

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452

NAAS Rating (2026): 4.49

Maths 2026; 11(3): 130-135

© 2026 Stats & Maths

<https://www.mathsjournal.com>

Received: 20-12-2025

Accepted: 18-02-2026

Anju A Bhagat

Assistant Professor, Division of
Economics, College of
Agriculture, Pune, MPKV,
Rahuri, Maharashtra, India

Kaveri Y Khairnar

Junior Research Assistant,
Department of Agricultural
Economics, MPKV, Rahuri,
Maharashtra, India

VD Pimparkar

Junior Research Assistant,
College of Agriculture, Pune,
MPKV, Rahuri, India,
Maharashtra, India

YC Sale

Associate Professor, Department
of Agriculture Economics,
Division of Economics, College of
Agriculture, Pune, MPKV,
Rahuri, Maharashtra, India

Rashmi Bhoge

Assistant Professor, Department
of Entomology, Seed Technology
Research Unit, Seed Cell,
MPKV, Rahuri, Maharashtra,
India

Corresponding Author:

Anju A Bhagat

Assistant Professor, Division of
Economics, College of
Agriculture, Pune, MPKV,
Rahuri, Maharashtra, India

Statistical modeling on production of papaya in Maharashtra state

Anju A Bhagat, Kaveri Y Khairnar, VD Pimparkar, YC Sale and Rashmi Bhoge

DOI: <https://www.doi.org/10.22271/math.2026.v11.i3b.2288>

Abstract

The main objective of the study is to identify the statistical model for production of papaya, to compare the efficiency of different statistical models and to forecast the papaya production of Maharashtra state by using best fitted model. The results revealed that, mean area, production and productivity of papaya were 4.19 thousand hectares, 130.15 thousand tons and 36.72 MT/ha respectively in Maharashtra. The results of skewness and kurtosis of area, production and productivity of papaya were positively skewed for area and production while negatively skewed for productivity, and were not normally distributed. The production of papaya has shown an increasing trend of 2.71% per annum during the study period. ARIMA model proved to be an improvement over the conventional models and are more appropriate to visual the temporal trend of production of papaya in Maharashtra state.

Keywords: Papaya, production, trend, non-linear models, ARIMA, forecasting

Introduction

India is known as the basket of fruits and vegetables in the world. Its diverse climate ensures availability of all varieties of fresh fruits & vegetables round the year. India produced 99069.2 thousand MT of fruits and 191769.11 thousand MT of vegetables (source: NHB database 2019-20). Among the fruits, papaya occupies a special place. Like banana it is available throughout the year and it is easy to cultivate (Biswas, 2010) [3]. Papaya is considered as one of the most important fruit crops in terms of production, productivity and export potential in India. In India, around 24 states are growing papaya, the major producing states in India are Andhra Pradesh, Maharashtra, Chhattisgarh, West Bengal, Assam, Tamil Nadu, Jharkhand. Hadolika *et al.* (2022) [4] studied status of papaya cultivation in India and showed that the area, production and yield of Papaya at all India level shown significant positive growth. The instability in production of Papaya was more compared to area and yield. Sagar *et al.* (2012) [6] studied the costs, returns and constraints in production and marketing of papaya (*Carica papaya* L.) in Middle Gujarat. Results of the study indicated that papaya was highly capital intensive crop and average cost of cultivation (Cost C₂) per hectare was Rs. 176660. On an average cost A (paid out cost) formed 61.64%, while cost B accounted for 87.38% of total cost.

The main objective of the study is to identify the statistical model for production of papaya, to compare the efficiency of different statistical models and to forecast the papaya production of Maharashtra state by using best fitted model.

Materials and Methods

The secondary data on area, production and productivity of papaya of Maharashtra state (1987-88 to 2022-23) were used for descriptive statistics and production time series data is considered for finding best fit statistical model.

The Agricultural production may vary from year to year due to nature of cultivation, environment changes and government policies.

Sometimes production of crop can increase or decrease with a certain rate. To investigate such trend, we consider eight following different statistical models in the present research study i.e.

