Scientiarum: A Multidisciplinary Journal
Volume 1, Issue 6, December 2025, pp. 1–6
DOI: 10.54646/SAPARS.2025.26

SAPAR SERVICES LLP
Scholarly Articles Publication And Repository Services

RESEARCH

Forecasting agricultural production using ARIMAX model

R. Mekha* **and M. Vijayakumar***

Department of Statistics, Annamalai University, Cuddalore, India

***Correspondence:**
R. Mekha,
mekhastat@gmail.com
M. Vijayakumar,
mvkumar1975@gmail.com

Received: 04 August 2025; **Accepted:** 01 November 2025; **Published:** 13 November 2025

This study explores the dynamic relationship between rainfall and agricultural production in Tamil Nadu, with a specific focus on maize, a staple crop critical to the region's food security and economy. Employing time series analysis, the research utilizes both the Autoregressive Integrated Moving Average (ARIMA) and ARIMA with exogenous variables (ARIMAX) models to understand and forecast maize production trends. Initially, an ARIMA (p,d,q) model is developed based on historical maize production data to identify and forecast future production trends. In parallel, an ARIMAX (p,d,q) model is constructed by integrating rainfall data as an exogenous variable, enabling the study to assess the impact of climatic factors particularly precipitation on maize productivity. This dual-model approach facilitates a comparative analysis of forecast accuracy. The study not only helps in understanding the underlying patterns in maize production but also emphasizes the significance of incorporating climatic variables into predictive modelling for more realistic forecasts. Such insights are useful for policymakers, agricultural planners, and farmers which helps them make decisions on crop planning, irrigation strategies, and risk management. Ultimately, this research contributes to the development of climate-resilient agricultural strategies in Tamil Nadu, where rainfall anomalies increasingly pose a challenge to stable food production.

Keywords: maize production, ARIMA, ARIMAX, rainfall forecasting, time series analysis, agricultural forecasting, exogenous variables, Tamil Nadu agriculture

Introduction

Maize (*Zea mays*), widely known as corn, is one of the most significant cereal crops globally, serving as a cornerstone of food security, animal feed, and industrial raw material. Renowned for its adaptability, maize is cultivated in a broad range of climatic zones, from temperate to tropical regions, making it a globally favoured crop. It ranks just after rice and wheat in terms of importance and area under cultivation in India. The major maize-producing countries include the United States, China, India, Brazil, and Mexico, reflecting its global demand and versatility. In India, maize is grown across various agro-climatic regions and serves diverse purposes, such as human consumption, poultry and cattle feed, starch extraction, and ethanol production.

Tamil Nadu has witnessed remarkable growth in maize cultivation in recent years, owing to the crop's compatibility with both monsoon (kharif) and post-monsoon (rabi) seasons. Districts such as Perambalur, Namakkal, Salem, Dindigul, Theni, Coimbatore, Madurai, and Virudhunagar are key maize-growing regions in the state. Favorable soil conditions especially well-drained red and black soils and moderate rainfall ranging from 600 mm to 900 mm support maize cultivation. The crop is particularly sensitive to fluctuations in rainfall and irrigation availability, making it vulnerable to climate variability.

Maize farming in Tamil Nadu faces several challenges, including dependence on the southwest monsoon, limited irrigation coverage, pest infestations, soil nutrient depletion, and price volatility. Additionally, unsustainable farming practices may lead to long-term soil degradation and water stress. However, interventions such as the introduction of high-yielding drought-tolerant hybrids, improved agronomic techniques, government support schemes, and precision agriculture are gradually improving maize productivity and farmer resilience. Strengthening post-harvest infrastructure and ensuring stable market access can further enhance the economic viability of maize farming in Tamil Nadu.

Data description

The dataset used in this paper includes maize production (in tonnes) and rainfall data of Tamil Nadu from 1965 to 2022 from the online portal of Department of Economics and Statistics. The maize production and rainfall data is recorded annually. The analysis is conducted using the R programming language, leveraging the auto.arima () function from the forecast package to select the optimal model parameters. By using the auto.arima () function the best Autoregressive Integrated Moving Average (ARIMA) model is automatically selected based on a given time series dataset.

