Journal of Tropical Resources and Sustainable Science
Website : http://journal.umk.edu.my/index.php/jtrss/
eISSN : 2462-2389
Vol: 14, Issue 1, 2026, 199-
DOI: https://doi.org/10.47253/jtrss.v14i1.2155

RESEARCH ARTICLE

Impact of exogenous variables on paddy productivity modeling: an ARIMAX model perspective

Kausalyaa A/P R. Thibagaran[1], Siti Mariam Norrulashikin[1*], Siti Rohani Mohd Nor[1], Aimi Athirah Ahmad[2], and Nur Arina Bazilah Kamisan[1]

[1]Faculty of Science, Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, Malaysia
[2]Malaysia Agriculture Research and Development Institute (MARDI), Persiaran MARDI-UPM, 43400, Serdang, Selangor, Malaysia

ARTICLE HISTORY
Received : 5 November 2025
Accepted : 10 January 2026
Online : 31 March 2026

KEYWORDS
paddy,
paddy production,
ARIMAX,
exogenous,
forecasting

☒ * **CORRESPONDING AUTHOR**
Dr. Siti Mariam Norrulashikin
Faculty of Science,
Universiti Teknologi Malaysia,
81310 Johor Bahru, Johor,
Malaysia.
Email: sitimariam@utm.my

ABSTRACT

Paddy productivity in Malaysia faces escalating difficulties in 2024–2025 due to increasing input costs, unpredictable weather patterns, and rising dependence on rice imports. These problems have intensified the necessity for precise forecasting instruments to inform food security policies and agricultural strategies. The present study examined the effectiveness of ARIMA and ARIMAX models in predicting yearly paddy productivity in the states in the northern region of Peninsular Malaysia, which were Kedah, Perak, Pulau Pinang, and Perlis, based on data from 1981 to 2022. The study investigated whether the inclusion of an exogenous factor, which was planted area, may improve forecasting accuracy beyond conventional univariate models. A systematic approach was employed involving stationarity assessment, model identification, residual diagnostics, and performance evaluation. The results indicated that ARIMAX modeling's performance consistently surpassed that of ARIMA modeling, with Kedah achieving the best accuracy, closely followed by Perlis, Perak, and Pulau Pinang. The incorporation of exogenous variables markedly enhanced model responsiveness and accounted for structural changes in paddy production. The residuals from the final model for each state exhibited no indications of autocorrelation, hence confirming statistical validity. The study finds that ARIMAX modeling offers a dependable and comprehensible forecasting framework for paddy output and can function as an essential decision-support instrument for agricultural policies, particularly during times of supply instability. The methodology is versatile for many crops and geographies, facilitating extensive applications in agricultural forecasting and food security strategies.

1. INTRODUCTION

Paddy productivity is an essential element of global food security, especially in Asia, which accounts for around 90% of the world's rice, while exporting merely 7%, highlighting its significance in domestic food systems (Firdaus et al., 2020). In Malaysia, where rice is a basic dietary component, an area such as Kedah, which is designated as the country's “rice bowl”, depends significantly on paddy productivity to maintain national food security, reduce import dependency, and protect the nation from fluctuations in world prices (Omar et al., 2019). In addition to producing food, the rice industry significantly contributes to social stability by maintaining rural jobs and sustaining local economies. Recent data indicate a decline in Malaysia’s paddy production from 2.28 million metric tons in 2022 to 2.17 million in 2023, which emphasizes the need to tackle various issues, such as those related to climate change, water shortages, and the expansion of market needs, for enhanced yield-prediction strategies to facilitate decision-making despite increasing population pressures and dietary changes (Muslim, 2025).

The Malaysian government has used policies, subsidies, and technological innovations to mitigate production issues and improve agricultural productivity and sustainability (Makhtar et al., 2022; Dorairaj & Govender, 2023). Malaysia's self-sufficiency in rice persists at 60% to 70%, reflecting a continued dependence on imports that may threaten food security during global supply disruptions. Emphasizing climate-resilient and sustainable agricultural methods is essential for ensuring long-term productivity; nevertheless, precise forecasting is made difficult by external variables, such as inconsistent planting areas. Current research is lacking thorough studies that evaluate predictive models designed for local situations, as the focus is either on advanced machine learning or conventional methods. The present study advocates the development of robust forecasting models, including ARIMA and ARIMAX models, which utilize local agricultural data to enhance the precision of regional paddy production predictions and optimize strategic agricultural planning.

An ARIMA model is widely used for time series forecasting in agriculture due to its ability to identify trends and seasonality using historical data; however, its dependence on linearity assumptions and past observations limits its accuracy in dynamic agricultural environments (Pandit et al., 2023; Hamjah, 2014; Sivapathasundaram & Bogahawatte, 2015). Although it offers short-term forecasting benefits under stationary conditions via pre-processing techniques, such as differencing to achieve stationarity, this process can lead to loss of information, which causes misleading interpretations of the model (Yogarajah et al., 2013). The exclusion of key external variables, such as climate factors or planted area, reduces its effectiveness in regions with fluctuating conditions, such as Kedah (Pandit et al., 2022; Md Yusof et al., 2019). While the ARIMA model has applications beyond agriculture, such as wind speed forecasting (Elsaraiti & Merabet, 2020), its inability to account for nonlinear environmental influences makes it insufficient for today's climate-sensitive agricultural forecasting needs (Dorairaj & Govender, 2023). Hence, there is a pressing need to adopt more advanced models that can integrate external factors or variables to improve prediction accuracy in paddy productivity forecasting.

