A Comparative Study of the Performance of Stepwise Ridge Regression, Box–Jenkins Method, and Holt's Method in Forecasting Khao Dawk Mali 105 Rice Production

Pheerach Rajapitaks[1*], Teeratas Sirisattayanupong[1]
Pirapat Pumkong

[1]*Princess Chulabhorn Science High School Chonburi , Nong Chak, Ban Bueng, Chonburi, 20170 Thailand*
**e-mail:6706390@pccchon.ac.th*

Abstract

This study aimed to: (1) examine the relationships between the independent variables and the production quantity of Khao Dawk Mali 105 rice, and develop an appropriate forecasting equation using stepwise multiple regression and ridge regression models; (2) develop time series forecasting models for the production quantity of Khao Dawk Mali 105 rice using the Box–Jenkins and Holt's methods; and (3) compare the forecasting performance of the regression models and time series models using the Mean Absolute Percentage Error (MAPE). The data used in this study consisted of annual secondary data from Ubon Ratchathani Province covering the period 2006–2023 (B.E. 2549–2566), comprising 18 years of observations. The independent variables included average annual rainfall, average annual temperature, average relative humidity, and average wind speed, while the dependent variable was the production quantity of Khao Dawk Mali 105 rice. Pearson's correlation analysis was employed to examine the relationships between the independent variables and the production quantity of Khao Dawk Mali 105 rice.

The results of the study revealed that: (1) average temperature and average rainfall were moderately positively correlated with the production quantity of Khao Dawk Mali 105 rice, whereas average wind speed exhibited a moderately negative correlation and average relative humidity showed a weak negative correlation with the production quantity of Khao Dawk Mali 105 rice. The most appropriate forecasting equation was obtained from the stepwise multiple regression model, The resulting forecasting equation is as follows:

$$\hat{Y} = -2{,}227{,}142.4 + 100{,}489.2(temp) + 1945.1(rain) - 64{,}371.3(wind)$$

(2) The time series model developed for forecasting the production quantity of Khao Dawk Mali 105 rice using the Box–Jenkins method was

$$(1 - 0.2503B + 0.5627B^2)(1 - B)Y_t = 28375.913 + \varepsilon_t$$

while the forecasting model developed using Holt's method was

$$\hat{Y}_{t+h} = 0.0083Y_t + 0.9917(L_{t-1} + T_{t-1}) + h\left[0.0001(L_t - L_{t-1}) + 0.9999T_{t-1}\right].$$

(3) The stepwise multiple regression model yielded a MAPE of 13.01%, whereas the ridge regression model produced a MAPE of 14.85%. The ARIMA(2,1,0) with drift model developed using the Box–Jenkins method achieved the lowest MAPE of 4.30%, indicating that it was the most appropriate forecasting model. In comparison, Holt's method yielded a MAPE of 5.30%.

Keywords: Stepwise Multiple Regression, Ridge Regression, Box–Jenkins Method, Holt's Method

