Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Implementation of Time Series Stochastic Modelling for Zea mays Production in India

T. Jai Sankar *and* P.Pushpa

Department of Statistics, Bharathidasan University, Tiruchirappalli, Tamil Nadu, INDIA

tjaisankar@gmail.com*and*pushbkr@gmail.com

Article Info
Page Number: 611-621
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract

This study deals with the implementation of time series stochastic modelling for Zea mays (Maize) production in India during the years from 1951 to 2018. The demand for maize is spiralling in India. Maize can be grown in all seasons viz., Kharif (monsoon), post monsoon, Rabi (winter) and spring. The study considers Autoregressive (AR), Moving Average (MA) and ARIMA processes to select the appropriate ARIMA model forZea mays production in India. Based on ARIMA (p,d,q) and its components Autocorrelation Function (ACF), Partial Autocorrelation Function(PACF), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Normalized BIC and Box-Ljung Q statistics estimated, ARIMA (0,1,1) was selected. Based on the chosen model, it could be predicted that Zea mays production would increase to 32.12million tons in 2025 from 27.02million tons in 2019 in India.

Keywords: ARIMA, BIC, Forecasting, MAPE, Maize Production, RMSE.

Introduction

In the maize growing countries, India rank 4th in area and 7th in production, representing around 4% of the world maize area and 2% of total production. During 2018-19 in India, the maize area has reached to 9.2 million ha (DACNET, 2020). During 1950-51 India used to produce 1.73 million MT maize, which has increased to 27.8 million MT by 2018-19, recording close to 16 times increase in production. The average productivity during the period has increased by 5.42 times from 547 kg/ha to 2965 kg/ha, while the area increased nearly by three times. In India, maize is principally grown in two seasons, rainy (kharif) and winter (rabi). Kharif maize represents around 83% of maize area in India, while rabi maize correspond to 17% maize area. The stress prone ecology contributes towards lower productivity of kharif maize (2706 kg/ha) as compared to rabi maize (4436 kg/ha), which is predominantly grown under assured ecosystem. In recent past spring maize area is also growing quite fast in North Western parts of the country, in the states of Punjab, Haryana and Western Uttar Pradesh.

In Indian states, Madhya Pradesh and Karnataka has highest area under maize (15% each) followed by Maharashtra (10%), Rajasthan (9%), Uttar Pradesh (8%) and others with health benefits (Figure 1). After Karnataka and Madhya Pradesh Bihar is the highest maize producer. Andhra Pradesh is having the highest state productivity. Some districts like Krishna, West Godavari etc. records as high as 12 t/ha productivity.

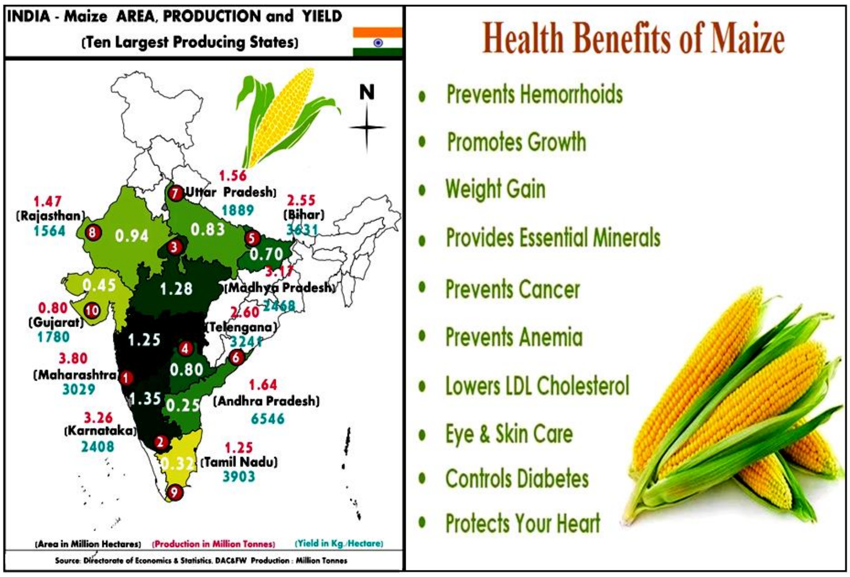

Figure 1. India's Maize Production States and Health Benefits

Bulk of the maize production in India, approximately 47%, is used as poultry feed. Of the rest of the produce, 13% is used as livestock feed and food purpose each, 12% for industrial purposes, 14% in starch industry, 7% as processed food, and 6% for export and other purposes.In the poultry feed industry maize constitutes about 60% of the feed and therefore is a critical raw material. The nutritional, minerals and vitamins values of maize per 100 grams are given in Figure 2:

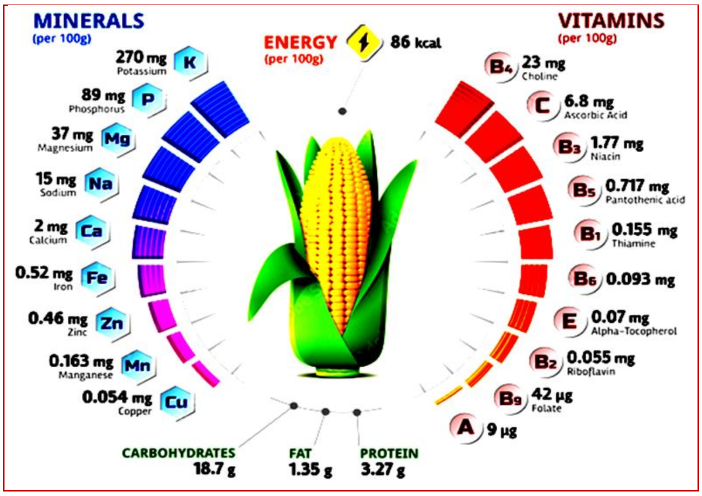

Figure 2. Nutritional, Minerals and Vitamins of Maize

Material and Methods

As the aim of the study was to design and development of stochastic modelling for *Zea mays* (Maize) production in India, various forecasting techniques were considered for use. ARIMA model, introduced by Box and Jenkins (1976), was frequently applied for discovering the pattern and predicting the future values of the time series data. Box and Pierce (1970) measured the distribution of residual autocorrelations in ARIMA. Akaike (1970) found the stationary time series by an AR (p), where p is finite and bounded by the same integer. Moving Average (MA) models were applied by Slutzky (1973). Md.Moyazzem Hossain andFaruq Abdulla (2016) applied ARIMA (0,2,1) model for yearly potato production in Bangladesh for the period from 1971 to 2013. Borkar et al. (2016) found that ARIMA (2,1,1) is the appropriate model for forecasting the production of cotton in India. Kour et al., (2017) applied ARIMA model for forecasting of productivity of pearl millet of Gujarat for the period from 1960-61 to 2011-12 and validated ARIMA (0,1,1) model performs quite satisfactorily as the RMAPE value is less than 6 percent. Hemavathi and Prabakaran (2018) found rice production data during 1990-2015 and applied ARIMA (0,1,1) model up to 2020. BholaNath et al. (2019) discovered ARIMA (1,1,0) model for wheat production in India for the period from 1949-50 to 2016-17 and forecasted up to 2026-27.

Stochastic time-series ARIMA models were widely used in time series data which are having the characteristics (Alan Pankratz, 1983) of parsimonious, stationary, invertible, significant estimated coefficients and statistically independent and normally distributed residuals. ARIMA model was used in this study, which required a sufficiently large data set and involved four steps: identification, estimation, diagnostic checking and forecasting. Model parameters were estimated to fit the ARIMA models.

Autoregressive process of order (p) is, $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$;

Moving Average process of order (q) is, $Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$; and

The general form of ARIMA model of order (p, d, q) is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where Y_t is maizeproduction, ε_t 's are independently and normally distributed with zero mean and constant variance σ^2 for t = 1,2,..., n; d is the fraction differenced while interpreting AR and MA and ϕs and θs are coefficients to be estimated.

