Prognosis of Sugarcane Yield in Coimbatore District

S R Krishna Priya[1], N Kausalya[2]
Assistant Professor[1], Research Scholar[2]
Department of Statistics, PSG College of Arts & Science, Coimbatore-641014.

Abstract:-Employment and the Indian economy are heavily reliant on agriculture. Accurate yield forecasting is essential for developing and structuring agricultural plans. This paper aims to evaluate the stability of sugarcane yield and project future yields. Secondary data was obtained from the Season and Crop Report, Directorate of Economics and Statistics, Government of Tamil Nadu, on sugarcane yields from 1960-1961 to 2018-2019. After examining two models, it was determined that ARIMA (2,1,1) is the most effective time series model for predicting sugarcane yield. The selected ARIMA model shows better results in terms of Root Mean Square Error, Mean Absolute ScaledError and Mean Absolute Percentage Error. There is a significant increase between the period 2022-2024. The remaining years had a moderate level of yield. This study has the potential to assist farmers, agriculturalists, and government sectors in making informed decisions to proactively address agricultural challenges.

Keywords: *Sugarcane Yield, Univariate data, R software.*

1. Introduction

Sugarcane cultivation in India dates back to 5000 BC and underwent genetic improvement through breeding in the 19th century. It is a valuable cash crop due to its strategic location and multiple domestic and industrial uses, promoting nutritional and economic sustainability. In India, varietal development was initiated to breed varieties suitable for the sub-tropical climate. India is the second-largest sugar producer globally, with a 17% market share. It is currently a sugar surplus country and is expected to produce a record 38 million tonnes of sugar in 2022. The sugarcane agriculture sector employs approximately 6 million farmers and a significant number of agricultural labourers. Additionally, the sugar industry is India's largest agro-processing industry, employing nearly half a million skilled and semi-skilled workers, most of whom are from rural areas. Sugarcane yields can be affected by various factors. Accurate information on crop yield history is crucial for forecasting and managing agricultural risks, as well as for ensuring national food security. This includes early determination of import/export schedules and prices. This paper aims to forecast the sugarcane yield in Coimbatore district for the following ten years (2019-2029) using time series univariate data analysis applied to R software. The forecast will assist researchers, government and non-government agencies in monitoring and analysing the potential yield of Coimbatore, enabling them to take adequate measures to improve the economy.

There are a few evidences available related to sugarcane yield and its forecast, but a vast research has been conducted with this regard in other parts of the world. Here are some discussions:

Iqbal *et al.*,(2005)forecast the future trend for the period 2021-2022 using data on wheat production and area in Pakistan for the period 2000-2020. The ARIMA (1,1,1) and (2,1,2) models were identified as the most appropriate to predict area and wheat productivity in Pakistan on the basis of ACF, PACF and differencing patterns. The time series graph of the residuals of the yield and area data exhibited a scattering pattern, validating the selection of an adequate model and confirming that the models were well fitted. In addition, the residuals plotted over the fitted measures of wheat area and productivity did not show any regular patterns indicating goodness of fit. A forecast of wheat production for 2000-2001 was made with lower and upper bounds of 19006.8 and 22334.0 thousand tonnes respectively. The above model was used to estimate the future trend of wheat production. With lower and upper bounds of 27542.1 and 32007.6 thousand tonnes, respectively, the projected wheat production for 2022 would be about 29774.8 thousand tonnes. The availability of appropriate inputs, farmer education and training, soil reclamation and conservation and, most importantly, government policies that promote wheat cultivation in the country are the keystones for increasing area and productivity.