Introduction

Jasmine rice, particularly the Khao Dawk Mali 105 (KDML 105) variety, is one of Thailand's most important economic crops. It serves not only as the staple food for the domestic population but also as a major export commodity that generates substantial revenue for the country's economy each year. The northeastern region of Thailand is the largest production area for jasmine rice, with Ubon Ratchathani Province being one of the leading provinces in terms of both cultivation area and rice production. Consequently, the annual yield of KDML 105 directly affects farmers' income and economic stability in the region, as well as the nation's food security and the management of the rice supply chain. However, Thailand's agricultural sector is currently facing significant challenges and risks resulting from climate change, an uncontrollable factor that has increasingly affected agricultural production. The cultivation of KDML 105 continues to rely heavily on natural weather conditions, making meteorological variables such as average rainfall, average wind speed, average temperature, and average relative humidity critical determinants of plant growth, panicle development, and final grain yield. Annual fluctuations in weather conditions lead to considerable uncertainty in rice production. The absence of accurate yield forecasting information makes it difficult for farmers and government agencies to effectively plan agricultural production, manage storage facilities, and formulate market price intervention policies. Therefore, the development of accurate mathematical and statistical forecasting models for predicting rice yield has become an essential strategic tool for reducing agricultural risks and improving decision-making. In the past, regression analysis has been one of the most widely used statistical methods for explaining the relationship between climatic variables and rice yield However, conventional regression analysis often encounters the problem of multicollinearity among independent variables because climatic factors are typically correlated with one another. This issue increases the estimation error of the model and reduces its reliability. To overcome this limitation, advanced statistical techniques such as Ridge Regression have been developed. In addition, time series forecasting methods, including the Box–Jenkins method and Holt's Exponential Smoothing, are recognized as highly effective approaches for analyzing historical data patterns, trends, and seasonal or cyclical behaviors to generate systematic and reliable forecasts. Recognizing the importance of integrating applied mathematics with data science to support the agricultural sector, this study aims to develop and compare the forecasting performance of four mathematical models for predicting the production of Khao Dawk Mali 105 (KDML 105) rice in Ubon Ratchathani Province. The four models consist of two regression-based approaches Stepwise Multiple Regression and Ridge Regression and two time series forecasting methods the Box–Jenkins method and Holt's Exponential Smoothing. Secondary annual data covering an 18-year period (2006–2023) were analyzed using the R statistical software. The forecasting performance of each model was evaluated using the Mean Absolute Percentage Error (MAPE). The findings of this study are expected not only to identify the most appropriate mathematical model for this dataset but also to provide a policy reference for government agencies and farmers in planning and managing rice production more sustainably.

Methodology

Research Methodology Part 1: Data Collection Data covering a period of 18 years were collected for both the independent and dependent variables. 1. Independent variables: Annual average rainfall, annual average air temperature, annual average relative humidity, and annual average wind speed, obtained from the Thai Meteorological Department (TMD). 2. Dependent variable: Annual production of Khao Dawk Mali 105 (KDML 105) rice, obtained from the Office of Agricultural Economics (OAE), Thailand. The details are as follows:

ตารางที่ 3 ข้อมูลตัวแปรอิสระและตัวแปรควบคุม

ปี	ตัวแปรอิสระ				ตัวแปรตาม
	ปริมาณน้ำฝนเฉลี่ย	ความชื้นสัมพัทธ์เฉลี่ย	ความเร็วลมเฉลี่ย	อุณหภูมิเฉลี่ย	ปริมาณข้าวดอกมะลิ 105
2549	127.20	72.00	3.40	27.80	442,898.00
2550	169.60	72.00	4.00	27.10	459,747.00
2551	116.30	73.00	3.80	27.40	460,852.00
2552	164.70	73.00	3.60	27.10	513,995.00
2553	130.30	71.00	3.40	28.00	576,707.00
2554	163.10	72.00	3.70	26.90	703,984.00
2555	138.20	71.00	3.20	28.10	724,686.00
2556	149.90	72.00	3.70	28.30	629,826.00
2557	209.00	72.00	3.70	28.30	614,314.00
2558	133.00	70.00	3.60	28.80	687,958.00
2559	173.30	70.00	4.80	28.60	775,175.00
2560	149.60	71.00	3.90	28.00	774,893.00
2561	173.20	71.00	2.10	27.80	794,083.00
2562	188.60	69.00	2.30	28.90	787,765.00
2563	145.00	70.00	2.30	28.50	835,492.00
2564	175.00	71.00	2.90	28.10	873,492.00
2565	222.80	74.00	2.90	27.60	836,132.00
2566	200.30	74.00	2.90	28.70	888,342.00

