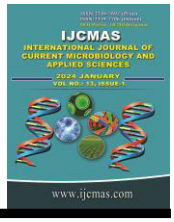
Int.J.Curr.Microbiol.App.Sci (2024) *13*(01): 53-60

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 13 Number 1 (2024)**
Journal homepage: http://www.ijcmas.com



Original Research Article https://doi.org/10.20546/ijcmas.2024.1301.006

Application of Stochastic Model in the Production of Sugarcane in India

S. R. Krishna Priya and Kausalya Nataraj*

Department of Statistics, PSG College of Arts and Science, Coimbatore, India

**Corresponding author*

Keywords

Sugarcane Production, ACF, PACF, ARIMA, Box test

Article Info

Received:
20 November 2023
Accepted:
26 December 2023
Available Online:
10 January 2024

A B S T R A C T

Forecasting is an essential tool to estimate the future trend of any crop shortly. There are various techniques in the present scenario for predicting future figures and Auto Regressive Integrated Moving Average (ARIMA) is one among them. Sugarcane is an imperative crop in India, keeping in view its importance for many areas of the country and its diverse uses. The present study was intended to check and identify the best forecasting model of sugarcane production in India using historical data between the years 2001 to 2020, based on the estimation of a suitable ARIMA model. The analysis of ACF & PACF of different series revealed that ARIMA was the most suitable model for forecasting based on diagnostics, such as ACF, PACF, and AIC. The selected ARIMA model predicted the sugarcane production for the immediate 10 years from 2021.

Introduction

Sugarcane is a giant grass full of sugar originated on the island of Papua New Guinea and followed the movements of people within the Pacific Ocean region, reaching Oceania, Southern Asia, Southern China and the Indus Valley in India, then it was in India that the history of sugar began. It belongs to the *Graminaceae* family and to the botanical genus Saccharum, which comprises three sugar-bearing species are *S. officinarum*, known as "Noble Cane", *S. sinense*, and *S. barberi* and three non-sugar-bearing species are *S. robustum, S. spontaneum* and *S. edule*. In the 1880s, agronomists began creating hybrids between noble cane and the other species.

The modern varieties are all derived from those crosses. India is the second largest sugar producer in the world, the first one being Brazil. Among Indian states, Uttar Pradesh is the top producer followed by Maharashtra and Karnataka standing on second and third positions respectively. In India, sugarcane is produced in both tropical and subtropical regions.

Sankar *et al.,* (2012) forecasted average weekly temperature of Tricomalee station in Sri Lanka, the ARIMA model is highly recommended. ARIMA (1,1,1) is identified as the best fitting time series model for temperature forecasting, and found to be highly effective by producing better results due to Root Mean Square Error and Mean Absolute Error. Diagnostic tests was performed for the verification of developed model.

Hossain *et al.,* (2016) focused on fitting an ideal model to predict potato production in Bangladesh between 1971 and 2013 using current model selection criteria such as AIC, BIC, etc. Box-Jenkins ARIMA (0,2,1) is a

particularly suitable model for evaluating the potato market in the entire Bangladesh. In addition, the histogram with normal curve probably indicates that the residuals of the fitted ARIMA(0,2,1) model are normally distributed, indicating that the predicted model is a more adequate fit for predicting potato production in Bangladesh.

Gopinath *et al.,* (2021) made the future pattern for the commercial crop groundnut from 2019 to 2040 by using ARIMA (1, 1, 15) model by involving observation in between 1951-2018 in India and concluded that there will be gradual increase in yield of commercial crops by taking the necessary steps.

Umar *et al.,* (2021) identified the best prediction model of millet yield in Andhra Pradesh using ARIMA. 56 years of time series data have been used as input for prediction. The ARIMA (1, 1, 1) model was selected as the best fitted model for sugarcane yield up to 2024 and observed that there was a growing production.

Materials and Methods

Source

Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare

Of the time series data on production of sugarcane for a period of 20 years (2001-2020) was used for the model. The time series model were applied on the data. In this study, Box and Jenkin's Methodology ARIMA model and systematically computed Auto ARIMA was used and finally observed the best appropriate fitting model.

For this study, detection of outlier and residuals analysis were used for the process of building time series models. The statistical software R programming language is used for modelling and forecasting sugarcane production of India.

Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF)

To check the stationarity of the raw data, the ACF and PACF were highly recommended. The autocorrelation function (ACF) and partial autocorrelation function (PACF) is a statistical technique that can use to identify how correlated the values in a time series are with each other. The ACF and PACF plots the correlation coefficient against the lag, which is measured in terms of a number of periods or units.

Augumented-Dickey Fulleer (ADF) Test

The unit root test is carried out by using the Augumented-Dickey Fuller test. The hypothesis of ADF test was time series is stationary and alternative hypothesis is non stationary.

$$\Delta\lambda_t = \alpha_0 + \alpha_2 t + \beta\Delta\lambda_{t-1}k_{i-1} + \varepsilon_t$$

Where λ_t indicates the yearly sugarcane production at time *t*, β is the coefficient to be estimated; *k* is number of lagged terms, α_2 is the estimated coefficient of the trend, α_0 is the constant and ε is the error term.

Box-Jenkins Methodology

In an Auto Regressive Integrated Moving Average (ARIMA) the future value of a time series assumed to be a linearly and past observations are random errors. The time series has the form,

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \ldots + \theta_p Y_{t-p} - \emptyset_1 \epsilon_{t-1} - \emptyset_2 \epsilon_{t-2} \cdots - \emptyset_q \epsilon_{t-q} + \epsilon_t$$

Y_t is the time series, p is number of autoregressive parameters, q is number of moving averages. The Box-Jenkins (1976) methodologies were applied forecasting model. For this study, sugarcane production be a discrete time series variables which takes different variables over a period of time.

Box-Ljung Test

The Box-Ljung test (1978) is a diagnostic tool used to testing the lack of goodness of fit of a time series model. It is applied to the residuals of a time series after fitting an ARMA(*p,q*) model to the data.

The test examines autocorrelations of the residuals. In this case, the significance value of the test is less than 0.05, suggesting that the assumptions are not met and the model results are therefore suspect. Box's M is robust to large datasets, which implies that it can spot small

departures from homogeneity when there are large numbers of cases.

Results and Discussion

The first and foremost step is plotting the Raw Data from the year 2001 to 2020given in the figure 1

For the study, Autocorrelation function and Partial autocorrelation functions was used for checking the time series data for stationarity is shown in figure 2.

The study here employs the unit root test (ADF test) to examine the time series data is stationary or not.

From the results, the *p* value is higher than the 0.05, hence it is inferred that the data is non-stationary. To achieve the non-stationary time series data, differencing process can be attempted in this stage. The time series data after first differencing is shown in figure 3.

After first differencing, ACF, PACF, ADF tests were employed to test the stationarity.

For a second time, the Autocorrelation and Partial autocorrelation functions was used for checking the first differenced time series data for stationarity is shown in figure 4.

The ACF and PACF plots of first differenced time series data are shown and it is concluded that the data is Non-Stationary. The diagnosis test is shown in the below table 2.

After first differencing, yet again the predicted *p*-value in greater than 0.05 and it represents the present form of India sugarcane data is non-stationary, hence the second differencing is needed to achieve the stationarity and is shown in figure 5.

Again, the Autocorrelation function and Partial autocorrelation functions was used for checking the second differenced time series data for stationarity is shown in figure 6.

The diagnosis test is for the second differenced time series data shown in the below table 3.

After second differencing the *p*-value is significantly reduced to 0.03 from 0.10 which means the present form of time series data is stationary and equipped for forecasting.

ARIMA Model Identification

The model can be identified through ACF and PACF plots of second differenced India's sugarcane production data as shown in figure 6. It gives the idea to find out the Autoregressive AR (*p*) and Moving Average (*q*) parts of the model for ARIMA forecasting.

It is considered that the autoregressive part and moving average part may be 1 and 1 that will implement in the further forecasting process. It is already known that to attain the stationarity, the data have been transformed by second differencing method so, the integrated part of the model would be 2. Finally, it is identified that the *p, d, q* value of the ARIMA model is 1, 2, 1 and it is suitable for forecasting the future trend.

The predicted ARIMA is shown in below table 4.

The ARIMA (1,2,1) results such as, AIC, MAPE, RMSE and MAE values are given in the above table.

There is a high density value bar near the value zero and from this result it is confirmed that the predicted ARIMA model is suitable for forecasting. The Residual Histogram Plot for the second differenced data for the predicted ARIMA (1,2,1) model is existing in Figure 7.