Badmus *et al.*,(2011)applied the Autoregressive Integrated Moving Average (ARIMA) model to forecast maize area and production in Nigeria. The study used time series data from 1970 to 2005. The data originated from the National Bureau of Statistics (NBS), International Financial Statistics (IFS) reports and the Central Bank of Nigeria (CBN). The ARIMA (1,1,1) and (2,1,2) fitted with the area and maize production models were determined using the S-plus statistical difference. Further, the result of the study also disclosed that the total quantity of maize produced in 2020 is expected to be about 9952.72 tonnes, with 6479.8 and 13425.64 thousand tonnes as the upper and lower bounds, respectively. In addition, the area under maize in 2020 is estimated to be 9229.74 thousand hectares, with lower and upper bounds of 7087.67 and 11371.81 thousand hectares, respectively. The importance of this forecast is that it provides valuable insights into the relative production, cost structure and consumption of maize in the country. The study concludes that with the adoption of land reclamation and conservation measures, the total area under cultivation could be expanded in the future.

Sankar*et al.*, (2012)focused on predicting milk production by using 1978 to 2008 data processes such as AR, MA and ARIMA have been considered into the study for choosing the better stochastic model for the prediction. ARIMA (1,1,0) was found to be the most appropriate model for predicting milk production based on the estimation of ARIMA involving components such as Auto Correlation Function, Partial Auto Correlation Function, Normalized BIC, Box-LJung Q statistics and Residuals. By the end of the study, it shows that the milk production would increase to 7.15 million tonnes in 2015 and 5.96 million tonnes in 2008 in Tamil Nadu and this could be useful for policymakers to formulating strategies for augmenting and sustaining milk production in state.

Biswas *et al.*,(2013)used data on gross area and annual production of rice in West Bengal from 1947-48 to 2007-08 to predict the future trend. The suitable ACF, PACF and number of differencing procedures were used to identify the more appropriate AR, MA and I terms. The results indicate that the ARIMA (2,1,3) model best fitted the gross area series, while the ARIMA (2,1,1) model best fitted the production series. The model shows a good level of accuracy in projecting the state's rice production and area in the future. Furthermore, the predicted gross area and production have a MAPE of 2.53 and 7.6 respectively, indicating a higher level of prediction accuracy. In conclusion, the model shows a reasonable level of accuracy for predicting rice production and area in the future across West Bengal.

Tripathi*et al.*, (2014) used Univariate ARIMA model for forecasting rice production, area and productivity of Odisha from the historical data of 1950-1951 to 2008-2009. The significant spikes in the plots of PACF and ACF of different time series were used for identifying the Autoregressive (*p*) and Moving Average (*q*). For all Indian rice productivity and production ARIMA (2,1,0) was found better, and in particular Odisha ARIMA (1,1,1) was best fitted for forecasting. Best fitted ARIMA models were used for predicting the immediate upcoming three years with minimum values of Akaike Information Criteria (AIC) and Schewartz-Bayesian Information Criteria (SBC). The final validation of the performance has been done by comparing with percentage deviation from actual vale and Mean Absolute Percentage Error (MAPE).

Priya*et al.*, (2015) made a comparative study between Double Exponential Smoothing model and ARIMA for Coimbatore data on sugarcane yield. To identify the best ARIMA, the AIC and BIC are sed. Among all the models in the ARIMA family, the performance of ARIMA (1,1,1) seems to be better by comparing the results with Double Exponential Smoothing model (based on Root Mean Square Error, Mean Absolute Error and mean Square Error values). The percentage of deviation, Root Mean Square Error, Mean Absolute Error calculated values for ARIMA are much compared to ARIMA model. By having the above results, with no second opinion, it is concluded that ARIMA (1,1,1) is appropriate for Coimbatore data on sugarcane yield and the performance is highly satisfactory in predicting the sugarcane yield.

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. ARIMA(0,2,1) is a particularly suitable Box-Jenkins ARIMA 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. In this regard, a comparison of the predicted series with the original series has similar patterns, suggesting that the fitted model is theoretically good for predicting Bangladesh's potato production. This implies that the models are accurate to a reasonable standard both during the estimation period and after the forecasting period. As a result, the fitted models can be used to forecast potato production in Bangladesh for policy-making purposes.

