Journal of Pharmacognosy and Phytochemistry
Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2019; 8(3): 4383-4387
Received: 13-03-2019
Accepted: 15-04-2019

SK Mahapatra
Department of Agricultural Statistics, College of Agriculture, OUAT, Bhubaneswar, Odisha, India

A Satapathy
Assistant Professor, M.S.Swaminathan School of Agriculture, CUTM, Paralakhemundi, Odisha, India

Correspondence
A Satapathy
Assistant Professor, M.S.Swaminathan School of Agriculture, CUTM, Paralakhemundi, Odisha, India

An application of box-jenkins methodology for forecasting of green gram productivity in Odisha

SK Mahapatra and A Satapathy

Abstract
A study was conducted on forecasting of green gram productivity of Odisha. Box-Jenkins Autoregressive integrated moving average (ARIMA) time-series methodology was considered for forecasting. The different ARIMA models are judged on the basis of Autocorrelation Function (ACF) and Partial autocorrelation Function (PACF) at various lags The data from 1971-72 to 2006-07 are used for model building and from 2007-08 to 2015-16 used for successful cross-validation of the selected model on the basis of the absolute percentage error. The ARIMA models are fitted to the original time series data as well as the first difference data to check the stationarity. The possible ARIMA models are identified on the basis of significant coefficient of autoregressive and moving average components. The best fitted models are selected on the basis of low value of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). ARIMA (0,1,1) without constant found to be best fitted for green gram productivity having absolute percentage error ranging from 15.89% to 43.60% in cross-validation of model.The best fitted Box-jenkins ARIMA model has been used to forecast the productivity of green gram for the year 2016-17 to 2018-19. The model showed the forecast in productivity for the year 2018-19 to be about 231.53 kg per hectare with lower and upper limit 42.76 and 420.30 kg per hectare respectively.

Keywords: Box-Jenkins ARIMA, green gram, productivity, ACF, PACF, RMSE, MAPE

Introduction

Pulses are an important commodity group of crops that provide high quality protein complementing cereal proteins for pre-dominantly substantial vegetarian population of the country. Pulses are the major sources of protein in the diet. Green gram is the important pulse crop in the state of Odisha. Green gram or mung bean (*Vigna radiata* L.) is the major pulse crop of the state Odisha covering total area of 8.36 lakh ha with average productivity 434 kg per ha. The contribution of green gram to the total pulse area in Odisha is 41.23%, during the year 2015-16. The share of green gram towards total production is 38.57%, during 2015-16.

Green gram is the important pulse crop in the state of Odisha. Green gram or mung bean (*Vigna radiata* L.) is the major pulse crop of the state Odisha covering total area of 8.36 lakh ha with average productivity 434 kg per ha. The contribution of green gram to the total pulse area in Odisha is 41.23%, during the year 2015-16. The share of green gram towards total production is 38.57%, during 2015-16.Pulses are grown in all the thirty districts of Odisha. At present, pulses are grown in 18.7 lakh ha area with production of 9.4 lakh tonnes and productivity of 502 kg/ha. The Mahanadi Delta, the Rusikulya plains and the Hirakud & Badimula regions are favorable for the cultivation of pulses. The Panpose plateau is dominated by pulses. The Rusikulya plain is the most important agricultural region of Orissa and is dominated by pulse

Crop area estimation, forecasting of production and crop yield are an essential procedure in supporting policy decision regarding land use allocation, food security and environmental issues. Various approaches have been used for forecasting such agricultural systems. Several studies have been carried out on the univariate time series models known as Auto regressive integrated moving average (ARIMA) models. The popularity of ARIMA model is due to its statistical properties as well as known Box and Jenkins methodology. This study helps to the policy makers to get an idea about the future requirements, enabling to take appropriate measures like selection of high yielding varieties, conducting trainings to farmers to improve cultural practices, adequate supply of inputs and use of latest technologies. Import and export of these crops can also be planned.

Materials and methods

The secondary data on area, production green gram for the period from 1971-72 to 2015-16 were collected from Directorate of Agriculture and Food Production, Government of Odisha.

In time series analysis, the Box–Jenkins method applies autoregressive moving average ARMA or ARIMA models to find the best fit of a time-series model to past values of a time series.

