JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Foreseeing Saccharum Officinarum Yield in Andhra Pradesh

[1]S R KRISHNA PRIYA AND [2]N KAUSALYA
[1]Assistant Professor, [2]Research Scholar
Department of Statistics,
PSG College of Arts & Science, Coimbatore, India.

Abstract: This study used a time series modeling approach (Box-Jenkins' ARIMA model) to forecast sugarcane production in Andhra Pradesh. The best ARIMA model order was found to be (2,0,0). Efforts were then made to forecast, as accurately as possible, the future production of sugarcane for a period of up to ten years by fitting the ARIMA (2,0,0) model to time series data. The forecast results showed that the annual sugarcane yield will grow from 2018–2023 and then will continuously follow a moderate spectrum the year 2030, neither up nor down. Demonstrating that the fitted model is statistically well-behaved to anticipate Saccharum Officinarum yield in Andhra Pradesh are the forecasted series, which are also depicted.

Keywords: Saccharum Officinarum yield, Box-Pierce Test, Mean Absolute Percentage Error, Box test, ARIMA

I. INTRODUCTION

Sugarcane is known by its botanical name, Saccharum officinarum, and this grass plant belongs to the Poaceae family and Saccharum genus. In addition to making a substantial financial contribution to the national coffers, it directly or indirectly supports the employment of over a million people. The countries in the world that grow sugarcane are located in tropical and subtropical zones, stretching from latitude 36.7° north to 31.0° south of the equator. The origins of sugar cane can be traced back thousands of years to New Guinea. Along the paths of human migration, sugar cane plants expanded across Asia and the Indian subcontinent. To create the commercial sugar cane which exists today, it crossed-bred with some wild relatives in this area. 196 000 hectares of sugarcane (composite) are grown in Andhra Pradesh, yielding 156 lakh tonnes of crop. This portion of the production is useful for making sugar, for making seeds, and for making jaggery. The average yield per acre in Andhra Pradesh (Composite) is only 28 tons. Sugarcane data from Andhra Pradesh indicated that the area's yield in 2022 was 0.050 hectares. Compared to the earlier figure of 0.060 ha mn for 2021, this is a reduction.

Varieties: The ruling varieties of sugarcane in Andhra Pradesh.

- Early varieties: Co.6907, 84A125, 81A99, 83A30, 85A261, 87A298, Co.8014, 86V96, 91V83.
- Mid-late Varieties: COA7607, CO8021, COT.8201, Co7805, COV92102 (83V15), 83V288.
- Late varieties: Co.7219, CoR8001, 87A380, Co7706

II. REVIEW OF LITERATURE

Numerous studies have been conducted by researchers to adapt an ARIMA model for various types of agricultural crops in the agriculture sector worldwide. Several agricultural sectors use the ARIMA model to anticipate agricultural production. The pertinent research on utilizing the Box-Jenkins (1970) ARMA model for forecasting, which provides insight into forecasting methodologies are listed below,

- Badmus et al., (2011) employs the Autoregressive Integrated Moving Average (ARIMA) model to forecast the area under cultivation and maize production in Nigeria. The study made use of time series data spanning from 1970 to 2005. The National Bureau of Statistics (NBS), International Financial Statistics (IFS) reports, and the Central Bank of Nigeria (CBN) were the sources of the data. The outcome also indicates that the amount of maize produced in 2020 is expected to be approximately 9952.72 tons, with 6479.8 and 13425.64 thousand tons being the upper and lower bounds, respectively. Additionally, according to the estimate, by 2020, the area dedicated to maize would be 9229.74 thousand hectares, with lower and maximum bounds of 7087.67 and 11371.81 thousand hectares, respectively. This projection is significant because it contributes to the development of sound policies regarding the nation's relative production, price structure, and consumption of maize. The result from the study is that, total cropped area can be expanded in future, if land reclamation and conservation measures are applied.

- Padhan (2012) used the ARIMA model in this work to predict the annual productivity of a particular agricultural product. A collection of thirty-four distinct goods have been considered for empirical research, pending the availability of the necessary data. Forecasted values for an additional five years since 2011 have been derived by applying annual data from 1950 to 2010. The model's

validity is confirmed using a number of model selection criteria, including lowest MAPE values, minimum AIC, and adj R2. Cardamom has the lowest AIC values while tea has the lowest MAPE values among the chosen crops.

- Biswas et al., (2014) was used the time-series approach in this study to assess the area, productivity, and production statistics of wheat for Punjab. Data on annual wheat yield from 1950–1951 to 2009–2010 was utilized as an input to project the yield through 2020–21.The yield was predicted using the Box-Jenkins ARIMA technique. Standard statistical methods were used to assess the model's validity. Data spanning the previous 60 years showed that wheat yield rose, rising from 0.8 t ha-1 in 1950–1951 to 4.3 t ha-1 in 2009–2010. According to the estimate, Punjab's production would rise by 15.3% between 2020 and 21. Punjab's production is projected to rise from 15844.7 thousand tons in 2010–11 to 18271.7 thousand tons in 2020–21, based on ARIMA.