Mishra*et al.*, (2022)applied ARIMA/GARCH and Holt’s Linear Approach by taking the yearly data between 1961-2018 for predicting milk production in South Asian countries including China. MAPE of Holt’s Linear model holds higher error when comparing with MAPE of ARIMA. ARIMA technique reveals that India, Pakistan and China would occupies the first three positions in milk production with 91 MMT, 26 MMT and 3 MMT among the South Asian countries including China. Then Bangladesh and Sri Lanka captured the place of lowest production of milk. The GARCH model fits better to Bangladesh and forecasts abundance in milk production with the reach up to 105 MMT, 58 MMT, and 4 MMT by the year 2024-2025. The study has policy implications as it can be helpful for proper planning of dairy products in South Asian countries to safeguard nutritional security.

2. Objective

The aim of the paper is to forecast sugarcane yield in Coimbatore district for the ten years from 2020-2029.

3. Methods

In order to meet the objective of the paper, secondary data has been collected from the Season and Crop Report, issued by Directorate of Economics and Statistics, Government of Tamil Nadu. Sugarcane yield data was collected year wise from 1960-2019 and this univariate data was used for forecasting the sugarcane yield in Coimbatore for next ten years. To attempt forecasting technique on this dataset, two models such as the exponential smoothing model and the ARIMA model were employed. The results and findings are discussed below.

4. Results and Discussion

Since the dataset is a Univariate time series data, it was essential to plot the dataset over time to detect the kind of trend. Figure 1 shows the time series is more fluctuating with respect to time which signifies the data is not stationary. The series data plot is displayed in **Figure 1.**

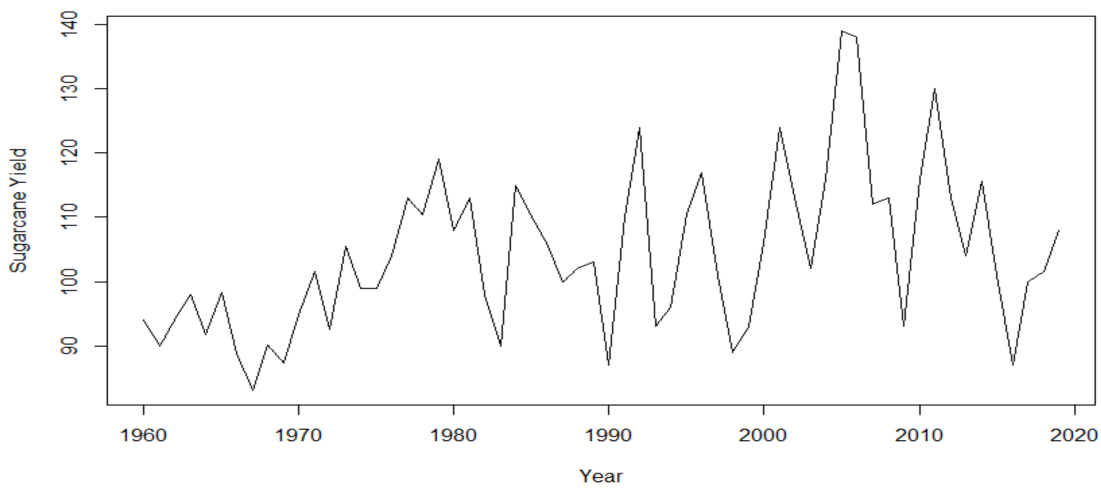

Figure 1. The raw data plot of sugarcane yield from January 1960 to December 2019.

The non-stationary data is not suitable for performing the forecasting technique. To remove the non-stationarity and to obtain clarity, differencing method was carried out. The dataset appears as a stationary trend in the plot below **Figure 2.** There should be no seasonality in the dataset's seasonal plot. Thus, the present form is used.

