EP-JOURNALS

GLOBAL JOURNAL OF ENGINEERING AND TECHNOLOGY REVIEW

Available on: http://gjetr.ep-journals.org/ E-ISSN: 3051-3782, P-ISSN: 3051-3774

Global Journal of Engineering and Technology Review

Volume: 02 Issue: 03 March 2026

# Forecasting Rice Production in Northern Samar: A Multi-Model Analysis Using Cyclone Frequency and Harvested Area as Predictors

**Kyle Arthur P. Sabilao**
Don Juan F. Avalon National High School, Northern Samar, Philippines

**Elvin L. Jarito, PhD**
Don Juan F. Avalon National High School, Northern Samar, Philippines

**Published Date: March 13, 2026** **Article DOI : 10.65150/EP-gjetr/V2E3/2026-01**

**ABSTRACT:** This study forecasted rice production in Northern Samar using three predictive models: Multiple Linear Regression (MLR), Random Forest Regression (RF), and Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX). The analysis utilized cleaned datasets from 2016 to 2024, covering 23 municipalities after excluding San Vicente due to data insufficiency. Two explanatory variables—area harvested and tropical cyclone frequency—were used, with the former projected through Random Forest (SMAPE = 12.75%) and the latter using a constant mean approach. For rice production forecasts, MLR demonstrated steady upward trends and yielded the lowest MAPE in 4 municipalities, notably Catarman (5.48%) and Las Navas (3.91%). RF achieved the lowest MAPE in 17 municipalities, with exceptional accuracy in Lapinig (1.63%), Laoang (3.58%), and Lavezares (3.40%); however, it failed to capture long-term growth as it produced identical forecasts for 2030, 2035, and 2040—indicating limitations in trend sensitivity due to capped inputs. SARIMAX, while incorporating time-series structure and external factors, displayed inconsistent performance with the lowest MAPE in only 2 municipalities (Allen: 6.56%, Biri: 25.43%). Its trend outputs were often unstable in areas with irregular patterns, making it less suited for long-range forecasting. Among the three models, MLR proved most practical for long-term planning due to its interpretability and ability to reflect realistic growth. RF was best suited for short-term accuracy and baseline projections, while SARIMAX offered mixed results depending on data structure. These forecasts serve as valuable tools for local agricultural planning, guiding food security initiatives and resource allocation in Northern Samar.

**KEYWORDS:** Rice production, forecasting models, Random Forest, Multiple Linear Regression, SARIMAX, Northern Samar

## 1. INTRODUCTION

Rice cultivation is a vital component of Northern Samar's agricultural economy, serving as both a primary source of livelihood for farmers and a staple food for the population (*Northern Samar Farmers to Enrol in Rice-to-Rise Program for Sustainable Farming | Province of Northern Samar*, 2024). The province's dependence on agriculture makes it highly vulnerable to fluctuations in rice production caused by climate variability, land degradation, pest outbreaks, and government policy changes (Marticio, 2024). The increasing occurrence of typhoons and droughts has further intensified production instability, threatening local food security and economic resilience (Cuyco et al., 2023). Addressing these challenges requires a data-driven approach that can provide accurate, evidence-based insights to guide sustainable agricultural planning and resource management in the province.

Accurate forecasting of rice production trends plays a crucial role in helping farmers, policymakers, and agricultural stakeholders make informed decisions (Parreño, 2023). Predictive analytics and time series forecasting (TSF) models such as Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA with Exogenous Variables (SARIMAX), and machine learning algorithms like Random Forest (RF) have been widely used to identify patterns, detect seasonality, and forecast yield fluctuations (Ensafi et al., 2022). These tools enable stakeholders to anticipate potential production shortfalls, optimize planting schedules, and allocate resources effectively (Yadav, 2024). In an agricultural region like Northern Samar, where external factors such as climate and land use significantly influence yield, forecasting serves as a key strategy for strengthening food security, promoting economic resilience, and enhancing the efficiency of agricultural operations.

Despite numerous studies on agricultural forecasting in the Philippines, limited research has explored multi-model forecasting of rice production in Northern Samar that integrates both climatic and agricultural predictors. Previous works have often focused solely on historical yield trends, overlooking vital factors such as tropical cyclone frequency and harvested area, which are particularly relevant to this climate-sensitive province

(Parreño & Anter, 2024). This research narrows the existing gap through the application and comparison of Multiple Linear Regression (MLR), Random Forest (RF), and SARIMAX forecasting models. The forecasted outputs for 2030, 2035, and 2040 are intended to guide policymakers and farmers toward data-informed planning and resilient rice production systems.

## 2. LITERATURE REVIEW

### *A. Time Series Forecasting Models in Rice Production Studies*

Time series forecasting has long been recognized as an effective approach in predicting agricultural production, particularly in staple crops such as rice. Traditional statistical models like the Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and Exponential Smoothing methods have been widely applied to capture patterns in historical yield data (Ibañez & Monterola, 2023). These models identify seasonality, trends, and irregular fluctuations that affect production levels over time. In rice-producing regions, such models have been instrumental in determining planting schedules, managing supply chains, and ensuring food stability, especially in countries with climates prone to variability (Kolambe, 2024). Their strength lies in their interpretability and capacity to represent temporal dependencies within agricultural datasets.