1. Linear

$$y_i = a + bt + \varepsilon_i \quad (1)$$

2. Polynomial 2nd degree

$$y_i = a + bt + ct^2 + \varepsilon_i \quad (2)$$

3. Polynomial 3rd degree

$$y_i = a + bt + ct^2 + dt^3 + \varepsilon_i \quad (3)$$

4. Compound

$$y_i = ab^t + \varepsilon_i \quad (4)$$

Where, y_i is the production of papaya, t is the time and ε_i is the random error. Taking logarithm on both sides of eq. (3), we get

$$\log y_i = \log a + t \log b + \varepsilon_i$$

5. Exponential

$$y_i = ae^{bt} + \varepsilon_i \quad (5)$$

6. Logarithmic

$$y_i = a + b \ln(t) + \varepsilon_i \quad (6)$$

7. Power

$$y_i = at^b + \varepsilon_i \quad (7)$$

Durbin Watson test

$$DW(d) = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

Where $e_t = (y_t - \hat{y}_t)$, y_t and $\hat{y}_t$ are observed and estimated values of response variable for individual t . $DW(d)$ becomes smaller as the serial correlations increase.

To test the autocorrelation at significance level α , the test statistic d is compared to lower and upper critical values ($d_{L,\alpha}$ and $d_{U,\alpha}$):

- If $d < d_{L,\alpha}$, there is statistical evidence that the error terms are positively autocorrelated.
- If $d > d_{U,\alpha}$, there is no statistical evidence that the error terms are positively autocorrelated.
- If $d_{L,\alpha} < d < d_{U,\alpha}$, the test is inconclusive.

8. ARIMA (Autoregressive Integrated Moving Average) Model

Autoregressive Integrated Moving Average (ARIMA) also known as Box-Jenkins methodology was used for the time series modeling. It includes AR and MA terms with their respective order p and q . Differencing eliminates trends and seasonality and stabilizes mean of the time series. The differencing was used in order to make the series stationary. The non seasonal ARIMA (p, d, q) model for papaya

production (y) in period t based on data up to period $t-1$ was used as follows

$$y_t = \mu + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

Where, μ is the constant, φ_p is the autoregressive (AR) coefficient at lag p , θ_q is the moving average (MA) coefficient at lag q and $e_{t-q} = y_{t-q} - \hat{y}_{t-q}$ is the forecast error that was made at period $t-p$.

The main steps of ARIMA modeling are model identification, parameter estimation, diagnostic checking and forecasting.

Results and Discussion

The descriptive statistics of area, production and productivity of papaya in Maharashtra state is presented in Table 1. The results revealed that, mean area, production and productivity of papaya were 4.19 thousand hectares, 130.15 thousand tons and 36.72 MT/ha respectively. The median for area, production and productivity was 1.90 thousand ha., 60.10 thousand MT and 36.30 MT/ha respectively. The maximum area under cultivation was 16.18 thousand hectares; production was 552.50 thousand MT and productivity was 85.00 MT/ha. Whereas, minimum area under cultivation was 0.58 thousand ha with production 7.80 thousand MT and minimum productivity was 6.20 MT/ha. The results of skewness and kurtosis of area, production and productivity of papaya data positively skewed for area and production while negatively skewed for productivity and was not normally distributed.

Table 1: Descriptive Statistics of area, production and productivity of papaya in Maharashtra State.

Particular	Area (In ' 000 ha)	Production (In ' 000 Mt)	Productivity (In Mt/ha)
Mean	4.19	130.15	36.72
Standard Error	0.53	18.08	2.70
Median	1.90	60.10	36.30
Mode	2.00	12.40	61.90
Standard Deviation	4.52	154.47	23.08
Kurtosis	0.52	0.98	-1.41
Skewness	1.31	1.51	-0.02
Range	15.60	544.70	78.80
Minimum	0.58	7.80	6.20
Maximum	16.18	552.50	85.00

Table 2: Compound growth rate of papaya production of Maharashtra (1950-51 to 2022-23)

Particular	Compound growth rate	S.E.
Papaya production of Maharashtra ('000MT)	2.71**	0.47

The results of compound growth rate of papaya production of Maharashtra state is presented in Table 2. The production has increased significantly at a compound growth rate of 2.71% per annum with standard error 0.47. The similar results was reported by Bhagat *et al* (2024) [2] concluded that the export performance of Indian milk and milk products quantity and its value exhibited positive annual growth rate of 17.20 and 24.46% during study period.

