Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.7148

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **6 (6):** 1121-1126 (2018)

Research Article

International Journal of Pure & Applied Bioscience

OPEN ACCESS

Arima Model for Forecasting Sunflower Production in India

B. Ramana Murthy[1*], Sk. Nafeez Umar[2] and O. Hari Babu[3]
[1]Assistant Professor, Department of Statistics and Computer Applications,
Acharya N. G. Ranga Agricultural University, S.V. Agricultural College, Tirupati, Andhra Pradesh
[2]Assistant Professor, Department of Statistics and Computer Applications,
ANGRAU, Agricultural College, Bapatla, Andhra Pradesh
[3]Associate Professor, Spirits College of MCA and MBA, Kadapa (Dt), Andhra Pradesh
*Corresponding Author E-mail: ramanastats@gmail.com

Received: 6.11.2018 | Revised: 14.12.201 | Accepted: 19.12.2018

ABSTRACT

The present research study was carried out to identify the appropriate Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) model for forecasting sunflower production in India. The validity of the model was tested using standard statistical techniques R^2, RMSE, and MAPE. ARIMA (4, 1, 4) model was found to be a best fitted model to forecast sunflower production in India for further five years. The important assumption of randomness of residuals was tested using one sample run test. The forecasted results showed for production of sunflower in India for the year 2017-18 to 2021-22 to be 220, 150, 114, 121 and 141 thousand tonnes respectively. And also it is showed downward and upward trend on production of sunflower in India for forecasted years.

Key words: *Sunflower, R^2, RMSE, MAPE and ARIMA.*

INTRODUCTION

Sunflower (Helianthus annuus L.) belongs to the family compositae and it is native of North and South America. It is one of the important oilseed crop of the country and it is being called "Champion of oilseed crop". It has got greater economic importance among the oilseed crops in the world. Sunflower is a major oilseed crop gaining paramount importance in the world and it ranks next only to soybean and groundnut in the total oilseeds production of the world. Sunflower is the world's fourth largest oil-seed crop and its seeds are used as food and its dried stalk as fuel. It is already been used as ornamental plant and was used in ancient ceremonies[2,3]. Additionally, medical uses for pulmonary afflictions have been reported. Sunflower is used in industry for making paints and cosmetics. India is one of the largest producers of oilseeds in the world and occupies an important position in the Indian agricultural economy. In 2014, global production of sunflower seeds was 41.4 million tones, led by Ukraine with 24% and Russia with 21% of the world total. China, Romania and Argentina also contributed significant volumes.

Cite this article: Murthy, B.R., Umar, S.N. and Hari Babu, O., Arima Model for Forecasting Sunflower Production in India, *Int. J. Pure App. Biosci.* **6(6):** 1121-1126 (2018). doi: http://dx.doi.org/10.18782/2320-7051.7148

Prabakaran *et al*.[5] analysed pulses area and production in India during the period from 1950-51 to 2011-12 by using ARIMA model and he was found that ARIMA (1, 1, 0) and ARIMA (2, 1, 1) models were best fitted to forecast area and production in India for four leading years. Manoj and Madhu[4] attempted to study forecasting of sugarcane production in India by ARIMA model. They have found that ARIMA (2, 1, 0) model was best fitted model for forecasting of sugarcane production. Ramana Murthy *et al*.[6] attempted to study forecasting groundnut area, production and productivity of India using ARIMA model. 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 and also 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.

The objective of the present study was to develop appropriate Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) models for the time series of sunflower production in India and to make five year forecasts with appropriate ARIMA model.

MATERIAL AND METHODS

The data of study for a period of 47 years (1970-71 to 2016-17) in India pertaining to on Production ('000 tonnes) of sunflower was collected from the source of Ministry of Agriculture & Farmers Welfare, Govt. of India in indiastat.com website.

2.1 Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) Model

The period from1970-71 to 2016-17 was used for forecasting the future values using Auto Regressive Integrated Moving Average (ARIMA) models. The ARIMA methodology is also called as Box-Jenkins methodology [Box and Jenkins[1]]. 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 has 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, q) process is given by

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

Where y_t and ε_t are the actual value and random error at time period t, respectively; $\phi_i (i = 1,2,3,\dots\dots, p)$ and θ_j (j=1, 2, 3,….., 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

2.2 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.

$$\mu_r = \frac{2n_1n_2}{N} + 1 \text{ and standard deviation } \sigma_r = \sqrt{\frac{2n_1n_2\left(2n_1n_2 - n_1 - n_2\right)}{\left(n_1 + n_2\right)^2\left(n_1 + n_2 - 1\right)}}$$

$$\text{Then, } H_0 \text{ may be tested by } z = \frac{r - \mu_r}{\sigma_r}$$

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

RESULTS AND DISCUSSION

In the present study, the data for sunflower production for the period of 47 years (1970-71 to 2016-17) were used for the study.