Forecasting rice production through time series models provides valuable insights into short- and long-term yield dynamics (Rathod et al., 2021). These models are particularly useful in provinces such as Northern Samar, where rice production is influenced by both climatic and environmental factors. Analyzing historical production data enables time series models to estimate future outputs and identify deviations from expected trends, allowing decision-makers to prepare for potential production shortfalls (Gulati, 2025). They also allow policymakers and agricultural planners to identify recurring cycles in yield patterns, which can guide interventions such as crop diversification, irrigation scheduling, and disaster risk management. Reliable forecasts derived from these models ultimately contribute to evidence-based planning, ensuring more efficient use of agricultural resources (Meinke & Stone, 2005).

Although traditional time series models remain widely used, their effectiveness largely depends on data quality, the inclusion of relevant predictors, and the stability of historical patterns (D. Shah & Thaker, 2024). In the context of Northern Samar, relying solely on historical yield data may not sufficiently capture the external influences that affect production, such as changes in harvested area or the frequency of tropical cyclones. These variables can significantly alter yield outcomes and, when integrated into forecasting models, can enhance predictive performance (Jabed & Azmi Murad, 2024). As such, developing a forecasting framework that combines traditional time series approaches with context-specific predictors provides a more accurate and responsive tool for anticipating rice production trends in the province.

### *B. Machine Learning and Hybrid Approaches in Agricultural Forecasting*

Recent advancements in computational modeling have led to the increasing use of machine learning and hybrid approaches in agricultural forecasting (Aderele et al., 2025). Unlike traditional time series models that rely heavily on linear assumptions, machine learning algorithms are capable of capturing complex, nonlinear relationships between variables influencing crop production (Kontopoulou et al., 2023). Techniques such as Random Forest (RF), Support Vector Regression (SVR), Gradient Boosting, and deep learning models like Long Short-Term Memory (LSTM) networks have demonstrated strong predictive capabilities in various agricultural contexts (Hia et al., 2023). These models can process large datasets, detect intricate interactions between climatic and environmental factors, and produce more adaptable and accurate forecasts, particularly in dynamic agricultural systems affected by multiple external drivers (Ahmad Hamdan et al., 2024).

Hybrid models, which combine statistical and machine learning techniques, have further enhanced the precision of agricultural forecasts (Kumari et al., 2024). The integration of statistical and machine learning techniques enables hybrid models to capture both temporal patterns and complex nonlinear relationships within agricultural data (P. Shah et al., 2025). Combining exponential smoothing or ARIMA-based methods with neural networks has proven effective in improving the accuracy of yield and price predictions. In the context of rice production, hybrid forecasting enables researchers and policymakers to capture both the temporal continuity of agricultural processes and the irregular influences of climate variability, technological changes, and policy interventions (Slater et al., 2022).

In the case of Northern Samar, machine learning and hybrid forecasting models offer an opportunity to better understand and anticipate rice production trends under changing environmental conditions. The province's vulnerability to tropical cyclones and irregular rainfall patterns necessitates forecasting tools that can incorporate external variables such as cyclone frequency and harvested area—factors that directly affect yield variability. Incorporating these predictors within a multi-model framework that includes both statistical (e.g., Multiple Linear Regression and SARIMAX) and machine learning (e.g., Random Forest) models enhances the reliability and adaptability of forecasts. Such integration supports the development of a data-driven, context-sensitive approach to agricultural planning, ultimately contributing to sustainable rice production and long-term food security in Northern Samar.

### *C. Statement of the Problem*

This study on forecasting rice production in Northern Samar aims to address the following research questions:

1. What are the historical trends in rice production, area harvested, and tropical cyclone frequency in the municipalities of Northern Samar from 2016 to 2024?
2. How can missing values and outliers in the dataset be treated to ensure data quality and consistency for forecasting?
3. How can area harvested and tropical cyclone frequency be projected for future years (2030, 2035, and 2040) using appropriate modeling techniques?
4. To what extent can different models forecast rice production in Northern Samar using area harvested and cyclone frequency as predictors, specifically:

a) Multiple Linear Regression (MLR)?
b) Random Forest (RF)?
c) SARIMAX?

5. Which among the three forecasting models demonstrates the most accurate and reliable performance in projecting rice production across all municipalities based on standard error metric (MAPE)?

## 3. SCOPE AND LIMITATIONS

This study focused on forecasting rice production in Northern Samar using three predictive models—Multiple Linear Regression (MLR), Random Forest (RF), and SARIMAX—based on two key explanatory variables: area harvested and tropical cyclone frequency. Forecasts were generated for 2030, 2035, and 2040 at both municipal and provincial levels, utilizing historical data from 23 municipalities (excluding San Vicente due to insufficient records). To ensure data reliability, preprocessing techniques such as linear interpolation, outlier correction through the Interquartile Range (IQR) method, and local smoothing were applied to handle missing and anomalous values.

The study acknowledged several limitations, primarily related to data quality and availability, particularly in municipalities with incomplete records. The models assumed that historical patterns would persist, which might not have captured future shifts in agricultural practices, technology, or extreme weather conditions. Cyclone frequency was projected using historical averages, potentially oversimplifying future climate variability. Other potential predictors—such as irrigation coverage, soil quality, pest outbreaks, and policy interventions—were excluded due to data constraints. Furthermore, evaluation was limited to in-sample validation given the small dataset, reducing generalizability. Despite these constraints, the study provided a valuable data-driven foundation for agricultural planning and underscored the challenges of forecasting rice production in a climate-sensitive and geographically diverse province like Northern Samar.

