuction at hidden layer, f= activation fuction at output layer.

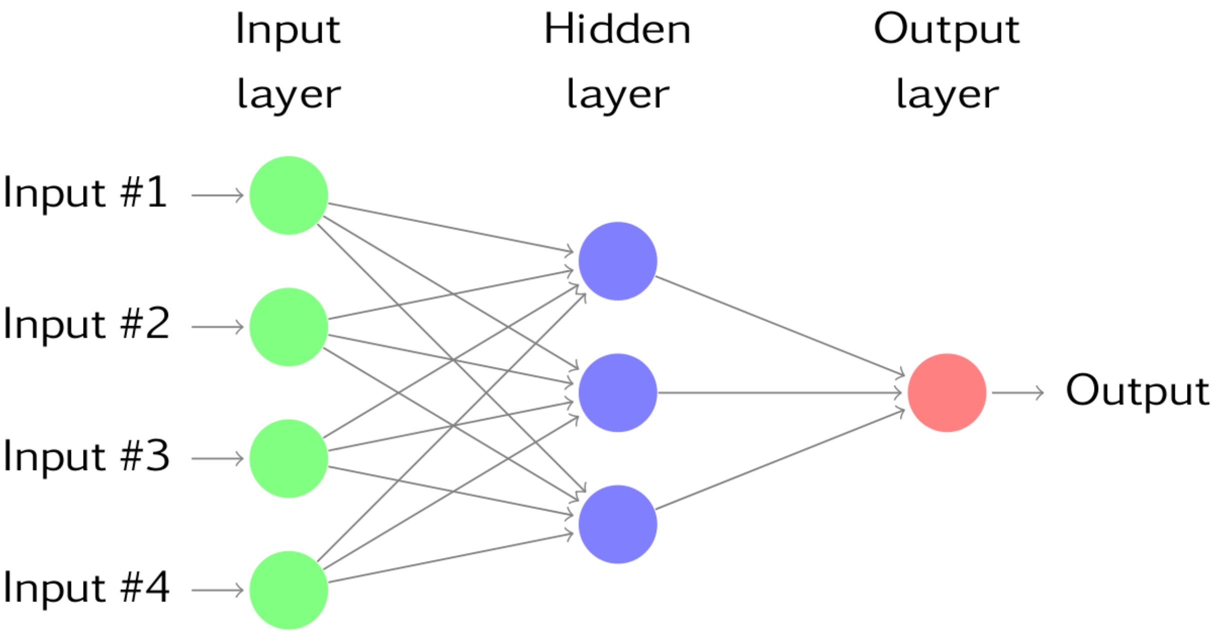


Fig 2 : A simple neural network architecture

Variety of architecture using different functions and varying in the number of hidden layers are used to find the model that optimises the performance of network, known as the best model. In forward propagation, they learn the data and assign weights to the nodes, where as in backward propagation, the network tries to reduce the error using gradient descent algorithm.

Hybrid ARIMA-ANN model: Generally agricultural data contain both linear and non-linear patterns, no single model is capable of identifying all the characteristics of timeseries in agriculture. Consequently, various types of parametric and non-parametric,linear and non-linear timeseries models are used for forecasting.(Fan and Yao, 2003, Ghosh et al., 2005)

A time series is composed of linear (Lt) and Non-linear (Nt) components, mathematically expressed as:

$$X_t = Lt + N_t \qquad (3) \quad \text{(Zhang,2023)}$$

To develop an effective hybrid model, two major steps are followed. First, the linear component is modelled by ARIMA and the residue(e_t) of the ARIMA model; the non-linearity of the time series to be modelled through an ANN model using p input nodes, where f is a non-linear determined by the neural networks and ε_t is the random error. (Ravichandran,2018)

$$e_t = f\left(e_{t-1,}e_{t-2}, \dots \dots \dots . e_{i-p}\right) + \varepsilon_t \qquad (4)$$

Thus the combined forecast provided by the hybrid model is given by, $\widehat{Xt} = \widehat{Lt} + \widehat{Nt}$ (5)

This combine approach leverages the strengths of both the models, where ARIMA efficiently captures linear structures and ANN addresses the non-linear patterns.

