Plant Archives Vol. 24, Special Issue (GABELS), 2024 pp.225-231 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org
DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2024.v24.SP-GABELS.034

TIME SERIES ANALYSIS FOR MUSTARD PRODUCTION, PRODUCTIVITY, AND AREA FORECASTING IN MADHYA PRADESH, INDIA

Lokesh Sharma, Umesh Singh, Kuldeep Rajpoot*, and Lokesh Choudhary
Department of Mathematics and Statistics, JNKVV, Jabalpur, M.P.-482004, India
*Corresponding author: kuldeep2rajpoot @gmail.com

ABSTRACT

This study is done with objective of forecasting of area, production and productivity of mustard using time series analysis in M.P. Secondary data are area, production and productivity has collected from official web sites of Directorate of Economics and Statistics of Madhya Pradesh. To achieve this objective, data from 1976 to 2021 are analysed. This study uses time series analysis model ARIMA to forecast mustard area, production, and productivity. The statistical software R and SPSS, were employed for analysis purpose. To validate the fitted models and testing the accuracy of the fitted models different indices like R-squared, RMSE, MAPE, MAE and were used. The results showed that the ARIMA (0, 1, 0), (0,1,1), (0,1,1) and models are suitable for forecasting mustard area, production, and productivity in M.P. We found that models are well fitted with MAPE value for area(12.851), production(23.510) and productivity(13.110).Thus the study concluded that the, over the time data of area, production and productivity of mustard over all it is found best ARIMA model for Madhya Pradesh.
Keywords: Mustard, productivity, forecasting, ARIMA

Introduction

The most significant edible oil crop is Indian Mustard (*Brassica juncea* L.), which belongs to the Cruciferae family. Oilseeds are broadly divided into two groups, primary groups consists of the edible group viz. Groundnut, Rapeseed (Toria, Mustard and Sarson), Soybean, Sunflower, Sesame, Safflower and Niger and secondary group consists of non-edible group viz. Castor and Linseed. Among oilseed mustard crops is the second most important edible oilseed crop and it constitutes one fourth area of oilseeds grown in India (Rajpoot and Kushwaha, 2017). In India, the states of Rajasthan, Uttar Pradesh, Madhya Pradesh, Haryana, and Punjab are the key locations for its cultivation. M.P.'s total area and production was 717.90 ("000 ha) and 1056.23 ("000" tonnes of output) and Madhya Pradesh is also one of the major mustard producing state in India and occupies 2nd rank in the production of mustard in India. (Source: name of web https://eands.dacnet.nic.in)

There have been several studies on the forecasting of area, production and productivity of different crops conducted by various researchers. Padhan (2012) applied ARIMA model for Prediction of mustard yield in India. Alam et *al.* (2018) studied method for long-term yield forecasting in Aligarh district of Uttar Pradesh. Hemavathi and Prabakaran (2018) predicted production, area and productivity of rice in Thanjavur district of Tamil Nadu. Kumar and Verma (2020) forecasted yield of mustard in Bhiwani and Hisar districts of Haryana using ARIMA model. Iqbal and Khan, (2022) used ARIMA model, Normalised Bayesian criteria (NBIC), Akaike information criterion (AIC), Hannan-Quinn and Schwarz criterion to forecast rapeseed and mustard production in Pakistan. Rajpoot *et al.* (2021) used ARIMA and ANN models to predict prices of potato and onion crop during the Covid-19 lockdown period in metropolitan cities of India. Mishra *et al.* (2022) used time series model ARIMA to predict the yield of mustard in Udaipur district of Rajasthan. Bhusanar and Meena (2023) used auto regressing integrated moving average (ARIMA) model to predict area, production and productivity of groundnut crop in Rajasthan.

In this study researchers used time series method ARIMA model to predict area, production and yield of mustard in Madhya Pradesh (M.P.) state of India.

Materials and Methods

The secondary data on area, production and productivity of mustard are collected for the state of Madhya Pradesh for the period 1976-77 to 2021-22 from Directorate of Economics and Statistics, Ministry of Agriculture. An Autoregressive Integrated Moving Average model is a statistical model which is used to predict the future trends. The ARMA models, which includes the order of differencing (which is done to station arise the data) is known as Autoregressive integrated moving average (ARIMA) model. A non-stationary ARIMA model is classified as an "ARIMA (p,d,q)" model, where, the parameters p, d, q are the non-negative integers where p is the number of autoregressive terms, d is the number of non stationary differences necessary for station arising the data, and q is the number of moving average terms.