## 4. METHODOLOGY

### *A. Research Design*

The methodology followed a multi-stage process. First, historical data on annual rice production, area harvested, and typhoon counts from 2016 to 2024 were collected from official government sources. These datasets were subjected to thorough preprocessing to ensure completeness and consistency. Missing values were estimated using linear interpolation, while outliers were detected and treated using the Interquartile Range (IQR) method. Following data cleaning, the explanatory variables were projected independently: Random Forest Regression was used to forecast the area harvested, while the number of typhoons was projected using a historical mean-based method. These forecasted variables served as inputs to three different models used for rice production forecasting: Multiple Linear Regression (MLR), Random Forest (RF), and SARIMAX. Each model was trained per municipality and evaluated using a consistent set of statistical error metrics—Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination ($R^2$), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE). The final phase involved a comparative analysis of the three models' outputs across all municipalities using MAPE. This enabled the identification of the most reliable forecasting approach for rice production in the region. The overall methodology is summarized in Figure 1.

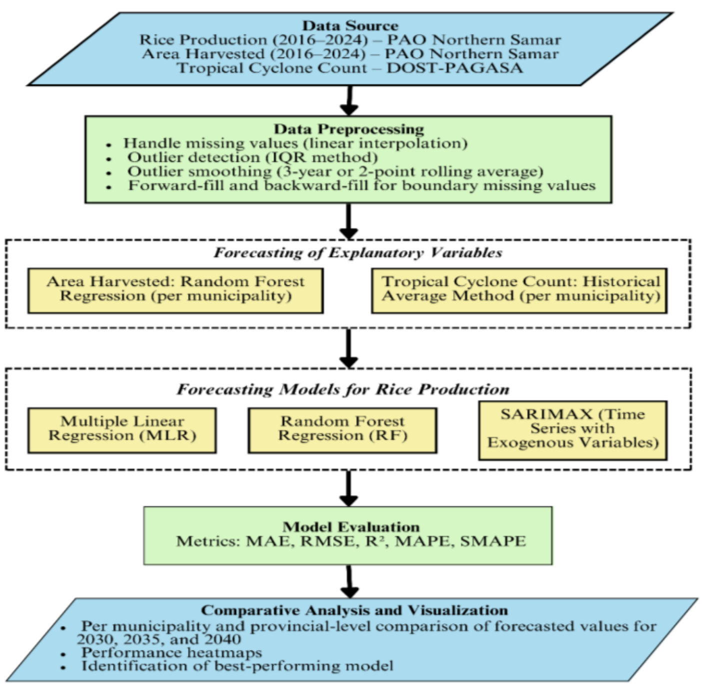


**Fig. 1. Overview of Methodology**

***B. Data Source and Preprocessing***

The rice production data used in this study came from the Provincial Agriculture Office (PAO) of Northern Samar, covering annual outputs from 2016 to 2024 across its municipalities. San Vicente was excluded due to missing records, leaving 23 municipalities with consistent data for reliable analysis. Two explanatory variables—area harvested (from PAO) and tropical cyclone count (from DOST-PAGASA)—were used to capture both agricultural capacity and climate-related disruptions affecting rice production.

To ensure accurate forecasting, the rice production dataset underwent several stages of data pre-processing. Initial inspection revealed missing values between 2017 and 2020 in some municipalities, which were addressed using linear interpolation—a method that estimates missing data based on the trend between adjacent years. Outliers were then detected using the Interquartile Range (IQR) method, which identifies values that deviate significantly from the normal range of the dataset. These anomalous values, possibly caused by reporting errors or unusual events, were replaced with nulls and corrected through local smoothing using a rolling average from nearby years. Any remaining gaps at the start or end of the data were filled using forward fill and backward fill techniques to ensure continuity. This process, implemented in Python with pandas and NumPy, produced a fully cleaned and consistent time series for each municipality, ready for forecasting.

The area harvested dataset, obtained from the same source, underwent the same pre-processing procedures to maintain consistency and accuracy. Missing values were interpolated to fill data gaps, while extreme values detected through the IQR method were treated as anomalies and corrected using local smoothing. Forward and backward filling were again applied to handle any remaining boundary gaps. The resulting dataset provided a complete, uniform, and reliable time series for use in forecasting rice production in Northern Samar.

***C. Forecasting of Explanatory Variables***

To enable long-term projections of rice production, two key explanatory variables were forecasted independently: (1) area harvested and (2) number of tropical cyclones per municipality. Each variable was modeled using techniques best suited to its data behavior and statistical characteristics.

The annual area harvested per municipality was forecasted using Random Forest Regression, an ensemble learning technique that combines multiple decision trees to produce accurate and stable predictions. Historical data from 2016 to 2024 served as the training set, after which the model was used to generate forecasts for 2030, 2035, and 2040. To prevent unrealistic results, the predicted values were capped at the maximum historical area harvested for each municipality. Model performance was evaluated using standard error metrics—Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination ($R^2$), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE)—to ensure the accuracy and reliability of the predictions.