Methodology

ARIMAX model

The Auto-Regressive Integrated Moving Average with Exogenous Variables (ARIMAX) model extends ARIMA by incorporating external variables, such as rainfall, to enhance prediction accuracy (1). This model accounts for additional predictors that influence the target variable, making it particularly useful in agricultural forecasting where climatic factors play a crucial role (2).

The ARIMAX model combines the following components

- Autoregressive Component (AR): Uses past values of the dependent variable to predict future values.
- Integrated Component (I): Applies differencing to make the time series stationary.
- Moving Average Component (MA): Accounts for past forecasting errors to refine predictions.
- Exogenous Variables (X): Incorporates independent variables that can explain variations in the dependent variable.

It is generally expressed as:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \gamma X_t + \ldots + \varepsilon_t$$

Where,

1. Y_t is the dependent variable.
2. α is a constant.
3. $Y_{t-1}, Y_{t-2}, \ldots\ Y_{t-p}$ are the past values of the dependent variable.
4. γ is the coefficient of exogenous variable.

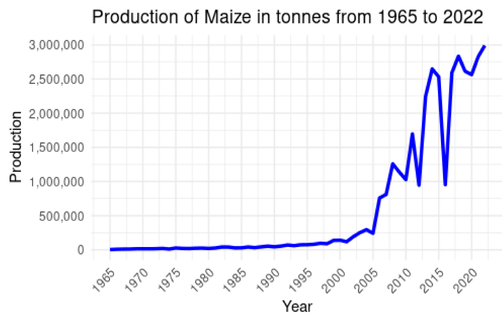

FIGURE 1 | Maize production from 1965 to 2022.

5. X_t are the exogenous variables. The trend in maize production in Tamil Nadu from 1965 to 2022 is presented in **Figure 1**.
6. ε_t is the error term.

The ARIMAX model is a powerful tool for time series forecasting when external factors play a significant role in influencing the behavior of the dependent variable. It combines the strengths of both time series analysis and regression models, allowing for more informed and accurate predictions that account for both historical patterns and external influences. This makes it particularly useful in applications such as economic forecasting, sales prediction, climate modeling and demand estimation.

ARIMA model

Time series forecasting plays an important role in various fields, including agriculture, finance, economics, and climate science. One of the most widely used statistical models for time series forecasting is the ARIMA model, which is effective in capturing temporal dependencies and patterns in historical data (3).

The ARIMA model is a combination of three components:

- Auto-Regressive (AR) Component: This captures the relationship between an observation and its previous values.
- Integrated (I) Component: This represents the differencing required to make a non-stationary time series stationary.
- Moving Average (MA) Component: This accounts for the dependency between an observation and past forecast errors.

Mathematically, an ARIMA (p, d, q) model is expressed as:

$$Y_t = \alpha + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots + \emptyset_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Where,

1. Y_t is the dependent variable (e.g., maize production).

TABLE 1 | Forecasted maize production using ARIMAX model for 20 years [2023–2042].

Year	Forecasted maize production (in 1,000 MT)
2023	2,862,077
2024	2,912,359
2025	2,962,641
2026	3,012,923
2027	3,063,205
2028	3,113,487
2029	3,163,768
2030	3,214,050
2031	3,264,332
2032	3,314,614
2033	3,364,896
2034	3,415,178
2035	3,465,460
2036	3,515,742
2037	3,566,023
2038	3,616,305
2039	3,666,587
2040	3,716,869
2041	3,767,151
2042	3,817,433

2. p is the number of AR terms.
3. d is the degree of differencing to make the series stationary.
4. q is the number of MA terms.
5. α is a constant.
6. $\theta_1, \theta_2, \ldots, \theta_q$ are the coefficients for the MA part of the model, which represent the relationship between the current value and past errors (lags).
7. $\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-q}$ are the past forecast errors (lags of the residuals).
8. ε_t is the white noise (random error term).

In the agricultural domain, ARIMA models are extensively used for forecasting crop production, yield estimation, and analyzing the impact of climatic variables on agriculture. By leveraging historical production and rainfall data, ARIMA helps in predicting future trends and making data-driven decisions for sustainable agricultural planning (4).

Results and discussion

Both ARIMAX and ARIMA models were applied to the maize production data. The statistical analysis is done using R software.