Autoregressive processes

Autoregressive processes are as their name suggests- regressions on themselves. Specifically, a p^{th} order autoregressive process viz. AR (p), $\{Y_t\}$ satisfies the equation

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots\dots\dots + \phi_p Y_{t-p} + e_t$$

Where we now use the symbols ϕ_1, ϕ_2, ………., ϕ_p for the finite set of weight parameters. The current value of the series Y_t is a linear combination of the p most recent past values of itself plus an “innovation” term e_t that incorporate everything new in the series at time t that is not explained by the past values.

Moving Average Processes

In the case where only finite numbers of the Ψ-weights are nonzero, we have what is called a moving average process. In this case, we change notation somewhat and write….

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots\dots\dots\dots - \theta_q e_{t-q}$$

We call such a series a moving average of order q and abbreviate the name to MA (q).

Where we now use the symbols $-\theta_1$, $-\theta_2$,………… $-\theta_q$ for the finite set of weight parameters.

ARIMA Model

A time series $\{Y_t\}$ is said to follow an Integrated autoregressive moving average model if the d^{th} difference $W_t = \nabla^d Y_t$ is a stationary ARMA process. If $\{W_t\}$ follows an ARMA (p, q) model, we say that $\{Y_t\}$ is an ARIMA (p, d, q) process.

Let us consider the model

$$\varphi(B) Y_t = \theta(B) e_t$$

Where $\varphi(B)$ is a nonstationary autoregressive operator such that d of the roots of $\varphi(B) = 0$ are unity and the remainder lie outside the unit circle.

Model identification

The ARIMA model is fitted to stationary data i.e. having constant mean and variance. Stationarity of data can be tested by using Augmented Dickey-Fuller test. If it is not stationary then it should be converted into stationary series by differencing the data at suitable lag. Usually, the data is stationarized after 1 or 2 differencing. After stationarizing the data, the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots are used to identify tentative Auto Regression (AR) and Moving Average (MA) orders. The orders of AR and MA are denoted by p and q respectively. Various tentative models based on identified AR and MA orders are fitted and parameters are estimated. After fitting the tentative models for a variable (area/production/productivity) the estimated coefficients are tested for the significance and the normality and independency of the residuals of the fitted models are checked by using Box-Ljung test statistic. The models having all the estimated coefficients significant and satisfying the normality and independency of the errors are now compared on the basis of model fit statistics like R-squared, Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean absolute error (MAE) and Schwarz's Bayesian Information criterion (SC, BIC, or SBC). Then the model having the lowest value of these model fit statistics is considered to be the best fit model for the variable.

The model fit statistics like R-squared, MAPE, RMSE, MAE and BIC are mathematically as follows:

R-squared:-

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{X}_i - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Root Mean Square Error (RMSE):-

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}}$$

Mean Absolute Percentage Error (MAPE) :-

$$MAPE = \frac{\sum_{i=1}^{n} \frac{X_i - \hat{X}_i}{X_i}}{n} \times 100$$

Mean absolute error (MAE):-

$$MAE = \frac{\sum_{i=1}^{n} X_i - \hat{X}_i}{n}$$

Schwarz's Bayesian Information criterion (SC, BIC, or SBC):-

$$BIC = \log_{\hat{\sigma}}{}^2 + 2\frac{p+q}{n}\log(n)$$

Where BIC is the Schwarz's Bayesian Information criterion, p+q denotes the number of parameters and n denotes the sample size i.e. no. of observations.

The model with lowest, R-squared, RMSE, MAE, MAPE and BIC values is selected as the best fit ARIMA model among selected tentative models and it is taken for forecasting.

Results and Discussion

The data on area, production and productivity of mustard crop was tested model diagnostics test and model fit statistics for the fitted ARIMA models. ARIMA for Area (0,1,0), Production (0,1,1), Productivity (0,1,1) model satisfies both the test of normality and independency of residuals. Thus this model is selected to be the best fit model for area under mustard crop presented in Table 1. The test results confirmed that the data was not stationary.