For the number of tropical cyclones, preliminary models such as linear, polynomial, and ARIMA yielded poor accuracy due to the absence of a clear trend or seasonality. As a result, a constant mean forecast was adopted as a more reliable alternative. The average number of cyclones recorded from 2016 to 2024 was computed per municipality and used as the projected value for future years (2030, 2035, and 2040). This approach assumes that cyclone frequency follows a stationary mean pattern over time, reflecting the inherent variability yet relative stability of climatic disturbances in the region.

***D. Forecasting of Rice Production***

The forecasting of rice production in Northern Samar employed three modeling approaches:

*1) Multiple Linear Regression (MLR)*

The Multiple Linear Regression (MLR) model was used to forecast rice production in Northern Samar by examining its linear relationship with the year, area harvested, and number of typhoons. For each municipality, a separate MLR model was fitted using annual data from 2016 to 2024, and forecasts were generated for 2030, 2035, and 2040 using projected explanatory variables. The model is expressed as:

$$Y_t = \beta_0 + \beta_1 \times Year_t + \beta_2 \times Area_t + \beta_3 \times CycloneCount_t + \varepsilon_t \quad \text{Eq. 1}$$

Where $Y_t$ is the rice production at year *t*; $\beta_0$ is the intercept; $\beta_1$, $\beta_2$, and $\beta_3$ are the regression coefficients for the corresponding predictors; and $\varepsilon_t$ is the error term.

Due to the limited number of observations, no train-test split was applied. Model accuracy was assessed using in-sample fitted values and evaluated with Mean Absolute Error, Root Mean Squared Error, Coefficient of Determination, Mean Absolute Percentage Error, and Symmetric Mean Absolute Percentage Error.

*2) Random Forest (RF)*

The area harvested dataset, obtained from the same source, underwent the same pre-processing procedures to maintain consistency and accuracy.

The Random Forest (RF) Regression model was used to forecast rice production in Northern Samar by capturing complex, nonlinear relationships between variables. It constructs multiple decision trees and averages their outputs, making it robust to noise and overfitting—ideal for agricultural data with limited observations.

The response variable was annual rice production (in metric tons), while predictors included the year, area harvested (in hectares), and number of typhoons. Separate models were trained for each municipality using data from 2016–2024 to account for local variations in land use and climate. The model can be mathematically expressed as:

$$\hat{Y} = \frac{1}{N}\sum_{i=1}^{N} f_i(X) \quad \text{Eq. 2}$$



Corresponding Author: Kyle Arthur P. Sabilao

Where $\hat{Y}$ is the predicted rice production, $N$ is the number of decision trees in the ensemble, $f_i(X)$ is the prediction of the $i^{th}$ tree, and $X$ represents the input features (Year, Area Harvested, Number of Cyclones). Each tree learns a different subset of the data and variables, and the aggregation of these predictions improves the model's generalization performance.

Model performance was evaluated using in-sample fitted values compared with actual observations. The same evaluation metrics as in the MLR model—MAE, RMSE, $R^2$, MAPE, and SMAPE—were used to assess accuracy.

*3) SARIMAX Model*

The Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX) model was applied to forecast rice production in Northern Samar. This model extends the ARIMA framework by integrating external variables that affect the target variable. In this study, rice production (in metric tons) served as the dependent variable, while the area harvested (in hectares) and the number of typhoons per year were used as exogenous regressors. It is represented as:

$$SARIMAX\ (p, d, q)\ (P, D, Q)_s$$

Where: $p$ is the non-seasonal autoregressive order (AR); $d$ is the degree of differencing (I); $q$ is the non-seasonal moving average order (MA); $P$, $D$, $Q$ are the seasonal components; and $s$ is the seasonality length. The model equation is given by:

$$Y_t = \mu + \emptyset_1 Y_{t-1} + \theta_1 \varepsilon_{t-1} + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t \qquad \text{Eq. 3}$$

Where: $Y_t$ is the rice production at year $t$; $\mu$ is the model intercept; $\emptyset_1$ is the autoregressive coefficient; $\theta_1$ is the moving average coefficient; $X_{1t}$ is the area harvested at time $t$, $X_{2t}$ is the number of cyclones at time $t$; $\beta_1$ and $\beta_2$ are the coefficients of the exogenous regressors; and $\varepsilon_t$ is the error term at time $t$.

Each municipality's model was trained on annual data from 2016–2024 and used to forecast rice production for 2030, 2035, and 2040 based on projected explanatory variables. To assess the model's performance, the same error metrics were used as in the evaluation of the other forecasting models. These metrics were computed for each municipality and aggregated to assess overall model performance.

***E. Comparative Analysis Framework***

To assess the performance of the forecasting models—SARIMAX, Random Forest, and Multiple Linear Regression (MLR)—a comparative analysis framework was established to determine which model provided the most reliable and accurate rice production forecasts across Northern Samar. The evaluation used a standardized error metric, the Mean Absolute Percentage Error (MAPE), ensuring consistent comparison of model accuracy and predictive stability. For each municipality, the computed MAPE values from all three models were compiled and normalized to highlight performance variations across the province. Aggregated results from combined in-sample fitted and actual values were also analyzed to evaluate overall provincial-level accuracy. The uniform evaluation process ensured fairness, consistency, and interpretability, ultimately identifying the most suitable forecasting approach for data-driven agricultural planning in Northern Samar.