The ARIMAX model, which is an extension of the ARIMA model, incorporates external variables, such as planted area, rainfall, and weather conditions, making it a more accurate and reliable forecasting tool for agricultural productivity as compared with the ARIMA model (Pandit et al., 2023; Pokhrel & Adhikari, 2023). Studies on spice crops in Bangladesh and Tamil Nadu, as well as rice production in Nepal, confirmed that an ARIMAX model improves forecasting accuracy when external influences are considered (Hamjah, 2014; Pokhrel & Adhikari, 2023; Sujatha & Sivasankari, 2023). Unlike the ARIMA model, an ARIMAX model can handle dynamic data and still provide accurate predictions even with missing data (Hyndman & Athanasopoulos, 2018; Montgomery et al., 2015). Its ability to reflect the impact of environmental factors enhances its suitability for agricultural forecasting by offering insights into how external variables influence yield fluctuations (Ray & Bhattacharyya, 2020; Ahmad et al., 2020). This model supports data-driven planning to mitigate risks, such as weather disruptions and land mismanagement, as shown in its practical application in Sri Lanka's Trincomalee district in Yogarajah et al. (2013).

While an ARIMAX model can enhance forecasting accuracy by incorporating external variables, its predictors need to be carefully selected to avoid overfitting (Pandit et al., 2023). Some important exogenous factors, such as planted areas, rainfall, temperature, irrigation, and fertilizer use, significantly influence agricultural yield. By integrating these variables, the ARIMAX model may outperform the ARIMA model in forecasting scenarios where external influences are critical, as shown in Sudipa et al.'s (2024) case study on rice price fluctuations in East Nusa Tenggara. The inclusion of planted area data is significantly important in capturing the supply-side dynamics, as increased cultivation can lead to production surpluses and price drops under stable demand (Ray & Bhattacharyya, 2020; Ghosh et al., 2014). This variable is shaped by government policies and market, seasonal, and environmental conditions, such as subsidies or weather extremes, which makes its inclusion crucial for adapting to real-world changes (Ahmad et al., 2020). Ultimately, the ability of the ARIMAX model to reflect these dynamics enhances its accuracy and reliability in agricultural forecasting (Hyndman & Athanasopoulos, 2018).

While prior studies have predominantly relied on machine learning models or national-level data for agricultural forecasting, relatively few have explored statistical time-series approaches at the regional level. This creates a gap in understanding localized agricultural dynamics, which are crucial for targeted policies and adaptive planning. The ARIMAX model, as a transparent and interpretable model, offers a valuable framework for integrating exogenous variables into forecasting, making it particularly relevant for applications in sustainable agriculture and disaster risk reduction. The present study evaluated and compared the forecasting performances of ARIMA and ARIMAX modeling in the Malaysian context, incorporating planted area as the key external factor. By focusing on state-level data, this research provides context-specific insights that can aid farmers and policymakers in enhancing planning precision and agricultural resilience (Pandit et al., 2023; Elsaraiti & Merabet, 2020).

2. MATERIALS AND METHODS

This study presents a systematic methodology for forecasting paddy production using ARIMA and ARIMAX models, beginning with data collection and exploratory data analysis to understand data behavior. Various steps, such as testing for normality and ensuring data stationarity, were used for accurate modeling. Univariate time series data were applied in the ARIMA model, while the ARIMAX model incorporated exogenous variables to capture external influences. The study also conducted diagnostic checks and performance evaluations to validate the accuracy of the models. This structured approach aimed to ensure reliable forecasting outcomes that indirectly can support the agricultural planning and decision-making process (Makhtar et al., 2022; Dorairaj & Govender, 2023).

2.1. Pre-processing analysis

Pre-processing analysis transforms raw data into a format that is suitable for statistical and machine learning models. This encompasses necessary normality tests, stationarity checks, and trend analysis in order to have a dataset that does not violate the assumptions of the model and is best positioned for accurate forecasting.

Statistical models mostly rely on the normality of a dataset. Normality testing is typically conducted using the Anderson-Darling (AD) method to ensure the validity of the analysis, as the method is mainly sensitive to deviations from the normal distribution and thus can identify possible issues that might affect the result. It also helps to identify if a dataset follows the normal distribution by comparing the *p*-value with the significance level. A significance level of 0.05 indicates that there is 5% risk of a Type I error occurring. The hypotheses are as follows:

H_0: Data is normally distributed

H_1: Data is not normally distributed

The null hypothesis (H_0) is rejected if the *p*-value is significantly less than the significance level, indicating that the data follows a normal distribution. The formula for the AD test is as follows:

$$A^2 = -n - S$$

where *n* is the sample size, while *S* is defined as:

$$S = \frac{1}{n}\sum_{i=1}^{n}[(2i-1)(ln(F(X_i)) + ln(1 - F(X_n + 1 - i)))]$$

Stationarity testing is an important process in determining whether a time series is stationary or non-stationary. The underlying stationarity assumption of ARIMA and ARIMAX models relies on a stable mean and constant variance over time. The stationarity of a dataset is analyzed by using the Augmented Dickey-Fuller (ADF) test, which is the standard statistical test applied to detect the presence of a unit root, indicating that the data series is non-stationary. The formula for the ADF test is as follows:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p}\delta_i \Delta y_{t-i} + \epsilon_t$$

The null hypothesis is rejected if the *p*-value is below the 5% significance level. Thus, the time series is stationary, and its statistical properties remain constant over the period of time. On the other hand, transformation or differencing techniques are required for non-stationary data to stabilize a time series and improve model accuracy.