Evaluation methods for forecasting Performance:

The most popular forecasting evaluation methods like root mean squared error (RMSE), Mean Absolute Error(MAE) and Mean Absolute Percentage Error (MAPE) were used to evaluate above models.

MAPE(Mean absolute percentage Error)

Here, the error is measured in percentage terms.

$$\text{MAPE=}\frac{\frac{\sum_t \left|P_t - A_t\right|}{A_t}}{n} * 100$$

RMSE(Root mean square error):

$$\text{RMSE=}\sqrt{\frac{\sum(P_t-A_t)^2}{n}}$$

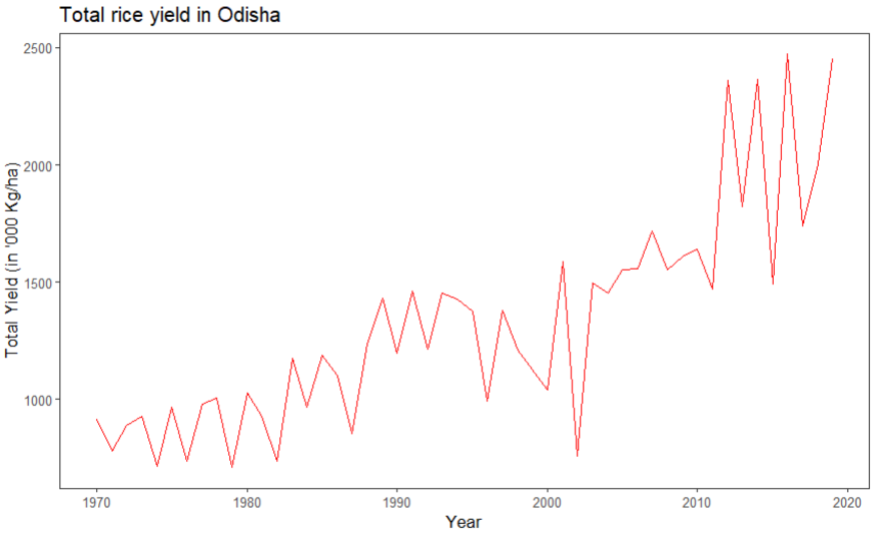
Where, P_t= Predicted Value
A_t= Actual value
n= no. of observations

RESULTS AND DISCUSSION

To understand the time series better summary statistics of yield and production of rice are presented in table 1, explains the series is highly heterogeneous as CV is high.The first step in time series analysis is to visualize the data understudy to inspect it's behavior. Figure 2 shows the time series plot of annual yield and production of rice in Odisha for the period 1970-71 to 2019-20 indicating non-stationarity and a positive trend in both of the series. To confirm the presence of non-stationarity, the Augmented Dickey Fuller(ADF) test was done in the original timeseries, the results are given in the table 2 . The table includes the results of ADF test after the first differencing indicating the absence of unit root in the differenced series. The Brock, Dechert and Scheinkman (BDS) test was employed to test the existence of non-linearity pattern in the data presented in table 3, shows the presence of both linear and non-linear pattern in the dataset.

Table 1 : Descriptive statistics of Yield and production of rice in Odisha

Statistics	Yield	Production
Mean	1325.26	5701.66
Standard Deviation	456.91	1799.97
Kurtosis	0.50	-0.09
Skewness	0.89	0.56
Minimum	709.00	2918.00
Maximum	2472.00	9845.00
CV(%)	34.48	31.57