Trend Fitting **:** The Box-Ljung Q statistics was used to transform the non-stationary data into stationarity data and also to check the adequacy for the residuals. For evaluating the adequacy of AR, MA and ARIMA processes, various reliability statistics like R^2, Stationary R^2, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and BIC were used. The reliability statistics viz. RMSE, MAPE, BIC and Q statistics were computed as below:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \right]^{1/2} ; \; MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{(Y_i - \hat{Y}_i)}{Y_i} \right| \textit{and } \mathrm{BIC(p,q) = ln\ v^*(p,q)+(p+q)\ [\ ln(n)\ /\ n]}$$

where p and q are the order of AR and MA processes respectively and n is the number of observations in the time series and v* is the estimate of white noise variance σ^2.

$$Q = \frac{n(n+2)\sum_{i=1}^{k} rk^2}{(n-k)}$$

where n is the number of residuals and rk is the residuals autocorrelation at lag k.

In this study, the data on *Zea mays* (Maize) production in India were collected from the Agricultural Statistics (2020)*, Directorate of Economics and Statistics, Department of Agriculture, Government of India for the period from 1951 to 2018 and were used to fit the ARIMA model to predict the future production.

Table 1. Actual MaizeProduction (million tons) in India

Year	**Production**	***Year***	**Production**	***Year***	**Production**	***Year***	**Production**
1951	1.73	*1969*	5.70	*1987*	7.59	*2005*	14.17
1952	2.08	*1970*	5.67	*1988*	5.72	*2006*	14.71
1953	2.87	*1971*	7.49	*1989*	8.23	*2007*	15.10
1954	3.04	*1972*	5.10	*1990*	9.65	*2008*	18.96
1955	2.98	*1973*	6.39	*1991*	8.96	*2009*	19.73
1956	2.60	*1974*	5.80	*1992*	8.06	*2010*	16.72
1957	3.08	*1975*	5.56	*1993*	9.99	*2011*	21.73
1958	3.15	*1976*	7.26	*1994*	9.60	*2012*	21.76
1959	3.46	*1977*	6.36	*1995*	8.88	*2013*	22.26
1960	4.07	*1978*	5.97	*1996*	9.53	*2014*	24.26
1961	4.08	*1979*	6.20	*1997*	10.77	*2015*	24.17
1962	4.31	*1980*	5.60	*1998*	10.82	*2016*	22.57
1963	4.61	*1981*	6.96	*1999*	11.15	*2017*	25.90
1964	4.56	*1982*	6.90	*2000*	11.51	*2018*	28.75
1965	4.66	*1983*	6.55	*2001*	12.04		
1966	4.82	*1984*	7.92	*2002*	13.16		
1967	4.89	*1985*	8.44	*2003*	11.15		
1968	6.27	*1986*	6.64	*2004*	14.98		

*Source: Directorate of Economics and Statistics, Department of Agriculture, Government of India

Results and Discussion

In this study, the data for *Zea mays* (Maize) production in India is collected from the period 1951 to 2018is given in Table 1. To fit an Autoregressive model, Autoregressive process for any variable involves four steps: identification, estimation, diagnostic and forecasting. ARIMA (p,d,q) model is fitted to check stationarity through examining the graph or time plot of the data. Figure 3 reveals that the data is non-stationary. The autocorrelation and partial autocorrelation coefficients of various orders of $Y_t$are computed Table 2. The graphs of ACF and PACF are given in Figure 4.