To overcome the difficulty in describing the dynamic structure of the data by fitting Autoregressive (AR) and Moving average (MA) models, the autoregressive moving average (ARMA) models was introduced. The ARMA models, which includes the order of differencing (which is done to stationarise the data) is known as Autoregressive integrated moving average (ARIMA) models. The ARIMA model with parameter (p,d,q) is fitted by univariate Box-jenkins techniques (Box and Jenkins, 1976) [2]. This model includes Autoregressive of order *p*, differencing to make stationary series of degree *d* and moving average of order *q*.

The time series data is said to be stationary, if it has constant mean and variances over the time. The original data plotted first and verified for stationarity, if the data is found to be non-stationary from the graph, then the first difference of the data is plotted and checked for stationarity (Dash, et.al. (2017)) [3] In this manner we proceed till the data become stationary. The maximum order of differencing (d) is usually 2.

The adequacy of the selected model is checked using Box-Ljung test. A formal test of the fitness of the model is also done by using Box-Ljung test of the residuals (Ljung and Box (1978)) [6] is done in following manner:

Null hypothesis is set as H_0: The errors are distributed randomly.

And the alternate hypothesis H_1: The errors are non-random.

The Box-Ljung test statistic,

$$Q = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k},$$

Where n is the number of observations,

$r_k$is the estimated autocorrelation of the series at lag k = 1,2,…,m and *m* is the number of lags being considered. Here the null hypothesis is rejected i.e., the errors are not independent if $Q \geq \chi^2_{1-\alpha, h}$

The null hypothesis is accepted i.e., the errors are independent if $Q < \chi^2_{1-\alpha, h}$

Where,$\chi^2_{1-\alpha, h}$is the chi-square distribution table value with 'h' degrees of freedom and level of significance α such that $P(\chi^2_h > \chi^2_{1-\alpha, h=}) = 1-\alpha$

Here the degrees of freedom, h = (m – p – q); p and q are the numbers of AR and MA terms, respectively. The Box-Ljung test is done by help of forecasting tool of SPSS 20.0.

The following model fit statistics are used to select the best fit model:

The following model fit statistics are used to select the best fit model:

1. Root mean square error (RMSE)

$$RMSE = \frac{\sum_t e_t^2}{n-2}$$

2. Mean absolute percentage error (MAPE)

$$MAPE = \frac{\frac{\sum_t |Y_t - \hat{Y}_t|}{yt}}{n} \times 100$$

The models among the selected ARIMA models have lowest value of RMSE and MAPE is considered to be the best-fit model from the respective data set.

Nearly 20% of the data at the end period is not used for model building and left out for cross-validation of the selected model. The actual value of the left out period and the forecasted value of the left out period from the selected model are used for cross-validation. Here the data from 1971-72 to 2006-07 are used for model building and data from 2007-08 to 2015-16 are used for cross-validation.

For this, the percentage error is calculated.

$$\% \text{ of Forecasting Error} = \left(\frac{Y-\hat{Y}}{Y}\right) \times 100$$

Where, Y is the observed value of remaining eight years is the forecast values of remaining eight years

Lower the value of forecasting error percentage, better is the prediction by the selected model.

After successful cross-validation of the selected best fit ARIMA model, it is used for forecasting; generally short term forecasting is used in case of ARIMA techniques. This is because if we go on predicting for longer periods than the error associated with the prediction will increase. So, ARIMA should be used for short term forecasting (Sarika *et al.* 2011) [13].

Results and discussion

Appropriate ARIMA model is fitted to the data on productivity of Green gram for the purpose of forecasting. Data used for model building is from the year 1971-72 to 2006-07. The data from 2007-08 to 2015-16 is used for cross validation of the selected model and the forecasting is done for the years 2016-17, 2017-18 and 2018-19 by using the selected best fit model.

The original data plot of productivity of green gram as shown in Fig. 1 data is non-stationary. Thus, the first difference of the data on area is plotted and shown in Fig. 2 This plot shows that the first differences data is found to be stationary.