2.2. ARIMA modeling

One popular statistical method for univariate time series forecasting is the ARIMA model. The three basic elements of the ARIMA model are the Autoregressive (AR), Integrated (I), and Moving Average (MA) processes. The AR component captures the influence of a past value on the current value, effectively utilizing lagged relationships in the dataset to predict future outcomes. The Integrated component takes care of non-stationarity by applying transformation or differencing techniques to the data series to achieve stationarity. Finally, the MA component accounts for dependence between current values and past forecast errors, providing better predictions because of the allowance of residual variations. All these put together allow the ARIMA model to predict linear trends and patterns in time series data. The general equation is as follows:

$$y_t = c + \sum_{i=1}^{p}\phi_i y_{t-i} + \sum_{j=1}^{q}\theta_j \epsilon_{t-j} + \epsilon_t$$

Parameters *p*, *d*, and *q* stand for the AR, differencing, and MA orders, respectively. Normally, these are determined using diagnostic tools, such as the autocorrelation function (ACF) plot, partial autocorrelation function (PACF) plot, and Akaike information criterion (AIC). An ARIMA model is fitted when the data are stationary to capture linear patterns and seasonal trends.

2.3. ARIMAX modeling

The ARIMAX model, which is an advanced ARIMA model, integrates with exogenous variables that influence the dependent variable over time. This model is capable of forecasting annual paddy productivity data with the inclusion of exogenous variables, such as planted area, rainfall, temperature, irrigation, or fertilizer usage, which have significant impacts on paddy productivity in Malaysia. The general equation is as follows:

$$y_t = c + \sum_{i=1}^{p}\phi_i y_{t-i} + \sum_{j=1}^{q}\theta_j \epsilon_{t-j} + \beta X_t + \epsilon_t$$

2.4. Diagnostic testing

To validate the accuracy and reliability of forecasting models, diagnostic checking needs to be conducted using residual analysis. Residual analysis is computed to check whether models capture all the patterns in the data and leave no missing patterns. Residuals are the difference between actual and predicted values at time *t*.

Randomness in residuals is checked to make sure that models capture all patterns. So, residuals are independent of each other and no remaining patterns are left based on the null hypothesis. The Ljung-Box Q-test is used to confirm whether residuals are independent or not. The formula for the Ljung-Box Q-test is as follows:

$$Q = n(n+2)\sum_{k=1}^{m}\frac{\hat{p}_k^2}{n-k}$$

For homoscedasticity, the Breusch-Pagan test is used to make sure that a model's errors are consistent and sized evenly over time. The test helps in determining whether errors are spread out equally. Otherwise, the model is not valid. The formula for the Breusch-Pagan test is as follows:

$$BP = \frac{nR^2}{2}$$

A significant *p*-value suggests heteroscedasticity, indicating that the residual variance is not constant over time and may require model adjustments. By confirming these assumptions, residual analysis ensures the robustness and reliability of a forecasting model without systematic errors. This enables the model to predict accurately and give more valid results.

2.5. Model comparison

In this study, the performance metrics of the ARIMA and ARIMAX forecasting models were evaluated based on the mean absolute percentage error (MAPE) between actual and predicted data. The evaluation of performance was guided by the principles outlined by Hamiane et al. (2024), where MAPE was used to measure the percentage deviation from actual values.

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right|$$

Metrics are evaluated to reduce forecasting error and enhance forecasting accuracy. The accuracy of a forecast can be measured without being affected by how big or small the data are. Metrics with lower values show that a model has a better performance in order to be chosen as the most suitable for forecasting. Yet, the conclusion might vary when external factors are incorporated. So, to verify the accuracy of the chosen model, actual and predicted values needed to be visually compared. To contextualize the forecasting performance of models, an interpretive scale is used, where a MAPE of less than 10% is considered highly accurate, 10%–20% is good, 20%–50% is reasonable, and greater than 50% is considered inaccurate (Lewis, 1982). This benchmark allows for meaningful comparison across different models and regions.

3. RESULT AND DISCUSSION

3.1. Data Collection

For this study, the data source used was a paddy dataset gathered from 13 states in Malaysia. The dataset underwent several refinement processes to maintain consistency and streamline for efficient analysis. Thus, data selection was focused on Peninsular Malaysia, and the scope of selection was further narrowed down to four states in the northern region of Peninsular Malaysia: Perak, Kedah, Pulau Pinang, and Perlis. These states are chosen due to their substantial role in paddy production and their variability in external factors, thereby enhancing the generalizability of findings for the research.

3.2. Descriptive analysis

The descriptive analysis for the dataset provided measures of central tendency, dispersion, and distribution that were calculated for paddy productivity and planted area in the selected states, as presented in Table 1. The dataset comprised yearly data from 1981 to 2022 for Perak, Kedah, Pulau Pinang, and Perlis. These measures provided valuable insights into data variability and central values, which were crucial for subsequent forecasting analysis.

Table 1: Descriptive statistics of paddy productivity (unit).

State	Mean	Standard Deviation	Minimum Value	Maximum Value
Kedah	3752.69	638.808	2002	4878
Pulau Pinang	4090.29	1202.21	1776	5872
Perak	3277.48	490.583	2461	4483
Perlis	3946.64	525.085	2471	5116

From Table 1, Pulau Pinang exhibited the highest mean productivity and also displayed a substantial standard deviation, denoting the greatest variability in productivity yield. Among the states analyzed, Perak recorded the lowest mean productivity, followed by Kedah and Perlis. Perak had the smallest standard deviation, which indicated that Perak exhibited the most stable paddy productivity level and less variability in yield.

Table 2 shows the overview of the planted area for the four states using the descriptive statistics, which were the mean, standard deviation, and minimum and maximum values. Kedah had the largest average planted area at 180774 hectares and the highest standard deviation of 64230.20. Perak had the second-largest average planted area at 77639.5 hectares and a standard deviation of 5916.87. The other two states had the smallest average planted areas, where Pulau Pinang and Perlis showed standard deviations of 3397.79 and 15280.3, respectively. The planted areas ranged between 46855 and 217053 hectares in Kedah, which shows that Kedah had the widest range among the four states. In contrast, Pulau Pinang had a narrower range of between 15190 and 28591 hectares.