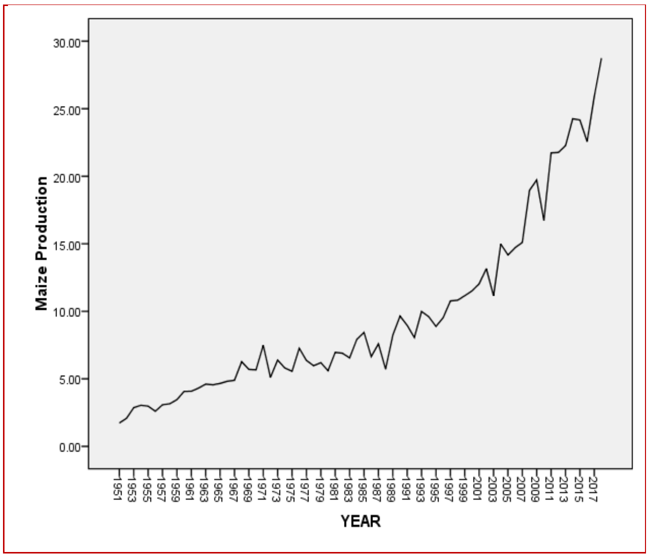

Figure 3 - Time plot of Maize Production

Table 2 - ACF and PACF of MaizeProduction

Lag	AC	Std. Error [a]	Box-Ljung Statistic	PAC	Std. Error	Lag	AC	Std. Error [a]	Box-Ljung Statistic	PAC	Std. Error
	Value	**Df**	**Sig.[b]**	**Value**	**Df**		**Value**	**Df**	**Sig.[b]**	**Value**	**Df**
1	0.902	0.119	57.792	0.902	0.121	***17***	0.147	0.104	366.763	-0.054	0.121
2	0.837	0.118	108.272	0.124	0.121	***18***	0.124	0.102	368.236	0.018	0.121
3	0.798	0.117	154.898	0.136	0.121	***19***	0.100	0.101	369.205	-0.038	0.121
4	0.735	0.116	195.074	-0.103	0.121	***20***	0.075	0.100	369.758	-0.003	0.121
5	0.668	0.115	228.825	-0.069	0.121	***21***	0.059	0.099	370.107	0.022	0.121
6	0.612	0.114	257.540	-0.024	0.121	***22***	0.033	0.098	370.222	-0.062	0.121
7	0.552	0.113	281.354	-0.036	0.121	***23***	0.015	0.097	370.246	-0.008	0.121
8	0.485	0.112	300.042	-0.070	0.121	***24***	0.003	0.096	370.247	0.052	0.121
9	0.453	0.111	316.575	0.131	0.121	***25***	-0.022	0.095	370.299	-0.098	0.121

10	0.395	0.110	329.401	-0.117	0.121	***26***	-0.044	0.094	370.521	-0.001	0.121
11	0.339	0.109	339.012	-0.013	0.121	***27***	-0.053	0.093	370.848	0.016	0.121
12	0.302	0.108	346.779	0.011	0.121	***28***	-0.071	0.092	371.450	-0.034	0.121
13	0.272	0.107	353.163	0.043	0.121	***29***	-0.099	0.091	372.638	-0.051	0.121
14	0.240	0.107	358.259	0.018	0.121	***30***	-0.118	0.089	374.376	-0.052	0.121
15	0.199	0.106	361.810	-0.084	0.121	***31***	-0.119	0.088	376.211	0.110	0.121
16	0.179	0.105	364.752	0.053	0.121	***32***	-0.134	0.087	378.580	-0.046	0.121

[a]The underlying process assumed is independence (white noise).

[b] Based on the asymptotic chi-square approximation.

The models and corresponding BIC values are given in Table 3. The value of normalized BIC is 0.526and R Squared value is 0.971. So the most suitable model for *Zea mays* (Maize) production is ARIMA(2,1,0) as this model has the lowest BIC value.

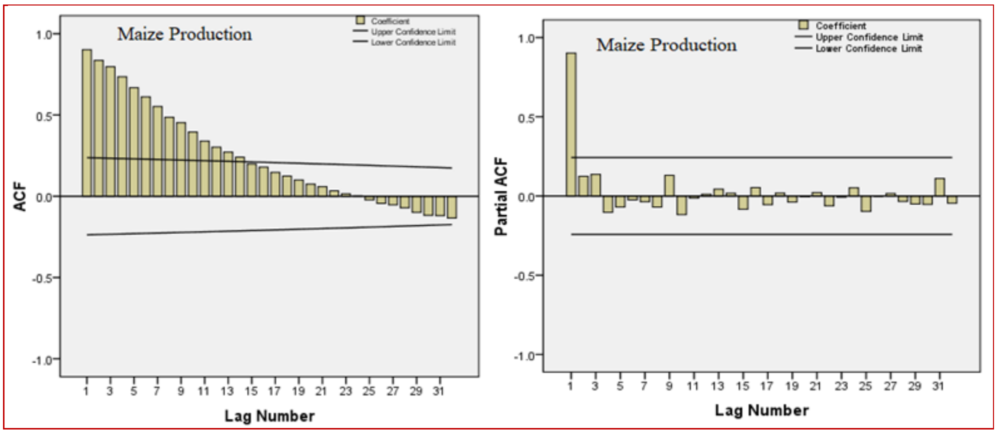

Figure 4 ACF and PACF of differenced data

Table - 3 BIC values of ARIMA(p,d,q)