It can be accurately quantify the future forecasted values for the next 10 years Table 5.

Box Test for Validation

The box test is the final validation for the predicted model. The value of *p* is gradually decreased by increasing the lag order. This confirms our predicted model a good fit for forecasting is shown in Table 6.

In our study, the developed model for production of sugarcane was found to be ARIMA (1,2,1). From the forecast available using the developed model, it can be seen that forecasted production of sugarcane keeps decreasing in the upcoming years.

The model can be helpful for researchers for predicting production of sugarcane in India. The government sectors, farmers should concentrate in taking proper steps to avoid lack of production and get benefited by the study. However, it should be updated from time to time with incorporation of current data.

Table.1 Diagnosis test of the model

Diagnosis Test	Value	*p*-value
Augumented-Dickey Fuller test (Non-Stationary)	-2.5592	0.3594 (p<0.05)

Table.2 Diagnosis test of the model after first differencing

Diagnosis Test	Value	*p*-value
Augumented-Dickey Fuller test (Non-Stationary)	-3.2142	0.1098 (p<0.05)

Table.3 Diagnosis test of the model after second differencing

Diagnosis Test	Value	*p*-value
Augumented-Dickey Fuller test (Stationary)	-3.8057	**0.03531**(p<0.05)

Table.4 Predicted Arima (1,2,1)

Model	AIC	MAPE	RMSE	MAE
ARIMA (1,2,1)	320.4	9.336456	37575.81	28636.92

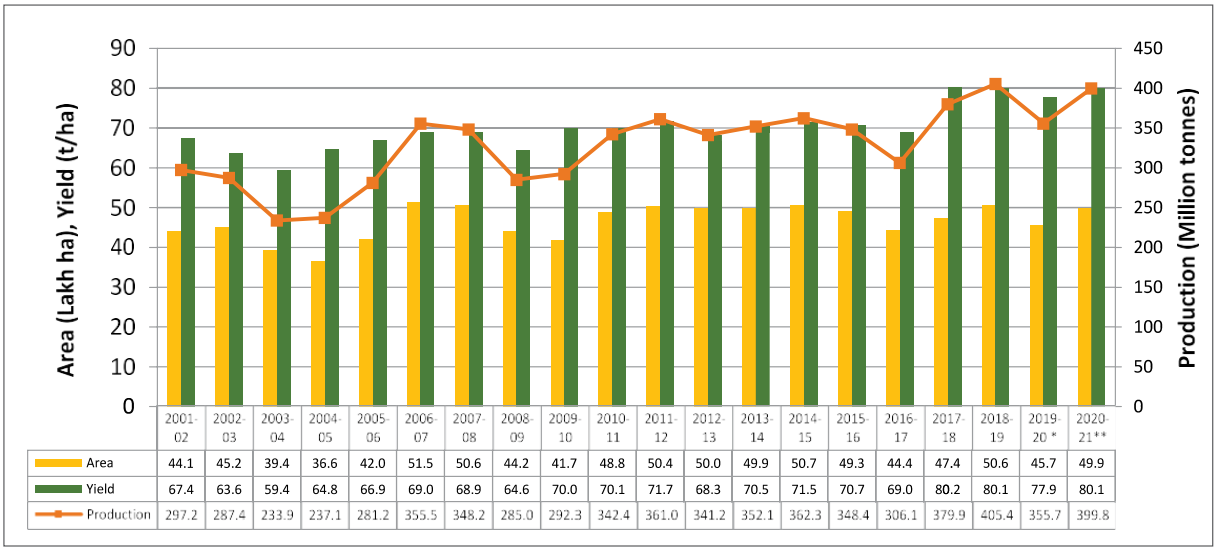
Table.5 The Forecasted India's sugarcane production for upcoming 10 years.

