

Original Research Article

<https://doi.org/10.20546/ijcmas.2020.907.135>

Statistical Model for Forecasting Area, Production and Productivity of Sesame Crop (*Sesamum indicum* L.) in Andhra Pradesh, India

N. Priyanka Evangilin^{1*}, B. Ramana Murthy¹, G. Mohan Naidu¹ and B. Aparna²

Department of Statistics and Computer Applications, ⁴Department of Agricultural Economics
Acharya N.G. Ranga Agricultural University, S.V. Agricultural College, Tirupati, India

*Corresponding author

ABSTRACT

This research study was carried out to fit different Linear, Non – Linear and time series ARIMA models on Area, Production and Productivity of Sesame (*Sesamum indicum* L.) in Andhra Pradesh for the period 1965-66 to 2017-18. The statistically best fitted model was chosen on the basis of goodness of fit criteria viz. R^2 , Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Among all the models ARIMA (3, 0, 0), ARIMA (3, 0, 3) and ARIMA (3, 1, 1) models, were found to be the best fitted models and these models are used to forecast area, production and productivity of Sesame crop in Andhra Pradesh for further five years. The forecasted results showed for area, production and productivity of Sesame crop for the year 2020-21 to be 51.66 thousand hectare, 17.64 thousand tonnes and 323.43 in kg/hectare respectively. And also it is showed, there is a fluctuated trend on Area and Production and increasing trend on Productivity from the period 2018-19 to 2022-23.

Keywords

Sesame crop, Area, Production, Productivity, Forecast and ARIMA Model

Article Info

Accepted:

11 June 2020

Available Online:

10 July 2020

Introduction

Sesame (*Sesamum indicum* L.) is the oldest indigenous oilseed crop, with longest history of cultivation in India. Sesame oil is an edible vegetable oil derived from sesame seeds. Sesame or Gingelly is commonly known as til (Hindi, Punjabi, Assamese, Bengali, Marathi), tal (Gujarati), nuvvulu, manchinuvvulu (Telugu), ellu (Tamil, Malayalam, Kannada), tila/pitratarpa (Sanskrit) and rasi (Odia) in

different parts of India. Indian people revere sesame and both the oil and seeds are used in traditional cooking methods, religious rituals, Ayurvedic medicine, and topically for skin nourishment. India ranks first in world with 19.47 Lakh ha area and 8.66 Lakh tonnes production. The average yield of sesame (413 kg/ha) in India is low as compared with other countries in the world (535 kg / ha). The main reasons for low productivity of sesame are its rainfed cultivation in marginal and

submarginal lands under poor management and input starved conditions. However, improved varieties and agro production technologies capable of increasing the productivity levels of sesame are now developed for different agro ecological situations in the country. A well-managed crop of sesame can yield 1200 - 1500 kg/ha under irrigated and 800 - 1000 kg/ha under rainfed conditions. The crop is grown in almost all parts of the country. More than 85% production of sesame comes from West Bengal, Madhya Pradesh, Rajasthan, Uttar Pradesh, Gujarat, Andhra Pradesh and Telangana.

Narayanaswamy *et al.* (2012) were fitted statistical models for growth pattern of root and shoot morphological traits in sesame. They have concluded that, the shoot morphological traits growth like mean inter nodal length and shoot dry weight was best explained by the quadratic function, plant height and shoot fresh weight was best explained by the linear function. In root morphological traits growth like root length was best explained by the linear functional form, while root volume, root fresh weight and root dry weight was best explained by the quadratic functional form.

Ramana Murthy *et al.*, (2018) studied the trends of area, production and productivity of Mango crop in Andhra Pradesh from 1992-93 to 2016-17 based on linear and Non-linear statistical models. The results reveal that there is decreasing trend on area and production and gradually increasing trend on productivity of Mango crop in Andhra Pradesh state in above study period.