For analyzing the pattern of papaya production various linear and non-linear models were fitted. The goodness of fit explains how well the given standard model fits a dataset. Measures of goodness of fit typically summarize the discrepancy between observed and predictive values under the model. In this research the efficiency of different statistical

models were checked by R^2 and Root Mean Square Error (RMSE) was determined by Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Akaike Information Criterion (AIC). Uzma *et al* (2023) [8] compared the performance of five nonlinear growth models, namely, Monomolecular, Logistics, Gompertz, Richards, and Weibull and studied the area and production of apple crop in Kashmir. The best model was chosen based on

various model adequacy test criteria including MAE, RMSE, MSE, and R^2 . Therefore, Richards and Weibull models were found to be the best fit for the area and production of apple, respectively. The area and production of the apple crop were forecasted for a few years using the best fit models. The estimated parameters along with their Durbin Watson Statistic (DW), R^2 , different errors and AIC values are presented in Table 3.

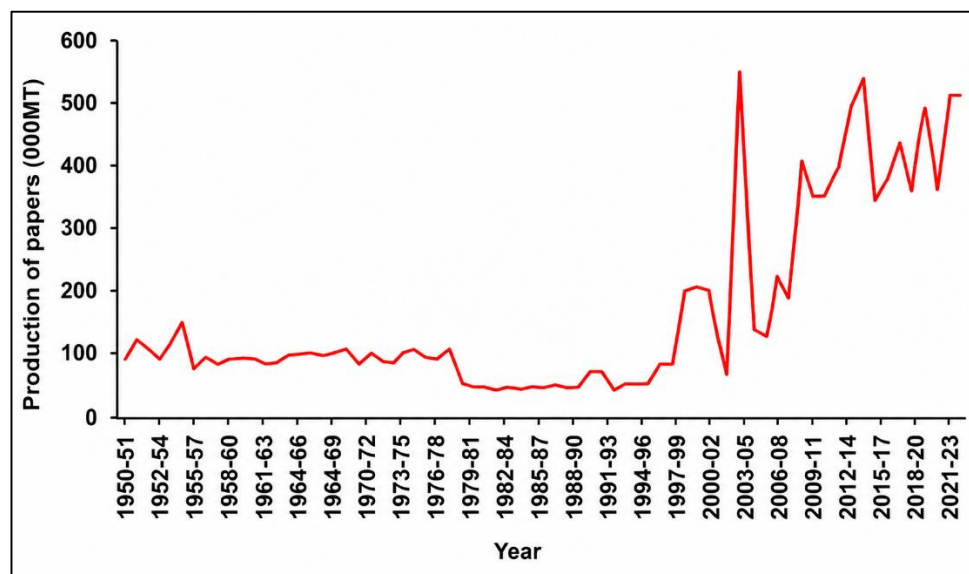


Fig 1: Time series plot of papaya production of Maharashtra State (1950-51 to 2022-23)

Table 3: DW statistic and estimates of parameters of papaya production of Maharashtra.