Table 1 : ARIMA Model fit Statistics of area, production and productivity in M.P.

	Model Fit Statistics							Ljung-Box Q(18)		
	R-square	**RMSE**	**MAPE**	**MAE**	**Max.APE**	**Max.AE**	**Normalized BIC**	**Statistics**	**DF**	**Sig.**
A(0,1,0)	0.807	82.348	12.851	64.416	64.642	270.656	8.907	8.749	18	0.965
P(0,1,1)	0.834	128.154	23.510	94.618	159.028	334.190	9.876	12.626	17	0.761
Y(0,1,1)	0.823	134.415	13.110	102.447	57.295	339.378	9.971	8.393	17	0.957

Table 2: ACF and PACF of mustard crop area, production and productivity in M.P.

ACF and PACF of mustard crop area

Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
area-Model_1	ACF	-0.143	-0.056	-0.13	0.041	-0.082	0.035	-0.007	0.047	-0.07	-0.207	-0.074	0.071	0.119	-0.056	0.123	0.032
	SE	0.149	0.152	0.153	0.155	0.155	0.156	0.156	0.156	0.157	0.157	0.163	0.164	0.165	0.167	0.167	0.169

Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
area-Model_1	PACF	-0.143	-0.078	-0.15	-0.008	-0.103	-0.012	-0.015	0.023	-0.057	-0.246	-0.17	-0.046	0.049	-0.073	0.093	0.002
	SE	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149

ACF and PACF of mustard crop production

Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
production-Model_1	ACF	0.098	-0.031	-0.145	-0.155	-0.037	-0.02	0.027	-0.01	-0.187	-0.166	-0.069	-0.003	0.135	0.068	0.224	-0.021
	SE	0.149	0.151	0.151	0.154	0.157	0.157	0.157	0.158	0.158	0.162	0.166	0.167	0.167	0.169	0.17	0.176

Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
production	PACF	0.098	-0.041	-0.14	-0.132	-0.021	-0.045	-0.009	-0.044	-0.21	-0.165	-0.086	-0.089	0.021	-0.042	0.179	-0.057
model-1	SE	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149

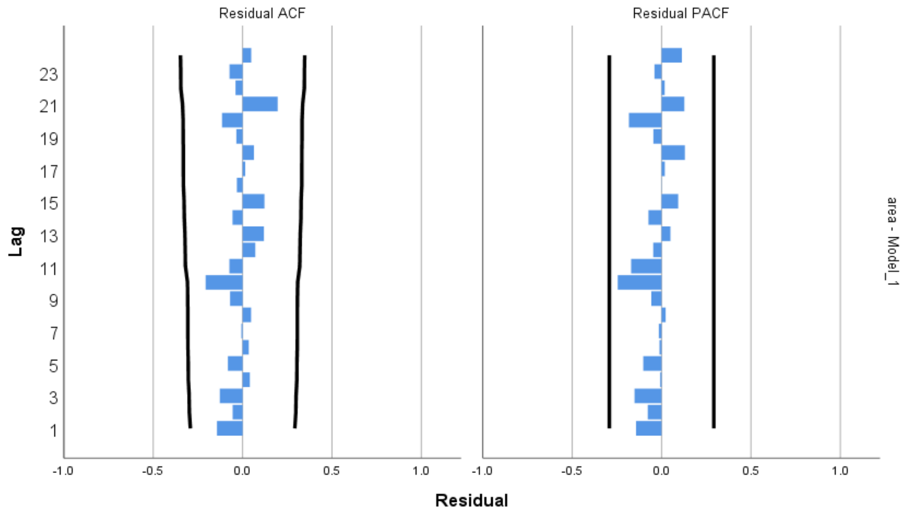
ACF and PACF of mustard crop productivity

Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
productivity-Model_1	ACF	-0.041	0.036	0.04	-0.094	0.09	-0.065	0.094	0.059	-0.128	0.066	-0.069	-0.04	0.052	-0.181	0.14	-0.013
	SE	0.149	0.149	0.15	0.15	0.151	0.152	0.153	0.154	0.155	0.157	0.158	0.158	0.158	0.159	0.163	0.166