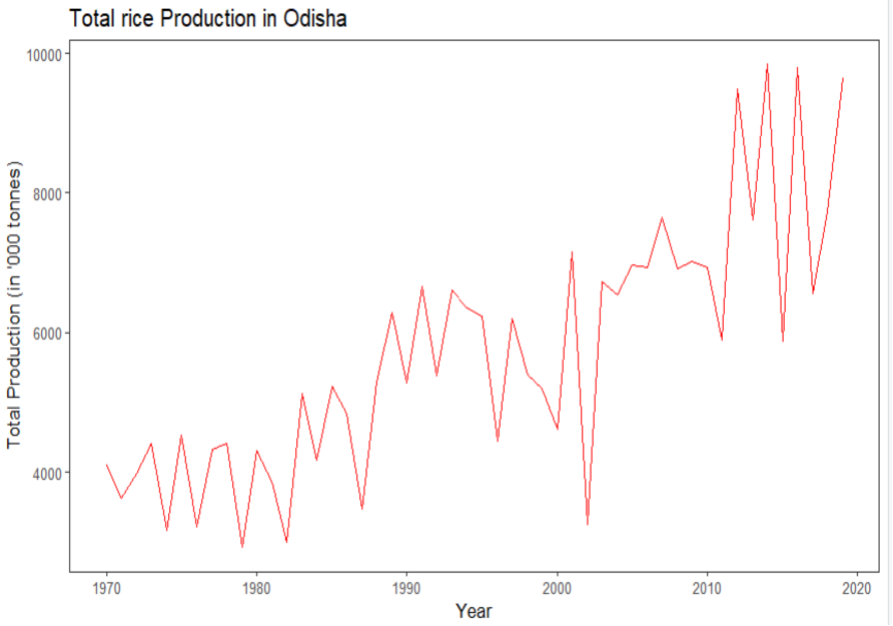
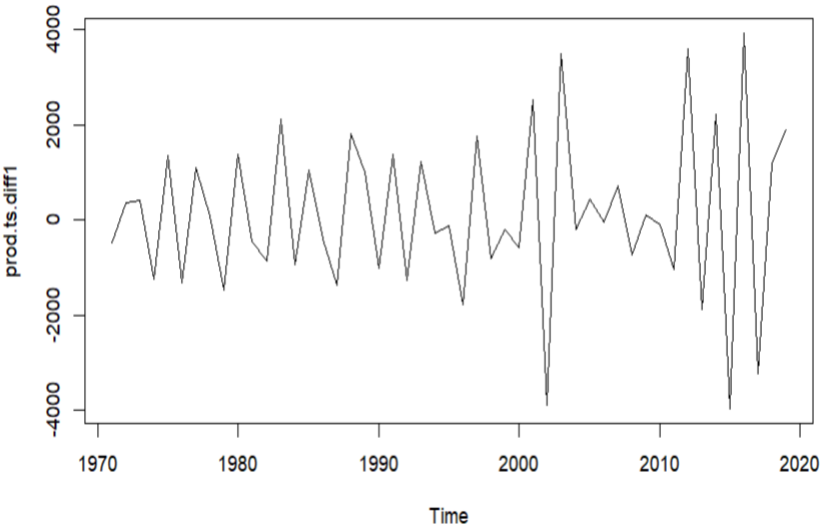


