Available Online at http://www.journalajst.com

ISSN: 0976-3376

ASIAN JOURNAL OF SCIENCE AND TECHNOLOGY

*Asian Journal of Science and Technology*
*Vol. 08, Issue, 05, pp.4736-4739, May, 2017*

# RESEARCH ARTICLE

# TIME SERIES MODELS FOR AREA, PRODUCTION AND PRODUCTIVITY FORECASTING IN COTTON CROP OF INDIA

***[1]Sundar Rajan, M. and [2]Palanivel, M.**

[1]PhD Research Scholar, Govt. Arts College, Udumalpet, Tamil Nadu, India-642 126
[2]Assistant Professor, Govt. Arts College, Coimbatore, Tamil Nadu, India-641 018

## ARTICLE INFO

***Article History:***
Received 22nd February, 2017
Received in revised form
04th March, 2017
Accepted 19th April, 2017
Published online 17th May, 2017

***Key words:***

Cotton (*Gossypiumhirsutum),*
Forecasting,
Time series data,
ARIMA.

## ABSTRACT

This study aims at presenting models for the forecasting time-series data of cotton (*Gossypiumhirsutum)* Area, Production and Productivity of India by using Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) models of time-series forecasting. The Time series data covering the period of 1951–2014 was used for the study. Presence of trend was checked through time series data were forecast for the staring from 2015-2021 the cultivated area, production and productivity of cotton in India using Autoregressive Integrated Moving Average (ARIMA) are calculated based on using this selected model. A comparison of these forecasts with observed values over this time period indicated that the model was highly accurate. The performances of the models are of ARIMA family for modeling as well as forecasting purpose.



## INTRODUCTION

Cotton is an important cash crop and India cultivates the highest acreage in the world. It provides the basic raw material (cotton fibre) to cotton textile industry. Known also as 'White Gold' Cotton enjoys a predominant position amongst all cash in India. World cotton production is estimated at 118.95 million bales of 480 lb in 2012-13 (USDA, February, 2013) which is 4 % less than the previous year 2011-12 and cotton area to the tune of 4.62%. India continued to maintain the largest area under cotton and second largest producer of cotton next to china with 34 % of world area and 21% of world production. Significant drop in the production level in Brazil about 43% due to 40% reduction in area under cotton compared to previous year. Though there is reduction of 3.41% of area under cotton in United States, there is increase of 7.85% in cotton production compared to 2011-12. China is going to be the largest importer around 14 million bales of Cotton this year and share of 34% of the world total cotton imports. China also likely to emerge as a leader to stock large cotton reserve, and it is estimated to around 52% followed by India with 10% of the world cotton reserve.

****Corresponding author: Sundar Rajan***
PhD Research Scholar, Govt. Arts College, Udumalpet, Tamil Nadu, India-642 126

United States continues to be the largest exporter of raw cotton and expected to export around 12.5 million bales which are 31% of the world total exports, followed by Brazil, India and Australia (11% each). China, India and Pakistan continue to be the largest consumer of raw cotton this year also, around 65% of the total world raw cotton consumption by these countries. Among the major cotton growing countries, Australia tops the productivity level of 2055 kg/ha followed by Brazil and Turkey (1415 kg/ha). The major cotton growing states are Punjab, Haryana, Rajasthan, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Tamil Nadu and Karnataka. With an objective to improve the quality of cotton, enhance/ha productivity, enhance the income of cotton growers by reducing the cost of cultivation, to improve the processing facilities etc., In our present study cotton crop of India have been forecasted using Auto Regressive Integrated Moving Average (ARIMA) models.

Lot of research was done on forecasting by using ARIMA models on various crops. Forecasting of area production and productivity of different crops for Tamilnadu State studied (Balanagammal *et al.*2000), forecasting on fish catches were studied by Tsitsika *et al*., 2007 andVenugopalan&Srinath1998, forecasting on milk production in Pakistan using ARIMA model were done Ahmed *et al*.(2011), Saeed *et al*. 2000 and Boken 2000 were forecasted wheat production in Pakistan and Canada respectively using ARIMA models, Mandal 2005 forecasted Sugarcane production in India,(Pal *et al.*2007) Milk production in India was forecasted using time series modeling techniques, Sahu 2006 Forecasting irrigated crops like Potato, Mustard and Wheat using ARIMA models,(Rahman, S.A., Noor, M.A 2005) Comparison of time series methods for electricity forecasting. An objective of the study was to forecast Cotton area production and productivity using the best fitted model of trend analysis. An ARIMA models has provided successful forecasts in many other applications and the basic tools use for short term forecasts to make eight years with appropriate prediction interval.