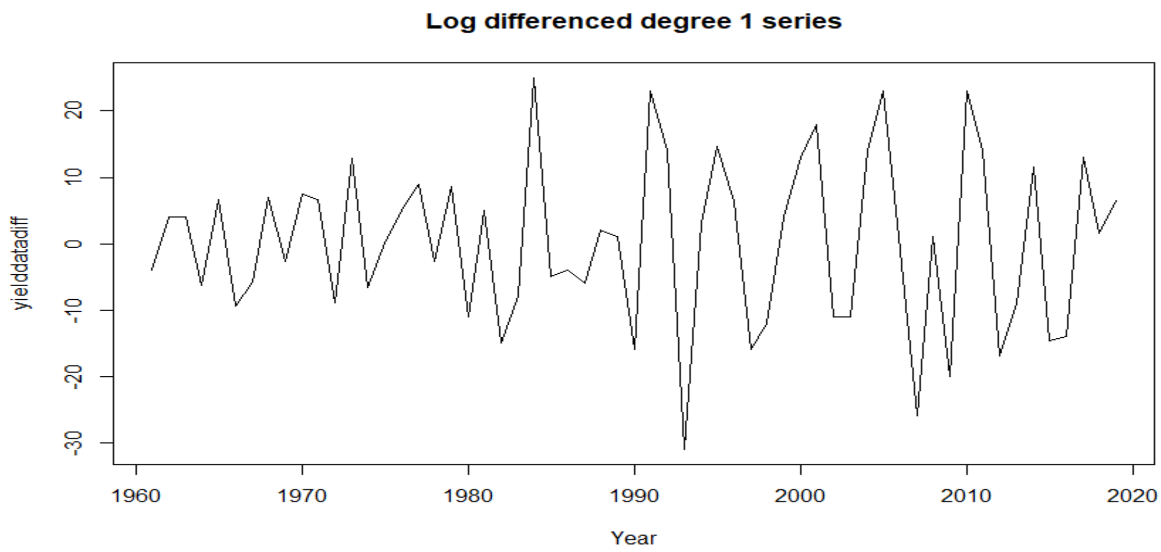

Figure 2. First-differenced time series data from January 1960 to December 2019.

ARIMA Model Identification

The model can be identified through the ACF and PACF plots of differenced sugarcane yield data as shown in **Figure 3.** 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 2 and 1 that will implement in the further forecasting process. It is previously known that to attain the stationarity, the data have been transformed by first differencing method so, the integrated part of the model would be 1. Finally, it is identified that the *p, d, q* value of the ARIMA model is 2, 1, 1 and it is suitable for forecasting the future trend.

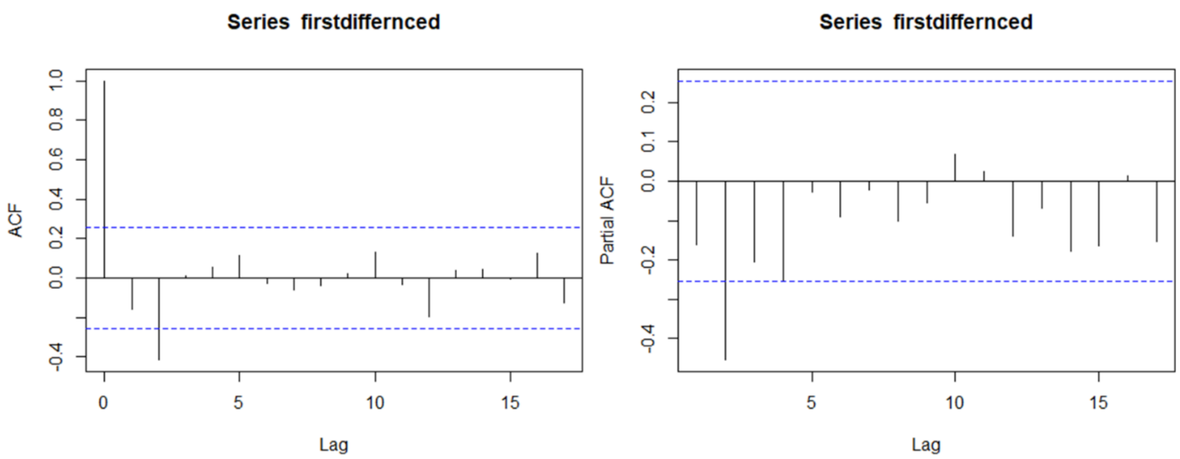

Figure 3. The ACF and PACF plots of first differenced sugarcane yield data.