Characteristics of the Data The data used in this study are annual time series data covering the period from 2006 to 2023, comprising a total of 18 observations. All variables were collected over the same time period to appropriately represent the relationship between climatic conditions and the production of Khao Dawk Mali 105 (KDML 105) rice. Part 2: Correlation Analysis The variables obtained in Part 1 were analyzed using correlation analysis to identify the independent variables significantly associated with the production of Khao Dawk Mali 105 (KDML 105) rice. The analysis was performed using RStudio. Part 3: Forecasting Model Analysis After the relevant variables had been selected, they were analyzed using four forecasting models to determine the most appropriate model for predicting the production of Khao Dawk Mali 105 (KDML 105) rice. All analyses were conducted using RStudio. Part 4: Development of Forecasting Equations Based on the results obtained in Part 3, forecasting equations were developed for each of the four models using RStudio. Part 5: Comparison of

Forecasting Models The forecasts generated from the four models in Part 4 were compared by evaluating their prediction errors using the Mean Absolute Percentage Error (MAPE). The comparison was conducted using RStudio to identify the model with the highest forecasting accuracy.

Result and Discussion

Results According to the Research Objectives

Objective 1 To examine the relationships between the independent variables and the production of Khao Dawk Mali 105 (KDML 105) rice and to develop appropriate forecasting equations using regression models, namely Stepwise Multiple Regression and Ridge Regression.

To ensure the suitability of the regression models, the relationships between the independent variables and the dependent variable were first examined using correlation analysis. The correlation coefficient was employed to determine both the strength and the direction of the relationships between each pair of variables. The results of this analysis provided the basis for selecting the variables to be included in the regression models for forecasting rice production.

1.1 Correlation Analysis of the Variables Used in the Study

Table 1. Correlation Coefficients of the Study Variables

	rain	temp	humidity	wind	rice
Rain	1.00	0.11	0.28	-0.22	0.50
temp	0.11	1.00	-0.52	-0.23	0.52
humidity	0.28	-0.52	1.00	0.12	-0.22
wind	-0.21	-0.23	0.12	1.00	-0.48
rice	0.50	0.52	-0.22	-0.48	1.00

As shown in Table 1, the climatic factors exhibited both positive and negative correlations with rice production. Average temperature (Temp) showed the strongest positive correlation with rice production ($r = 0.52$), followed by annual rainfall (Rain) ($r = 0.50$). In contrast, average wind speed (Wind) was negatively correlated with rice production ($r = -0.48$), while average relative humidity (Humidity) showed a weak negative correlation ($r = -0.22$). Regarding the relationships among the independent variables, the correlation coefficients ranged from low to moderate. The strongest correlation was observed between average temperature and relative humidity ($r = -0.52$), which remained well below the commonly accepted threshold of 0.80. Therefore, no evidence of multicollinearity was found, indicating that all independent variables were suitable for inclusion in the multiple regression analysis.

1.2 Results of Stepwise Multiple Regression Analysis

1.2.1 Variable Selection Using the Stepwise Regression Method

The Stepwise Regression method was employed to identify the most appropriate independent variables for constructing the forecasting model. Variables were selected based on their ability to explain variations in the dependent variable while maintaining the statistical adequacy of the model. This procedure helps eliminate unnecessary variables, resulting in a more parsimonious and efficient forecasting model. The initial set of independent variables included annual rainfall (Rain), average temperature (Temp), average relative humidity (Humidity), and average wind speed (Wind). The analysis indicated that the model selected the most appropriate variables for the development of the forecasting model. The analysis identified three independent variables as significant predictors of rice production: annual rainfall (Rain), average temperature (Temp), and average wind speed (Wind). In contrast, relative humidity (Humidity) was not selected for inclusion in the forecasting model. These findings indicate that relative humidity contributed less to explaining the variation in rice production than the other climatic variables in this dataset. Furthermore, when considered together with the other weather-related variables, it did not significantly improve the predictive performance of the model and was therefore excluded during the variable selection process. Accordingly, the Stepwise Regression model selected annual rainfall, average temperature, and average wind speed as the primary predictors for constructing the rice production forecasting equation. This variable selection is consistent with the results of the refined multiple regression analysis and demonstrates that these three variables are the most appropriate predictors for forecasting rice production in this study.