### Analytic Techniques

To find the best fitted model for area, production and productivity of cotton crop in India. Such as Linear, Quadratic and Exponential models. These measures include Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation(MAD) and Mean Squared Deviation (MSD). The aim and purpose of the present study is twofold one is to check the past trends of cotton area, production and productivity in India and Second is to Forecast to cotton crop in the next eight values are very close to actual values.

## METHODOLOGY

The study was carried out on the basis of Cotton area, production and productivity during the years 1951-52 to 2014-15 collected from secondary sources, (Agricultural at Glance, India 2014). The data was model using Autoregressive Integrated Moving Average (ARIMA) stochastic model as proposed by Box and Jenkins (1970). The model is called AR model if it contains only the autoregressive terms. Model is known as MA model if it involves only the moving average terms. It is known as ARMA models when both autoregressive and moving average terms are involved. Finally when a non-stationary series is made stationary by differencing method, it is known as ARIMA model. The results showed that the prediction is quite accurate using method of ARIMA in the study area, production and productivity.

### Techniques of Prediction using ARIMA model

ARIMA time series models traditionally expressed as ARIMA (p,d,q) combine as many as three types of processes, viz.

AR: $p$ = order of the autoregressive part

I: $d$ = degree of first differencing involved

MA: $q$ = order of the moving average part

The form of the ARIMA (p,d,q) respectively,

A p[th]- order autoregressive model: AR(p), which has the general form:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$$

Where, $Y_t$= Response (dependent) variable at time t

$Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ =Response variable at time

Lags $t-1, t-2, \ldots, t-p$, respectively

$\varphi_0, \varphi_1, \varphi_{2,\ldots,} \varphi_p$ =Coefficient to be estimated

$\varepsilon_t$ = Error term at time t.

A q[th]- order moving average model: MA(q), which has the general form:

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

Where, $Y_t$= Response (dependent) variable at time t

$\mu$ = Constant mean of the process

$\theta_1, \theta_2, \ldots, \theta_q$ = Coefficients to be estimated

$\varepsilon_t$ = Error term at time t

$\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \epsilon_{t-q}$ = Error in previous time.

Periods that is incorporated in the response $Y_t$

Finally, the general form of ARIMA model of order (p,d,q) has the representation

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} + \varepsilon_t.$$

Where $Y_t$ is the response (dependent),$\varepsilon_t$'s are independently and normally distributed with zero mean and constant variance $\delta^2$for $t = 1, 2, \ldots, n$ and $d$ is the fraction differenced while interpreting AR and MA and $\varphi_s$and $\theta_s$ are the coefficients to be estimated.

### Trend fitting

For evaluating the adequacy of AR, MA and ARIMA processes, maximum value of coefficient of determination ($R^2$), Among the competitive models, best models are selected based on minimum value of Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Schwartz by Normalized Bayesian Information Criteria (BIC) [as suggested 1978] and Q statistics have been used. Best fitted models are put under diagnostic checks through Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) of the residuals. The lower the values of above statistics, the better are the models which are given by,

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2\right]^{1/2} \quad MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(Y_i - \hat{Y}_i)}{Y_i}\right|$$

$$MAE = \frac{\sum_{i=1}^{n}\left|X_i - \hat{X}_i\right|}{n} \quad R^2 = \frac{\sum_{i=1}^{n}\left(\hat{X}_i - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

$$\text{Adjusted } \overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Adjusted $\overline{R}^2$ is a modification of $\overline{R}^2$ that adjusts for the number of explanatory terms in a model. The adjusted $\overline{R}^2$ can be negative, and will always be less than or equal to $\overline{R}^2$.

$$BIC\ (p,q) = In\ v^*(p,q) + (p,q)[In\ (n)/n]$$

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 $\sigma^2$. The significance level of individual coefficients is measured by Box-Jenkins Q-statistics were computed Fig1,

$$Q^* = n(n+2)\sum_{k=1}^{h}(n-k)^{-1}r_k^2$$

Where $n$ is the number of residuals, $h$ is the maximum lag and $rk$ is the residuals autocorrelation at $lag^k$.