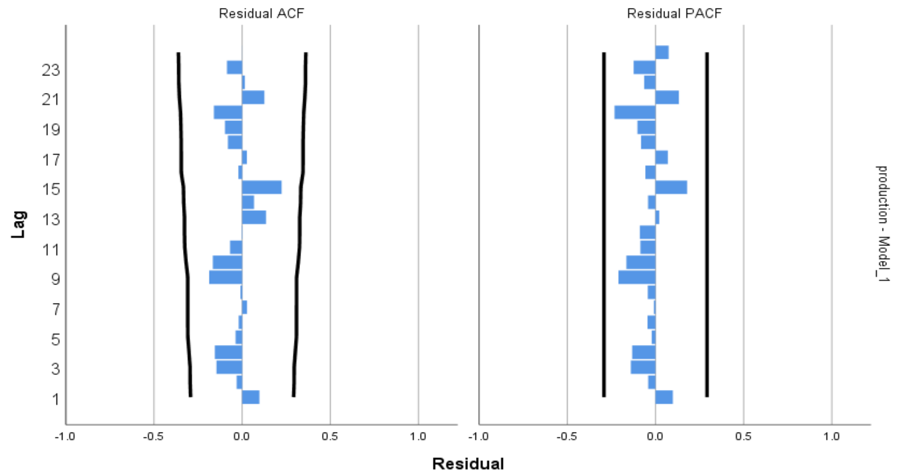
Model		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Productivity	PACF	-0.041	0.035	0.043	-0.093	0.081	-0.055	0.093	0.054	-0.114	0.034	-0.039	-0.047	0.035	-0.158	0.106	0.024
model-1	SE	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149

In the next step the order of AR and MA terms such as p and q were identified using the ACF and PACF plots shown in Fig. 1. Different tentative models were identified using the orders of AR and MA terms. Fig. 2. Observed and Forecast value of area, production, productivity in M.P.

ACF and PACF of observed time series at different time lag for area



ACF and PACF of observed time series at different time lag for production



ACF and PACF of observed time series at different time lag for productivity

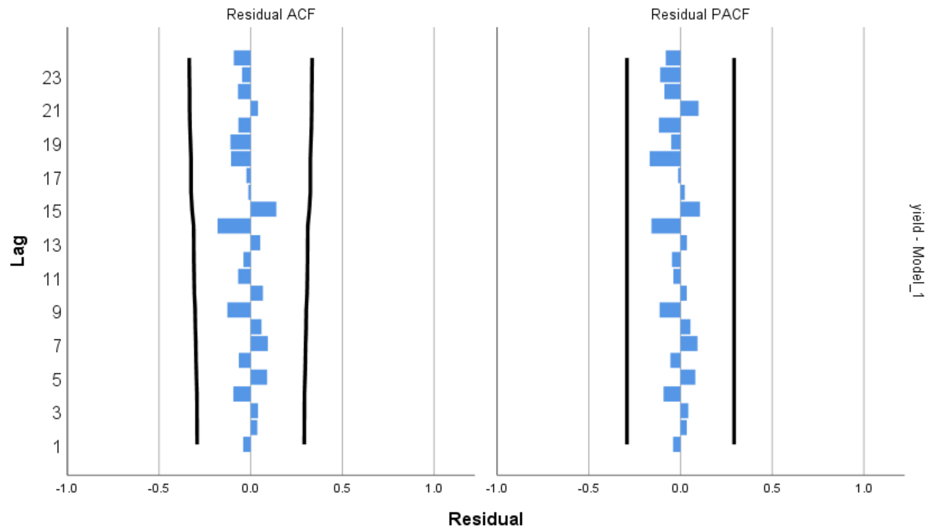


Fig. 1 : Identify the model by using ACF and PACF of observed time series at different time lag for mustard crop area, production and productivity in M.P.