Table 2: Descriptive statistics of planted area (hectare).

State	Mean	Standard Deviation	Minimum Value	Maximum Value
Kedah	180774	64230.2	46855	217053
Pulau Pinang	24389.3	3397.79	15190	28591
Perak	77639.5	5916.87	56998	87191
Perlis	43192.9	15280.3	8570	60478

Table 3 illustrates the relationship between paddy productivity and planted area across the four states, revealing varying correlations. Kedah demonstrated a strong and positive correlation, which highlights a strong positive relationship between paddy productivity and planted area. This indicates a suggestion for expanding the size of planted areas in Kedah, which can lead to higher productivity gains. Additionally, Kedah's status as Malaysia's "rice bowl", where advanced agricultural practices, including widespread mechanization, structured irrigation systems, and targeted government support, enhance the efficiency and responsiveness of cultivation efforts. Perlis presented a moderate correlation, while Pulau Pinang and Perak showed weaker correlations, suggesting that focusing on expanding planted areas may not be the most effective way. Instead, these other states with weaker or moderate correlations are required to address other influential factors, such as farming practices, soil quality, and other tailored approaches that are more beneficial.

Table 3: Correlation between paddy productivity and planted area.

State	Correlation Coefficient
Kedah	0.8478
Pulau Pinang	0.3961
Perak	0.3787
Perlis	0.5115

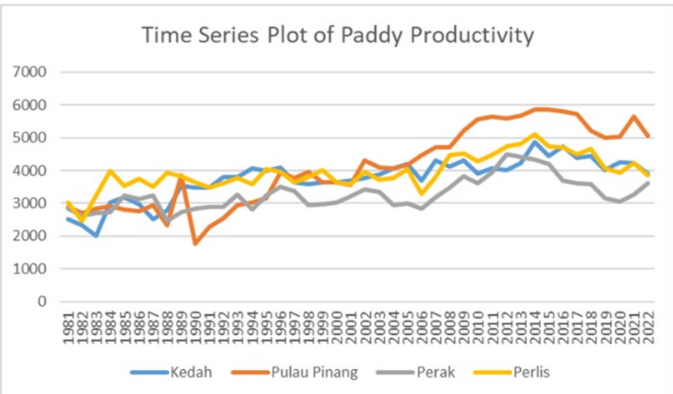

Figure 1: Time series plot.

The time series plot in Figure 1 provides an overview of the trend patterns of paddy productivity in the four states. The data revealed that Pulau Pinang consistently outperformed other states and maintained the highest productivity levels. The plot illustrates a significant upward trend of productivity in Pulau Pinang that reached the maximum point around the 30th index. However, a slight decrease occurring after the peak is observed. This suggests that there are potential factors that might impact the fluctuations of productivity in the state.

Perlis displayed a moderate level of productivity, with fluctuations around its constant mean. In contrast, Perak exhibited a moderate upward trend of productivity, followed by considerable variability, and the trend started to decline after the 30th index. Among the four states, Kedah consistently demonstrated the lowest level of productivity. Based on the figure, the overall trend remained in lower values with minor fluctuations.

3.3. Pre-processing analysis

Table 4 illustrates the result of the stationarity test, which used the Augmented Dickey-Fuller test to check whether the paddy productivity data remained statistically stable over time in Kedah, Pulau Pinang, Perak, and Perlis. Based on the result above, the *p*-values for all the states exceeded the 0.05 significance level, which indicates that the data series were not stationary. Therefore, the data in each region had trends over time that did not fluctuate around a constant mean and variance. It was possibly caused by changing weather, technology, or farming policies.

Table 4: Stationarity from Augmented Dickey-Fuller Test.

State	*p*-value	Stationarity
Kedah	0.8045	Non-stationary
Pulau Pinang	0.5835	Non-stationary
Perak	0.3849	Non-stationary
Perlis	0.7626	Non-stationary

Figure 2 illustrates the ACF and PACF plots for all the states, showing a gradual and slow decay in the plots across multiple lags. The ACF and PACF plots helped in identifying patterns in the data that can guide model selection. The ACF plot showed that the current year's productivity values were influenced by the previous year's values, which is common in agricultural data, where past conditions often affect present outcomes.

This pattern indicates the appearance of a non-stationary behavior of the time series in paddy productivity for each state. In contrast, a rapid decline was exhibited in the ACF plot for the stationary time series. The ACF declined rapidly after a few lags, as past observations caused future values to diminish quickly for stationary data. Yet, it is observed that autocorrelation was consistently present at early lags, which suggests that there existed a trend pattern or seasonality in the data that fulfilled the characteristics of non-stationary data.

Therefore, it was necessary to do the differencing of the time series to achieve stationarity in order to proceed to the subsequent analysis. After differencing, stationarity was successfully achieved in the time series data. Differencing is essential to ensure the accuracy of time series forecasting methods. Without this step, a forecast will be misleading due to underlying trends that have not been accounted for.