3.1 Model Identification

Among the several methods studied the goodness of fit of fitted models is examined by highest R^2 (Coefficient of determination) value, minimum Root Mean Square Error (RMSE) and minimum Mean Absolute Percentage Error (MAPE) value. According to these criterions, it was found that ARIMA (4, 1, 4) is the best fitted models for forecasting sunflower production.

The Coefficient of determination (R^2), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are given by

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

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2}{n}} \qquad (3)$$

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \qquad (4)$$

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 different ARIMA models were tested and the models which had Maximum R^2 Minimum RMSE and Minimum MAPE were chosen. The models and the corresponding values are as under

Sunflower	Model ARIMA (p,d,q)	R^2	RMSE	MAPE
Production	(1, 0, 0)	0.863	167.104	35.885
	(2, 0, 0)	0.866	167.453	35.205
	(2, 2, 1)	0.835	185.523	29.633
	(2, 1, 4)	0.895	154.919	29.378
	(2, 3, 1)	0.813	196.085	31.308
	(3, 1, 4)	0.910	145.342	29.017
	(3, 2, 4)	0.850	186.404	28.229
	(3, 3, 3)	0.816	202.776	31.208

	(4, 1, 1)	0.899	150.288	30.690
	(4, 1, 4)	**0.914**	**144.314**	**28.082**
	(1, 1, 4)	0.893	154.246	28.863
	(4, 2, 1)	0.867	171.043	29.225
	(4, 3, 4)	0.826	202.486	29.876
	(4, 1, 5)	0.907	151.672	28.572
	(5, 1, 4)	0.905	153.684	28.888
	(5, 2, 4)	0.885	168.521	28.228

3.2 Model Parameters Estimation and Verification

The parameters of the model were estimated by using SPSS 20 package. The ARIMA (4, 1, 4) model is the best fitted model for production of sunflower of India. The model verification (or) diagnosed was checked by using 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

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

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 $(p-value>\alpha)$. The results of estimation are reported in Table 1. And the estimated of parameters of ARIMA (4, 1, 4) model are reported in Table. 2.

3.3. Forecasting with ARIMA model

After the identification of the model and its adequacy check, it is used to forecast the production of sunflower in the next periods. Hence we used the identified ARIMA model to forecast the production of sunflower for the years 2017-18 to 2021-22. The forecasting results are presented Table. 3. And also the diagrams of actual and forecasted values are presented in Fig.1.

3.4 Test for randomness and normality of residuals for Sunflower Production.

Examination of assumptions about residuals of ARIMA (4, 1, 4) is presented in table. 4. Table revealed that $(p-value>\alpha)$ the residuals are randomly distributed as the run test statistic.

Table-1: Estimates of the fitted ARIMA (4, 1, 4) model for sunflower production

	Model fit Statistics		**Ljung-Box Q (18)**	
R-Square	RMSE	MAPE	Statistic	p-value
0.914	144.314	28.082	13.273	0.209

RMSE: Root Mean Square Error.
MAPE: Mean Absolute Percentage Error

Table2. Estimates of parameters along with their SE for fitted ARIMA (4, 1, 4) model

ARIMA (4, 1, 4)		**Estimate**	**SE**
	Constant	1.828	0.812
AR	Lag1	1.011	0.286
	Lag2	-0.037	0.337
	Lag3	-0.556	0.336
	Lag4	0.296	0.273
	Difference	1	
MA	Lag1	1.056	5.515
	Lag2	-0.003	9.823
	Lag3	-1.051	6.113
	Lag4	0.993	9.041

Table 3: Forecasted values of Sunflower Production with 95% Confidence Level (CL)

Year	Production		
	Forecasted values	**UCL**	**LCL**
2017-18	220	410	78
2018-19	150	382	13
2019-20	114	363	0
2020-21	121	452	7
2021-22	141	541	12