ARIMA (p,d,q)	BIC Values
0,1,0	0.801
0,1,1	0.535
0,1,2	0.586
1,1,0	0.685
1,1,1	0.606
1,1,2	0.666
2,1,0	**0.526**

2,1,1	0.593
2,1,2	0.668
3,1,0	0.591
3,1,1	0.67
3,1,2	0.674

Model Estimation: Model parameters were estimated and reported in Table 4and Table 5. The model verification is concerned with checking the residuals of the model to improve on the chosen ARIMA (p,d,q). This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders, up to 32 lags were computed and the same along with their significance which is tested by Box-Ljung test are provided in Table 6. This proves that the selected ARIMA model is an appropriate model.

Table 4 - Estimated ARIMA Model of Maize Production

	Estimate	**SE**	**T**	**Sig.**
Constant	-26.424	6.954	-3.800	0.000
AR 1	-0.655	0.118	-5.563	0.000
AR 2	-0.487	0.121	-4.031	0.000

The ACF and PACF of the residuals are given in Figure 5. It also indicates 'good fit' of the model. So the fitted ARIMA model for the maize production data is

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

$$Y_t = -26.424 - 0.655Y_{t-1} - 0.487Y_{t-2} + \varepsilon_t$$

Table 5 - Estimated ARIMA Model Fit Statistics

ARIMA (p,d,q)	**Stationary R^2**	**R^2**	**RMSE**	**MAPE**	**MaxAPE**	**MAE**	**MaxAE**	**Normalized BIC**
0,1,0	0.043	0.955	1.402	11.282	50.905	0.987	4.212	0.801
0,1,1	0.322	0.968	1.189	10.975	46.063	0.881	2.866	0.535
0,1,2	0.34	0.969	1.182	10.75	47.546	0.85	2.925	0.586
1,1,0	0.212	0.963	1.282	10.697	37.372	0.919	3.754	0.685
1,1,1	0.327	0.327	0.327	0.327	0.327	0.327	0.327	0.327
1,1,2	0.339	0.969	1.193	10.964	49.221	0.874	3.077	0.666

2,1,0	**0.379**	**0.971**	**1.147**	**10.524**	**52.213**	**0.836**	**2.987**	**0.526**
2,1,1	0.386	0.971	1.15	10.626	54.012	0.834	3.09	0.593
2,1,2	0.388	0.971	1.157	10.55	54.6	0.828	3.123	0.668
3,1,0	0.387	0.971	1.149	10.632	54.401	0.833	3.112	0.591
3,1,1	0.387	0.971	1.158	10.638	54.422	0.834	3.113	0.67
3,1,2	0.432	0.973	1.125	9.629	52.716	0.764	3.015	0.674