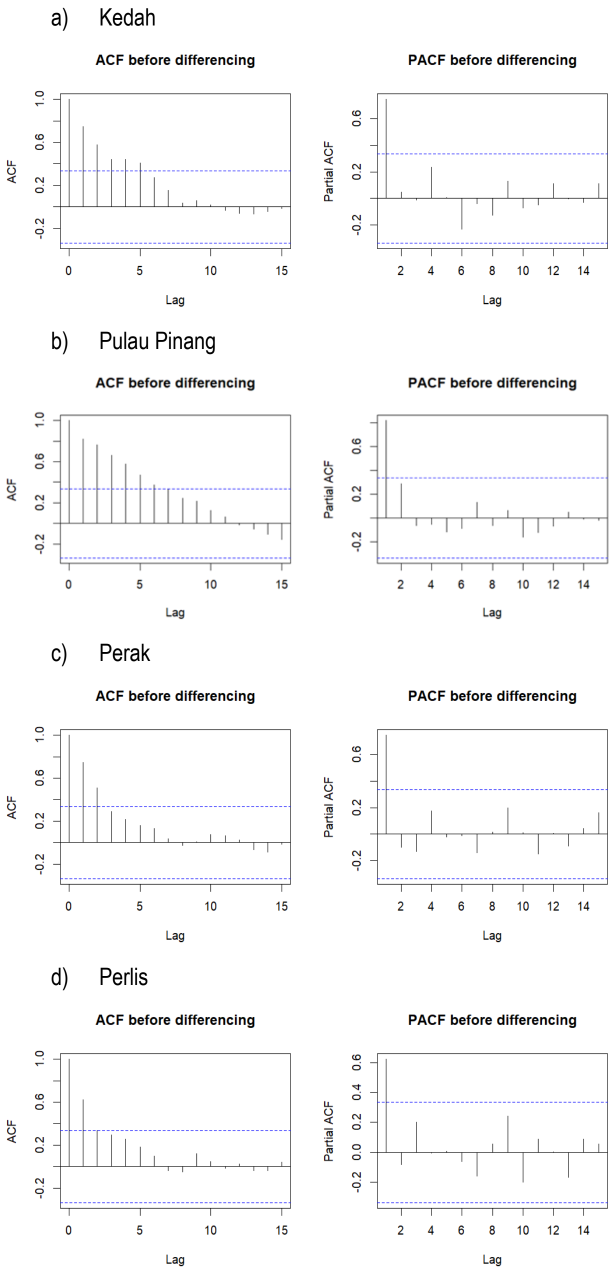

Figure 2: ACF and PACF plots for each state before differencing.

After the first differencing, the ACF plots for Kedah and Perlis displayed a significant spike at Lag 1, followed by a rapid decline, with subsequent lags falling within the confidence intervals. This pattern indicates the successful removal of long-term trends and the presence of short-term autocorrelation, which are characteristics of stationary data. However, the ACF plots for Pulau Pinang and Perak demonstrated slowed decay, which confirmed the presence of remaining non-stationary components and the need for more differencing.

A second differencing step was required for Pulau Pinang and Perak, since both states did not achieve stationarity after the first differencing. The second differencing successfully led to stationarity. Since both states had *p*-values below 0.05 in the ADF test, this allowed the rejection of the null hypothesis of a unit root. This outcome was significant because the second differencing effectively removed any residual errors and made the data suitable for robust ARIMA modeling.

3.4. ARIMA modeling

The next phase of the time series analysis was fitting ARIMA models to the differenced paddy productivity data for the four Malaysian states. The primary objective was to select the best-fit ARIMA model for each state. This was accomplished by comparing the various model configurations by using AIC and BIC criteria, with lower values indicating a better fit, and then performing diagnostic tests to ensure all model assumptions were met.

The selected ARIMA models for all four states are presented in Table 5, where the models featured autoregressive, and moving average orders ranging from 0 to 2. The table included model parameters and AIC and BIC values, as well as the outcomes of all diagnostic tests needed to validate the models. These tests were essential for assessing whether the residuals of the fitted models would exhibit any autocorrelations, non-normality, or heteroscedasticity. If there were remaining autocorrelations in residuals, this would indicate that there were uncaptured temporal dependencies.

These selected models were evaluated on whether the residuals passed the Ljung-Box test and the Breusch-Godfrey test by the *p*-values indicating no significant autocorrelation. This ensured the models were a good fit for the data, where they effectively captured the patterns of the dataset without unnecessary complexity. Moreover, each model met crucial criteria, as verified by the Shapiro-Wilk test for normality and the Breusch-Pagan test for homoscedasticity. These outcomes underlined model adequacy and reliability for precise forecasting.

The best-fitted model chosen for Kedah was ARIMA (0,1,1), as it had the lowest AIC value of 537.75 and the lowest BIC value of 540.98 among the models. Also, for Kedah, the *p*-value from the Ljung-Box test was 0.3538 and the *p*-value from the Breusch-Godfrey test was 0.5762, both indicating the absence of significant autocorrelation in the residuals, which suggests that the model was an excellent fit. The Shapiro-Wilk test for Kedah yielded a *p*-value of 0.1490, indicating no significant deviation from a normal distribution. Conversely, the model presented a *p*-value of 0.01457 from the Breusch-Pagan test, which suggests the presence of heteroscedasticity. This means that the residual variability for Kedah might change over time, which may affect the residual precision and accuracy of forecasting. Hence, variance-stabilizing transformations or robust standard errors might need to be applied to improve model reliability.

For the other states, the result validates the use of ARIMA modeling, with each meeting key statistical assumptions for reliable forecasting. The strong diagnostic outcomes, such as the absence of autocorrelation, normal distribution of residuals, and stable variance, demonstrated the robustness and reliability of the selected models, supporting their application in short-term productivity forecasting and enabling data-driven agricultural planning at the state level.

Table 5: Selected ARIMA models and diagnostics.

State	Best-fitted ARIMA model	AIC	BIC	Ljung-Box Test (*p*-value)	Breusch-Godfrey Test (*p*-value)	Shapiro-Wilk Test (*p*-value)	Breusch-Pagan Test (*p*-value)
Kedah	(0,1,1)	537.75	540.98	0.3538	0.5762	0.1490	0.01457
Pulau Pinang	(1,2,1)	550.60	555.35	0.9204	0.9193	0.1200	0.1941
Perak	(0,2,1)	518.85	522.02	0.0634	0.1522	0.3477	0.7741
Perlis	(2,1,0)	537.90	542.73	0.3868	0.6506	0.9864	0.1558

Table 6: Selected ARIMAX models and diagnostics.

