Agro Economist - An International Journal
Citation: AE: 5(1): 01-06, June 2018
DOI: 10.30954/2394-8159.01.2018.1
©2018 Renu Publishers. All rights reserved

Forecasting Maize Production in India using ARIMA Model

Pawan Kumar Sharma[1*], Sudhakar Dwivedi[1], Lyaqat Ali[2] and R.K. Arora[1]

[1]*Sher-e-Kashmir University of Agricultural Sciences & Technology of Jammu (J&K), India*
[2]*Sher-e- Kashmir University of Agricultural Sciences & Technology of Kashmir (J&K), India*

*Corresponding author: pawanvatsya@gmail.com

ABSTRACT

Forecasting is an important tool to estimate the area, production and productivity of any crop in near future. There are several methods available for foresting the future figures and autoregressive integrated moving average (ARIMA) is one of them. Maize is an important cereal of India, keeping in view its importance for rainfed areas of the country and diverse uses. The present study was conducted to forecast maize production for the year 2018 to 2022 based on the estimation of suitable ARIMA model. The analysis of ACF & PACF of differenced series revealed that ARIMA (2,1,0) was the most suitable model for forecasting based on the diagnostics, such as ACF, PACF, AIC, SBC etc. The selected ARIMA model predicted an increase of 13.76 percent increase in maize production in next five years w.e.f. 2017 to 2022.

Keywords: Maize, forecasting, ARIMA model

Maize is the third most important cereal In India after rice and wheat, however globally, maize is known as queen of cereals due to its highest genetic yield potential and the second most important cereal crop in terms of acreage. It also accounts for the major share of the coarse grain production. Maize can be grown in wider range of agro-climatic zones, ranging from extreme semi-arid to sub-humid and humid regions (Joshi *et al.* 2005). The area, production and productivity of maize in India have shown an increasing trend in recent years, especially with the introduction of single cross hybrids. The area under Maize has gone up with 3 percent Compound Annual Growth Rate from 6600’000 hectares in the year 2000-01 to 10200’000 hectares in the year 2016-17 (FICCI, 2018).Crop yield forecasting is a crucial activity for ensuring food security, managing import/export and implementing price policy. All over the world, scientists applied different types of models to obtain accurate forecasts for area, production and productivity of different field crops. Martin *et al.* (2000) applied agroclimatological model to create a historical record of maize water stress as a function of evapotranspiration. Rathod (2017) forecasted maize yield in India using an autoregressive integrated moving average (ARIMA) model and genetic algorithm (GA) approach. Horie *et al.* (1992) forecasted rice yield using Simulation Model for Rice-Weather Relationships (SIMRIW). Manatsa *et al.* (2011) demonstrated the possibility of exploiting the complementary roles that remote sensing, modeling, and geospatial data analysis can play in forecasting maize yield. Tahir and Habib (2013) checked the trend analysis of area and production for Maize in Pakistan, whereas Bussay *et al.* (2015) applied regression-based crop yield forecasting methodology currently employed in the MARS-Crop Yield Forecasting System (M-CYFS) for maize in Hungary. The present paper also attempts to forecast maize yield in India for the year 2018 to 2022, using ARIMA model.

MATERIALS AND METHODS

The present study was based on secondary data on maize production in India, collected from different publications and websites, including Open Government Data (OGD) Platform India and Department of Agriculture Cooperation & Farmers Welfare. The data was analysed for forecasting maize production using ARIMA model as discussed below:

Auto Regressive Integrated Moving Average (ARIMA) Model

Box and Jenkins, 1976 suggested a method for identifying, estimating, and checking models for a specific time series dataset popularly known as ARIMA (p,d,q) model. A non-seasonal ARIMA model is classified as an "ARIMA (p,d,q)" model, whereas 'p' is the number of autoregressive terms, 'd' is the number of non-seasonal differences needed for stationarity, and 'q' is the number of lagged forecast errors in the prediction equation. The lags of the stationarized series in the forecasting equation are called "autoregressive (AR)" terms, lags of the forecast errors are called "moving average (MA)" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series.