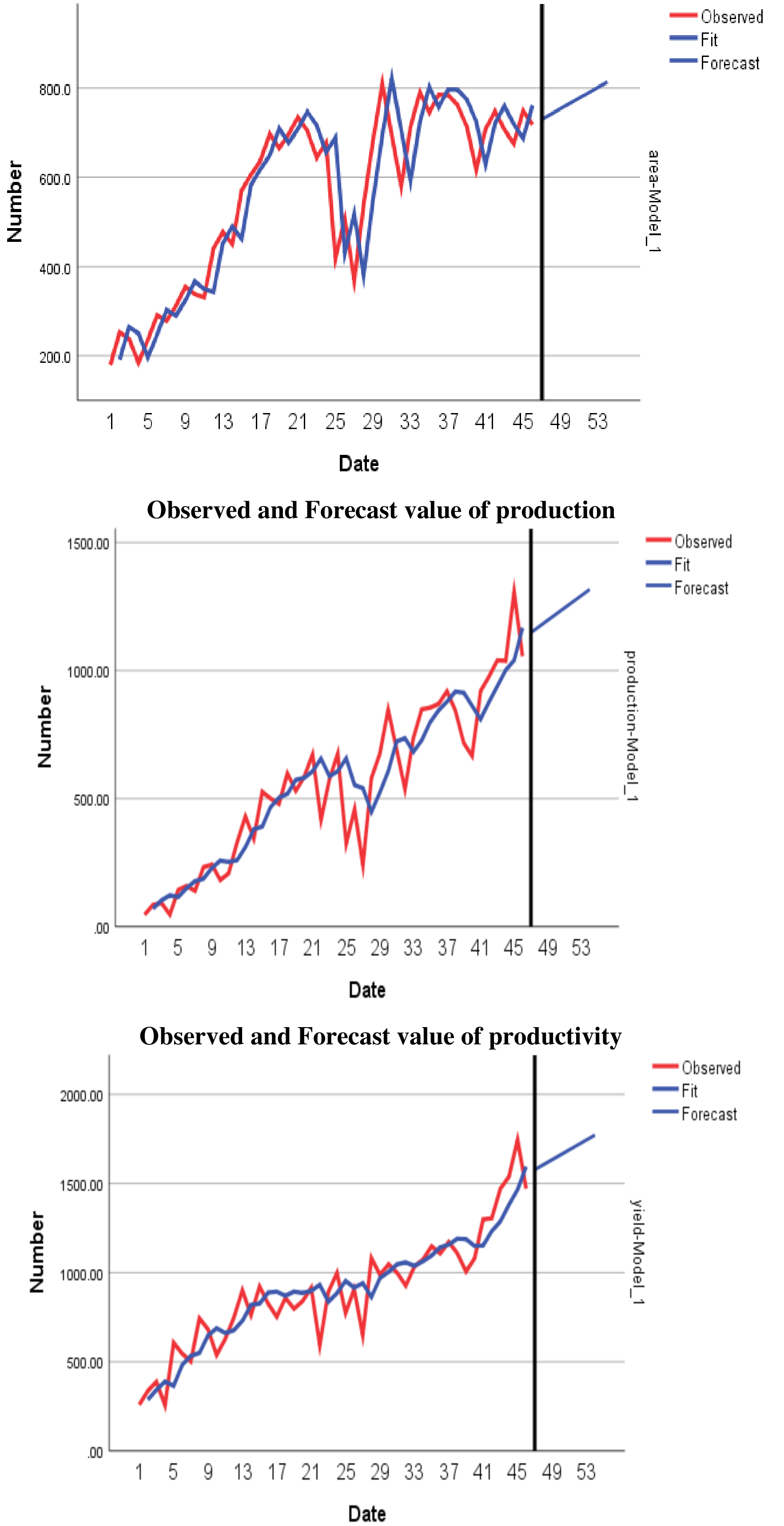


Fig. 2 : Observed and Forecast value of area, production, productivity in M.P.

Observed and Forecast value of area

The tentative models of area, production and productivity their estimated coefficients along with error measures are shown in the Table 3. The study of the table reveals that ARIMA for Area (0,1,0), Production (0,1,1), Productivity(0,1,1) with constant model has all the estimated coefficients significant.

Table 3 : Estimates of ARIMA Model of area, production and productivity in M.P.