State	Best-fitted ARIMAX model	AIC	BIC	Ljung-Box Test (*p*-value)	Breusch-Godfrey Test (*p*-value)	Shapiro-Wilk Test (*p*-value)	Breusch-Pagan Test (*p*-value)
Kedah	(0,1,1)	531.21	536.04	0.3101	0.222	0.1533	0.1215
Pulau Pinang	(1,2,1)	551.56	557.89	0.8016	0.7269	0.2083	0.1318
Perak	(0,2,1)	520.12	524.87	0.0746	0.1845	0.4122	0.7682
Perlis	(2,1,0)	539.38	545.83	0.1897	0.5141	0.9178	0.1686

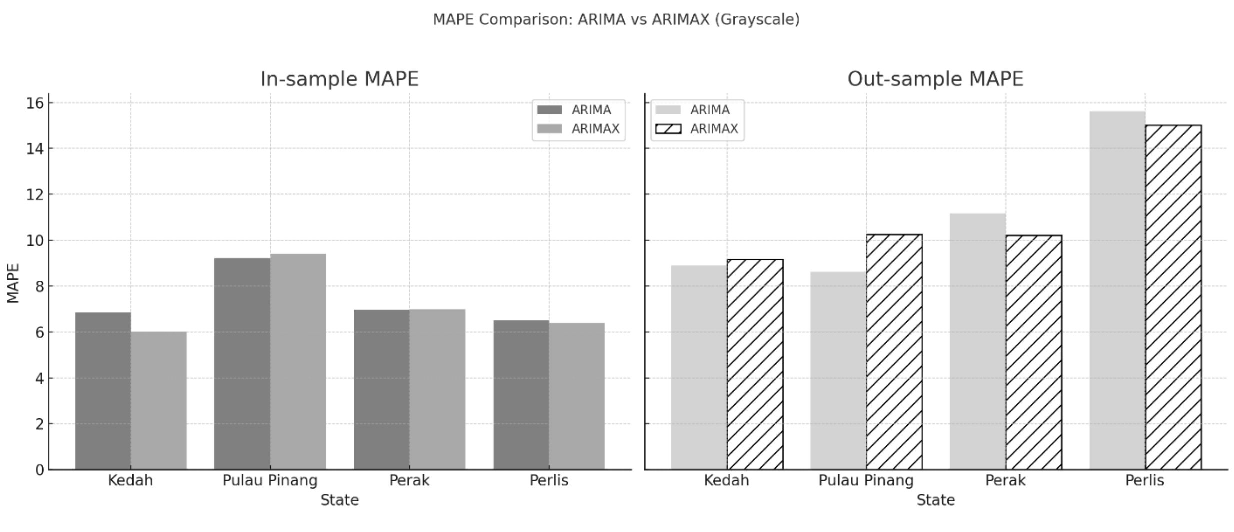

Figure 3. Performance metric's error.

3.5. ARIMAX modeling

ARIMAX models were developed by including planted area as the external factor. The inclusion of planted area was assumed to improve accuracy and provide a clearer understanding of how cultivation areas affect paddy production over time. The models were evaluated to select the most reliable model. This ensured reliable forecasting performance for both inherent temporal patterns and external factors.

Table 6 outlines the comprehensive evaluation of the selected ARIMAX models for each state, showing the AIC and BIC values and the outcomes of all necessary diagnostic tests. The selection process involved first testing the ARIMAX models for each state with AIC and BIC values. As with the ARIMA models, the chosen ARIMAX models featured autoregressive and moving average orders ranging from 0 to

2. The final selection of best-fitted models not only relied on ACF and PACF plots but also prioritized overall model performance and diagnostic checks. The final ARIMAX models were chosen based on the lowest AIC and BIC values.

These selected models were evaluated on whether the residuals passed the Ljung-Box test and the Breusch-Godfrey test by the *p*-values indicating no significant autocorrelation. This ensured that the models were a good fit for the data, where they effectively captured the patterns of the dataset without unnecessary complexity. Moreover, each model met crucial criteria, as verified by the Shapiro-Wilk test for normality and the Breusch-Pagan test for homoscedasticity. These outcomes highlighted the accuracy and reliability of the models for precise forecasting.

3.6. Performance metric

The comparison of the performances between ARIMA and ARIMAX modeling across the four northern states, as displayed in Table 7 and Figure 3, reveals different outcomes regarding the effectiveness of the inclusion of the external factor, which was planted area, in forecasting paddy production. For Kedah and Pulau Pinang, ARIMA modeling provided more accurate out-of-sample forecasts, despite ARIMAX modeling exhibiting a better in-sample fit. In contrast, ARIMAX modeling proved more effective in out-of-sample forecasting, resulting in better outcomes for Perak and Perlis by including planted area as the external factor to improve the accuracy of forecasting.

Table 7. Performance metric's error.

State	In-sample		Out-of-sample	
	ARIMA	ARIMAX	ARIMA	ARIMAX
Kedah	6.84	6.01	8.90	9.15
Pulau Pinang	9.23	9.40	8.61	10.24
Perak	6.97	6.99	11.16	10.22
Perlis	6.50	6.39	15.62	15.01

This suggests that in these states, fluctuations in land use had a noticeable impact on how efficiently rice was produced. This finding is useful for policymakers and agricultural planners, as it highlights that models should not rely only on past productivity trends but also consider external factors, such as planted area, especially in less consistent or more vulnerable regions.

Overall, ARIMAX modeling's effectiveness for paddy production forecasting was found to be regionally dependent. This proves more beneficial in states where external factors have a significant impact on paddy production and align closely with historical data trends.