In order to choose the most suitable model to carry out the forecasting process, the dataset was tested with two different forecasting models such as, Exponential Smoothing model and ARIMA model is given in below table.

Table 1. The computed forecasting results for exponential smoothing model and ARIMA model.

Model	RMSE	MAPE	MASE	AIC	AICc
Exponential Smoothing	11.09388	8.187404	0.8551261	489.1066	489.3374
ARIMA	10.19157	7.828849	0.7996529	445.42	445.49

The Root Mean Square Error, the Mean Absolute Percent Error (MAPE), the Mean Absolute Scaled Error (MASE), the Akaike Information Criterion for the two models have been compared and given in **Table 1**. The MAPE and MASE, values of exponential smoothing are higher than the ARIMA model. Also, the Ljung-Box test in the exponential smoothing model is not stable by changing the lag order which indicates the presence of autocorrelation. Hence, it is obvious that exponential smoothing model is not appropriate for forecasting the present univariate time series dataset. When making predictions, the ARIMA model performs better than the exponential smoothing approach. Also the Ljung-box test for the exponential smoothing model demonstrates the dependence of residuals on each other, i.e., the presence of autocorrelation. The MAPE value, since it is a percentage, indicates that the prediction is about 7% off. Nonetheless, the MAPE value for the ARIMA model is substantially better than the exponential smoothing model.

Residual Plot

Using the residuals produced by ARIMA (2,1,1), **Figure 4** presents three graphs. The initial graph illustrates that, with the exception of a few outliers, the variance of the residual series remains rather constant throughout the sugarcane yield data. Based on this, it appears that the residual series variance would almost remain constant. The ACF plot falls within the boundaries, suggesting no autocorrelation. The model appears to have a good fit for the data set, suggesting that the forecast generated by this model would be fairly accurate. The residuals in the histogram have normally distributed, which implies that the mean is quite close to zero. Since the Ljung-Box test result is less than 0.05 (=0.012), the residuals are likely independent.