Figure 3: time plots of yield and production of Rice in Odisha



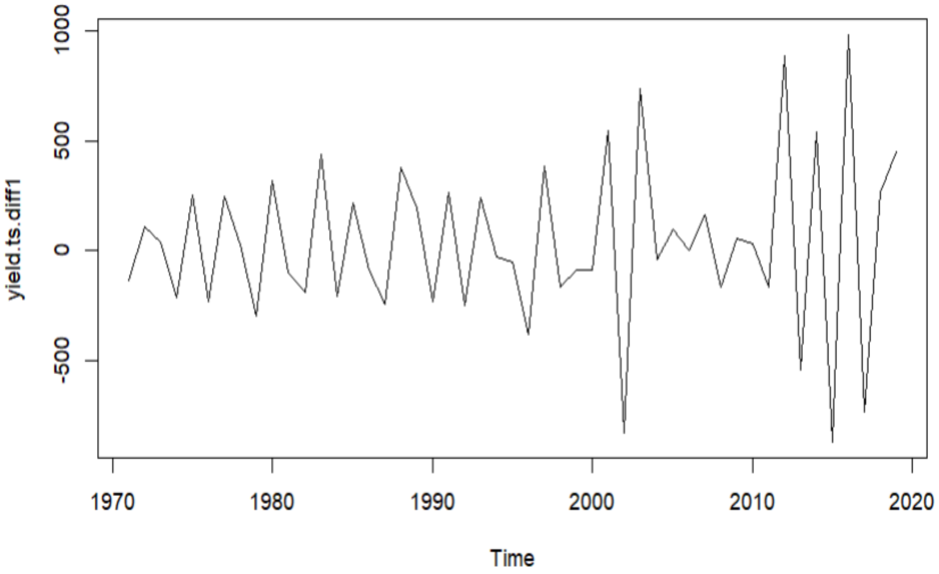


Figure 4: Time plots of differenced series of Yield and Production of Rice in Odisha

Table 2: ADF test of stationarity of data on yield and production of rice in Odisha

Variable	Original series		Remark	First order differenced series		Remark
	Adf test statistic	P value		Adf test statistic	P value	
yield	-1.8968	0.6157	Non-stationary	-4.5058	0.01	Stationary
Production	-2.9587	= 0.1897	Non-stationary	-4.4636	0.01	Stationary

Table 3: BDS test result on the yield and production of Rice in Odisha

Yield	Epsilon	Embedded Dimension			
		2	p-value	3	p-value
	228.456	20.657	0	36.179	0.000
	456.911	7.011	0	9.851	0.000
	685.367	2.016	0.044	3.817	0
	913.822	-0.506	0.613	2.061	0.039
Production	899.984	14.815	0	21.763	0
	1799.969	4.071	0	7.517	0
	2699.953	0.278	0.781	2.689	0.007
	3599.937	-0.675	0.500	1.579	0.114

The potential ARIMA models were identified on the basis of the ACf and PACF plots (fig 5 and 6) indicating that the maximum order AR is 1 and for MA is 1 in both of the time series. After considering the lowest AIC of the potential ARIMA models ARIMA(1,1,1) with constant were found to be best fitted ARIMA model for both yield and production of Rice in Odisha (Table 4). To further clarify the credibility of the models normality and autocorrelation assumptions of the residuals were done with the help ofShapiro wilk's test and L-jung Box test. The results are displayed in table 3. Subsequently, both of the series were modelled using neural networks. The NNAR (3,2) model was found best for modelling both yield and production.