Sudha *et al.*, (2013) studied to measure the growth trends of area, production and productivity of maize between 1970-71 to 2008-09 and to estimate the future projections up to 2015 AD by using the growth functions like linear, logarithmic, inverse, quadratic,

cubic, compound, power and exponential. Based on highest coefficient of determination (R^2) and its adjusted R^2 . They concluded that among all the models cubic function was found to be best fitted model for future projections of maize area, production and productivity.

Ramana Murthy *et al.*, (2018) was made an attempt to develop an appropriate ARIMA model for forecasting groundnut area, production and productivity of India. They have concluded that ARIMA (2, 1, 3), ARIMA (3, 0, 3) and ARIMA (2, 1, 3) models were best fitted to forecast area, production and productivity of groundnut in India. For four leading years, they have found that there was a decreasing trend on area and fluctuations on production and productivity from the period 2016-17 to 2019-2020.

Prabakaran *et al.*, (2014) analyzed the Pulses Area and Production in India during the period from 1950-51 to 2011-12 by using ARIMA model and he found that ARIMA (1, 1, 0) and ARIMA (2, 1, 1) models were best fitted to forecast Pulses Area and Production in India.

The objective of the present study was to fit different linear, non-linear and appropriate Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) models on area, production and productivity of Sesame crop and to forecast next five years future values based on selected model.

Materials and Methods

The data of study for a period of 53years (1965-66 to 2017-18) in Andhra Pradesh pertaining to Area ('000 Hectare), Production ('000 tonnes) and Productivity (in Kg/Hectare) of Sesame crop were collected from the source of EPWRF (Economic and Political Weekly Research Foundation) India Time series, Directorate of Economics and

Statistical and Ministry of Agriculture, Govt. of India in www.indiastat.com. In order to examine the nature of change and degree of relationship in area, production and productivity of Sesame crop in Andhra Pradesh by various linear, non-linear and ARIMA statistical models were worked out by using SPSS 22 version.

Linear and non-linear growth models

The linear and non-linear growth models for the crop characteristic i.e., Area, Production and Productivity of Sesame crop in Andhra Pradesh are estimated by fitting the following functions.

Auto Regressive Integrated Moving Average (ARIMA)

The ARIMA methodology is also called as Box-Jenkins methodology (Box and Jenkins 1976). The Box-Jenkins procedure is concerned with fitting a mixed ARIMA model to a given set of data. The main objective in fitting ARIMA model is to identify the stochastic process of the time series and predict the future values accurately. This method have also been useful in many types of situations which involve the building of models for discrete time series and dynamic systems. However the optimal forecast of future values of a time series are determined by the stochastic model for that series. A stochastic process is either stationary or non-stationary. The first thing to note is that most time series are non-stationary and the ARIMA models refer only to a stationary time series. Since the ARIMA models refer only to a stationary time series the first stage of Box-Jenkins model is for reducing non-stationary series to a stationary series by taking the differences.

The ARIMA (p, d, q) process is given by

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} - \dots - \theta_q \varepsilon_{t-q} \tag{1}$$

Where y_t and ε_t are the actual value and random error at time period t , respectively. ϕ_i ($i = 1, 2, 3, \dots, p$) and θ_j ($j = 1, 2, 3, \dots, q$) are model parameters. p and q are integers and often referred to as orders of the model. Random errors ε_t are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2 .

The main stages in setting up a Box-Jenkins forecasting model are as follows:

1. Identification
2. Estimating the parameters
3. Diagnostic checking
4. Forecasting

Results and Discussion

In the present study, the data for Area, production and Productivity of Sesame crop in Andhra Pradesh for the period of 53 years (1965-66 to 2017-18) were used for the study.