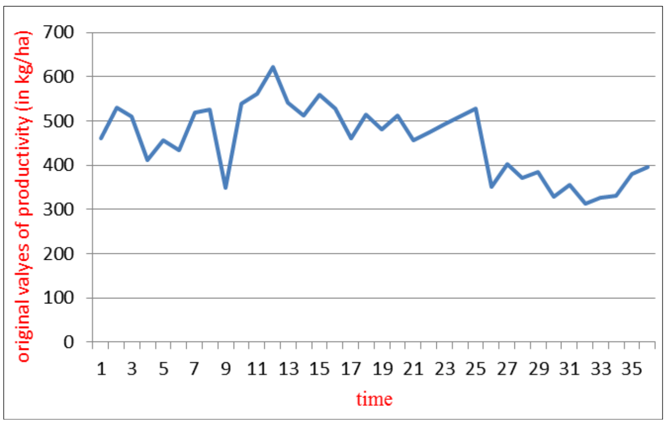

Fig 1: Plot of original values of productivity of green gram vs time

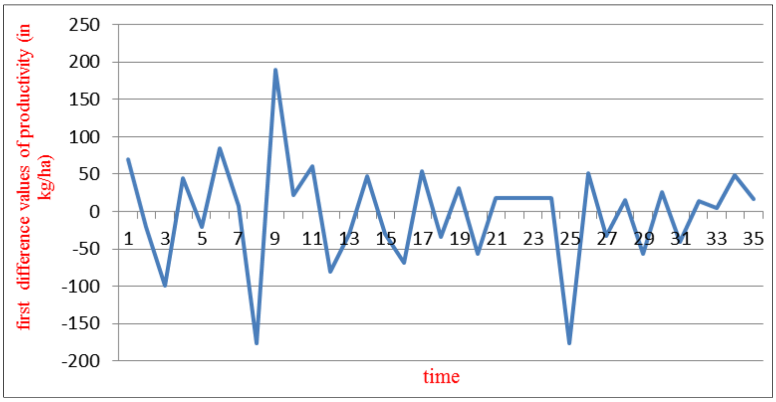

Fig 2: Plot of first difference values of productivity of green gram vs time

The ACF and PACF plot of first difference values of green gram productivity is shown in Fig. 3., which suggests that the tentative value of *q* and *p* that would be suitable for production of green gram is *q*=1 and *p*=0, Thus the ARIMA model that found to be best fitted for productivity of green gram is ARIMA(0,1,1)

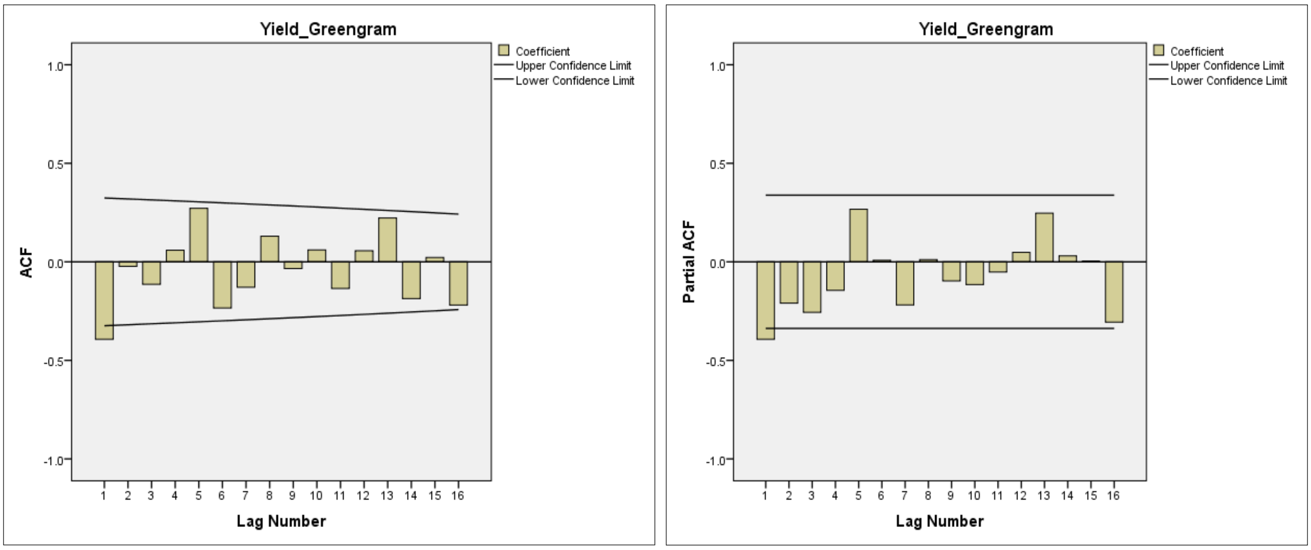

Fig 3: ACF and PACF plot of first difference values of yield of green gram

Table 1: Coefficients of the AR and MA components of the fitted ARIMA model considered for forecasting area, production and productivity of green gram in Odisha

	Best fit ARIMA model	Constant (μ)	Coefficient of autoregressive components		Coefficient of moving average components	
			α_1	α_2	θ_1	θ_2
Productivity	011	10.008 (5.173)	-	-	0.987 (0.783)	-
	011 (without constant)	-	-	-	0.651** (0.145)	-