Model Identification, diagnostic checking and forecasting

The application of ARIMA model involves three steps; Identification of model (stationarity tests for assessing need for differencing, computation of autocorrelations, inverse autocorrelations, partial autocorrelations, and cross correlations), Estimation and diagnostic checking (estimation of the parameters of identified model, significance tests for parameter estimates, goodness-of-fit statistics, diagnostic tests etc.) and Forecasting (to forecast future values of the time series with confidence intervals).The Auto Regressive Integrated Moving Average (ARIMA) models can assume any of the equations, such as Firstorder autoregressive model [ARIMA(1,0,0)], Random walk [ARIMA(0,1,0)], Differenced first-order autoregressive model [ARIMA(1,1,0)], Simple exponential smoothing [ARIMA(0,1,1) without constant], Simple exponential smoothing with growth [ARIMA(0,1,1) with constant], Linear exponential smoothing [ARIMA(0,2,1) or (0,2,2) without constant], Damped-trend linear exponential smoothing [ARIMA(1,1,2) without constant].

RESULTS AND DISCUSSION

Trends in area and production of maize in India

The area and production of maize in India has increased over the years from 4407'000 hectares and 4109'000 MT in the year 1960-61 to 9500'000 hectares and 18730'000 MT, respectively in the year 2016-17. The descriptive statistics of area and production of maize in India is presented in Table 1.

Table 1: Summary statistics of area and production of maize in India

Production		Area	
Mean	10383310	Mean	6352291
Standard Error	739043.7	Standard Error	201470
Standard Deviation	5628389	Standard Deviation	1534350
Sample Variance	3.17E+13	Sample Variance	2.35E+12
Kurtosis	-0.19843	Kurtosis	1.910944
Skewness	0.995436	Skewness	0.193879
Range	18891000	Range	8790454
Minimum	4109000	Minimum	1152300
Maximum	23000000	Maximum	9942754
Sum	6.02E+08	Sum	3.68E+08
Count	58	Count	58

Identification of time series model for maize

The first step in identifying the perfect model is to find out the trend and stationery of data, which can be assessed from the line diagram and statistically by Augmented Dickey-Fuller (ADF) Test. The line diagram of maize production (original series) as depicted in Fig. 1 indicates that the production shows increasing trend over the years and thus indicated non-stationarity of data.

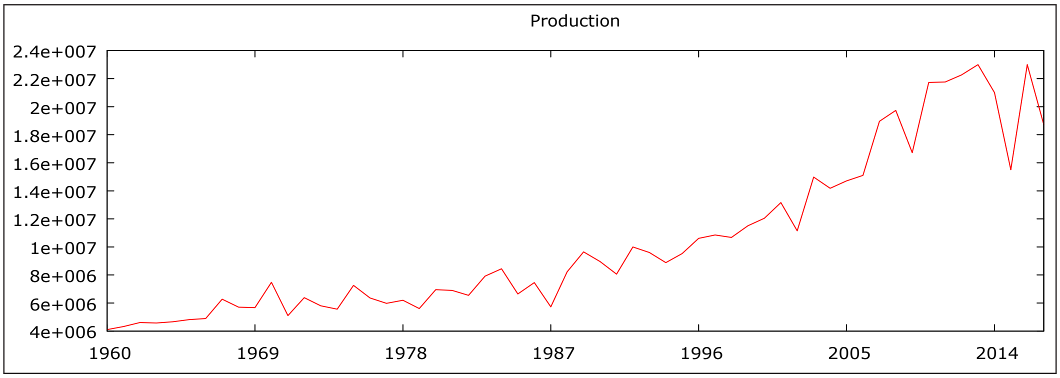

Fig. 1: Line diagram of maize production (original series)