Years	Point Forecast	Low 80%	High 80%	Low 95%	High 95%
2021	237102.24	-413475.23	887679.7	-757870.14	1232074.6
2022	219691.66	-591659.77	1031043.1	-1021163.28	1460546.6
2023	202278.08	-781230.77	1185786.9	-1301868.91	1706425.1
2024	184865.88	-981415.91	1351147.7	-1598808.20	1968540.0
2025	167453.04	-1191664.58	1526570.7	-1911137.98	2246044.1
2026	150040.50	-1411465.88	1711546.9	-2238077.42	2538158.4
2027	132627.82	-1640395.83	1905651.5	-2578977.86	2844233.5
2028	115215.20	-1878074.87	2108505.3	-2933258.90	3163689.3
2029	97802.56	-2124170.30	2319775.4	-3300411.68	3496016.8
2030	80389.92	-2378383.19	2539163.0	-3679979.06	3840758.9
2031	62977.29	-2640444.98	2766399.6	-4071550.29	4197504.9

Table.6 The Box Test for the forecasted ARIMA(1,2,1) model

Lags	5	6	7	8	9	10
X-squared	10.6	11.668	11.744	11.862	12.155	12.675
df	3	3	4	4	4	4
***p* - value**	0.0141	0.008611	0.01936	0.01841	0.01623	0.01298

Figure.1 The time series were obtained from www.indiastat.com website.



	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20 *	2020-21**
Area	44.1	45.2	39.4	36.6	42.0	51.5	50.6	44.2	41.7	48.8	50.4	50.0	49.9	50.7	49.3	44.4	47.4	50.6	45.7	49.9
Yield	67.4	63.6	59.4	64.8	66.9	69.0	68.9	64.6	70.0	70.1	71.7	68.3	70.5	71.5	70.7	69.0	80.2	80.1	77.9	80.1
Production	297.2	287.4	233.9	237.1	281.2	355.5	348.2	285.0	292.3	342.4	361.0	341.2	352.1	362.3	348.4	306.1	379.9	405.4	355.7	399.8

Figure.2 Raw data plot of time series data

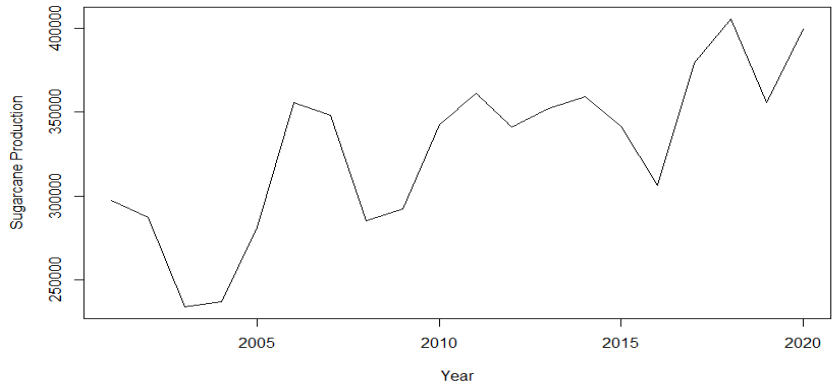
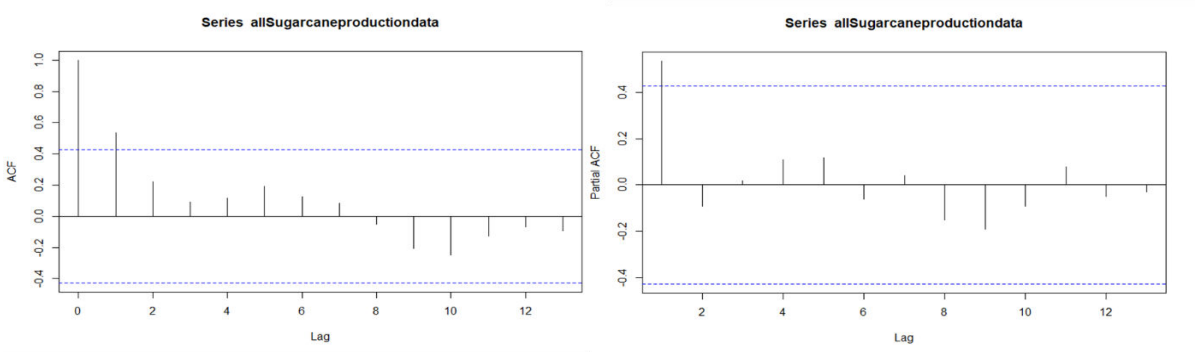


Figure.3 ACF and PACF of time series data



From the above plots, it is concluded that the data is Non-Stationary.