- Kumar et al., (2014) employed the Box-Jenkins' ARIMA model, a time series modeling technique, to forecast sugarcane production in India. The optimal ARIMA model was determined to be in the order (2,1,0). Additionally, by fitting the ARIMA (2,1,0) model to our time series data, an attempt was made to anticipate the future sugarcane production as accurately as possible for a maximum of five years. According to the forecast results, the annual production of sugarcane is expected to increase in 2013 before experiencing a significant decline in 2014. From 2015 through 2017, the production is expected to grow at an average annual growth rate of almost 3%.

III. THE MISSION OF THE STUDY

The primary goal of this research is to create an Autoregressive Integrated Moving Average (ARIMA) model for Saccharum Officinarum crop yield forecasting in Andhra Pradesh.

IV. DATA SOURCE AND USED SOFTWARE

The agricultural data sets can be accessed at the *www.indiastat.com* website. R, an open-source software based on statistical programming, was used to conduct this research in its entirety. *Forecast, tseries, and auto.arima* are other library packages utilized in the analysis.

V. METHODS AND METHODOLOGY

A time series is a set of numbers that measures the status of some activity over time. It is the historical record of some activity, with measurements taken at equally spaced intervals with a consistency in the activity and the method measurement.

Moving Average Process

Moving average models were first considered by Slutsky (1972) and Wold (1938). The moving average series can be written as

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} - \cdots - \theta_q e_{t-q} \quad (1)$$

Such a series a moving average of order q and abbreviate the name to **MA (q)**, where Y_t is the original series and the e_t is the series of errors.

Auto-Regressive Process

Yule (1926) carried out the original work on autoregressive process are as their name suggests regression on themselves. Specifically, a p[th] order autoregressive process $\{Yt\}$ satisfies the equation

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \Phi_3 Y_{t-3} + \cdots + \Phi_p Y_{t-p} + e_t \quad (2)$$

The current value of the series Y_t is a linear combination of the p most recent past values of itself plus an "innovation" term e_t that incorporates everything new in the series at time t that is not explained by the past values. Thus, for every t, it can be assumed that e_t is independent of Y_{t-1}, Y_{t-2}, Y_{t-3}, ... Y_{t-q}

Autoregressive Integrated Moving Average (ARIMA) Model

The Box-Jenkins (1970) procedure is the milestone of the modern approach to time series analysis. Given an observed time series, the aim of the Box and Jenkins procedure is to build an ARIMA model. In particular, passing by opportunate preliminary transformations of the data, the procedure focuses on stationary process.
Then, it is tried to fit the Box-Jenkins Autoregressive Integrated Moving Average model. This model is the generalized model of the non-stationary ARMA *(p,q)* can be written as,

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (3)$$

VI. RESULTS AND DISCUSSION

Plotting the dataset across time was necessary to identify the type of trend because it is a Univariate time series. In Figure 1, the series data plot is shown.

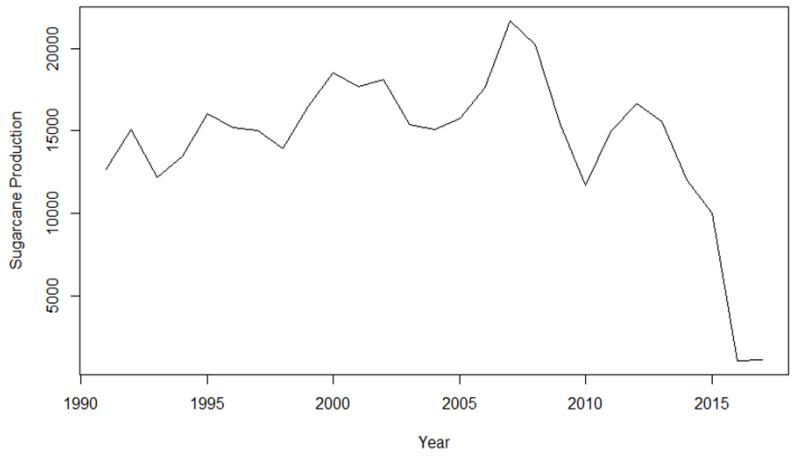

Figure 1. Raw Data Plot of Saccharum Officinarum yield from 1991-2017

The above plot doesn't clearly shows whether the data is not stationary or not. To obtain the decision, the Box-Pierce test, one of the stationarity tests has been employed to the data.