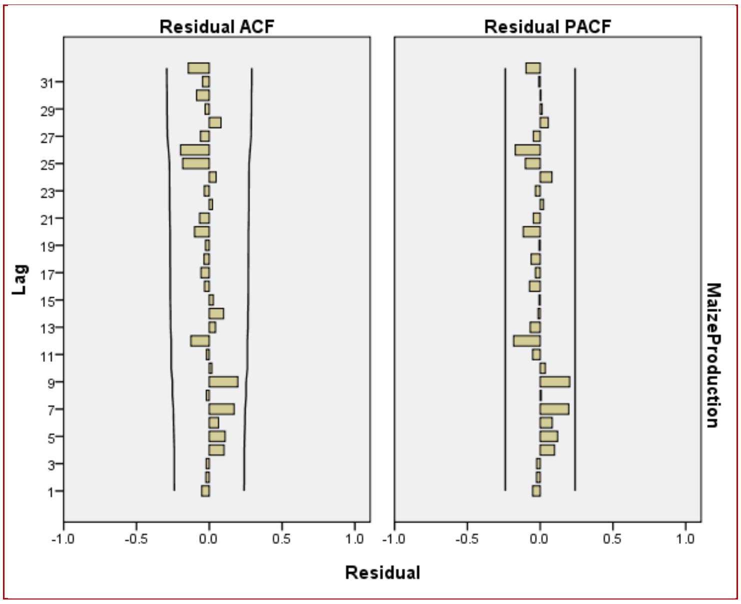

Figure 5 - Residuals of ACF and PACF

Forecasting: Forecasted value of maize production (quantity in million tons) for the year 2019 through 2025 respectively given by 27.02, 28.55, 30.21, 30.21, 31.28, 32.48 and 33.12 are given in Table 7. To assess the forecasting ability of the fitted ARIMA (p,d,q) model, important measures of the sample period forecasts' accuracy were computed. This measure indicates that the forecasting inaccuracy is low. Figure 6 shows that the actual and forecasted value of maize production data with 95% confidence limits.

Table 6 - Residual of ACF and PACF of MaizeProduction

Lag	ACF		PACF		*Lag*	ACF		PACF	
	Mean	**SE**	**Mean**	**SE**		**Mean**	**SE**	**Mean**	**SE**
1	-0.052	0.122	-0.052	0.122	***17***	-0.057	0.137	-0.032	0.122
2	-	0.122	-	0.122	***18***	-	0.137	-	0.122

	0.023		0.026			0.035		0.062	
3	-0.021	0.123	-0.024	0.122	***19***	-0.026	0.137	-0.008	0.122
4	0.101	0.123	0.099	0.122	***20***	-0.100	0.137	-0.116	0.122
5	0.108	0.124	0.119	0.122	***21***	-0.066	0.138	-0.047	0.122
6	0.064	0.125	0.083	0.122	***22***	0.022	0.139	0.023	0.122
7	0.170	0.126	0.196	0.122	***23***	-0.034	0.139	-0.033	0.122
8	-0.020	0.129	0.007	0.122	***24***	0.047	0.139	0.081	0.122
9	0.196	0.129	0.202	0.122	***25***	-0.182	0.139	-0.101	0.122
10	0.017	0.134	0.035	0.122	***26***	-0.197	0.143	-0.171	0.122
11	-0.019	0.134	-0.051	0.122	***27***	-0.060	0.147	-0.047	0.122
12	-0.126	0.134	-0.182	0.122	***28***	0.080	0.147	0.055	0.122
13	0.043	0.135	-0.068	0.122	***29***	-0.029	0.148	0.014	0.122
14	0.098	0.136	-0.015	0.122	***30***	-0.088	0.148	0.005	0.122
15	0.027	0.137	-0.007	0.122	***31***	-0.046	0.149	-0.010	0.122
16	-0.032	0.137	-0.073	0.122	***32***	-0.144	0.149	-0.098	0.122

Table 7 - Forecast of MaizeProduction

Year	Predicted	LCL	UCL
2019	27.02	24.73	29.31
2020	28.55	26.13	30.97
2021	30.21	27.70	32.72
2022	30.22	27.32	33.13
2023	31.28	28.21	34.36
2024	32.49	29.28	35.70
2025	33.12	29.69	36.54