Model identification

Among several models Linear, Non-linear and ARIMA (p, d, q) studies the goodness of fitted models were examined by highest R^2 value, lowest RMSE (Residual Mean Square Error) and lowest MAPE (Mean Absolute percentage Error) values. Based on these criterions, it was found that ARIMA (3, 0, 0), ARIMA (3, 0, 3) and ARIMA (3, 1, 1) are the best fitted models for forecasting Sesame crop area, production and productivity respectively. The Coefficient of determination (R^2), Mean Absolute Percentage Error (MAPE) and Residual Mean Square Error (RMSE) are given by

$$R^2 = 1 - \left[\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \right] \tag{2}$$

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (4)$$

Where y_t is the actual observation for time period 't' and \hat{y}_t is the predicted value for the same period and \bar{y} is the overall sample mean of observations. The models and the corresponding values are shown in table (1), table (2) and table (3).

Model estimation and verification

The parameters of the model were estimated by using SPSS 22 package. The

ARIMA (3, 0, 0), ARIMA (3, 0, 3) and ARIMA (3, 1, 1) were found to be best fitted models for area, production and productivity of Sesame. The model verification (or) diagnosed by the Ljung-Box Q statistic. The Ljung-Box Q statistic is to check the overall adequacy of the model. The test statistic Q is given by

$$Q_n = nr(nr + 2) \sum_{l=1}^n \frac{r_l^2(e)}{nr - l} \quad (5)$$

Where $r_l(e)$ is the residual autocorrelation at lag l , nr is the number of residual, n is the number of time lags included in the test for model to be adequate, p -value associated with Q statistics should be large ($p\text{-value} > \alpha$). The results of estimation are reported in Table 4.

Parametric Trend models

Model	Functional form
Linear function	$y_t = a + bt$
Logarithmic function	$y_t = a + b \ln(t)$
Inverse function	$y_t = a + b/t$
Quadratic function	$y_t = a + bt + ct^2$
Cubic function	$y_t = a + bt + ct^2 + dt^3$
Compound function	$y_t = ab^t$
Power function	$y_t = at^b$ (or) $\ln(y_t) = \ln(a) + b \ln(t)$
S- Curve function	$y_t = \text{Exp}(a + b/t)$ (or) $\ln(y_t) = a + b/t$
Growth function	$y_t = \text{Exp}(a + bt)$ (or) $\ln(y_t) = a + bt$
Exponential function	$y_t = ae^{bt}$ (or) $\ln(y_t) = \ln(a) + bt$

Table.1 Linear, Non-linear and Time series models for Area of Sesame crop in Andhra Pradesh

Model	Parameter				Criteria		
	a	b	c	d	R ²	RMSE	MAPE
Linear	240.658**	-3.113**			.675**	33.0220078	21.06546
Logarithmic	302.736**	-48.301**			.544**	39.1418826	29.85071
Inverse	141.570**	175.040**			.213**	51.4177387	41.863751
Quadratic	221.316**	-1.002	-.039		.695**	31.9943513	18.921215
Cubic	269.986**	-11.338**	.435**	-.006**	.775**	28.3455185	17.90037
Compound	276.277**	.976**			.629**	37.028656	22.675628
Power	408.354**	-.347**			.425**	48.6003197	31.016042
S-Curve	4.862**	1.156**			.140**	57.9152436	41.320534
Growth	5.621**	-.024**			.629**	37.028656	22.675628
Exponential	276.277**	-.024**			.629**	37.028656	22.675628
Time Series Models							
ARIMA(1,0,3)					0.790	28.203	17.533
ARIMA(1,1,1)					0.763	29.277	17.669
ARIMA(1,2,1)					0.659	34.285	19.790
ARIMA(3,0,0)					0.793	27.711	17.487
ARIMA(2,0,3)					0.780	29.194	18.712
ARIMA(2,2,2)					0.711	32.260	18.758
ARIMA(3,2,2)					0.722	31.984	17.834
ARIMA(3,1,3)					0.779	29.491	17.914

Table.2 Linear, Non-linear and Time series models for Production of Sesame crop in Andhra Pradesh