Figure.4 First-differenced time series of India's sugarcane production from 2001 to 2020

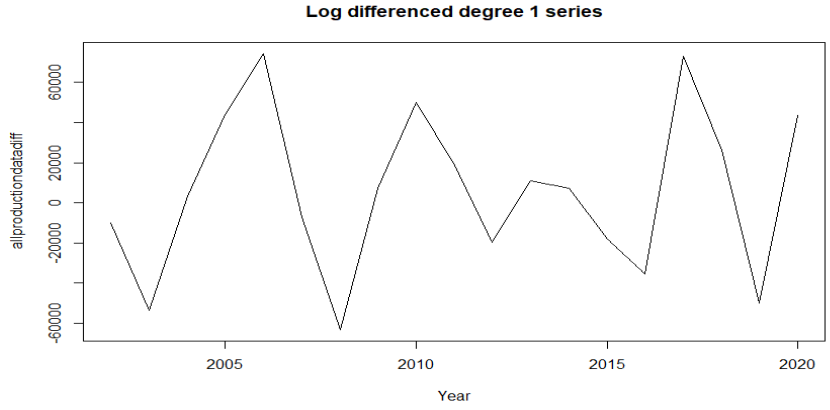


Figure.5 ACF and PACF of first differenced time series data

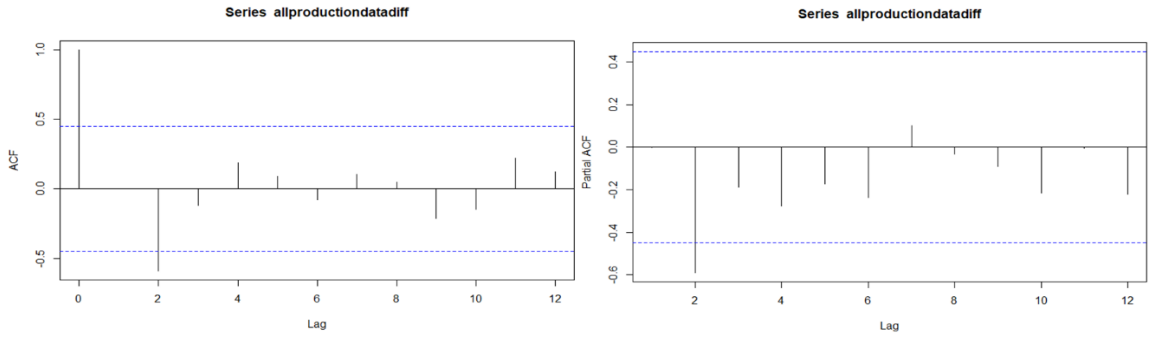


Figure.6 Second-differenced data of India sugarcane production from 2000-2020

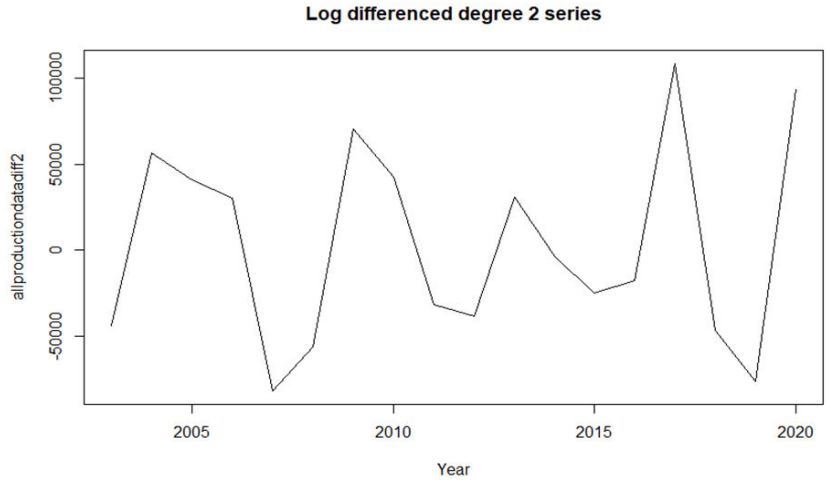


Figure.7 ACF and PACF of second differenced time series data

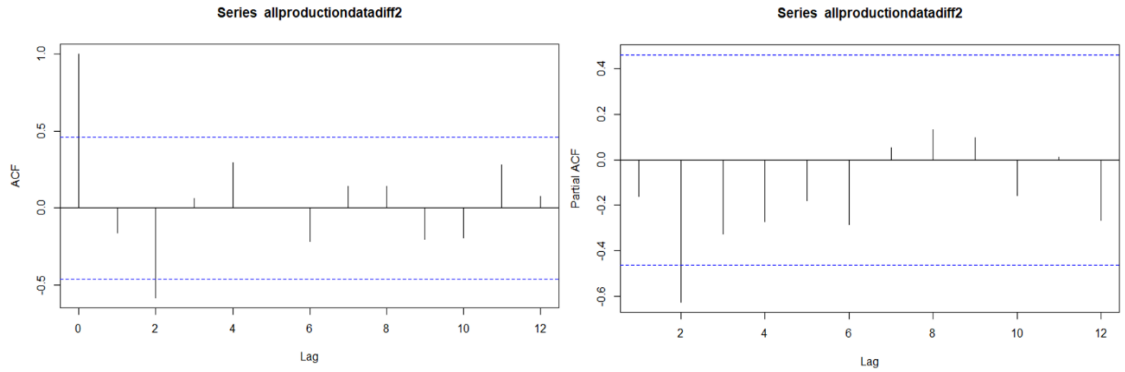
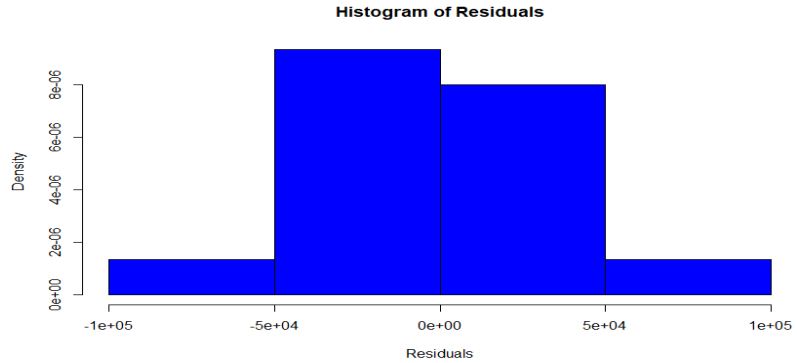


Figure.8 Residual Histogram Plot



Author Contribution

S. R. Krishna Priya: Investigation, formal analysis, writing—original draft. N. Kausalya: Validation, methodology, writing—reviewing.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Research Funding: Not applicable

Ethical Approval: Not applicable.

Consent to Participate: Not applicable.

Consent to Publish: Not applicable.

Conflict of Interest: The authors declare no competing interests.

References

Box, G. E. P., G. M. Jenkins. Time Series Analysis, Forecasting and Control. San Francisco, Holden-Day, California, USA, 1976.

Badmus, M., Ariyo, O., Badmus, M. A., Ariyo, O. S. Forecasting Cultivation Area and Production of Maize in Nigeria Using ARIMA Model. *Asian J. Agric. Sci.*(2011), *3* (3), 171–176.

Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare, Government of India.

Gopinath, M., Naveenapriyaa, M., Sindhu, T., Abinaya, K., &Prathiskaaarthi, S. (2021). Production of commericial crop prediction using arima model. *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal NVEO*, 4985-4998.

Hossain, M. M., & Abdulla, F. (2016). Forecasting potato production in Bangladesh by ARIMA model. *J Adv Stat*, *1*(4), 191-198.

Iqbal, N., Bakhsh, K., Maqbool, A., Abid Shohab, A. Use of the ARIMA Model for Forecasting Wheat Area and Production in Pakistan. *J. Agric. Soc. Sci.*(2005), *2* (1), 120–122.

Sankar, T. J., Prabakaran, R. Forecasting Milk Production in Tamilnadu. *Int. Multidiscip. Res. J.* (2012), *2* (1), 10–15.

ShaikNafeez Umar, T., Gangaram, O., HariBabu, G., Sathyanarayana Reddy, B., Ramana Murthy and Samreen Aalia, P. (2021). Application of Arima Models in Millet Production in Andhra Pradesh. *Int.J.Curr.Microbiol.App.Sci.* 10(7):525-531. https://doi.org/10.20546/ijcmas.2021.1007.057

How to cite this article:

Krishna Priya, S. R. and Kausalya Nataraj. 2024. Application of Stochastic Model in the Production of Sugarcane in India. *Int.J.Curr.Microbiol.App.Sci.* 13(01): 53-60. **doi:** https://doi.org/10.20546/ijcmas.2024.1301.006