Model	DW Statistic	Co-efficient	Estimate	R^2	RMSE	MSE	MAE	AIC
Linear	0.80	a	-51.20	0.45	115.02	13230.55	92.63	304.88
		b	4.90**					
Polynomial 2 nd degree	0.41	a	146.00**	0.76	76.58	5865.21	52.92	279.08
		b	-10.87**					
		c	0.27**					
Polynomial 3 rd degree	1.79	a	90.24*	0.77	74.01	5477.79	46.20	276.92
		b	-2.13					
		c	-0.08					
		d	0.003*					
Compound	0.41	a	24.54**	0.45	115.02	13230.55	92.63	304.88
		b	1.03**					
Exponential	0.41	a	-24.53	0.22	927.99	861160.55	803.34	437.26
		b	0.03					
Logarithmic	0.54	a	-119.39	0.19	139.52	19466.11	109.85	317.12
		b	74.88					
Power	0.34	a	23.91	0.05	160.50	25761.62	95.87	326.00
		b	0.03					

In order to test the autocorrelation among the residuals of different fitted models Durbin Watson test was used. The test statistic is presented in Table 3. After fitting models, residuals were estimated for all models. After applying Durbin Watson test to the residuals of different models, it has been observed that in models that are linear, polynomial 2nd and 3rd degree, compound, exponential, logarithmic and power residuals were positively auto correlated.

The results of ADF test are presented in Table 4. It indicated that the papaya production are non-stationary at level and stationary after first difference and after identifying the values of p and q from the ACF and PACF graph. The time series data of papaya production with 73 data points was divided into two sets. 73 data points out of 58 were used as the training set while 15 years data points were used as test set. The results of ARIMA are presented in Table 5, 6, 7 and 8

along with the R^2 , different errors and AIC value. As perusal into the goodness of fit values for papaya production indicated that among the different linear and non linear models fitted, the lowest RMSE (70.35) and AIC (270.71) were observed in ARIMA (2,1,6) model as compared to all other models. Fig. 3 shows the graphical representation of papaya production using best fitted ARIMA model. Sharma *et al* (2023) [7] stated that most appropriate ARIMA model was selected to capture the forecasting trends in the area and production of papaya for the period from 2021 to 2030 and projected the values for area under cultivation and production of papaya and reached 1538 hectares and 6,384,220 metric tons, respectively, by 2030. This study found that significance in defining the potential demand gap between the area and production for papaya in the future.

Table 4: Result of Augmented Dickey Fuller test of papaya production.

Data	Difference	Lag order	Dickey -Fuller	P value
Papaya production('000MT)	d = 0	4	-0.69	0.97
Papaya production (Diff)	d = 1	4	-4.14	0.01