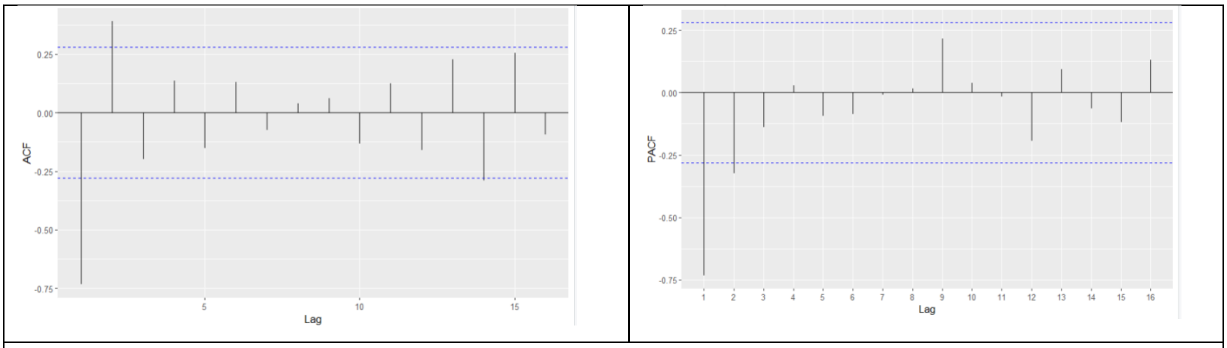


Figure 5: ACF and PACF plot for yield of rice in Odisha

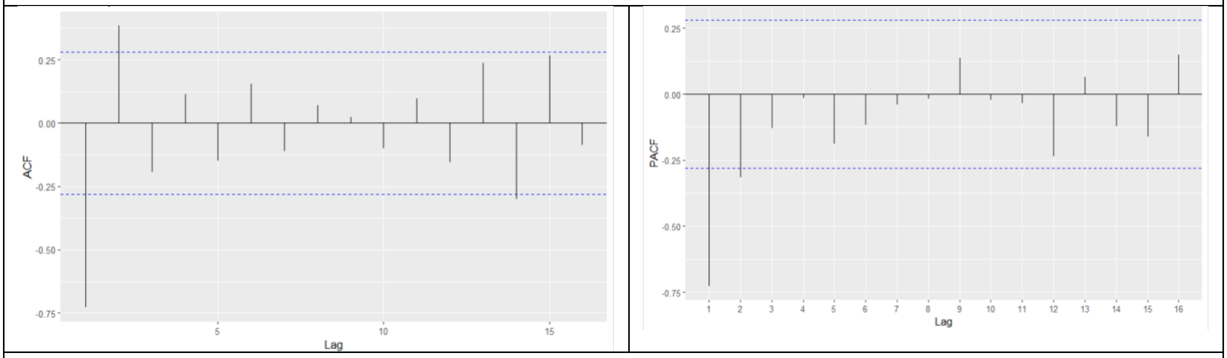


Figure 6 : ACF and PACF plot for production of rice in Odisha

Table 4: Best fitted ARIMA model for forecasting yield and production of Rice

	Yield	**Production**
Best fit ARIMA model	ARIMA(1,1,1)	ARIMA(1,1,1)
Constant(μ)	25.1954* (10.7839)	94.783* (43.539)
AR lag 1	-0.542** (0.173)	-0.50174 ** (0.18554)
MA lag 1	-0.5206** (0.1727)	-0.58858** (0.19078)

(Figures in the parantheses indicate the standard error)

Table 5: Model fit statistics and residual Diagnostic of best fitted ARIMA models for yield and production of rice

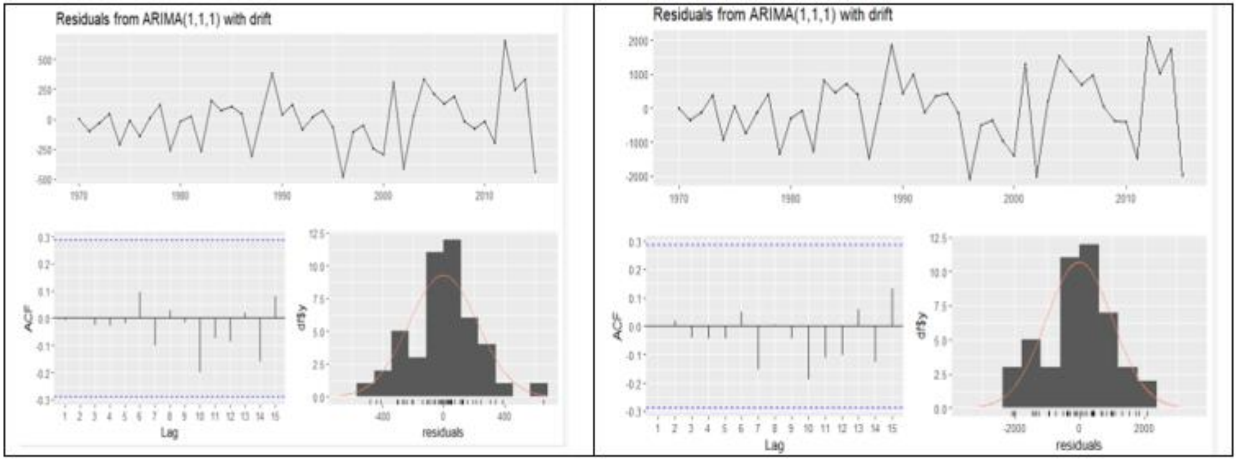


Figure 7: Residual plots for fitted ARIMA models for yield and production of Rice in Odisha