UCL: Upper Confidence Level, LCL: Lower Confidence Level,

Table 4: Test for randomness of the fitted ARIMA (4, 1, 4) model

Run Test	Residuals
Total cases	47
Number of runs	18
Z- value	-0.996
Sig (2-tailed)	0.319

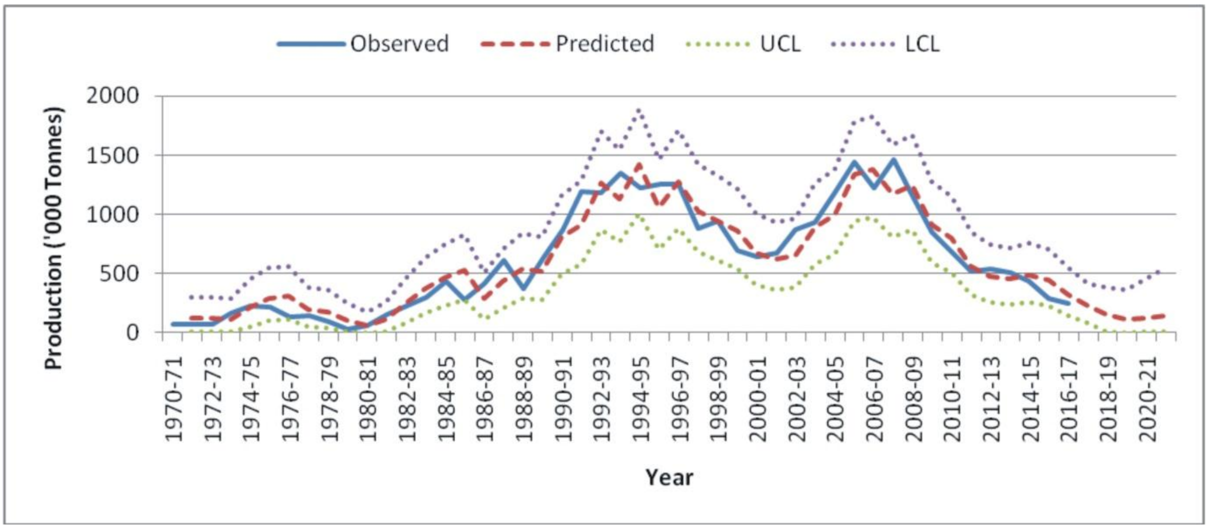

Fig. 1: Forecasted Sunflower production (1970-71 to 2021-22)

CONCLUSION

In the present study, the developed ARIMA (4, 1, 4), is the best model for forecasting the sunflower production of India based on R^2 , RMSE and MAPE criterions. The study revealed that in coming next five years the trend is decrease and then gradually increasing of sunflower production of India. Sunflowers are a permanent source of food, oilseed and biofuels. The sunflower, has several potential markets, it is a good choice for growers in both small and large scales. However, scientific, technical or agricultural projects linked with sunflower have to include environmental side effects such as pollution, greenhouse gases emissions, salinization, or energy consumption elsewhere in order to shape a sustainable future.

REFERENCES

1. Box, G. E. P and Jenkins, G. M., *Time series of analysis*, *Forecasting and Control*. San Fransisco, Holden Day, California, USA. (1976).
2. Harter, A. V., Gardner, K. A., Falush, D., Lentz, D. L., Bye, R. A. & Rieseberg, L. H., Origin of extant domesticated sunflower in eastern North America. *Nature* **430(6996):** 201-205 (2004).
3. Muller, M. H., Latreille, M. & Tollon, C., The origin and evolution of a recent agricultural weed: population genetic

diversity of weedy populations of sunflower (*Helianthus annuus* L.) in Spain and France. *Evol Appl* **4(3):** 499-514 (2011).

4. Manoj, K. and Madhu, A., An Application of Time Series ARIMA Forecasting Model for Predicting Sugarcane Production in India. *Studies in Business and Economics Journal*. **9(1):** 81-94 (2014).
5. Prabakaran, K., Nadhiya, P., Bharathi, S. and Isaivani, M., Forecasting of Pulses area and production in India – An ARIMA Approach. *Indian Streams Research Journal*. **4(3):** 1-8 (2014).
6. Ramana Murthy, B., Mohan Naidu, G., Ravindra Reddy, B. and Nafeez Umar, S. K., Forecasting Groundnut area, production and productivity of India using ARIMA Model. *Int. Journal of Agricultural and Statistical Sciences*, **14(1):** 153-156 (2018).
7. www.indiastat.com