Model	Parameter				Criteria		
	a	b	c	d	R ²	RMSE	MAPE
Linear	40.850**	-.395**			.298**	9.2792075	30.05092
Logarithmic	50.770**	-6.805**			.295**	9.29439455	30.95508
Inverse	27.791**	27.838**			.147**	10.2256399	34.175233
Quadratic	39.839**	-.285	-.002		.299**	9.26937772	29.71323
Cubic	53.814**	-3.252**	.134**	-.002**	.481**	7.97996468	24.78012
Compound	41.915**	.985**			.330**	9.42631694	28.277905
Power	57.646**	-.238**			.283**	9.63411592	29.267087
S-Curve	3.255**	.919**			.126**	10.8263891	32.498897
Growth	3.736**	-.015**			.330**	9.42631694	28.277905
Exponential	41.915**	-.015**			.330**	9.42631694	28.277905
Time Series Models							
ARIMA(1, 1, 2)					0.501	8.254	23.562
ARIMA(1, 2, 3)					0.383	8.934	24.011
ARIMA(2, 0, 3)					0.579	7.715	22.831
ARIMA(2, 1, 3)					0.514	8.327	23.350
ARIMA(2, 2, 3)					0.368	9.142	24.409
ARIMA(3, 0, 3)					0.621	7.400	20.876
ARIMA(3, 1, 2)					0.513	8.335	23.333
ARIMA(3, 2, 3)					0.388	9.104	24.091

Table.3 Linear, Non-linear and Time series models for Productivity of Sesame crop in Andhra Pradesh

Model	Parameter				Criteria		
	a	b	c	d	R2	RMSE	MAPE
Linear	146.748**	2.090**			.367**	41.9919318	17.9532
Logarithmic	128.546**	24.668**			.171**	48.0526083	21.13302
Inverse	208.208**	-58.592			.029	52.0114761	21.950139
Quadratic	195.068**	-3.182*	.098**		.517**	38.6928202	15.60092
Cubic	200.826**	-4.404	.154	-.001	.518**	38.6413231	15.59041
Compound	151.714**	1.010**			.310**	41.485376	17.438246
Power	141.165**	.109**			.135	48.0257121	20.33387
S-Curve	5.301**	-.237			.019**	52.4376734	21.173282
Growth	5.022**	.010**			.310**	41.485376	17.438246
Exponential	151.714**	.010**			.310**	41.485376	17.438246
Time Series Models							
ARIMA(1,0,1)					0.477	39.701	16.209
ARIMA(1,1,2)					0.532	38.218	14.772
ARIMA(1,2,1)					0.250	48.396	19.171
ARIMA(2,1,3)					0.519	39.601	15.360
ARIMA(3,0,3)					0.539	38.873	15.476
ARIMA(3,1,1)					0.551	37.842	14.343
ARIMA(3,2,1)					0.304	47.649	18.461
ARIMA(2,2,3)					0.319	47.672	18.291

1. The value of the criterion for a model with bold numbers shows that the model is better than the other models with respect to that criterion.
2. **, * indicates significant at 1% and 5% level of probability respectively.

Table.4 Estimates of the fitted ARIMA (3, 0, 0), ARIMA (3, 0, 3) and ARIMA (3, 1, 1) models for Sesame crop Area, Production and Productivity respectively

Sesame	Model fit Statistics			Ljung-Box Q (18)	
	R-Square	RMSE	MAPE	Statistic	p-value
Area	0.793	27.711	17.487	6.952	0.959
Production	0.621	7.400	20.876	14.950	0.244
Productivity	0.551	37.842	14.343	22.577	0.068

Table.5 Test for randomness of the residuals for fitted models of Sesame crop Area, Production and Productivity

Model	Run Test for Residuals			
	Total Cases	No. of Runs	Z- Value	Sig(2-tailed)
Area : ARIMA (3, 0, 0)	53	31	0.974	0.330
Production : ARIMA (3, 0, 3)	53	30	0.696	0.489
Productivity: ARIMA (3, 1, 1)	53	21	-0.969	0.333