1.2.2 Residual Analysis

Table 2. Residual Statistics of the Stepwise Regression Model

min	1Q	Median	3Q	Max
-170748	-64815	9976	65115	148889

As presented in Table 2, the residuals of the Stepwise Regression model ranged from –170,748 to 148,889, with a median value of 9,976, which is close to zero. This indicates that the residuals show no clear systematic bias. The first and third quartiles were –64,815 and 65,115, respectively, suggesting that most residual values were distributed within this range. In addition, no severe outliers were observed. Therefore, it can be concluded that the residuals are reasonably distributed, indicating that the forecasting model is statistically reliable and appropriate for predicting rice production.

1.2.3 Regression Coefficient Analysis

Table 3. Regression Coefficients of the Stepwise Regression Model

	Estimate	Std.Error	t value	Pr(>|t|)
Intercept	-2227142.4	1308529.2	-1.702	0.1108
Rain	1945.1	921.7	2.110	0.0533
temp	100489.2	45340.7	2.216	0.0437
humidity	-64371.3	40774.3	-1.579	0.1367

As shown in Table 3, the Stepwise Regression model selected three independent variables: annual rainfall (Rain), average temperature (Temp), and average wind speed (Wind). Among these variables, average temperature had a statistically significant positive effect on rice production (β = 100,489.2, p = 0.0437). Annual rainfall also showed a positive relationship with rice production (β = 1,945.1, p = 0.0533); however, its effect was not statistically significant at the 0.05 significance level. In contrast, average wind speed exhibited a negative relationship with rice production (β = –64,371.3, p = 0.1367) and was likewise not statistically significant. These findings are consistent with the results obtained from the multiple regression analysis.

1.2.4 Results of the Model Performance Analysis

Table 4. Performance Statistics of the Stepwise Regression Model

R^2	Adjusted R^2	F-statistic	p-value	MAPE
0.5482	0.4513	5.662	0.009385	13.01

As presented in Table 4, the Stepwise Regression model achieved a Multiple R-squared value of 0.5482 and an Adjusted R-squared value of 0.4513, indicating that the model explained

54.82% and 45.13% of the variation in rice production, respectively. The F-statistic was 5.662, with a p-value of 0.009385 ($p < 0.05$), demonstrating that the overall regression model was statistically significant. Furthermore, the model yielded a Mean Absolute Percentage Error (MAPE) of 13.01%, indicating an acceptable level of forecasting error. Therefore, it can be concluded that the Stepwise Regression model provides satisfactory performance in explaining and forecasting the production of Khao Dawk Mali 105 (KDML 105) rice.

1.2.5 Results of the Regression Coefficient Analysis and Forecasting Equation

Based on the variable selection results obtained from the Stepwise Regression analysis, the variables identified as the most appropriate predictors for constructing the rice production forecasting model were annual rainfall (Rain), average temperature (Temp), and average wind speed (Wind). Relative humidity (Humidity) was excluded from the final model, as it did not provide a statistically meaningful contribution to explaining the variability in rice production. Consequently, it was omitted from the final forecasting model. Based on the estimated regression coefficients shown in the Coefficients table, the final forecasting equation is expressed as follows:

$$\hat{Y} = -2,227,142.4 + 100,489.2(temp) + 1945.1(rain) - 64,371.3(wind)$$

Where:

$\hat{Y}$ = Predicted production of Khao Dawk Mali 105 (KDML 105) rice (tons)

rain = Average annual rainfall (millimeters)

temp = Average annual temperature (degrees Celsius)

wind = Average annual wind speed (knots)

Based on the regression equation, annual rainfall and average temperature are positively associated with rice production, whereas average wind speed is negatively associated with rice production. Specifically, holding all other variables constant, a one-unit increase in annual rainfall is associated with an increase of approximately 1,945.1 tons in the predicted rice production. Likewise, a one-unit increase in average temperature is associated with an increase of approximately 100,489.2 tons in rice production. In contrast, a one-unit increase in average wind speed is associated with a decrease of approximately 64,371.3 tons in the predicted rice production. The intercept (constant term) of the model is −2,227,142.4. The forecasting equation obtained from the Stepwise Regression analysis is identical to that derived from the Multiple Regression model, confirming that annual rainfall, average temperature, and average wind speed are the most appropriate predictors of rice production. These variables can therefore be used to forecast Khao Dawk Mali 105 (KDML 105) rice production while maintaining satisfactory model performance.