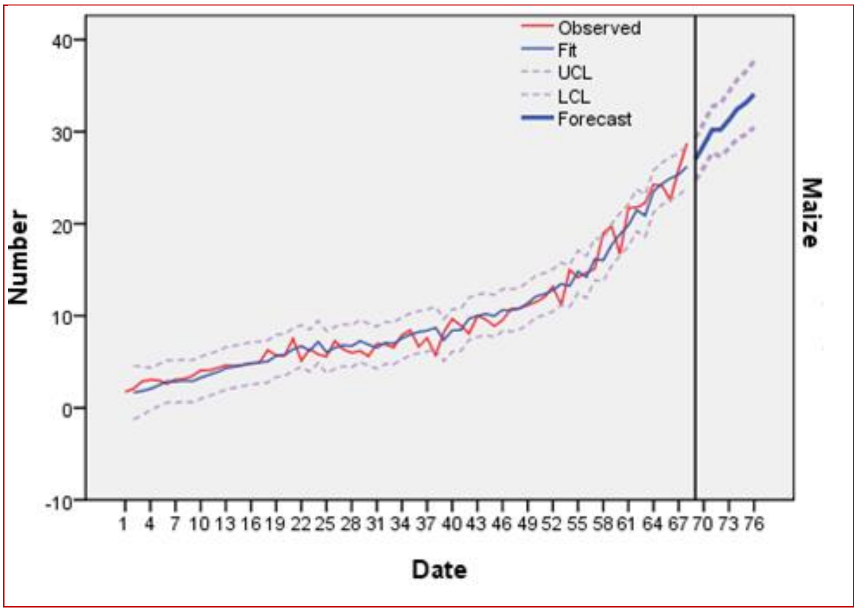

Figure 6 - Actual and Estimate of Maize Production

Conclusion

The most appropriate ARIMA model for *Zea mays* (Maize) production forecasting of data was found to be ARIMA (2,1,0). From the time series data, it can be found that forecasted production would increase to 33.12million tons in 2025 from 27.02 million tons in 2019 in India for using time series data from 1951 to 2018 on *Zea mays* (Maize) production. In India, during rabi and spring seasons to achieve higher yield at farmer's field assured irrigation facilities are required. this study provides an evidence on future *Zea mays* (Maize) production in the country, which can be considered for future policy making and formulating strategies for augmenting and sustaining *Zea mays* (Maize) production in India.

References

1. Agricultural Statistics (2020),Directorate of Economics and Statistics, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India.*https://eands.dacnet.nic.in/PDF/Pocket%202020-%20Final%20web%20file.pdf.*

2. Akaike, H. (1983). Statistical Predictor Identification. *Annals of Institute of Statistical Mathematics*, **22:** 203-270. https://link.springer.com/article/10.1007/BF02506337.

3. Alan Pankratz, (1983). Forecasting with Univariate Box-Jenkins Models: Concepts and Cases. John Wiley & Sons, New York. https://onlinelibrary.wiley.com/doi/book/ 10.1002/9780470316566.

4. BholaNath, Dhakre, D.S. and Debasis Bhattacharya, (2018). Forecasting wheat production in India: An ARIMA modelling approach. *Journal of Pharmacognosy and Phyto chemistry*, **8(1)**: 2158-2165. https://www.researchgate.net/publication/ 331471229.

5. Borkar, Prema and Tayade, P.M., (2016). Forecasting of Cotton Production in India Using ARIMA Model. *International Journal of Research in Economics and Social Sciences*, **6(5):** 1-7. https://euroasiapub.org/wp-content/uploads/2016/10/1ESSMay-3546-1.pdf.

6. Box, G.E.P. and Jenkins, G.M., (1976). Time Series Analysis: Forecasting and Control. San Francisco, Holden-Day, California, USA. https://link.springer.com/chapter/10.1057/9781137291264_6.

7. Box, G.E.P. and Pierce, D.A., (1970). Distribution of Residual Autocorrelations in ARIMA Models. *J. American Stat. Assoc.,***65**: 1509-1526. https://www.jstor.org/stable/2284333?seq=1.

8. DACNET, (2020). India maize scenario. *https://eands.dacnet.nic.in*.

9. 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. https://doi.org/10.20546/ijcmas.2018.702.019.

10. KourSatvinder, Pradhan, Ranjit Kumar Paul, U.K. and Vaishnav, P.R., (2017). Forecasting of Pearl Millet Productivity in Gujarat Under Time Series Framework. *Economic Affairs*, **62(1)**: 121-127. https://www.researchgate.net/publication/319122005.

11. Md. Moyazzem Hossain andFaruq Abdulla, (2016). Forecasting Potato Production in Bangladesh by ARIMA Model. *Journal of Advanced Statistics*, **1(4)**:191-198. https://www.academia.edu/30930688.

12. Slutzky, E., (1973). The summation of random causes as the source of cyclic processes. *Econometrica*, **5**: 105-146. https://www.scribd.com/document/388512380.