## 5. RESULTS AND DISCUSSION

***A. Data Cleaning and Forecasting Results of Area Harvested and Cyclone Frequency***

Before forecasting, the rice production dataset underwent preprocessing to address missing values (mainly from 2017–2020) and statistical outliers through linear interpolation, the Interquartile Range (IQR) method, and smoothing techniques such as local averaging and edge-value imputation. San Vicente was excluded from the analysis because it had only one recorded value (2016). The cleaned dataset revealed clear production trends across 23 municipalities—Catarman, Catubig, Las Navas, and Laoang showed consistently high outputs, while Biri, Silvino Lobos, and San Antonio had persistently low yields (below 500 metric tons), likely due to limited farmland or environmental constraints. Similarly, the preprocessed area harvested dataset showed complete and realistic temporal patterns, with Catarman, Catubig, Las Navas, and Laoang recording the largest cultivated areas, while Biri, San Antonio, and Silvino Lobos had the smallest. San Roque's sharp increase in 2019 suggested expansion in cultivation or improved reporting.

For forecasting, two explanatory variables—area harvested and tropical cyclone count—were projected independently. The area harvested was modeled using Random Forest Regression trained on 2016–2024 data, effectively capturing nonlinear temporal patterns and producing forecasts for 2030, 2035, and 2040 capped at each municipality's historical maximum. Model evaluation yielded strong results: $R^2 = 0.806$, MAE = 98.57, RMSE = 120.91, MAPE = 26.19%, and SMAPE = 12.75%, indicating reliable predictive accuracy. In contrast, cyclone frequency showed no clear trend under linear, polynomial, or ARIMA models, leading to the adoption of a constant mean forecast of 2.78 cyclones per year, representing stable long-term behavior from 2016 to 2024.

***B. Forecasting Results of Rice Production***

*1) Multiple Linear Regression (MLR)*

The Multiple Linear Regression (MLR) model exhibited strong predictive performance in forecasting rice production across several municipalities of Northern Samar. Model accuracy varied across localities, with high precision observed in municipalities characterized by stable production patterns and consistent explanatory variables, such as Lapinig (MAPE: 2.25%), Las Navas (MAPE: 3.91%), and Catarman (MAPE: 5.48%). In contrast, areas like Biri (MAPE: 187.21%) and San Roque (MAPE: 94.06%) produced higher errors, reflecting the effects of data volatility, irregular production behavior, or gaps in historical records.

A summary graph (Fig. 2) illustrates the province's total projected rice production, indicating an overall upward trajectory toward 2040. This trend implies that, assuming stable land use and climatic conditions, Northern Samar may continue to experience gradual increases in rice output in the coming years.



Corresponding Author: Kyle Arthur P. Sabilao

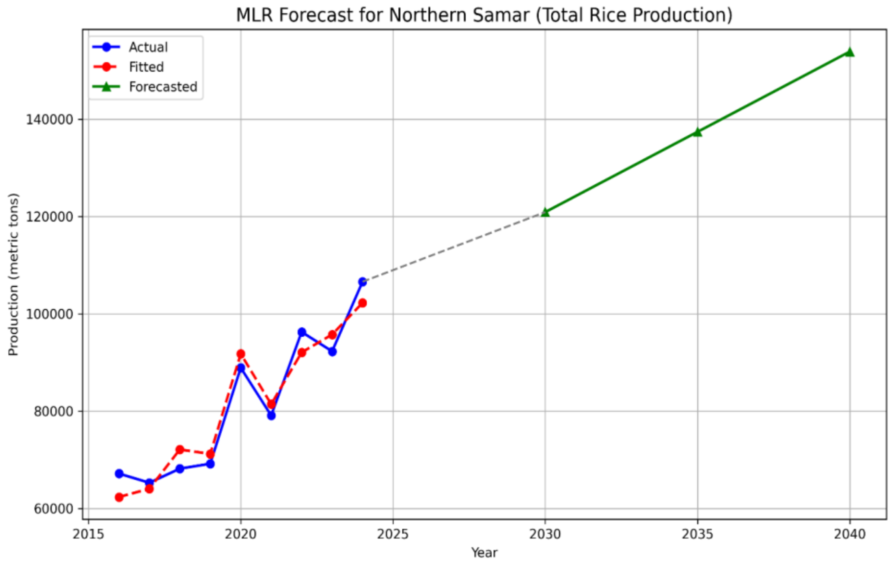


**Fig. 2. Rice Production Forecast using MLR in Northern Samar**

*2) Random Forest (RF)*

The evaluation of the three forecasting approaches—Multiple Linear Regression (MLR), Random Forest Regression (RF), and SARIMAX—

The Random Forest (RF) Regression model produced highly accurate forecasts of rice production across several municipalities in Northern Samar. Most areas exhibited low prediction errors, particularly Lapinig (MAPE: 1.63%), Laoang (MAPE: 3.58%), and Lavezares (MAPE: 3.40%), where consistent production trends and reliable input variables allowed the model to capture complex relationships effectively. In contrast, Biri (MAPE: 399.64%) and San Roque (MAPE: 157.04%) recorded higher errors, likely due to irregular data patterns and outlier values in historical records that affected the model's learning process.