Table 6: Forecasted values of Sesame crop Area, Production and Productivity with 95% Confidence Level (CL)

Year	Area			Production			Productivity		
	Forecasted values	LCL	UCL	Forecasted values	LCL	UCL	Forecasted values	LCL	UCL
2018-19	44.23	-11.35	99.8	10.5	-3.54	24.54	304.86	230.07	379.65
2019-20	54.19	-10.97	119.34	20.63	4.73	36.54	318.67	240.88	396.46
2020-21	51.66	-14	117.32	17.64	1.28	33.99	323.43	245.51	401.36
2021-22	47.87	-18.94	114.67	20.58	4.22	36.94	325.29	246.01	404.58
2022-23	47.72	-20.74	116.18	20.95	4.47	37.43	330.61	251.18	410.04

LCL: Lower Confidence Level, UCL: Upper Confidence Level

Fig.1 Forecasted Sesame crop Area (1965-66 to 2022-23)

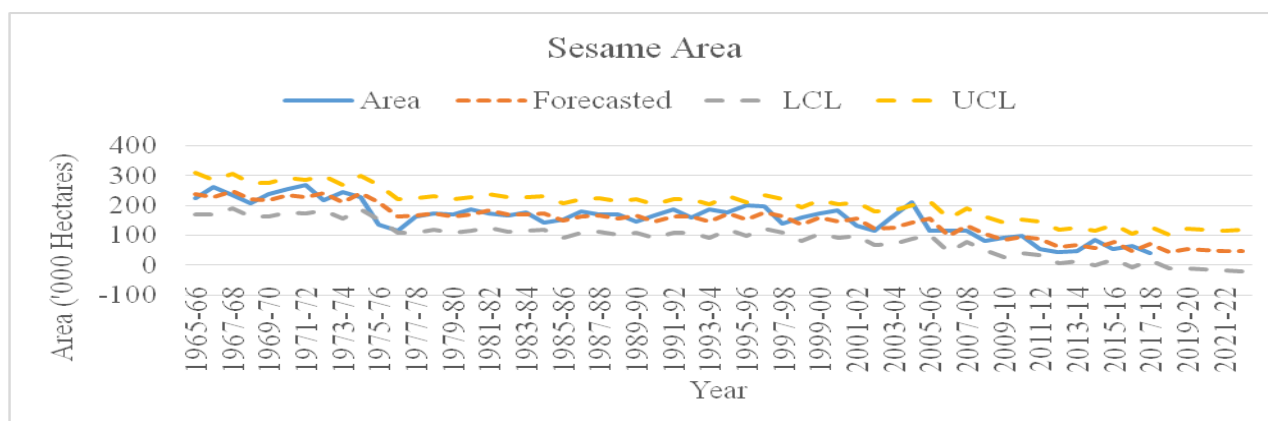


Fig.2 Forecasted Sesame crop Production (1965-66 to 2022-23)