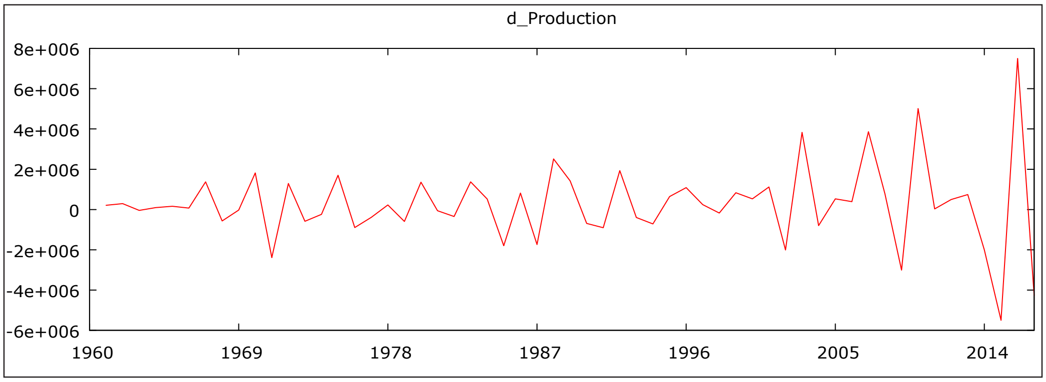

Fig. 2: Line diagram of differenced series

Table 2: Augmented Dickey-Fuller Test for original & differenced series

Type	Null hypothesis	Production model	(a-1)	Tau	Pr<Tau	F	Pr>F
ADF Test for original series (Lags 10)							
With Constant	a = 1	(1-L)y = b0 + (a-1)*y(-1) + ... + e	0.0105623	0.242965	0.9753	F(2, 51): 12.683	0.0000
With Constant & Trend	a = 1	(1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e	-0.171849	-1.60153	0.7928	F(2, 50): 8.017	0.0010
ADF Test for differenced series (Lag 10)							
With Constant	a = 1	(1-L)y = b0 + (a-1)*y(-1) + ... + e	-2.05898	-8.46369	0.0000	F(2, 51): 12.683	0.0000
With Constant & Trend	a = 1	(1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e	-2.08111	-8.50701	0.0000	F(2, 50): 8.017	0.0010

The first difference of the given data, however indicates thatthe series is following stationarity after differencing (Fig. 2).

Further, the results of Augmented Dickey-Fuller Unit Root Tests (Table 2) confirms that the probability of Tau in case of both the cases with constant mean and with constant and trend for differenced series data is significant, and thus indicates that after the first difference, the series became stationary. Thus, the value of 'd' was fixed as '1'. In order to identify the order of Autoregressive 'AR' for the value 'p' and 'the order of Moving Average 'MA' for the value of q', correlograms of partial autocorrelation functions (PACF) and autocorrelation functions (ACF), respectively were examined.

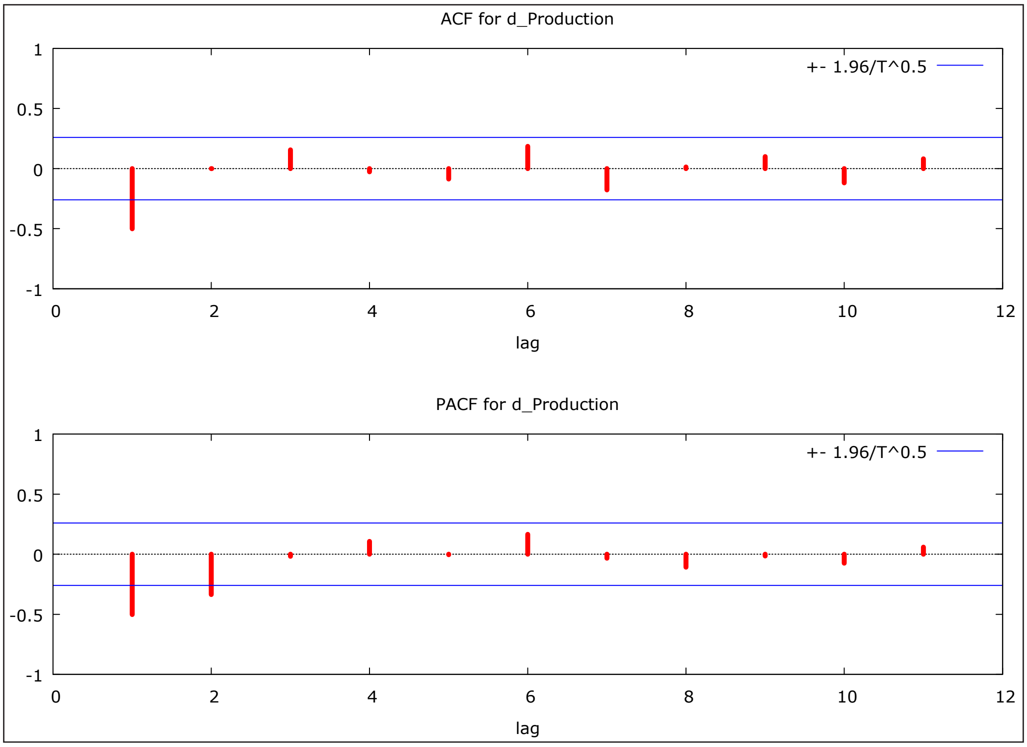

Fig. 3: Correlograms of ACF and PACF of differenced series

The correlogram of autocorrelation function (ACF) of differenced series (Fig. 3) indicates that the auto correlation function falls immediately after one lag, hence the value of 'q' was decided to be '0'. Further, the correlogram of partial autocorrelation function (PACF) of differenced series (Fig. 3) indicates that the auto correlation function falls quickly after the lag 2, hence the value of 'p' was decided to be '2'. Thus, the ARIMA (2,1,0) model was selected for parameter estimation, model validation and forecasting of maize production in India.