However, the projected production values for 2030, 2035, and 2040 appeared constant across these years, suggesting that the model's outputs were influenced by repetitive input variables and the applied capping strategy, which restricted predictions to the historical maximum. This constraint resulted in conservative forecasts that may underestimate potential growth under enhanced farming practices or climate adaptation measures. Despite this limitation, the RF model provided a reliable baseline scenario for production stability under current conditions. The overall provincial trend, as shown in Figure 3, illustrates a plateau in rice output extending toward 2040.

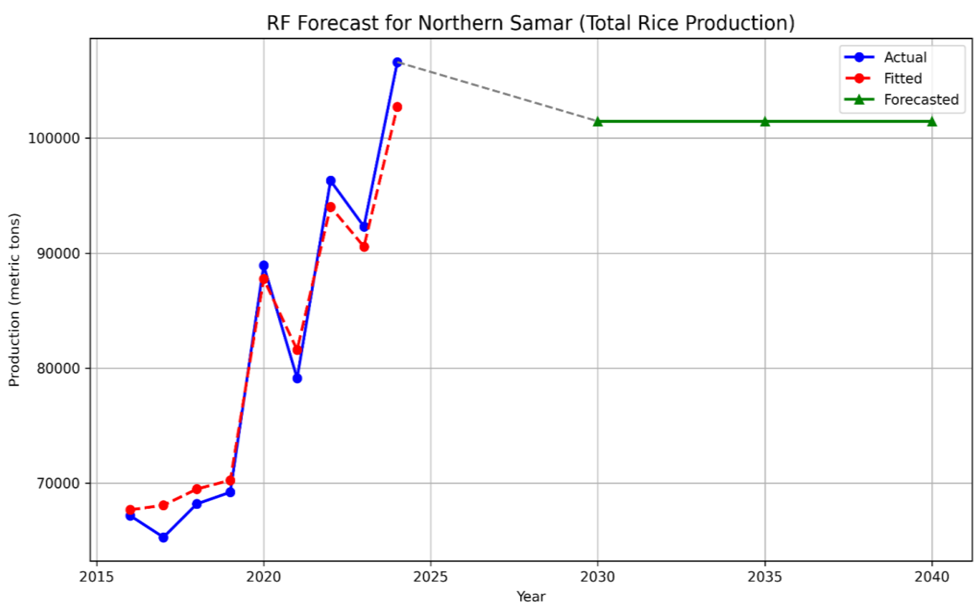


**Fig. 3. Rice Production Forecast using RF in Northern Samar**

*3) SARIMAX Model*

The SARIMAX model generated mixed results in forecasting rice production across the municipalities of Northern Samar, integrating time series dynamics with exogenous variables such as area harvested and typhoon count. Municipalities like Lapinig (MAPE: 8.63%), Lavezares (MAPE: 5.31%), Las Navas (MAPE: 4.63%), and Allen (MAPE: 6.56%) exhibited relatively low errors, suggesting that SARIMAX effectively captured their temporal dependencies and responses to the explanatory factors. These results indicate that in areas with stable year-to-year production and consistent environmental conditions, the SARIMAX model performs adequately in representing local agricultural trends.

Conversely, municipalities such as Mapanas (MAPE: 62.78%) and San Roque (MAPE: 168.58%) displayed substantial forecast errors, reflecting the model's difficulty in handling irregular and nonlinear production patterns. These discrepancies may stem from abrupt changes in output or weak correlations with the included predictors. Despite these limitations, the provincial-level forecast (Figure 4) shows mild fluctuations in projected rice production from 2030 to 2040, suggesting that SARIMAX retained some sensitivity to temporal variations while maintaining stable long-term projections.