**Box and Jenkins Methodology for time series modeling steps**

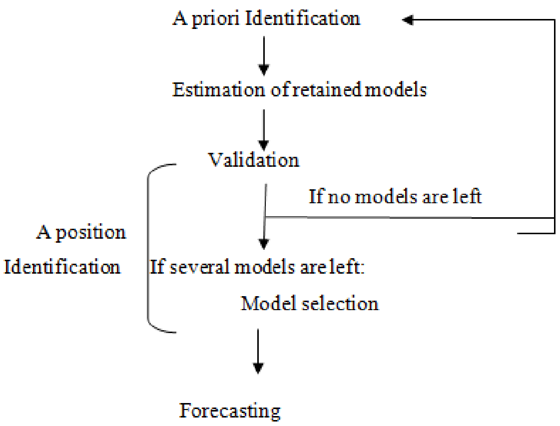


**Fig 1. Schematic representation of the Box - Jenkins Methodology for time series modeling**

The Box-Jenkins Methodology is regarded as the most efficient forecasting technique, used in analysis and forecasting and is used widely and extensively specially for univariate time series modeling. The modeling of the time series was applied according to four stages namely:

***Model identification***: Specify the order of AR and MA components were determined.

***Model estimation***: Linear model coefficients were estimated.

***Model validation***: certain diagnostic methods were used to test the suitability of the estimated model.

***Forecasting***: If the model passes the validation step, then it can be used best model chosen for forecasting. ARIMA methodology may be precisely visualized from Figure 1

## RESULTS AND DISCUSSION

This output contains information is useful in understanding the descriptive qualities of the data. Since 1951, the area under cotton has been increased from 5.58 Lakh ha to 12.18 Lakh ha till 2014. The average shows under cotton being 8.14 Lakh hectare. With the mere 3.04 Lakh bales of production it has reached to 36.59 Lakh bales during the year 2014. Starting with only 85 kg per hectare, it has reached to 532 kg ha during the year 2014. Positive values of skewness reveal that starting from the initial years of the period under investigation a continuous effort was there to increase the production of cotton these presented given in Table 1.

### Trends in Area Production and Productivity behavior of Cotton

To work out the trends in area production and productivity of cotton in India different parametric models like linear, quadratic and exponential were attempted to among the competitive models. The best model was selected on the basis of the maximum $R^2$ value, Adj $R^2$, MAPE, MAD, MSD and significance of the model and its coefficient shows Table 2. Table 2 revealed that all the values of accuracy measures for cotton area in India are smaller in Quadratic model and for production are smaller in Exponential model Last one for productivity are smaller in Exponential model. So this model best fitted was selected on the basis of the maximum $R^2$ value, Adj $R^2$and minimum MAPE, MAD, MSD and significance of the model and its coefficient.

### ARIMA Modeling of Cotton Area, Production and Productivity

The result is clear from the observed and predicted values on Cotton crop Area Production and Productivity. The stationary check of time series revealed that the time series Cotton crop. It was observed that on Cotton Area and Production and productivity was not stationary at level. So it was made stationary by using the first order differencing technique. The appropriate model was selected on highest $R^2$ and least RMSE,

**Tables 1. Descriptive Statistics of Cotton Crop in India**

| Variables | Minimum | Maximum | Mean | Std. Deviation | Skewness |
|---|---|---|---|---|---|
| Area | 5.58 | 12.18 | 8.15 | 1.23 | 1.52 |
| Production | 3.04 | 36.59 | 10.42 | 8.09 | 1.97 |
| Productivity | 85 | 532 | 203.25 | 117.01 | 1.41 |

**Table 2. Model summary of Trend model fitted**

| Variables | Model | MAPE | MAD | MSD | $R^2$ | *Adj* $R^2$ | Trend Model |
|---|---|---|---|---|---|---|---|
| Area | Quadratic | 7.83 | 0.63 | 0.59 | 0.601 | 0.588 | $Y_t = 7.85027 - 0.0591588^{*}t + 0.00158746^{*}t^{**}2$ |
| Production | Exponential | 17.33 | 2.21 | 16.32 | 0.863 | 0.861 | $Y_t = 3.12129^{*}(0.03105^{**}t)$ |
| Productivity | Exponential | 12.66 | 29.07 | 2169.75 | 0.894 | 0.892 | $Y_t = 77.2601^{*}(1.02594^{**}t)$ |

MAPE, MAE and Normalized BIC values have been presented in Table 3 for cotton area, production and productivity. From the study, it is revealed that ARIMA (1,1,8) model was found to be appropriate for cotton area and production but ARIMA (2,1,5) model was suitable for cotton productivity.