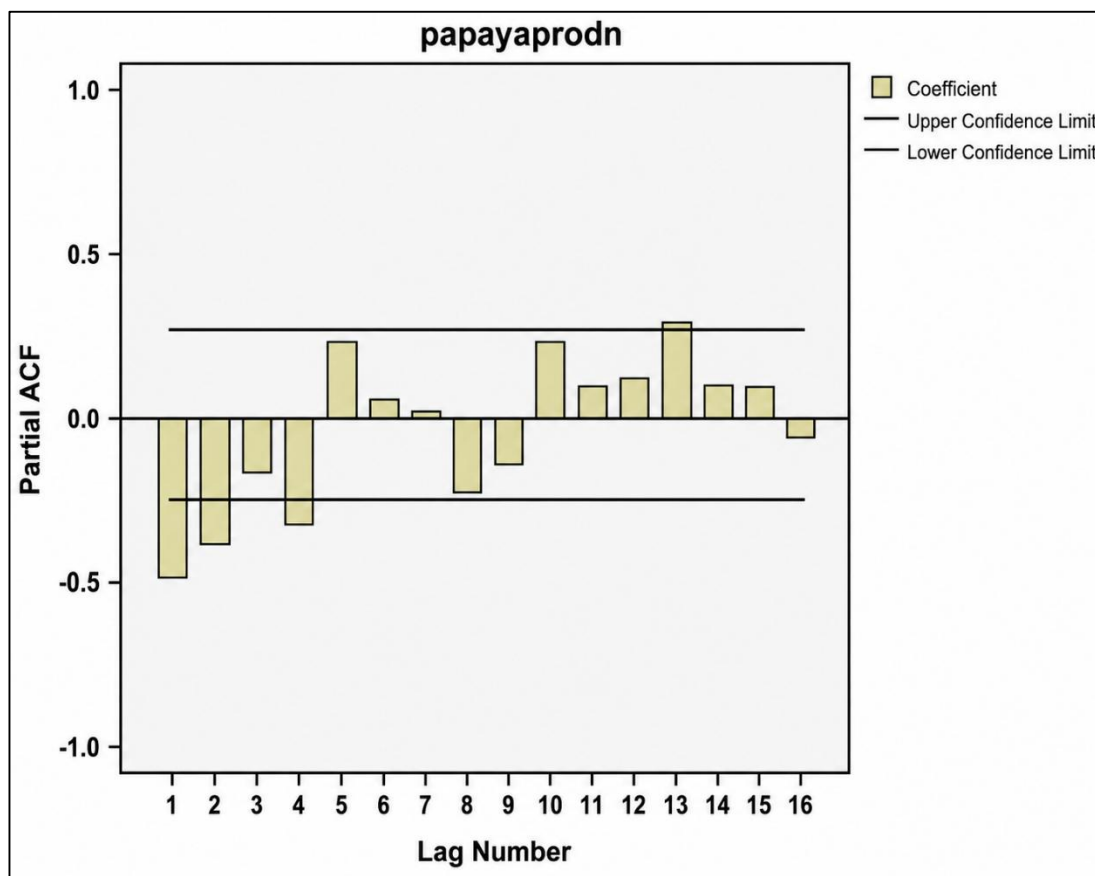
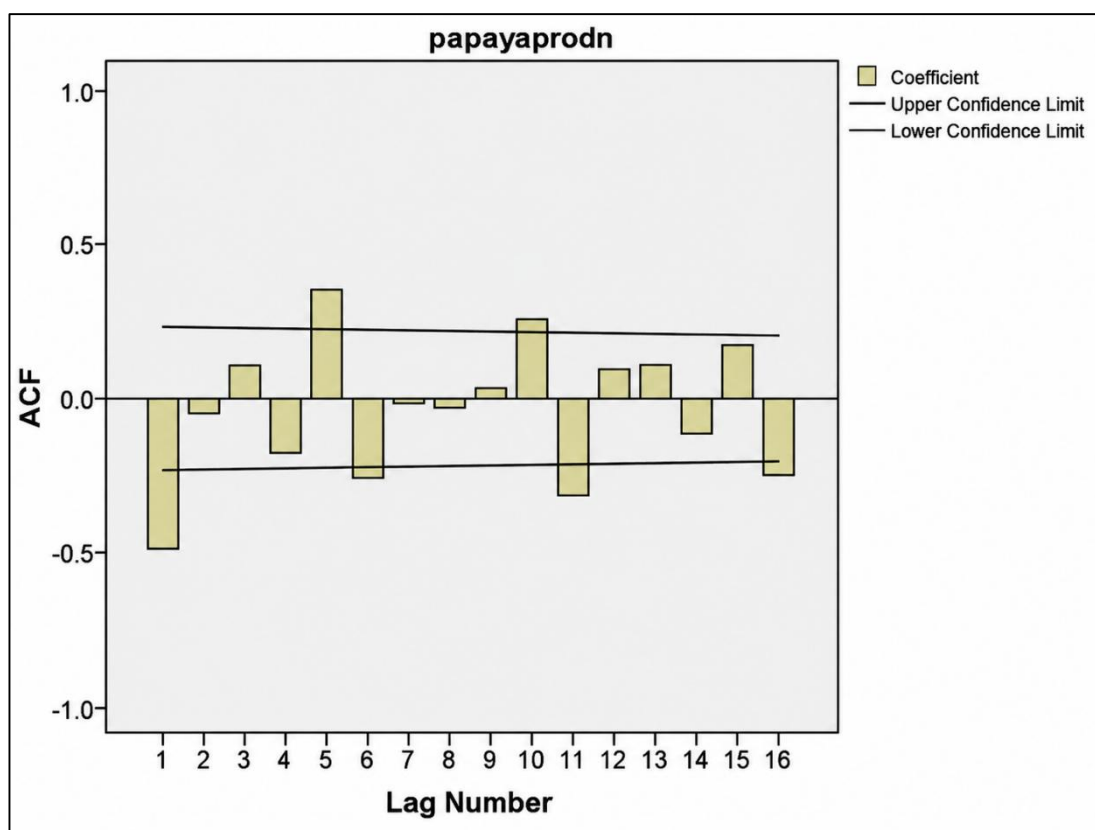
**Fig 2:** ACF and PACF of papaya production of Maharashtra State