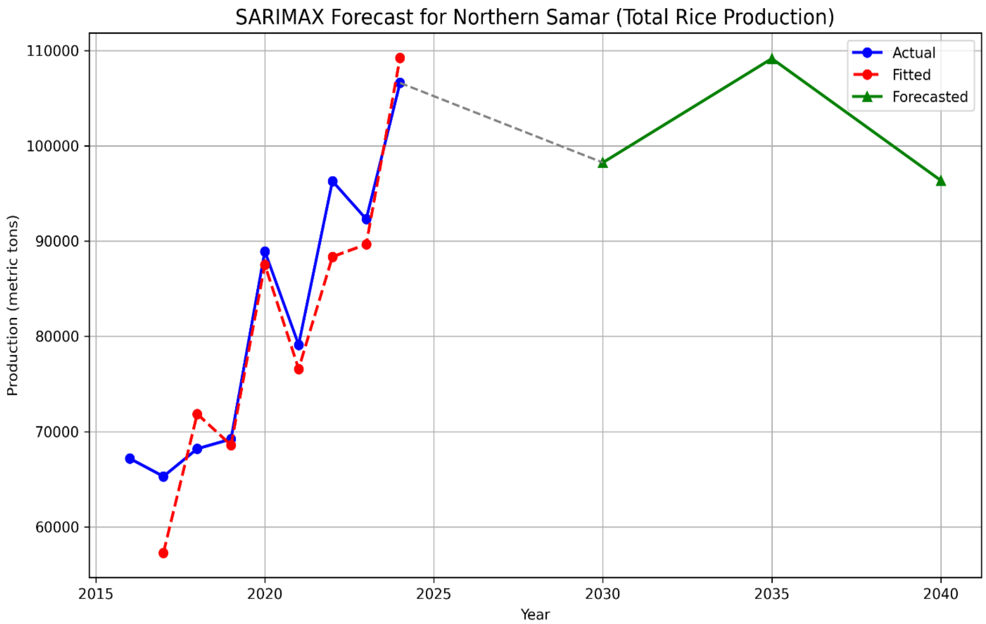


**Fig. 4. Rice Production Forecast using SARIMAX in Northern Samar**

***C. Comparative Summary of Model Accuracy Across Municipalities***

The evaluation of the three forecasting approaches—Multiple Linear Regression (MLR), Random Forest Regression (RF), and SARIMAX—highlighted notable differences in predictive accuracy and trend representation across the municipalities of Northern Samar. Using Mean Absolute Percentage Error (MAPE) as the primary evaluation metric ensured fair comparison across areas with varying production scales. The Random Forest model produced constant forecast values for 2030, 2035, and 2040 (Fig. 3), resulting in a flat projection pattern influenced by capped future inputs and a tendency to regress toward historical maximums. While effective for generating conservative baseline estimates, it was less capable of capturing dynamic growth. The MLR model, in contrast, showed an upward trend (Fig. 2), reflecting increases in area harvested and suggesting steady productivity improvements. Meanwhile, SARIMAX forecasts (Fig. 4) displayed minor fluctuations, revealing the model's sensitivity to temporal variations and external factors, though its accuracy differed across municipalities.

**Table 1 MAPE Comparison per Municipality Across Models**

| Municipality | MLR (%) | RF (%) | SARIMAX (%) |
|---|---|---|---|
| Allen | 10.59 | 7.74 | **6.56** |
| Biri | 187.21 | 399.64 | **25.43** |
| Bobon | 16.68 | **8.68** | 13.07 |
| Capul | 15.86 | **6.91** | 20.63 |
| Catarman | **5.48** | 10.91 | 10.27 |
| Catubig | 8.55 | **7.44** | 12.82 |
| Gamay | 19.64 | **11.85** | 18.92 |
| Laoang | 7.23 | **3.58** | 8.67 |
| Lapinig | 2.25 | **1.63** | 8.63 |
| Las Navas | **3.91** | 4 | 4.63 |
| Lavezares | 5.71 | **3.4** | 5.31 |
| Lope de Vega | 9.4 | **8.08** | 21.94 |
| Mapanas | 50.66 | **25.54** | 62.78 |
| Mondragon | 16.65 | **6.89** | 14.24 |
| Palapag | 6.97 | **5.86** | 13.01 |
| Pambujan | 14.58 | **7.92** | 19.82 |
| Rosario | 10.21 | **7.91** | 16.18 |
| San Antonio | 17.66 | **14.48** | 21.81 |
| San Isidro | 8.8 | **5.28** | 14.24 |
| San Jose | 15.31 | **12.44** | 27.85 |
| San Roque | **94.06** | 157.04 | 168.58 |
| Silvino Lobos | 17.13 | **8.17** | 17.19 |
| Victoria | **7** | 7.21 | 10.95 |

Comparative analysis of MAPE values across 23 municipalities showed that Random Forest Regression achieved the lowest errors in 17 municipalities, surpassing MLR and SARIMAX in overall consistency and precision. MLR yielded the best results in 4 municipalities, particularly where production patterns followed a linear and stable trajectory. SARIMAX performed best in only 2 municipalities—Allen (6.56%) and Biri (25.43%)—indicating its selective effectiveness under certain conditions. As summarized in Table 1, Random Forest emerged as the most accurate and stable short-term model, though limited in projecting long-term growth. MLR provided realistic upward trends suitable for long-range forecasting, while SARIMAX, despite effectively modeling temporal dependencies, displayed inconsistent accuracy across the province.

## 6. CONCLUSION

This study forecasted rice production across 23 municipalities in Northern Samar using cleaned historical data and two key predictors: area harvested and tropical cyclone frequency. MLR emerged as the most practical model for forward-looking agricultural planning, showing upward trends across the forecast period and performing best in 4 municipalities, making it suitable for scenarios assuming gradual improvements in agricultural capacity. RF achieved the highest short-term accuracy in 17 municipalities but produced flat forecasts for 2030, 2035, and 2040 due to repeated inputs and capped values, making it more appropriate for conservative or no-growth planning. SARIMAX showed limited reliability, performing well in only 2 municipalities. A careful match between each model, the data structure, and specific planning goals is essential to ensure accurate and meaningful forecasts that support sustainable agriculture and food security in the province.

## 7. RECOMMENDATIONS

To strengthen forecasting accuracy and support agricultural planning in Northern Samar, local agencies should standardize data collection, integrate additional predictive variables, and develop a centralized digital forecasting tool. Moreover, the proposed framework can be adapted for other crops and regions to enhance agricultural resilience and evidence-based policymaking.