Table 1. Box-Pierce Test

Data	X-Squared	df	*p*-value
Yield data	11.256	1	0.0007935

The *p*-value is less than 0.05 hence the present form of time series data is stationary it can be used for forecasting the future trend. Then, the auto.arima function has been used in the R software for predicting the proper ARIMA *(p,d,q)* model for forecasting.

Table 2. Different ARIMA models

Model	MAPE	RMSE
ARIMA (1,0,1)	45.24856	2668.247
ARIMA (1,1,1)	49.33914	2755.06
ARIMA (0,1,1)	49.26104	2757.506
ARIMA (1,1,0)	48.13093	2769.727
ARIMA (2,0,0)	**43.91877**	**2664.07**

From the tentative analysis, the best selected ARIMA model to forecast the sugarcane yield in Andhra Pradesh is ARIMA (2,0,0) with the MAPE = 43.91877, which refers as the low MAPE shows the best fit than other models. Validation Analysis for Predicted ARIMA (2,0,0)

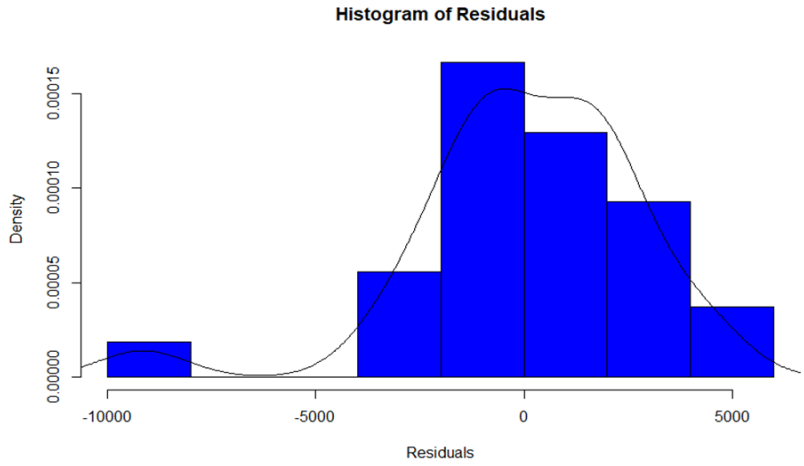

Figure 2. Histogram of Residuals for predicted ARIMA (2,0,0)

The above plot assured that the fitted model for forecasting is absolutely appropriate through the curve which follows normal distribution and the bar in the histogram near the value zero is high in density than the rest of the bars. The ARIMA (2,0,0) is best suitable for forecasting future values of sugarcane yield data, according to the present evaluation from the above.

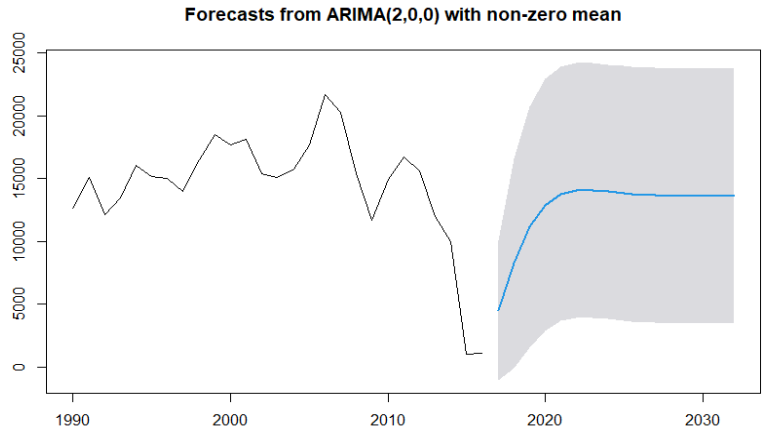

Figure 3. Forecasted ARIMA (2,0,0) Model for the Saccharum Officinarum yield from 2018-2030

The Summary Statistics of the predicted ARIMA (2, 0, 0) model is given in the table 3 & 4 and the forecasting criteria of the fitted model is shown in the table 5.

Table 3. Summary Statistics-I

Coefficients	Estimates	Std.Error
ar1	1.1146	0.1789
ar2	-0.3831	0.2317
Mean	13624.504	2065.062

Table 4. Summary Statistics-II

Model	Sigma^2	log likelihood	AIC	AICc	BIC
ARIMA (2,0,0)	7984429	-251.96	511.92	513.74	517.1

Table 5. Error Measures of the ARIMA (2,0,0)

Error Measures	ME	RMSE	MAE	MAPE	MASE
ARIMA (2,0,0)	45.16387	2664.07	1987.196	43.91877	0.9008768

From all the results, it can be precisely quantify the future forecasted values with the confidence interval 95% for the next 10 years and it is shown in Table 6.