Table 5: Results of ARIMA modeling.

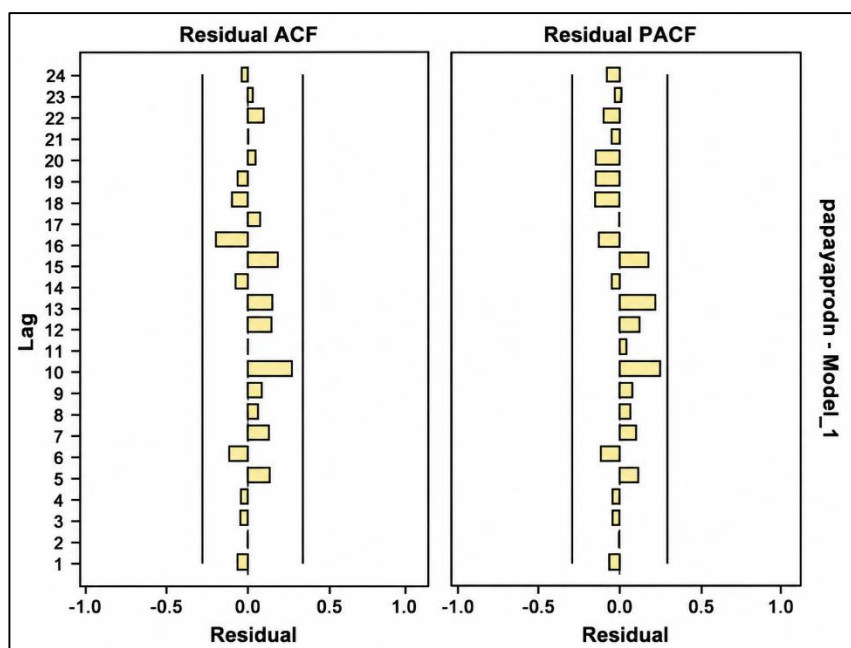
ARIMA model	R ²	RMSE	MAPE	Normalized BIC	Ljung Box Q(18)
ARIMA(1,1,1)	0.73	82.33	92.19	9.45	Sign.
ARIMA(1,1,5)	0.75	80.57	80.27	9.19	Sign.
ARIMA(1,1,6)	0.77	79.16	88.64	9.21	N.S.
ARIMA(1,1,10)	0.81	73.67	73.10	9.31	N.S.
ARIMA(1,1,11)	0.82	73.09	71.33	9.36	N.S.
ARIMA(1,1,16)	0.84	71.90	77.26	9.62	Sign.
ARIMA(2,1,1)	0.73	82.02	89.84	9.05	Sign.
ARIMA(2,1,5)	0.77	79.31	79.74	9.22	N.S.
ARIMA(2,1,6)	0.87	70.35	67.73	9.04	N.S.
ARIMA(2,1,10)	0.80	76.90	73.86	9.46	Sign.
ARIMA(2,1,11)	0.82	73.40	72.47	9.42	N.S.
ARIMA (2,1,16)	0.84	72.39	77.66	9.69	N.S.
ARIMA(4,1,1)	0.70	88.05	59.47	9.31	N.S.
ARIMA(4,1,5)	0.68	94.30	54.52	9.69	N.S.
ARIMA(4,1,6)	0.78	78.00	82.72	9.37	N.S.
ARIMA(4,1,10)	0.82	73.62	69.65	9.49	Sign.
ARIMA(4,1,11)	0.82	74.33	73.03	9.57	Sign.
ARIMA (4,1,16)	0.84	73.37	76.32	9.84	N.S.
ARIMA(13,1,1)	0.82	73.29	68.19	9.48	N.S.
ARIMA(13,1,5)	0.85	70.55	72.76	9.64	Sign.
ARIMA(13,1,6)	0.85	70.83	72.66	9.71	Sign.
ARIMA(13,1,10)	0.85	72.96	71.71	10.00	Sign.
ARIMA(13,1,11)	0.86	72.29	71.57	10.04	Sign.
ARIMA (13,1,16)	0.85	78.90	73.84	10.52	Sign.