**REFERENCES**


1) Aderele, M. O., Srivastava, A. K., Butterbach-Bahl, K., & Rahimi, J. (2025). Integrating machine learning with agroecosystem modelling: Current state and future challenges. *European Journal of Agronomy*, *168*, 127610. https://doi.org/10.1016/j.eja.2025.127610
2) Ahmad Hamdan, Kenneth Ifeanyi Ibekwe, Emmanuel Augustine Etukudoh, Aniekan Akpan Umoh, & Valentine Ikenna Ilojianya. (2024). AI and machine learning in climate change research: A review of predictive models and environmental impact. *World Journal of Advanced Research and Reviews*, *21*(1), 1999–2008. https://doi.org/10.30574/wjarr.2024.21.1.0257
3) Cuyco, R., Balag, J., Castillo, M., Garbo, S., & Staff, M.-L. (2023). *Climate and Disaster Risk Assessment (CDRA) Report 2021—Lavezares, Northern Samar*.
4) Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, *2*(1), 100058. https://doi.org/10.1016/j.jjimei.2022.100058
5) Gulati, J. (2025, March 12). Mastering Time Series Forecasting: From ARIMA to LSTM. *MachineLearningMastery.Com*. https://machinelearningmastery.com/mastering-time-series-forecasting-from-arima-to-lstm/
6) Hia, S., Kuswanto, H., & Prastyo, D. D. (2023). Robustness of Support Vector Regression and Random Forest Models: A Simulation Study. In Y. B. Wah, M. W. Berry, A. Mohamed, & D. Al-Jumeily (Eds.), *Data Science and Emerging Technologies* (pp. 465–479). Springer Nature. https://doi.org/10.1007/978-981-99-0741-0_33
7) Ibañez, S., & Monterola, C. (2023, September 21). *A Global Forecasting Approach to Large-Scale Crop Production Prediction with Time Series Transformers*. https://www.mdpi.com/2077-0472/13/9/1855
8) Jabed, Md. A., & Azmi Murad, M. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, *10*(24), e40836. https://doi.org/10.1016/j.heliyon.2024.e40836
9) Kolambe, M. (2024). Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques. *Journal of Electrical Systems*, *20*, 575–586. https://doi.org/10.52783/jes.1478
10) Kontopoulou, V., Panagopoulos, A., Kakkos, I., & Matsopoulos, G. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet*, *15*, 255. https://doi.org/10.3390/fi15080255
11) Kumari, P., Chandan, G., Agarwal, D., Joshi, S., & Vekariya, P. (2024). Combining predictive models: Hybrid approaches in crop recommendation systems. *International Journal of Research in Agronomy*, *7*(11), 108–115. https://doi.org/10.33545/2618060X.2024.v7.i11b.1951
12) Marticio, M. T. (2024, September 27). *Manila Bulletin—Northern Samar, firm partner to hike rice production*. https://mb.com.ph/2024/09/27/northern-samar-firm-partner-to-hike-rice-production
13) Meinke, H., & Stone, R. C. (2005). Seasonal and Inter-Annual Climate Forecasting: The New Tool for Increasing Preparedness to Climate Variability and Change In Agricultural Planning And Operations. *Climatic Change*, *70*(1–2), 221–253. https://doi.org/10.1007/s10584-005-5948-6
14) *Northern Samar Farmers to enrol in Rice-to-Rise Program for Sustainable Farming | Province of Northern Samar*. (2024, November 13). https://northernsamar.gov.ph/northern-samar-farmers-to-enrol-in-rice-to-rise-program-for-sustainable-farming/
15) Parreño, S. J. (2023). Forecasting quarterly rice and corn production in the Philippines: A comparative study of seasonal ARIMA and Holt-Winters models. *ICTACT Journal on Soft Computing*, *14*, 3224–3231. https://doi.org/10.21917/ijsc.2023.0449
16) Parreño, S. J., & Anter, M. (2024). New approach for forecasting rice and corn production in the Philippines through machine learning models. *Multidisciplinary Science Journal*, *6*, 2024168. https://doi.org/10.31893/multiscience.2024168
17) Rathod, S., Saha, A., Patil, R., Ondrasek, G., Gireesh, C., Anantha, M. S., Rao, D. V. K. N., Bandumula, N., Senguttuvel, P., Swarnaraj, A. K., Meera, S. N., Waris, A., Jeyakumar, P., Parmar, B., Muthuraman, P., & Sundaram, R. M. (2021). Two-Stage Spatiotemporal Time Series Modelling Approach for Rice Yield Prediction & Advanced Agroecosystem Management. *Agronomy*, *11*(12), 2502. https://doi.org/10.3390/agronomy11122502
18) Shah, D., & Thaker, M. (2024). A Review of Time Series Forecasting Methods. *INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS*, *11*, 749. https://doi.org/10.1729/Journal.38816
19) Shah, P., Pahari, S., Bhavsar, R., & Kwon, J. S.-I. (2025). Hybrid modeling of first-principles and machine learning: A step-by-step tutorial review for practical implementation. *Computers & Chemical Engineering*, *194*, 108926. https://doi.org/10.1016/j.compchemeng.2024.108926
20) Slater, L., Arnal, L., Boucher, M.-A., Chang, A. Y.-Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A., & Zappa, M. (2022). *Hybrid forecasting: Using statistics and machine learning to integrate predictions from dynamical models*. https://doi.org/10.5194/hess-2022-334
21) Yadav, A. S. (2024). *Unveiling Patterns in Time Series Forecasting Models—Arima, Arimax, Var*. *11*, 1335–1345.