Forecast using ARIMAX model

The ARIMAX model with ARIMA (0,1,1) residuals is identified as the best model, similar to findings reported in earlier agricultural forecasting studies (5). The ARIMAX model with rainfall as an exogenous variable shows that the moving average coefficient (ma1) is −0.5335, the drift term is 50,281.88, and the rainfall coefficient (xreg) is 290.8014, indicating a positive relationship between rainfall and agricultural production.

The training set error measures, including a negative mean error (ME) of −1,831.89 and a high Mean Absolute Percentage Error (MAPE) of 230.68%, indicate that the model tends to underestimate agricultural production and has significant prediction errors. However, the low

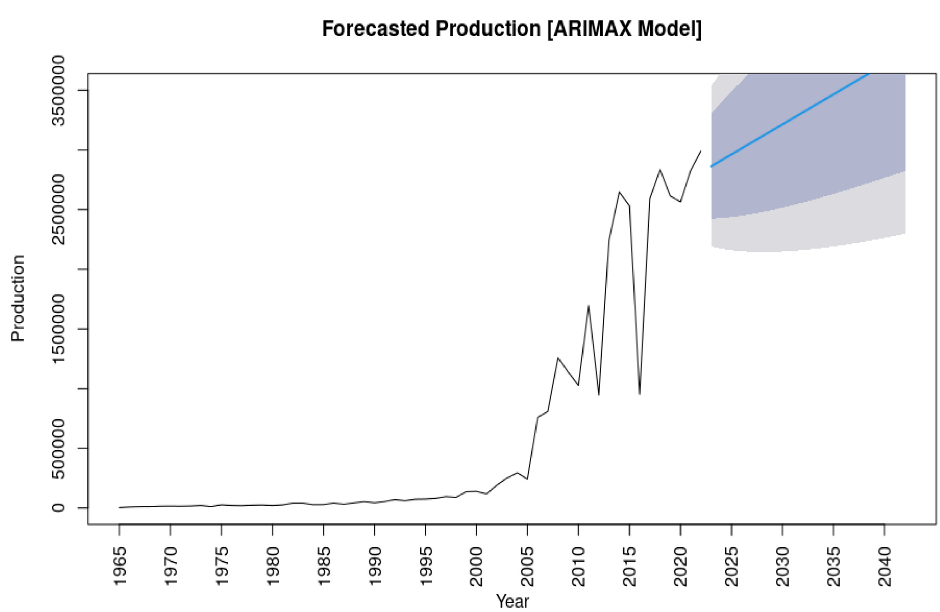

FIGURE 2 | The forecasted production using ARIMAX model.