3.7. Discussion

The comparative forecasting evaluation between ARIMA and ARIMAX models reveals key insights into how exogenous variables, specifically the planted area, affect paddy productivity trends in Malaysia's northern states. The results indicate that ARIMAX generally outperformed ARIMA in terms of in-sample performance, especially in Kedah and Perak, reinforcing the model's utility in environments where exogenous factors exhibit significant influence.

This aligns with Joseph et al. (2025) who demonstrated that ARIMAX models outperformed conventional ARIMA in maize productivity forecasting in Tanzania by accounting for rainfall variability and land use dynamics. In their findings, the inclusion of planted area and weather-related variables enhanced both interpretability and accuracy, particularly in regions prone to seasonal disruptions.

Notably, the strong correlation observed in Kedah between productivity and planted area underscores the state's efficiency in translating land allocation into agricultural output. This echoes the results of Pokhrel & Adhikari (2023) in Nepal, where land use was shown to be a significant predictor of paddy productivity when modeled through ARIMAX. Their work emphasized that land expansion coupled with appropriate inputs could drive productivity gains, provided the environmental and policy contexts are supportive.

However, the modest improvement in Pulau Pinang and Perak using ARIMAX suggests that external variables must be chosen carefully and validated statistically before inclusion. As Siddique et al. (2025) caution, introducing weak or indirectly correlated variables can lead to overfitting or model degradation, particularly in short out-of-sample windows. This observation highlights a broader statistical implication: while ARIMAX enhances model responsiveness, it also requires robust variable selection, possibly using techniques such as stepwise regression or principal component analysis to mitigate multicollinearity and ensure parsimony.

Furthermore, the variation in model performance across the four states supports the idea that agricultural forecasting must be context-specific, adapting to regional factors such as irrigation infrastructure, farmer behavior, and policy incentives. Ayyoob et al. (2025) found similar regional dependencies in Kerala's rubber forecasting models, where ARIMAX accuracy varied by district based on how closely the exogenous inputs (fertilizer usage and subsidies) mirrored local practices.

Additionally, this study supports the integration of hybrid approaches for future work. As Pandit et al. (2023) emphasized in their hybrid model for Rabi crops in India, combining ARIMAX with machine learning models such as LSTM or Support Vector Regression (SVR) can provide better long-range and nonlinear forecasting capabilities, especially important in agricultural systems impacted by climate change.

Finally, the model's operational value for policy design is evident. By capturing external dynamics like planted area, ARIMAX models can provide anticipatory signals for planning inputs, subsidies, and land allocation. Zheng (2025) highlighted how cloud-enabled ARIMAX forecasting tools are increasingly adopted for precision agriculture, allowing real-time updates and model re-training as new data becomes available.

Therefore, this study reaffirms the importance of incorporating well-selected exogenous variables in time-series forecasting for agriculture. While ARIMA provides a solid baseline, ARIMAX enhances sensitivity to external shifts and supports more responsive and resilient agricultural planning, particularly in regions like Malaysia where yield is influenced by dynamic socio-environmental factors.

4. CONCLUSION

This study aimed to forecast paddy productivity in four key rice-producing states in Malaysia, namely, Kedah, Pulau Pinang, Perak, and Perlis, by using time series models. Two forecasting models, ARIMA and ARIMAX, were applied and compared to assess their ability to capture historical productivity trends and improve predictive accuracy by including an external variable in ARIMAX modeling. While ARIMA modeling relied solely on past productivity data, ARIMAX modeling incorporated planted area as an exogenous variable, making it more comprehensive in modeling real-world agricultural influences.

To evaluate model performance, the MAPE was used, since expressing forecast errors as a percentage of actual values is particularly important, offering an easily interpretable measure of how accurate the forecast is. In general, a MAPE value below 10% is considered highly accurate and indicates that a model's predictions are close to actual observations. Although not all MAPE values in this study fell below 10% for high accuracy, there were values that stayed below 16%. This indicates that a good forecasting performance was achieved for all four northern states. ARIMAX modeling outperformed ARIMA modeling, as it enhanced the accuracy of forecasting by including planted area.

Although the ARIMA and ARIMAX models produced accurate forecasts of paddy productivity, there is room for improvement in future research. One potential enhancement is the inclusion of additional exogenous variables beyond planted area. Various factors, such as rainfall, temperature, fertilizer usage, irrigation systems, and government policy interventions, can significantly influence agricultural productivity. Including these factors in more advanced models could lead to a better understanding and more accurate predictions of paddy yield.

However, a key concern is an ARIMAX model's sensitivity to multicollinearity among exogenous variables, which can inflate variance estimates and obscure the individual contribution of predictors. Additionally, including multiple exogenous variables may lead to a higher risk of overfitting, especially when the sample size is limited or variables are weakly informative. The model also assumes linear relationships and may not fully capture complex nonlinear interactions inherent in agricultural systems. Future extensions could benefit from rigorous variable selection techniques, dimensionality reduction (e.g., PCA), or hybrid modeling approaches that mitigate these risks.

In addition, future research should consider the application of advanced machine learning techniques, such as Long Short-Term Memory (LSTM) networks. These deep learning models are well-suited for time series forecasting, as they are capable of recognizing complex and nonlinear relationships in data over time. LSTM models can be particularly beneficial in agricultural contexts, where external factors are highly variable and unpredictable.

ACKNOWLEDGEMENT

This research was supported by the Fundamental Research Grant Scheme (FRGS) from the Ministry of Higher Education Malaysia, with Proposal No: FRGS/1/2023/STG06/UTM/02/12 and vote number R.J130000.7854.5F620, and the authors thank Universiti Teknologi Malaysia for its financial support under the UTM Fundamental Research, Reference No: PY/2024/00951, vote number Q.J130000.3854.23H62.