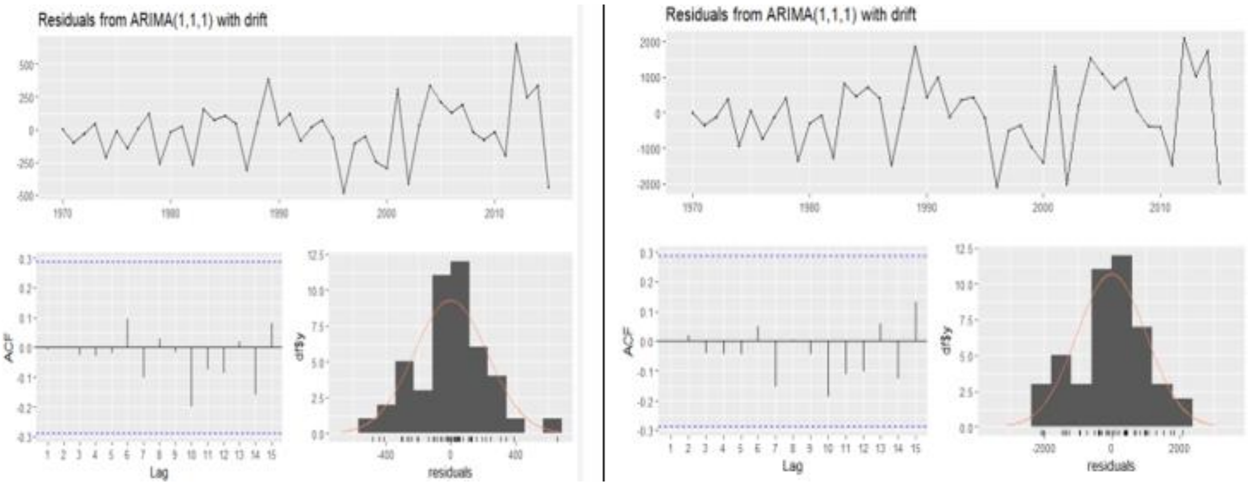


Figure 8: Residual plots for fitted ARIMA models for yield and production of Rice in Odisha

	Model fit Statistics		**Residual Diagnostics**			
	RMSE	**MAPE**	**Ljung-Box Q statistics**	**p-value**	**Shapiro-Wilk's statistics**	**p-value**
Yield	222.3074	13.94817	1.2516	0.9897	0.97801	0.5268
Production	1010.185	12.086	1.9027	0.965	0.9794	0.5815

The residuals were obtained from the applied ARIMA model for both the time series and then the residuals are modelled using neural networks. The NNAR(1,1) model was found suitable for modelling both the residual series. Later, both the ARIMA models for the data series and the NNAR models for the residual series are combined and used to forecast the yield and production, respectively. The performances of all the three models viz., ARIMA,ANN and hybrid model are compared using forecast using evaluation measures like RMSE and MAPE in table 6. The result indicated that the hybrid ARIMA-ANN models are best suited due to the lower RMSE and MAPE values for both the time series . The hybrid ARIMA-ANN methodology was used to forecast the production and yield of rice for 2024 is found to be 8860.610 thousand tons and 2237.562 kg/ha, respectively.