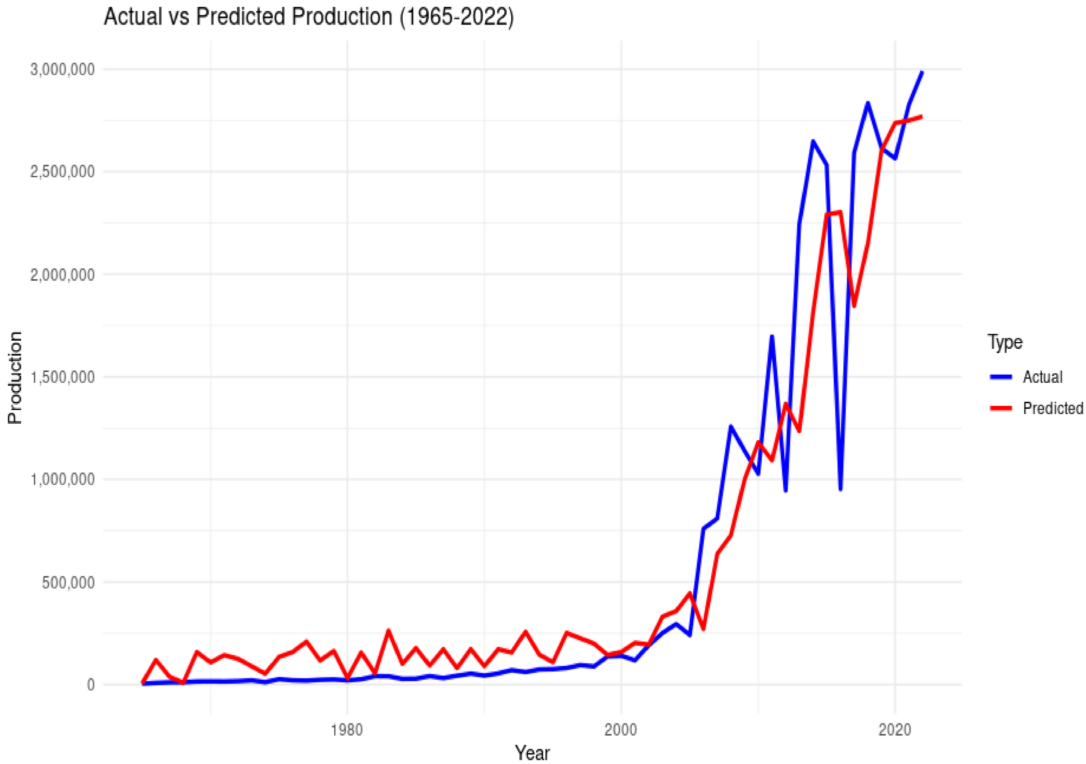

FIGURE 3 | The actual vs predicted maize production.

autocorrelation of the residuals (ACF1 = −0.017) suggests that the model effectively captured the temporal structure in the data. The model's Akaike Information Criterion (AIC) of 1,619.75 and Bayesian Information Criterion (BIC) of 1,627.92 suggest a reasonable fit, though there is room for improvement in predictive accuracy.

The maize production is forecasted for 20 years [2023–2042] using ARIMAX model. The forecasted maize production is computed and presented in **Table 1** and **Figure 2**.

The actual and ARIMAX predicted maize production is compared and presented in **Figure 3**.

Forecast using ARIMA model

The ARIMA (0,1,1) with drift is identified as the best model. The moving average coefficient of −0.5409 and drift term of 51,426.08 are well-estimated. The model shows a slight negative bias with a ME of −1,642.93. The low autocorrelation of the residuals (ACF1 = −0.0283) suggests that the model has effectively captured the underlying structure of the data. The error metrics such as Root Mean Square Error (RMSE) (335,221) and Mean Absolute Error (MAE) (202,918) highlight areas for potential improvement, but overall, the model demonstrates a solid foundation for forecasting agricultural production. An AIC value of 1,619.47, indicates an efficient model selection.

The maize production is forecasted for 20 years [2023–2042] using ARIMA model. The forecasted maize production is computed and presented in **Table 2** and **Figure 4**.

TABLE 2 | Forecasted maize production using ARIMA model for 20 years [2023–2042].

Year	Forecasted maize production (in 1,000 MT)
2023	2,929,800
2024	2,981,226
2025	3,032,652
2026	3,084,078
2027	3,135,504
2028	3,186,930
2029	3,238,356
2030	3,289,782
2031	3,341,208
2032	3,392,634
2033	3,444,061
2034	3,495,487
2035	3,546,913
2036	3,598,339
2037	3,649,765
2038	3,701,191
2039	3,752,617
2040	3,804,043
2041	3,855,469
2042	3,906,895