Table 6: Results of ARIMA model parameter for papaya of Maharashtra state.

Particulars	Estimate	S.E.	t	Sign.
Constant	5.638	3.574	1.577	0.120
AR				
Lag1	0.176	0.167	1.051	0.297
Lag2	-0.737	0.170	-1.322	0.195
Difference	1			
MA				
Lag1	0.887	0.182	0.872	0.560
Lag2	-0.895	0.232	-1.865	0.278
Lag3	0.261	0.216	1.205	0.233
Lag4	0.223	0.199	1.121	0.266
Lag5	-0.470	0.166	-1.839	0.623
Lag6	0.400	0.131	1.062	0.113

Table 7: Results of different errors of best fit ARIMA modeling.

ARIMA model	R ²	RMSE	MAE	MSE	AIC	BIC	Ljung Box Q(18)
ARIMA(2,1,6)	0.87	70.35	42.67	4504.05	270.71	9.04	N.S.

**Fig 3:** Residual plot ACF and PACF of papaya production of Maharashtra State

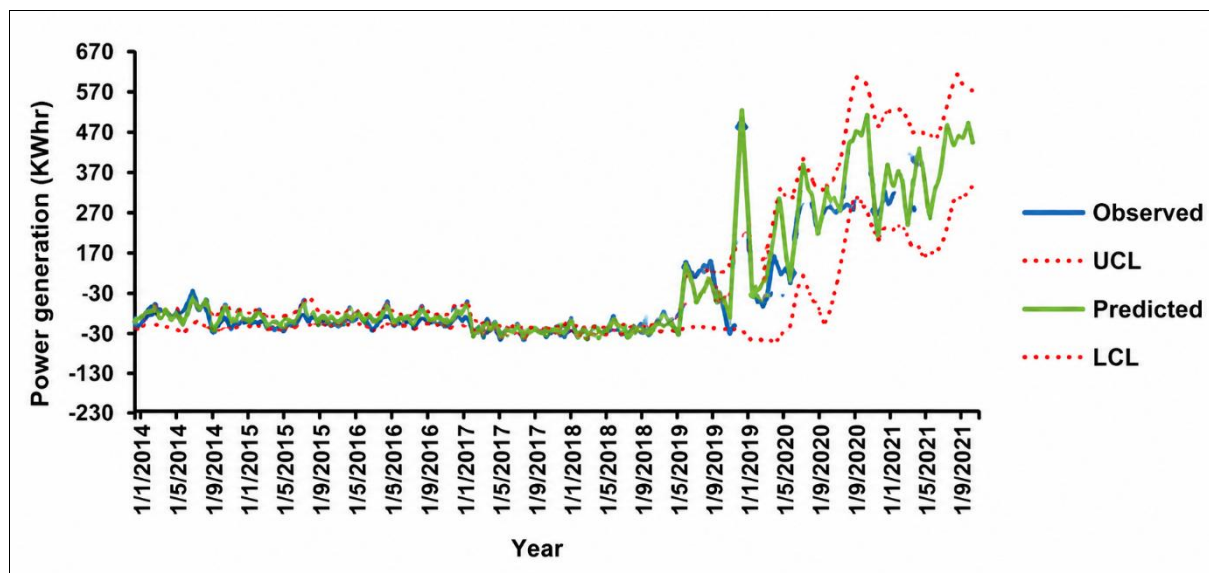


Fig 4: Observed and best fitted predicted ARIMA (2, 1, 6) model for papaya production of Maharashtra State (1950-51 to 2022-23)