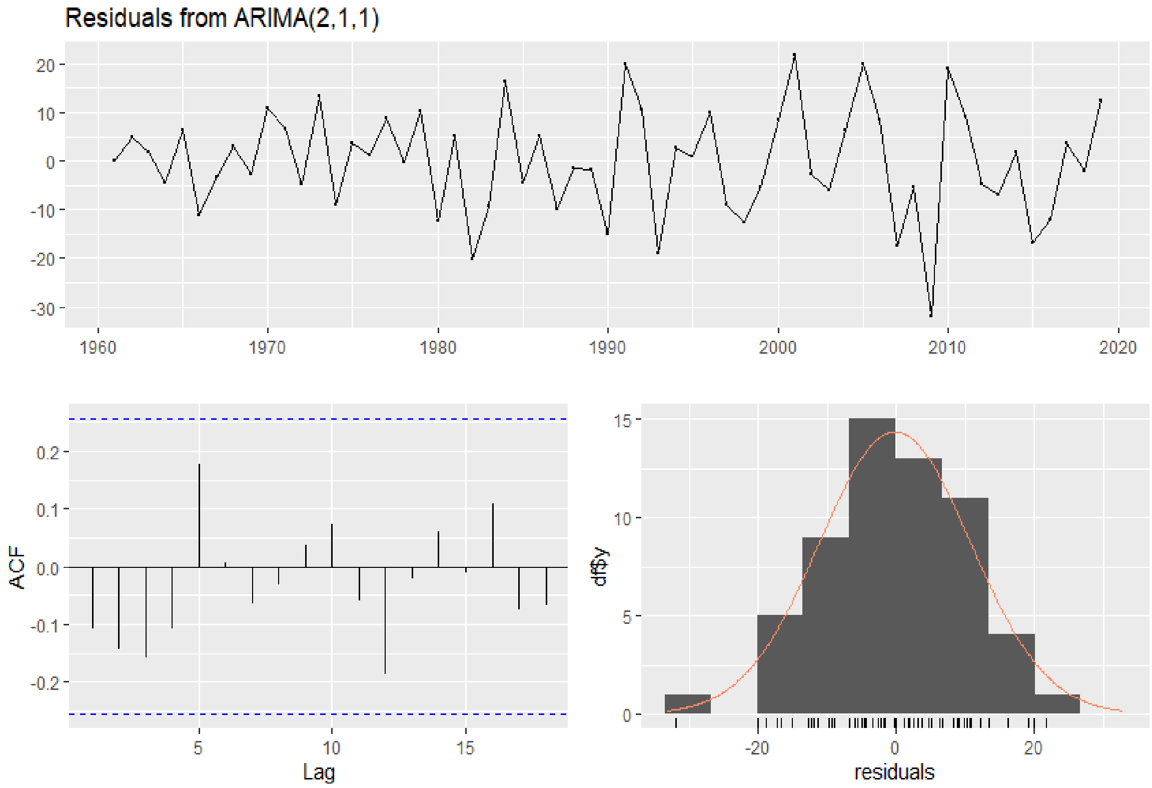

Figure 4. The residual plot for predicted ARIMA (2,1,1) model of sugarcane yield data.

ARIMA Forecasting

From the above findings, the ARIMA (2,1,1) is the best suitable for forecasting future trend of sugarcane yield for next 10 years from 2020–2029. The predicted ARIMA forecasted values are displayed in dotted lines in the **Figure 5**. The predicted values of 80% and 95% confidence intervals are provided in the light and dark shaded areas. The forecasted data shows that the trend of sugarcane yield increased from 2022 to 2024. Furthermore, the remaining years follows a moderate trend in sugarcane yield. It might be due to the increasing population and lack of land to cultivate the sugarcane crop. The expected values display a remarkably consistent pattern.

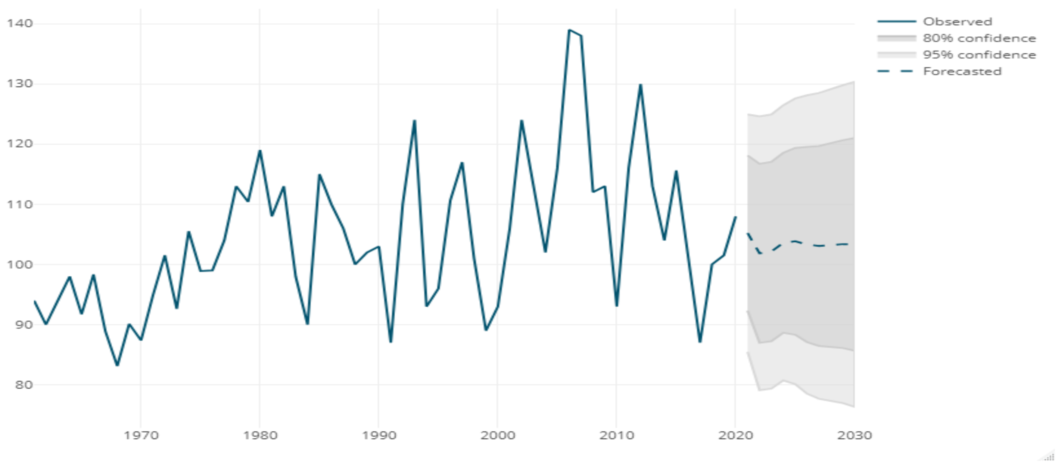

Figure 5. The Forecasted ARIMA (2,1,1) model for the sugarcane yield data.

From the below-forecasted results, it can accurately quantify the future forecasted values for the next 10 years is shown in**Table 2.**

Table 2. The forecasted sugarcane yield for next 10 years.