Table 3: Log likelihood &AIC Values for Different ARIMA Models

Model	AIC	SBC
ARIMA(011)	1799.590	1806.554
ARIMA(110)	1802.026	1808.156
ARIMA(111)	1799.833	1808.005
ARIMA(210)	1798.366	1806.539
ARIMA(012)	1799.069	1807.241
ARIMA(211)	1800.350	1810.565
ARIMA(112)	1800.735	1810.950
ARIMA(212)	1802.178	1814.436

The use of other diagnostics such as minimum value of AIC & SBC, significance of AR and MA parameters also confirms the selection of the ARIMA (2,1,0) model (Table 3).

Parameter Estimates for maize production in India

After identifying the suitable ARIMA (2, 0,1) structure, the parameters of identified model were assessed and are presented in Table 4.

Table 4: Estimates of ARIMA (2,1,0) model for maize production in India

Model Parameter	Estimate	Std. Error	Z	p-value
Intercept	275744***	105263	2.620	0.0088
Autoregressive, Lag 1	−0.682301***	0.124857	−5.465	<0.0001
Autoregressive, Lag 2	−0.348459**	0.141975	−2.454	0.0141

***significant at 1% level; **significant at 5% level.

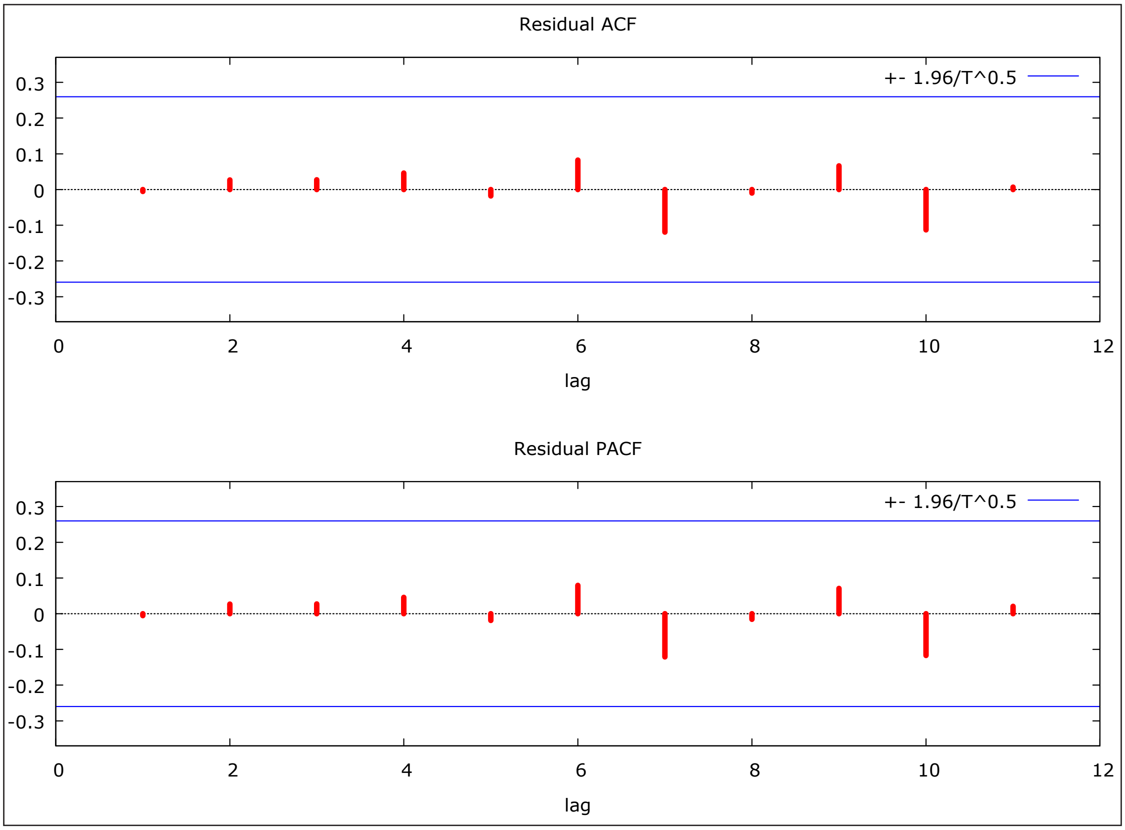

Fig. 4: Residual plots (correlograms) of ACF and PACF for identified ARIMA model (2,1,0)

Model Diagnostics

The goodness of fit for the identified model was checked by plotting the residuals any systematic pattern, as shown in Fig. 4. As the time series plots of ACF and PACF of the residuals of fitted ARIMA (2,1,0) model, exhibited a non-significant pattern, the given model was considered as valid for forecasting.