Thus, based on the goodness of fit criteria the ARIMA (2, 1, 6) model can be regarded as best fit for describing the trend in papaya production during 1950-51 to 2022-23. Hence, the ARIMA (2, 1, 6) model can be used for the prediction of papaya production in the Maharashtra state for the next year (fig 4). It was observed that there will be increase in papaya production of Maharashtra in next year and it will be 504.99 thousand metric tonnes and it will lie between LCL 345.32 to 664.60 UCL thousand metric tonnes. The similar results were reported by Murthy *et al* (2020) [5] and Balaga and Dash (2022) [1].

Table 8: Results of Validation and predicted value of papaya production for Maharashtra State.

Year	Observed	LCL	Predicted	UCL
2008-09	150.89	-24.8	194.49	382
2009-10	388.75	148.05	439.31	454.85
2010-11	319.00	66.37	275.02	373.17
2011-12	327.00	2.53	363.93	376.33
2012-13	363.00	137.81	347.21	444.61
2013-14	500.50	299.76	530.66	606.56
2014-15	542.00	341.17	536.07	647.97
2015-16	316.83	258.57	286.79	565.36
2016-17	344.12	241.25	421.93	548.04
2017-18	408.30	272.19	409.77	578.98
2018-19	330.89	232.74	308.73	539.54
2019-20	459.97	196.65	459.13	503.45
2020-21	327.24	200.22	320.88	507.01
2021-22	496.50	218.61	513.27	525.4
2022-23	500.06	341.14	498.10	647.94
Forecasted (2023-24)				
2023-24		345.32	504.99	664.60

Conclusion

The estimated compound annual growth rates revealed that the production of papaya has shown an increasing trend of 2.71% per annum during the study period. ARIMA model proved to be an improvement over the conventional models and are more appropriate to visualize the temporal trend of production of papaya in Maharashtra state. It is concluded from the study that on comparing different linear and non linear statistical models, the ARIMA (2, 1, 6) model performed better for describing papaya production in Maharashtra state. The predicted value of papaya production for the next year (2023-24) by using best fit model

ARIMA(2,1,6) will be 504.99 thousand metric tonnes and it will lie between Lower Control Limit (LCL) 345.32 and Upper Control Limit (UCL) 664.60 thousand metric tonnes.

References

- Balaga D, Das A. Using ARIMA model to forecast the area, yield and production of arhar in Odisha. *Biol Forum Int J*. 2022;14(3):1179-1185.
- Bhagat AA, Lad DB, Mahajan PJ, Gondhali BV. Stability analysis of milk and milk product export from India. *Int J Stat Appl Math*. 2024;9(6):220-227.
- Biswas BC. Success stories of papaya farmers. *Fertil Mark News*. 2010;41(2):15-18.
- Hadolikar SB, More SS, Jadhav VG. Export performances of fresh papaya in India: Markov chain approach. *Pharma Innov J*. 2021;SP-10(11):12-14.
- Murthy BR, Naidu M, Tamilselvi C, Evangilin NP. Validation of ARIMA model on production of papaya in India. *Indian J Pure Appl Biosci*. 2020;8(2):64-68.
- Sagar SB, Parmar HC, Darji VB. Economics of production of papaya in middle Gujarat region of Gujarat, India. *Glob J Biol Agric Health Sci*. 2012;1(2):10-17.
- Sharma S, Gupta R, Vishal R. Forecasting the future of papaya in India: predicting area and production through autoregressive integrated moving average. *J Appl Fruit Sci*. 2023;66(2):1-9.
- Uzma M, Nazir N, Mir SA, Shah IA, Bhat IA. Nonlinear statistical modelling of apple crop of Kashmir Valley. *J Indian Soc Agric Stat*. 2023;77(3):249-255.