Years	Point Forecast	Low 80%	High 80%	Low 95%	High 95%
2020	105.2140	92.31598	118.1120	85.42219	124.9398
2021	101.8577	86.98178	116.7335	79.10696	124.6084
2022	102.1513	87.25667	117.0458	79.37195	124.9306
2023	103.5995	88.64746	118.5516	80.73230	126.4668
2024	103.8449	88.31908	119.3707	80.10022	127.5895
2025	103.3138	87.09933	119.5283	78.51590	128.1117
2026	103.0793	86.46472	119.6939	77.66948	128.4892
2027	103.2376	86.32572	120.1494	77.37312	129.1020
2028	103.3735	86.10951	120.6376	76.97048	129.7766
2029	103.3430	85.67868	121.0074	76.32772	130.3584

Box Test

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 3.**

Table 3. The Box test for the forecasted ARIMA(2,1,1) model.

Lags	5	6	7	8	9
X-squared	4.1604	4.2468	4.3001	4.3065	4.32
df	1	1	1	1	1
p - value	0.04138	0.03932	0.03811	0.03797	0.03767

5. Conclusion

The study reveals that the ARIMA model was the best fit out of the two models used to forecast the future sugarcane yield in Coimbatore, and from the predicted ARIMA it is concluded that the model (2,1,1) is the best fit. The forecasted data indicates an increase in sugarcane yield from 2022 to 2024, followed by an up and down trend from 2025 to 2029. This trend may be due to the increasing population and lack of land for cultivating the sugar crop. Therefore, it is essential for the government to pay close attention when formulating rules and procedures that enable farmers to expand crop production. This will ultimately benefit the growth and economy of our nation in all aspects.

References

[1] Tripathi, R., Nayak, A. K., Raja, R., Shahid, M., Kumar, A., Mohanty, S., Panda, B. B., Lal, B., Gautam, P. (2014). Forecasting Rice Productivity and Production of Odisha, India, Using Autoregressive Integrated Moving Average Models. Adv. Agric. 2014, 9.

[2] Mishra, P., Matuka, A., Abotaleb, M. S. A., Weerasinghe, W. P. M. C. N., Karakaya, K., Das, S. S. (2022). Modeling and Forecasting of Milk Production in the SAARC Countries and China. Model. Earth Syst. Environ., 8(1), 947–959.

[3] Krishna Priya, S. R., Bajpai, P. K., Suresh, K. K. (2015). Stochastic Models for Sugarcane Yield Forecasting, Indian J. Sugarcane Technol., 30(01), 1-5.

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

[5] Biswas, R., Bhattacharyya, B. (2013).ARIMA Modeling to Forecast Area and Production of Rice in West Bengal, J. crop weed, 9(12), 26-31.

[6] Hossain, M. M., Abdulla, F. (2016). Forecasting Potato Production in Bangladesh by ARIMA Model. J. Adv. Stat., 1(4), 191-198.

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

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

[9] Hamjah, M. A. (2014). Forecasting Major Fruit Crops Productions in Bangladesh Using Box-Jenkins ARIMA Model. J. Econ. Sustain. Dev., 5(7), 96–107.

[10] Dash, A., Kumar Mahapatra, S. (2020). Using ARIMA Model for Yield Forecasting of Important Pulse Crops of Odisha, India. Amaz. J. Plant Res., 4(3), 646–659.

[11] Ahmadzai, M. K., Eliw, M. (2019). Using ARIMA Models to Forecasting of Economic Variables of Wheat Crop in Afghanistan. Asian J. Econ. Bus. Account., 13(4), 1–21.

[12] Dimri, T.; Ahmad, S.; Sharif, M. Statistical Analysis of Precipitation, Temperature and Snow Cover in Bhagirathi River Basin. *Mausam* (2022), *73* (2), 263–272.

[13] The Season and Crop Report, Directorate of Economics and Statistics, Government of Tamilnadu.