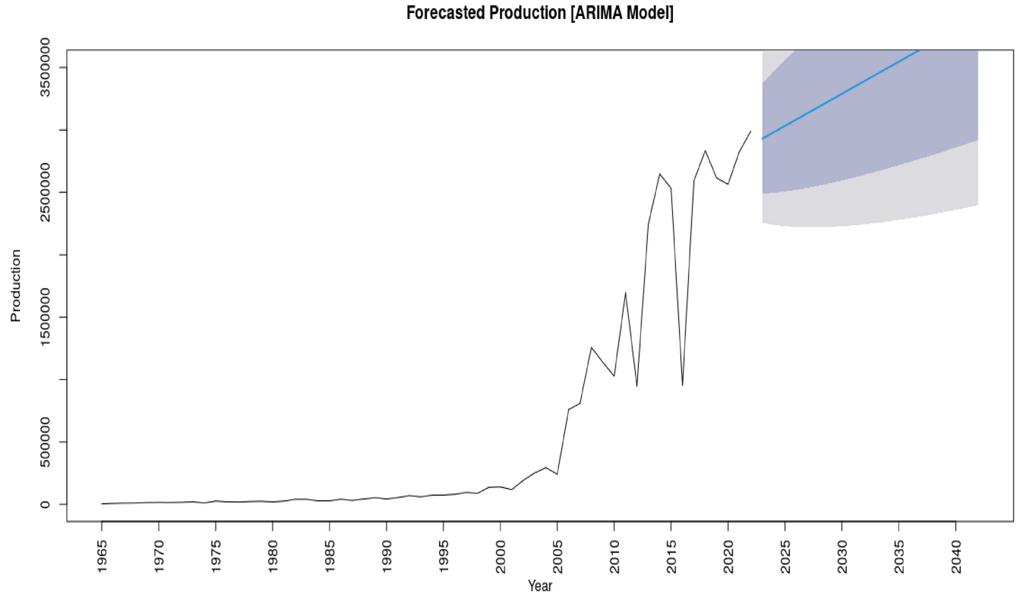

FIGURE 4 | The forecasted maize production using ARIMA model.

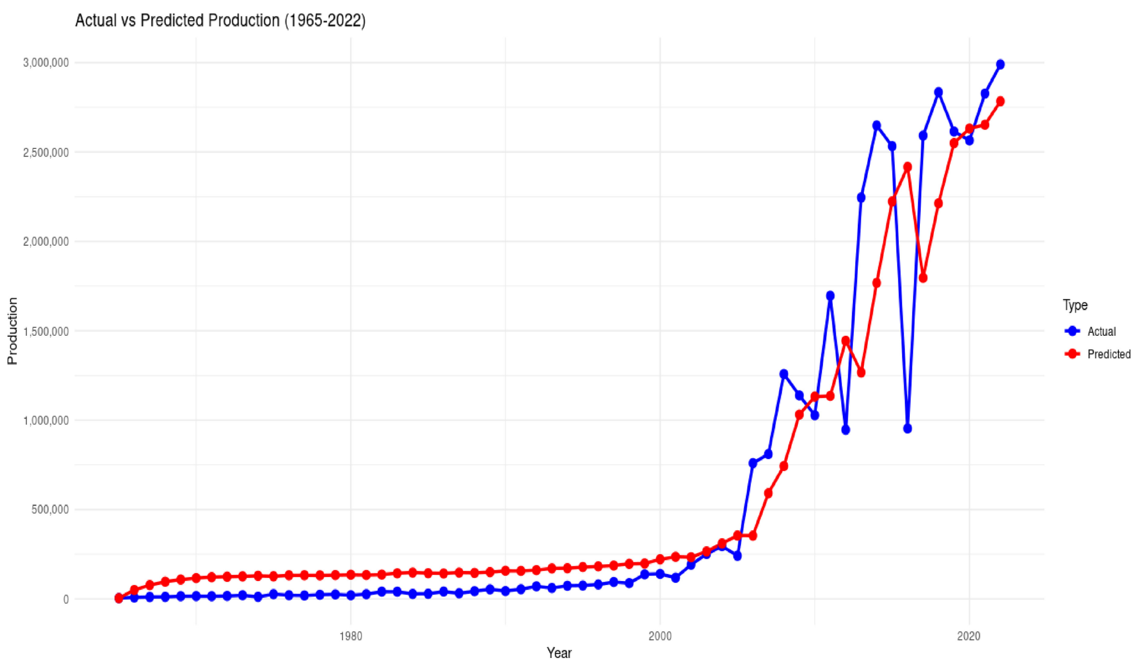

FIGURE 5 | The actual vs ARIMA predicted maize production.

The actual and the ARIMA predicted maize production is compared and presented in the **Figure 5**.

Comparison between ARIMA and ARIMAX models

The error metrics of ARIMA and ARIMAX models are computed and given in **Table 3**.

The actual vs ARIMA and ARIMAX predicted maize production is compared and presented in **Figure 6**.

The ARIMAX model performs slightly better than the ARIMA model, as seen by the reduction in both MAPE and MSE. While the ARIMAX model improves on the ARIMA model by incorporating rainfall as an additional predictor.

TABLE 3 | The error metrices of ARIMAX and ARIMA model.