Table 6. Forecasted values for the ten years from 2018-2030

Years	Point Forecast	Lo 95	Hi 95
2018	8257.757	-35.46697	16550.98
2019	11132.640	1570.96642	20694.31
2020	12902.856	2899.46024	22906.25
2021	13774.694	3666.22179	23883.17
2022	14068.348	3947.80167	24188.89
2023	14061.686	3941.13809	24182.23
2024	13941.772	3819.32561	24064.22
2025	13810.666	3685.91098	2335.42
2026	13710.468	3584.35436	23836.58
2027	13649.008	3522.38022	23775.64
2028	13618.886	3492.13421	23745.64
2029	13608.856	3482.08949	23735.62
2030	13609.241	3482.44770	23735.98

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 table 7.

Table 7. The Box Test for the forecasted ARIMA (2,0,0) model.

Lags	10	11	12	13	14	15
X-squared	6.3581	6.5451	6.8132	7.1894	9.5991	9.7224
df	1	1	1	1	1	1
p - value	0.01168	0.01052	0.009049	0.007333	0.001947	0.00182

VII. CONCLUSION

The current study's attempt to model and forecast Saccharum Officinarum yield in Andhra Pradesh was conducted using the Autoregressive integrated moving average (ARIMA) model, which is regarded as one of the finest models. The data's trend was determined by estimating the Box-Pierce test, and the auto.arima function was then used to build the ARIMA model (2,0,0). According to the fitted model, Saccharum Officinarum yield will rise throughout the following six years, from 2018 to 2023. Researchers, legislators, sugar growers, and businesspeople throughout the state of Andhra Pradesh could all benefit from using these models to inform their decisions.

VIII. REFERENCES

[1] Badmus, M. A., & Ariyo, O. S. (2011). Forecasting cultivated areas and production of maize in Nigerian using ARIMA Model. *Asian Journal of Agricultural Sciences*, *3*(3), 171-176.

[2] Padhan, P. C. (2012). Application of ARIMA model for forecasting agricultural productivity in India. *Journal of Agriculture and Social Sciences*, *8*(2), 50-56.

[3] Biswas, B., Dhaliwal, L. K., Singh, S. P., & Sandhu, S. K. (2014). Forecasting wheat production using ARIMA model in Punjab. *International Journal of Agricultural Sciences*, *10*(1), 158-161.

[4] Kumar, M., & Anand, M. (2014). An application of time series ARIMA forecasting model for predicting sugarcane production in India. *Studies in Business and Economics*, *9*(1), 81-94.

[5] Madlul, N. S., Al-Najjar, E. Y., Baker, Y. T., & Irhaim, F. I. (2020). USING THE ARIMA MODELS TO PREDICT WHEAT CROP PRODUCTION IN IRAQ. *International Journal of Agricultural & Statistical Sciences*, *16*(1).

[6] Prabakaran, K., Sivapragasam, C., Jeevapriya, C., & Narmatha, A. (2013). Forecasting cultivated areas and production of wheat in India using ARIMA model. *Golden Research Thoughts*, *3*(3), 2831-2831.

[7] Ajetomobi, J. O., & Olaleye, A. O. (2019). Auto-regressive integrated moving average (ARIMA) modeling of cocoa production in Nigeria: 1900-2025. *Journal of Crop Improvement*, *33*(4), 445-455.

[8] Hamjah, M. A. (2014). Forecasting major fruit crops productions in Bangladesh using Box-Jenkins ARIMA model. *J. Econ. Sustain. Dev*, *5*(7).

[9] Mishra, P., Yonar, A., Yonar, H., Kumari, B., Abotaleb, M., Das, S. S., & Patil, S. G. (2021). State of the art in total pulse production in major states of India using ARIMA techniques. *Current Research in Food Science*, *4*, 800-806.

[10] Rahman, N. M. F. (2010). Forecasting of boro rice production in Bangladesh: An ARIMA approach. *Journal of the Bangladesh Agricultural University*, *8*(452-2016-35761).

[11] Jadhav, V., CHINNAPPA, R. B., & Gaddi, G. M. (2017). Application of ARIMA model for forecasting agricultural prices.

[12] Singh, L. N., Darji, V. B., & Parmar, D. J. (2015). Forecasting of wheat production and productivity of Ahmedabad region of Gujarat state by using ARIMA models. *Indian Journal of Economics and Development*, *3*(6), 1-7.

[13] Lwaho, J., & Ilembo, B. (2023). Unfolding the potential of the ARIMA model in forecasting maize production in Tanzania. *Business Analyst Journal*, *44*(2), 128-139.

[14] Rahman, A. (2017). Modeling of Tea Production in Bangladesh Using Autoregressive Integrated Moving Average (ARIMA) Model. *Journal of Applied & Computational Mathematics*, *6*(2), 349-357.

[15] Darekar, A., & Reddy, A. A. (2017). Cotton price forecasting in major producing states. *Economic Affairs*, *62*(3), 373-378.