(Figures in the parentheses indicate the standard error)
* Significant at 5% level of significance; ** Significant at 1% level of significance best fit models)

The study of table 2 shows that all the fitted model satisfy the assumptions of normality of error as they all have non-significant S-W Statistic and also all the models are found to be adequate due to non-significant of Ljung-Box Q Statistic. For area under green gram, the ARIMA(1,1,0) without constant has low value of RMSE and MAPE. So the best fit model is ARIMA(1,1,0) without constant. In case of production of green gram, the ARIMA(2,1,0) with constant has low value of RMSE and MAPE. So the best fit model is ARIMA(2,1,0) with constant. For productivity of green gram, the ARIMA(0,1,1) without constant has low value of RMSE and MAPE. So the best fit model is ARIMA(0,1,1) without constant.

Table 2: Model Fit Statistics and Residual Diagnostics of the ARIMA models fitted for area, production and productivity of green gram in Odisha

	Model	Model fit statistics		Residual diagnostics	
		RMSE	MAPE	Ljung – Box Q Statistic	Shapiro-Wilk's Statistic
Productivity	011	60.709	10.752	20.592 (7.657)	0.951 (0.636)
	011 (without constant)	60.152	10.309	19.082 (6.435)	0.957 (0.754)

(Figures in the parentheses indicate the standard error)
(Models highlighted as bold are the best fit models)

The cross validation of the selected best fit ARIMA(0,1,1) without constant model for productivity of green gram presented on the table 3 shows that absolute percentage error are quite low, thus the selected model is successfully cross validated.

Table 3: Cross validation of the selected best fit ARIMA (0, 1, 1) without constant model for productivity of green gram Odisha

Year	Actual value (in kg/ha) (Y)	Forecasted value (in kg/ha) $(\hat{Y})$	Error $(Y-\hat{Y})$	Absolute Percentage Error $\left(\frac{\lvert Y-\hat{Y}\rvert}{Y}\right)$ X 100
2007-08	409	344	65	15.89242
2008-09	452	334.97	117.03	25.89159
2009-10	411	325.7	85.3	20.75426
2010-11	434	316.18	117.82	27.14747
2011-12	416	306.43	109.57	26.33894
2012-13	476	296.45	179.55	37.72059
2013-14	476	286.22	189.78	39.86975
2014-15	477	275.76	201.24	42.18868
2015-16	470	265.06	204.94	43.60426

The forecasted values for the area, production and productivity of green gram are obtained from respective best fit ARIMA model are presented in table 4 shows that both area and production shows an increase in their forecasted values but there is decrease in the value of productivity from 2016-17 to 2018-19.

Table 4: Forecasted values (with 95% confidence limits) for area, production and yield of green gram in Odisha by using the selected ARIMA model

	Year	Forecasted value	Lower Confidence Limit (95%)	Upper Confidence Limit (95%)
Productivity (in kg/ha)	2016-17	254.12	75.44	432.80
	2017-18	242.94	59.15	426.74
	2018-19	231.53	42.76	420.30

Figure 4.10,4.11 and 4.12 shows that the observed values and fit values of area, production and productivity of green gram along with their upper and lower limit as obtained from their last best fit ARIMA model.