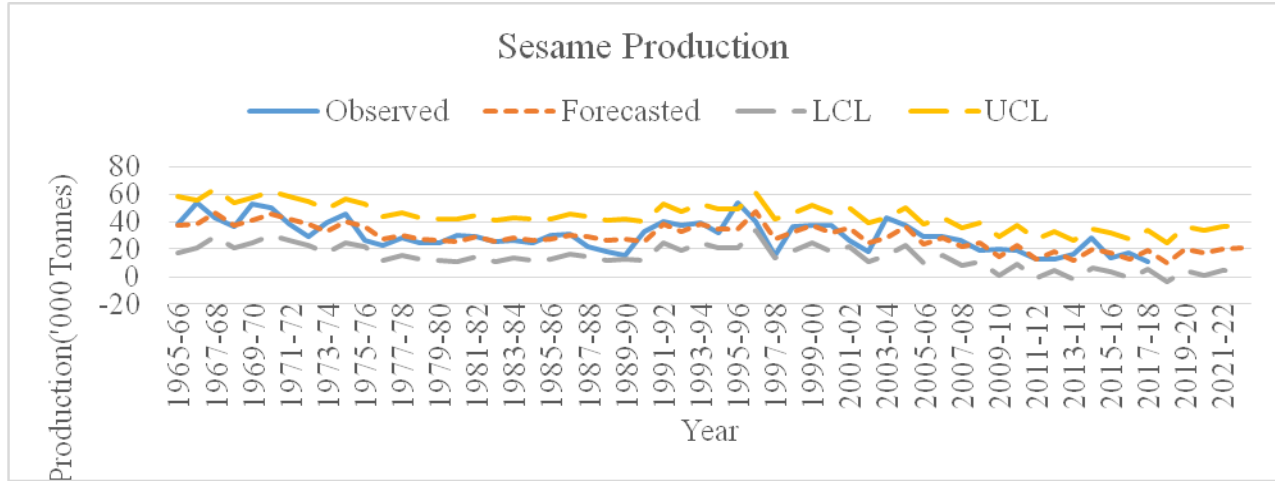
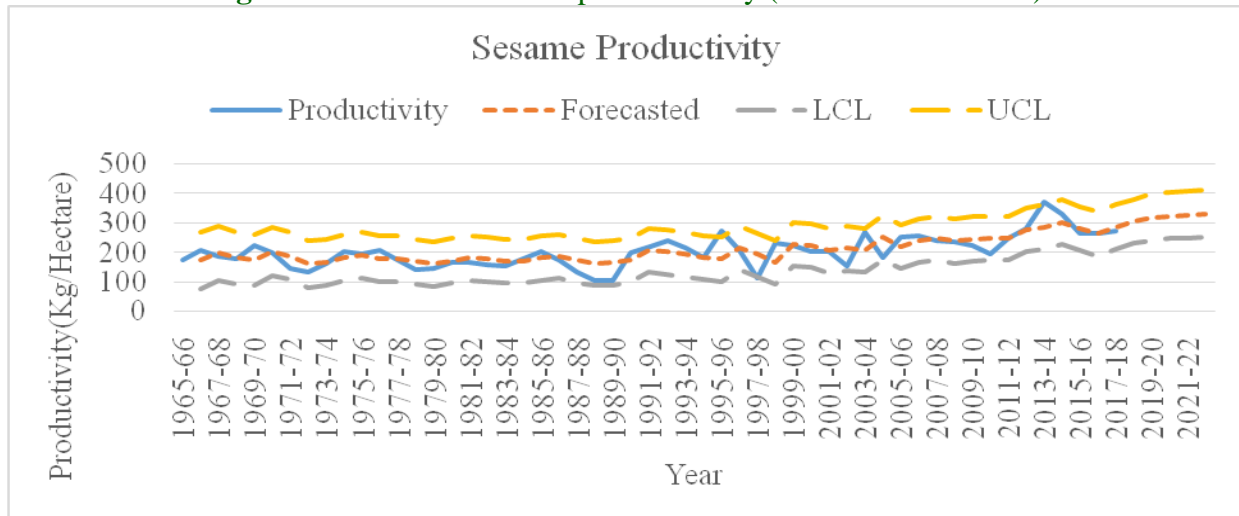


Fig.3 Forecasted Sesame crop Productivity (1965-66 to 2022-23)