Table 6: Forecast Comparison of ARIMA, ANN and Hybrid ARIMA-ANN

	Yield		Production	
Models	RMSE	MAPE	RMSE	MAPE
ARIMA	262.445	10.009	1027.479	15.786
ANN	433.441	17.309	1806.20	17.91
ARIMA-ANN	252.437	8.722	990.866	15.46

Table 7: Forecasting result of Hybrid ARIMA-ANN Model

Year	Yield in Kg/Ha	Production in '000 tones
2020	1977.247	7702.753
2021	2264.654	8734.490
2022	2147.170	8350.764
2023	2255.130	8686.144
2024	2237.562	8660.610

CONCLUSION

Timeseries analysis of agricultural data is crucial for better preparedness for forthcoming challenges, potential future scenarios, policies and scheme making for the farmers. ARIMA being the most widely used timeseries analysis techniques falls on short while capturing the non-linear data patterns, for which as an alternative approach ANN are used. In this current investigation, the suitability of hybrid ARIMA-ANN for forecasting the yield and production of rice in Odisha have been assessed. As bench mark models, ARIMA and ANN models have also been employed. The hybrid ARIMA-ANN methodology was used to forecast the production and yield of rice for 2024 is found to be 8860.610 thousand tons and 2237.562 kg/ha, respectively. The performances of ARIMA-ANN hybrid was also compared with individual ARIMA and ANN methods using forecast accuracy measures, where the hybrid methodology out-performed.

REFERENCES

[1] Agricultural Statistics at a glance,2020, Government Of Odisha, Department of Agriculture

[2] Bhardwaj, Nitin. "Time Series Prediction Using Hybrid ARIMA-ANN Models for Sugarcane." *International Journal of Plant & Soil Science* 34.23 (2022): 772-782.

[3] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994), Time Series Analysis: Forecasting and Control (3rd ed.). Holden-Day, San Francisco.

[4] Das S.R. "Rice in Odisha" 2012. IRRI Technical Bulletin Los Baños (Philippines): International Rice Research Institute, no. 16, Pp.31, 2012. http://books.irri.org/TechnicalBulletin16_content.pdf

[5] Fan, J. and Yao, Q. 2003. Nonlinear time series: nonparametric and parametric methods, Springer, New York.

[6] Ghosh, H., prajneshu and Paul, A, K. 2005. Study of nonlinear timeseries modelling in agriculture. IASRI. Project Report.

[7] Kakati, Nishigandha, et al. "Forecasting yield of rapeseed and mustard using multiple linear regression and ANN techniques in the Brahmaputra valley of Assam, North East India." *Theoretical and Applied Climatology* 150.3 (2022): 1201-1215.

[8] Mishra, G.C. & Singh, Abhishek. (2013). A Study on Forecasting Prices of Groundnut Oil in Delhi by Arima Methodology and Artificial Neural Networks. Agris On-line Papers in Economics and Informatics. 5. 25-34.

[9] Pradhan, Jayashree, and Abhiram Dash. "Using ARIMA Model to Forecast Production of kharif Sweet Potato in Odisha." *Environment and Ecology* 42.2 (2024): 383-390.

[10] Pradhan, S. & Rahman, Feroze & Sethy, S.K. & Pradhan, G. & Sen, J.. (2020). Evaluation of Short Duration Drought Tolerant Rice Varieties in Drought Prone Areas of Subarnapur District of Odisha. International Journal of Plant & Soil Science. 21-26. 10.9734/ijpss/2020/v32i830314.

[11] Rathod, Santosha & Singh, Kamalesh & Patil, Santosh & Naik, Ravindrakumar & Ray, Mrinmoy & Meena, Vikram Singh. (2018). Modeling and forecasting of oilseed production of India through artificial intelligence techniques. Indian Journal of Agricultural Sciences. 88. 10.56093/ijas.v88i1.79546.

[12] Ravichandran, S., B. S. Yashavanth, and K. Kareemulla. "Oilseeds production and yield forecasting using ARIMA-ANN modelling." *Journal of Oilseeds Research* 35.1 (2018): 57-62.

[13] Tripathi, Rahul & Nayak, A.K. & Raja, R. & Shahid, Mohammad & Kumar, Anjani & Mohanty, Sangita & Panda, Bipin & Lal, Dr. B. & Gautam, Priyanka. (2014). Forecasting Rice Productivity and Production of Odisha, India, Using Autoregressive Integrated Moving Average Models. Advances in Agriculture. 2014. 10.1155/2014/621313.

[14] Zhang G 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50: 159- 175.