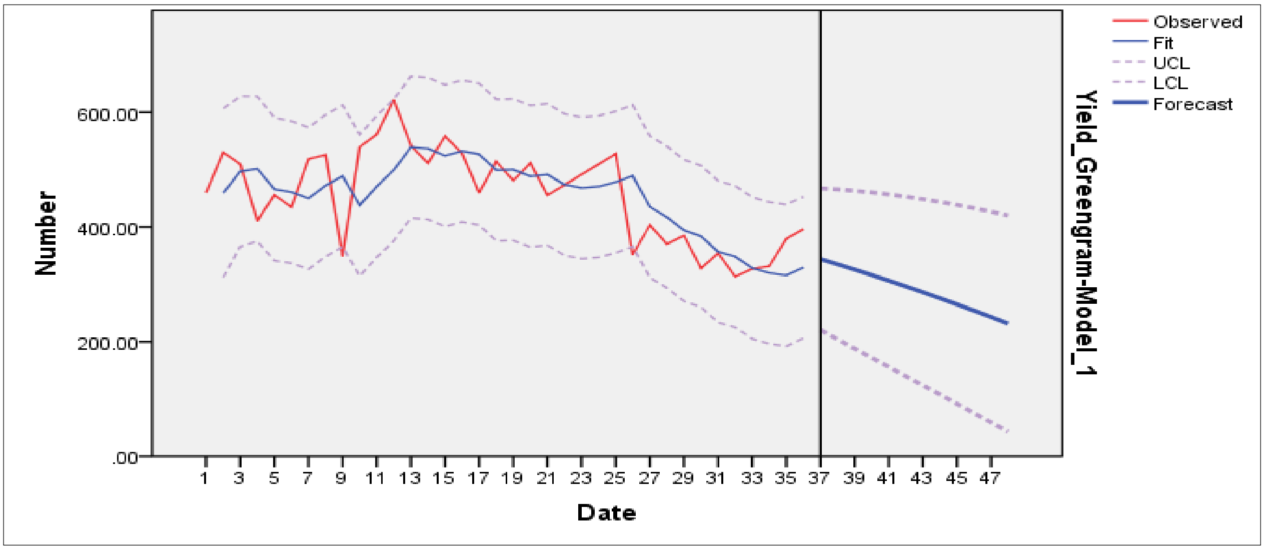

Fig 4: Observed and fit values of productivity (yield) along with upper and lower limit by using best fit ARIMA(0,1,1) without constant model

Conclusion

For productivity of green gram ARIMA (0,1,1) and ARIMA (0,1,1) without constant are the selected model. It is found that the constant is non-significant in ARIMA (0,1,1). So, that ARIMA (0,1,1) without constant is also fitted. Due to its significant coefficient of MA (1) and low value of RMSE and MAPE, ARIMA (0,1,1) is the best fitted model.

After successful cross validation of the best fitted model on productivity of green gram, it is found that the absolute percentage error is lowest (15.89%) during the year 2007-08 and highest absolute percentage error (43.60%) during the year 2015-16.

After successful forecasting it is found that the productivity of green gram is expected to be decrease for the future year i.e. from 2016-17 to 2018-19.

References

1. Amin M, Amanullah M, Akbar A. Time series modeling for forecasting wheat production of Pakistan. The journal of Animal and Plant Sciences. 2014; 24(5):1444-1451.
2. Box GEP, Jenkins GM. Time series analysis: forecasting and control, second edition, Holden day, 1976.
3. Dash A, Dhakre DS, Bhattacharjee D. Forecasting of food grain production in Odisha by fitting ARIMA

model. Journal of pharmacognosy and phytochemistry. 2017; 6(6):1126-1132.

4. Dasyam R, Bhattacharyya B, Mishra P. Statistical modeling to area, production and yield of potato in west Bengal. International Journal of Agriculture Sciences. 2016; 8(53):2782-2787.
5. Folarin SB, Iyiola OS. Causality and invariability of autoregressive integrated moving average (arima) model, Asian Journal of Current Engineering and Maths. 2013; 2(4):260-266.
6. Ljung GM, Box GEP. On a Measure of a Lack of Fit in Time Series Models. Biometrika. 1978; 65(2):297-303.
7. Odisha Agriculture Statistics. Directorate of Agriculture and Food Production, Government of Odisha, Bhubaneswar.
8. Padhan PC. Application of ARIMA Model for Forecasting Agricultural Productivity in India. Journal of Agriculture & Social Sciences. 2012; 8(2):50-56.
9. Prabakaran K, Sivapragasam C. Forecasting areas and production of rice in India using ARIMA model, International Journal of Farm Sciences. 2014; 4(1):99-106.
10. Prakash K, Muniyandi B. Application of ARIMA model for forecasting production of jasmine flower in Madurai district of Tamil Nadu, India. American International Journal of Research in Humanities, Arts and Social Sciences. 2014.
11. Rahman NAF, Aziz MA, Mukhlesur M. Modeling on Grass Pea and Mung Bean Pulse Production in Bangladesh Using ARIMA Model. IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS). 2013; 6(1):20-31.
12. Raman RK, Sathianandan TV, Sharma AP, Mohanty BP. Modeling and Modeling and Forecasting Marine Fish Production in Odisha Using Seasonal ARIMA Model. The National Academy of Sciences. 2015; 40(6):393-397.
13. Sarika Iquebal MA, Chattopadhyay C. Modeling and forecasting of pigeon pea (*Cajanus cajan*) production using autoregressive integrated moving average methodology. Indian Journal of Agricultural Sciences. 2011; 81(6):520-523.