		Estimate	SE	t	Sig.
Area (0,1,0)	Constant	11.956	12.276	.974	0.335
	AR Lag 1 (Φ)				
	MA Lag 1 (Φ)				
Production (0,1,1)	Constant	24.055	7.666	3.138	0.003
	AR Lag 1 (Φ)				
	MA Lag 1 (Φ)	0.613	0.123	4.978	0.000
Productivity (0,1,1)	Constant	27.689	7.549	3.668	0.001
	AR Lag 1 (Φ)				
	MA Lag 1 (Φ)	0.639	0.123	5.203	0.000

Table 4 : The result of cross validation of the selected best fit ARIMA model by one-step ahead forecasting from 2022-2029

	M.P.		
Year	**Area**	**Production**	**Productivity**
2022	729.90	1148.46	1577.57
2023	741.80	1172.52	1605.25
2024	753.80	1196.57	1632.94
2025	765.70	1220.63	1660.63
2026	777.70	1244.68	1688.32
2027	789.60	1268.74	1716.01
2028	801.60	1292.79	1743.70
2029	813.50	1316.85	1771.39

Conclusion

For M.P. ARIMA (0,1,0), (0,1,1) and (0,1,1) or we can say non-stationary model ARMA(1), (1,1) and (1,1).ARIMA (0,1,1) model found to be a good model for fitting on mustard crop of M.P. Graph of ACF of residuals and PACF of residuals indicates that choosing model is an appropriate model for forecasting of mustard crop area, production, productivity of M.P. Value of MAPE is found to be in range of good accuracy of forecasting by model.

Further Work

ARIMA modelling also performed for annual mustard area, production and productivity forecasting by using annual ARIMA models. ARIMA modelling also able to forecasting production and consumption of agricultural products. In order to analysis the time series data, SAS is good package. ANN is also a good analysis for mustard crop area, production and yield forecasting.

References

Alam, W., Sinha, K., Kumar, R.R., Ray, M., Rathod, S., Singh, K.N. And Arya, P. (2018). Hybrid linear time series approach for long term forecasting of crop yield. *Indian J. Agric. Sci,* (88), 1275-1279.

Hamjah, M.A. (2014). Rice production forecasting in Bangladesh, An application of Box-Jenkins ARIMA model. *Mathematical theory and modeling*, 4(4), 1-11.

Hemavathi, M. and Prabakaran, K. (2018). ARIMA model for forecasting of area, production and productivity of rice and its growth status in Thanjavur District of Tamil Nadu, India. *Int. J. Curr. Microbiol. App. Sci*, 7(2), 149-156.

Iqbal, Z. and Khan, F.U. (2022). Forecasting Rapeseed and Mustard Production In Pakistan, A Time Series Approach. *Sarhad Journal of Agriculture*, 38(2), 578-584.

Kumar, A. and Verma, U. (2020). Forecasting mustard yield in Haryana with ARIMA model. *The Pharma Innovation*, 9(4), 136-140.

Sharma, L. (2023). Forecasting of area, production and productivity of mustard using time series analysis in M.P. and India. M.Sc. (Ag.) Thesis, Jawahar lal Nehru Krishi Vishwa Vidyalaya, Jabalpur (Madhya Pradesh).

Mishra, A., Rawat, S., Gautam, S. and Mishra, E.P. (2022). Comparison between Different Mustard Yield Prediction Models Developed using Various Techniques for Udaipur

Regionof Rajasthan. *International Journal of Environment and Climate Change*, 12(11), 475-485.

Padhan (2012). Application of ARIMA model for forecasting agricultural productivity in India. *Journal of Agriculture and Social Sciences*, 8(2), 50-56.

Rathore, S., Babu, S., Shekhawat, K., Singh, K.V., Upadhyay, K.P., Singh, K.R.R.R., Singh, H. and Zaki, M.F. (2022). Oilseed Brassica Species Diversification and Crop Geometry Influence the Productivity, Economics and Environmental Footprints under Semi-Arid Regions. *Sustainability*, (14), 2-18.

Rajpoot, K. and Kushwaha, K.S. (2017). Prediction of mustard production in Madhya Pradesh through a multiple regression approach. *Trends in Biosciences*, 10(22), 4500-4506.

Rajpoot, K., Singla, S., Singh, A. and Shekhar, S. (2022). Impact of COVID-19 lockdown on prices of potato and onion in metropolitan cities of India. *Journal of Agribusiness in Developing and Emerging Economies*, 12(3), 386-399.

Sadat, N.N. (2016). Implementation of time series approaches to financial data. 1-92.