Model	MAPE	MSE
ARIMAX	230.6836	109,061,149,061
ARIMA	241.0660	112,373,129,496

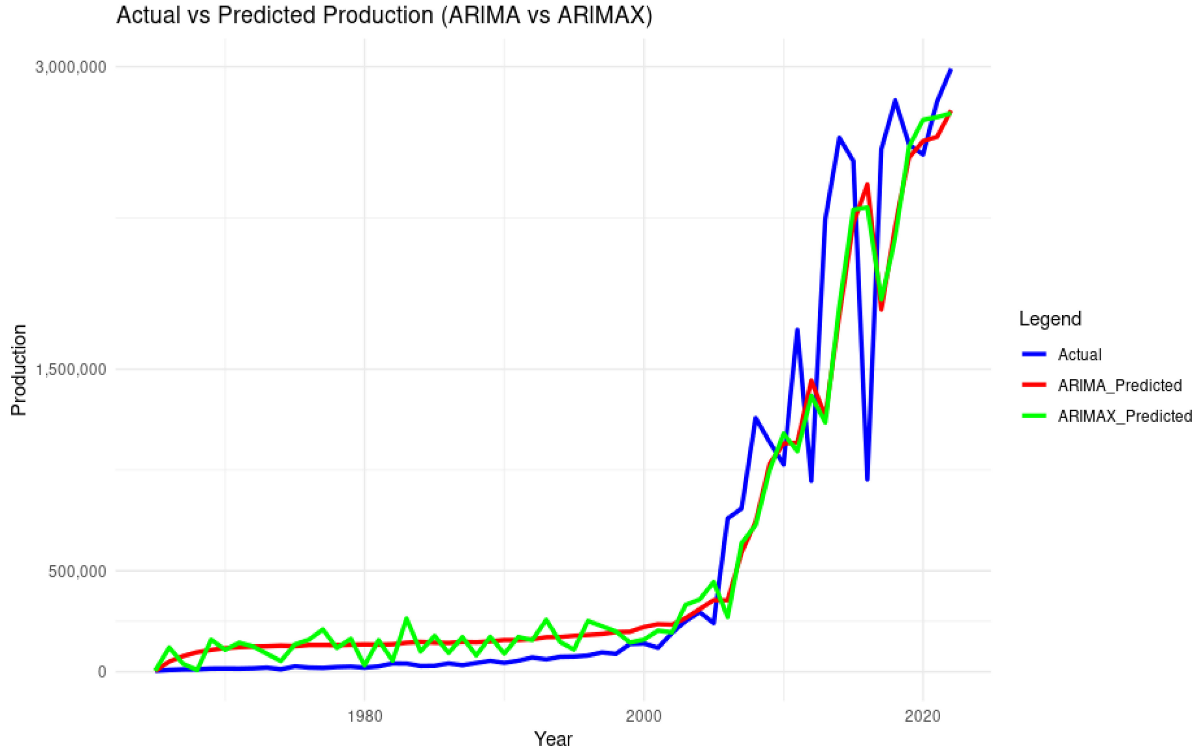

FIGURE 6 | The actual vs ARIMA and ARIMAX predicted maize production.

Conclusion

This study identifies the significance of using time series models for agricultural forecasting. The comparison between ARIMA and ARIMAX models explains about incorporating external variables like rainfall and shows how it improves the predictive accuracy. The automated model selection process using auto.arima () function provides a powerful and efficient approach to time series analysis, making it a valuable tool for researchers. This method simplified the analysis, allowing a focus on interpreting results and drawing meaningful conclusions. The results shows that the ARIMA model captures trends effectively but the ARIMAX model performs better by incorporating rainfall as an exogenous variable. This improvement is reflected in lower error values for ARIMAX compared to ARIMA, explaining the importance of exogenous factors in forecasting maize production.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Wangdi K, Singhasivanon P, Silawan T, Lawpoolsri S, White NJ, Kaewkungwal J. Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan. *Malar J.* (2010) 9:1–9.
2. Yogarajah B, Elankumaran C, Vigneswaran R. Application of ARIMAX model for forecasting paddy production in Trincomalee district in Sri Lanka. *Proceedings of the Third International Symposium* (2013).
3. Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis: Forecasting and Control.* New York: John Wiley & Sons Inc (2008).
4. Gopinath M, Kavithamani M. Production of sugarcane forecasting using ARIMAX model. *Int J Innovat Technol Explor Eng.* (2019) 8(12s):596–600.
5. Stellamani S, Vijayakumar M, Philip Arokiadoss A. Forecasting time series for paddy production. *Aryabhatta J Math Informat.* (2022) 14:113–20.