Test for randomness of residuals

Non-parametric one sample run test can be used to test the randomness of residuals. A *run* is defined as a succession of identical symbols in which the individual scores or observations originally were obtained. Let ‘ n_1 ’, be the number of elements of one kind and ‘ n_2 ’ be the number of elements of the other kind in a sequence of $N = n_1 + n_2$ binary events. For small samples i.e., both n_1 and n_2 are equal to or less than 20 if the number of runs ‘ r ’ fall between the critical values, we accept the H_0 (null hypothesis) that the

sequence of binary events is random otherwise, we reject the H_0 . For large samples i.e., if either n_1 or n_2 is larger than 20, a good approximation to the sampling distribution of r (runs) is the normal distribution, with mean

$$\mu_r = \frac{2n_1n_2}{N} + 1 \quad \text{and standard deviation}$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

$$z = \frac{r - \mu_r}{\sigma_r}$$

Then, H_0 may be tested by

The significance of any observed value of Z computed from the above formula may be determined by reference to the standard normal distribution table.

Forecasting with ARIMA model

After the identification of the model and its adequacy check, it is used to forecast the Area, Production and Productivity of Sesame crop for the next five periods. Hence we used the identified ARIMA model to forecast Area, Production and Productivity of Sesame crop for the years 2018-19 to 2022-23. The forecasting results are presented in Table. 6. And also the diagrams of actual and forecasted values are presented in Figs.1, 2 and 3.

It is concluded, in the present study the developed ARIMA (3, 0, 0), ARIMA (3, 0, 3) and ARIMA (3, 1, 1) were the best models for forecasting the Sesame area, production and productivity based on R^2 , RMSE and MAPE criteria in Andhra Pradesh. The study revealed that in coming next five years there is a fluctuations on area and production and increasing trend on productivity of sesame crop in Andhra Pradesh. Sesame seeds have many potential health benefits and have been used in folk medicine for thousands of years. They may protect against heart disease, diabetes, and arthritis. Most of Sesame seed is used for oil extraction which is mainly used for cooking purpose. Thus the agricultural scientist and farmers should take more attention to improve the production and productivity of sesame in Andhra Pradesh.

References

- Box GEP and Jenkin GM (1976), "Time series of analysis, Forecasting and Control", San Francisco, Holden Day, California, USA.
- Narayanaswami. T., Surendra. H.S., and Santosh Rathod. (2012). Fitting of Statistical Models for Growth of Root and Shoot Morphological Traits in Sesame. *Environment & Ecology*, ISSN: 0970-0420, 30 (4): 1362-1365
- Prabakaran, K., Nadhiya, P., Bharathi, S. and Isaivani, M., (2014). Forecasting of Pulses area and production in India – An ARIMA Approach. *Indian Streams Research Journal*. 4(3):1-8.
- Ramana Murthy. B. and Haribabu.O (2018). A Statistical trend analysis of Mango Area, Production and Productivity in Andhara Pradesh. *Int. Journal of Agricultural and Statistical Sciences*, 14 (1):337-342.
- Sudha, CH. K., Rao, V.S and Suresh, CH. (2013). Growth trends of maize crop in Guntur district of Andhra Pradesh. *International Journal of Agricultural Statistical Sciences*. 9 (1): 215-220.
- Ramana Murthy, B., Mohan Naidu, G., Ravindra Reddy, B., and Nafeez Umar, Sk. (2018). Forecasting Groundnut area, production and productivity of India using ARIMA Model. *Int. Journal of Agricultural and Statistical Sciences*, 14 (1):153-156.
- Ramana Murthy. B., Mohan Naidu. G., Tamilselvi.C and PriyankaEvangilin. N. (2020). Validation of ARIMA Model on Production of Papaya in India. *Indian Journal of Pure and Applied Biosciences*, 8(2), 64-68.

www.epwrfits.in

www.indiastat.com.

How to cite this article:

Priyanka Evangilin, N., B. Ramana Murthy, G. Mohan Naidu and Aparna, B. 2020. Statistical Model for Forecasting Area, Production and Productivity of Sesame Crop (*Sesamum indicum* L.) in Andhra Pradesh, India. *Int.J.Curr.Microbiol.App.Sci.* 9(07): 1156-1166.
doi: <https://doi.org/10.20546/ijcmas.2020.907.135>