Forecast of maize production in India

After the model identification, estimation of its parameters and diagnostic checks, the forecasting of was made to know the future values of maize production in India by using the selected and fitted ARIMA model (2,1,0). The forecasted values of maize production in India for the year 2018 to 2022 are presented in Table 5.

The graphical representation of actual and fitted values of maize production by ARIMA (2,1,0) model along with the forecasted values with 95 percent confidence interval is also presented in Fig. 5.

Table 5: Forecasting of maize production with confidence intervals

Year	Maize production forecast (MT)	95% confidence limits	
		Lower value	Upper value
2018	19589900	16467600	22712300
2019	21051100	17775000	24327200
2020	20314500	16768200	23860700
2021	20867900	16867900	24867900
2022	21306900	17073400	25540500

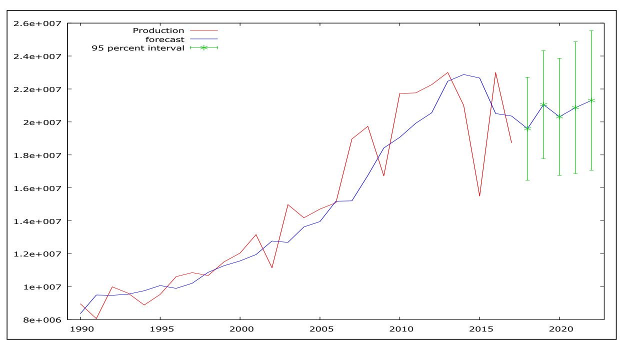

Fig. 5: Predicted & actual values of maize production & forecasts

CONCLUSION

The Autoregressive integrated moving average (ARIMA) model is considered to be one of the best model when the data consists if at least 50 observations. The present study attempts at modelling and forecasting of maize production in India was done using Autoregressive integrated moving average (ARIMA) model. Autocorrelation function (ACF) and partial autocorrelation function (PACF) functions were estimated, which led to the identification and construction of ARIMA model (2,1,0). The fitted model indicated an increase in maize production in the next five years from 19590’000 MT in the year 2018 to 21307’000 MT in the year 2022.

REFERENCES

Bussay, A. Velde, M., Fumagalli, D. and Seguini, L. 2015. Improving operational maize yield forecasting in Hungary. Agricultural Systems, **141**: 94-106.

FICCI. 2018. Website: http://www.ficci.in/events-page.asp?evid=23706, accessed on 10-07-2018.

GEP, B. and Jenkin, G.M. 1976. Time series of analysis. Forecasting and Control, Sam Franscico, Holden Day, California, USA.

Horie, T., Yajima, M. and Nakagawa, H. 1992. Yield forecasting. Agricultural Systems, **40**(1-3): 211-236.

Joshi, P.K., Singh, N.P., Singh, N.N., Gerpacio, R.V.and Pingali, P.L. 2005. Maize in India: Production Systems, Constraints, and Research Priorities. Mexico, D.F.: CIMMYT.

Manatsa, D., Nyakudya, I.W., Mukwada, G. and Matsikwa, H. 2011. Maize yield forecasting for Zimbabwe farming sectors using satellite rainfall estimates. *Natural Hazards,* 59(1): 447–463.

Martin, R.V., Washington, R. and Downing, T.E. 2000. Seasonal Maize Forecasting for South Africa and Zimbabwe Derived from an Agroclimatological Model. *Journal of Applied Meteorology,* **39**: 1473-1479.

Santosha Rathod, Singh, K.N., Prawin Arya, Mrinmoy Ray, Anirban Mukherjee, Kanchan Sinha, Prakash Kumar and Ravindra Singh Shekhawat. 2017. Forecasting maize yield using ARIMA-Genetic Algorithm approach. *Outlook on Agriculture*, **46**(4): 265-271.

Tahir, A. and Habib, N. 2013. Forecasting of maize area and production in Pakistan. *Crop Production,* **2**(2): 44-48.