REFERENCES

Ahmad, N., Kamal, M. A., & Hassan, R. (2020). Climate change impact on rice production and adaptation strategies in Malaysia. *Journal of Environmental Management*, *255*, 109877. https://doi.org/10.1016/j.jenvman.2019.109877

Ayyoob, K. C., Bhattacharya, D., & Sarkar, K. A. (2025). Comparison of statistical models and machine learning algorithms in forecasting rubber yield in Kerala. *Journal of Rubber Research*. https://doi.org/10.1007/s42464-025-00328-z

Dorairaj, D. & Govender, N. T. (2023). Rice and paddy industry in Malaysia: Governance and policies, research trends, technology adoption and resilience. *Frontiers in Sustainable Food Systems*, *7*. https://doi.org/10.3389/fsufs.2023.1093605

Elsaraiti, M. & Merabet, A. (2020). A comparative analysis of the ARIMA and LSTMs predictive models and their effectiveness for predicting wind speed. *Energies*, *14*(20), 6782. https://doi.org/10.3390/en14206782

Firdaus, R. B. R., Tan, M. L., Rahmat, S. R., & Gunaratne, M. S. (2020). Paddy, rice and food security in Malaysia: A review of climate change impacts. *Cogent Social Sciences, 6*(1). https://doi.org/10.1080/23311886.2020.1818373

Ghosh, K., et al. (2014). Development of crop yield forecast models under FASAL—a case study of kharif rice in West Bengal. *Journal of Agrometeorology*, *16*(1), 1–8.

Hamiane, S., Ray, S., & Bhattacharyya, B. (2024). Comparative analysis of LSTM, ARIMA, and hybrid models for forecasting future GDP. *Ingénierie des Systèmes d'Information*, *29*(3), 853–861. https://doi.org/10.18280/isi.290306

Hamjah, M. A. (2014). Temperature and rainfall effects on spice crops production and forecasting the production in Bangladesh: An application of Box-Jenkins ARIMAX model. *Mathematical Theory and Modeling*, *4*(10), 149–152.

Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts.

Joseph, A., Mwakisisile, A. J., Sinkwembe, E. E., & Mbare, N. S. (2025). Modeling weather variability effects on maize production in Tanzania using ARIMAX. *Scientific African, Elsevier*. https://doi.org/10.1016/j.sciaf.2025.XXXXXX

Lewis, C. D. (1982). *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. Butterworths.

Makhtar, N. S., Abidin, N. I. S. Z., & Islam, N. R. (2022). Reviewing food security on paddy production: A conceptual paper. *International Journal of Industrial Management*, *16*(1), 51–58. https://doi.org/10.15282/ijim.16.1.2022.9012

Md Yusof, Z., Misiran, M., Yaacob, M. F., Abdul Aziz, N. A., & Sanan, N. H. (2019). Projection of paddy production in Kedah, Malaysia: A case study. *Asian Journal of Advanced Agricultural Research*, *10*(3), 1–6.

Montgomery, D. C., Jennings, C. L., & Kulachi, M. (2015). *Introduction to Time Series Analysis and Forecasting*. Wiley.

Muslim, N. (2025, June 6). The future in a grain: Malaysia's bet on gene-edited rice. *Bernama*. https://www.bernama.com/en/news.php?id=2431314

Omar, S. C., Shaharudin, A., & Tumin, S. A. (2019). *The Status of the Paddy and Rice Industry in Malaysia*. Khazanah Research Institute.

Pandit, P., et al. (2022). Deep predictive models based on IoT and remote sensing for precision agriculture. *International Journal of Emerging Technology and Advanced Engineering, 12*(11), 79–81.

Pandit, P., et al. (2023). Hybrid time series models with exogenous variable for improved yield forecasting of major Rabi crops in India. *Scientific Reports*, *13*, 22240. https://doi.org/10.1038/s41598-023-49531-1

Pokhrel, A. & Adhikari, R. (2023). Leveraging exogenous insights: A comparative forecast of paddy production in Nepal using ARIMA and ARIMAX models. *Economic Review Nepal*, *6*(1), 52–69.

Ray, S. & Bhattacharyya, B. (2020). Statistical modeling and forecasting of ARIMA and ARIMAX models for food grains production and net availability of India. Unpublished manuscript.

Siddique, M. A. B., Mahalder, B., & Haque, M. M. (2025). Impact of climatic and water quality parameters on tilapia broodfish growth: Integrating ARIMA and ARIMAX. *PLOS ONE*, *20*(3), e0313846. https://doi.org/10.1371/journal.pone.0313846

Sivapathasundaram, V., & Bogahawatte, C. (2015). Forecasting of paddy production in Sri Lanka: A time series analysis using ARIMA model. *Tropical Agricultural Research*, *24*(1), 21–29. https://doi.org/10.4038/tar.v24i1.7986

Sudipa, I. G. I., et al. (2024). Predictive time-series modelling of rice price fluctuations in East Nusa Tenggara using ARIMAX: A data-driven case study of regional agricultural market dynamics. *Power System Technology*, *48*(4), 111–123.

Sujatha, P. & Sivasankari, B. (2023). Forecasting area, production and productivity of rice in Tamil Nadu using time series model. *International Journal of Statistics and Applied Mathematics*, *8*(Special Issue 5), 392–397.

Yogarajah, B., Elankumaran, C., & Vigneswaran, R. (2013). Application of ARIMAX model for forecasting paddy production in Trincomalee district in Sri Lanka. *Proceedings of the Third International Symposium*, 21–25.

Zheng, L. (2025). AEAPIM-CC: A cloud-enabled integrated model for agricultural economic forecasting via feature-matrix analysis and ARIMAX. *Informatica*, *49*(2), 317–328. https://informatica.si/index.php/informatica/article/view/8884