**Table 3.Time Series models in Area, Production and Productivity of Cotton in India**

| | Time Series Models | $R^2$ | RMSE | MAPE | MAE | Normalized BIC |
|---|---|---|---|---|---|---|
| Area | 1,1,8 | 0.87 | 0.46 | 4.31 | 0.34 | -.817 |
| Production | 1,1,8 | 0.97 | 1.61 | 13.67 | 1.16 | 1.675 |
| Productivity | 2,1,5 | 0.96 | 25.88 | 9.57 | 18.31 | 7.099 |

**Table 4. Forecasting Projections of Area, Production and Productivity of Cotton in India**

| Year | Area (million ha.) | Production (million bales. 170kg) | Productivity (Kg/ha.) |
|---|---|---|---|
| 2015 | 11.76 | 37.41 | 584 |
| 2016 | 11.63 | 35.27 | 585 |
| 2017 | 11.87 | 36.08 | 571 |
| 2018 | 11.94 | 38.99 | 624 |
| 2019 | 11.62 | 40.42 | 663 |
| 2020 | 11.53 | 40.56 | 643 |
| 2021 | 11.56 | 41.33 | 656 |

After model fitting, next step is diagnostic checking of the fitted model. For this purpose, we used ACF and PACF were plotted for residuals of the fitted model. In this study ACF and PACF were lying within the limits for cotton area, production and productivity which indicated that models were well fitted. The best selected fitted models were validated by comparing the observed values with predicted values of Cotton in India by 2021 were calculated and the results were presented in Table 4. The projected area, production and productivity under cotton by 2021 would be 11.56 million hectare, 41.33 million bales (for 170Kgs.each) and 656 (Kg/ha) respectively. The projections of cotton area would be decreased and production and productivity would be slightly by increasing than the actual values by 2021.

### Conclusion

In the study ARIMA model being stochastic in nature, it could be successfully used for modeling as well as forecasting the cotton area, production and productivity in India. From the forecast available by using the developed model it can be seen that forecasted cotton cultivated production and productivity were to increase in the next eight years. Cotton is the basic need of any country all over the world. Presented study, we developed time series models to forecasts cotton of India on the basis of historical data i.e. 1951-2014. We have developed different time series models on cotton area, production and productivity of India. Best model is selected on the basis of model selection criteria i.e. BIC. On the basis of these model selection criteria, we have found that best model for cotton forecasting of India is ARIMA (1,1,8)(1,1,8) and (2,1,5). On the basis of developed time series model, we have found that best time series model for forecasting cotton area, production and productivity of India ARIMA (1,1,8)(1,1,8) and (2,1,5) because this model has lower BIC as compared to other fitted time series models. The validity of the forecasted value can be checked when the data for the lead period become available. The model demonstrated a good performance in terms of explaining variability and predicting power. The forecasting of cotton yield can help to farmers as well as the planner and researchers for future planning.

## REFERENCES


All India Coordinated Cotton Improvement Project – Annual Report (2012-13).

Balanagammal, D., C. R. Ranganathan, and K. Sundaresan. 2000. Forecasting of agricultural scenario in Tamilnadu: A time series analysis. *Journal of Indian Society of Agricultural Statistics,* 53(3):273-286.

Box, G.E.P. and Jenkins, G.M. 1970. 'Time Series Analysis-Forecasting and Control'. Holden-Day Inc., San Francisco.

Ljunge, G,M. and Box, G.E.P. 1978. On a measure of lack of fit in time series models. Biometrika, 65:67-72.

Mandal B.N. 2005. Forecasting Sugarcane Production in India using ARIMA model, http://interstat.statjournals.net.

Rahman, S.A., Noor, M.A. and Khalid, K. 2005. Comparison of time series methods for electricity forecasting. A case study in Perlis. ICOQSIS.

Saeed, N., A.Saeed, M. Zakria and T.M. Bajwa. 2000. Forecasting of wheat production in Pakistan using ARIMA models. *International journal of Agricultural Biology*, 2(4):352-353.

Sahu, P.K. 2010. Forecasting production of major food crops in four major SAARC countries. International Journal of Statistical Sciences, Volume 10, pp. 71-92.

Tsitsika, E.V., C.D. Maravelias and J. Harala-bous, 2007. Modeling and forecasting pelagic fish production using univariate and multivariate ARIMA models. *Fisheries Science,*73:979-988.

Venugopalan, R., and M. Srinath, 1998. Modeling and forecasting fish catches: Comparison of regression, univariate and multivariate time series methods, *Indian Journal of Fisheries,* 45(3): 227-237.


*******