1.3 Results of Ridge Regression Analysis

Ridge Regression is a regression technique developed to address the problem of multicollinearity among independent variables by incorporating a regularization parameter (λ, lambda) into the coefficient estimation process. This regularization reduces the variability of the estimated coefficients, improves the stability of the regression model, and enhances its predictive performance. In this study, the independent variables selected from the previous analysis annual rainfall (Rain), average temperature (Temp), and average wind speed (Wind) were used to construct the Ridge Regression model. The model's forecasting performance was evaluated using the Mean Absolute Percentage Error (MAPE) and compared with those of the other forecasting models. The results of the Ridge Regression analysis include the estimated regression coefficients, the performance evaluation of the model, and the development of the forecasting equation for predicting the production of Khao Dawk Mali 105 (KDML 105) rice, which are presented in the following sections.

1.3 .1 Results of the Regression Coefficient Analysis for the Ridge Regression Model

Table 5. Regression Coefficients of the Ridge Regression Model

Intercrpt	-1212919.894
rain	1283.234
temp	66070.087
wind	-47261.554

As presented in Table 10, the Ridge Regression model included three independent variables: annual rainfall (Rain), average temperature (Temp), and average wind speed (Wind). A lambda (λ) value of 92,108.96 was selected using cross-validation to regularize the regression coefficients and reduce the effects of multicollinearity. The results showed that annual rainfall had a regression coefficient of 1,283.234, while average temperature had a coefficient of 66,070.087, indicating positive relationships with rice production. In contrast, average wind speed had a coefficient of –47,261.554, indicating a negative relationship with rice production. Compared with the Multiple Regression and Stepwise Regression models, the regression coefficients estimated by the Ridge Regression model were smaller in magnitude. This coefficient shrinkage enhanced the stability of the model and reduced the risk of overfitting. Overall, the findings are consistent with those obtained from the other forecasting models examined in this study.

1.3.2 Results of the Performance Analysis of the Ridge Regression Model

The Ridge Regression model achieved a Mean Absolute Percentage Error (MAPE) of 14.85%, indicating that the model's forecasts deviated from the actual rice production by an average of 14.85%. When compared with the Multiple Regression and Stepwise Regression models, both of which produced a MAPE of 13.01%, the Ridge Regression model exhibited a slightly higher forecasting error. This is likely because the dataset did not exhibit severe multicollinearity, limiting the advantages typically associated with Ridge Regression. Nevertheless, the Ridge Regression model demonstrated acceptable forecasting performance and can be considered a reliable approach for predicting the production of Khao Dawk Mali 105 (KDML 105) rice.

1.3 .3 Development of the Rice Production Forecasting Equation Using the Ridge Regression Model

The Ridge Regression model developed from the rice production data, using annual rainfall (Rain), average temperature (Temp), and average wind speed (Wind) as the independent variables, can be expressed as the following forecasting equation:

$$\hat{Y} = -1{,}212{,}919.894 + 1{,}283.234(rain) - 47{,}261.554(wind) + 66{,}070.087(temp)$$

Where:

$\hat{Y}$ = Predicted production of Khao Dawk Mali 105 (KDML 105) rice (tons) rain = Average annual rainfall (millimeters)

temp = Average annual temperature (degrees Celsius)

wind = Average annual wind speed (knots)

Annual rainfall and average temperature exhibited positive relationships with rice production, whereas average wind speed showed a negative relationship. The directions of the regression coefficients were consistent with those obtained from the Multiple Regression and Stepwise Regression models. However, the coefficient estimates in the Ridge Regression model were smaller in magnitude due to the application of the lambda (λ) regularization parameter, which improves model stability and reduces the risk of overfitting. Therefore, the Ridge Regression model can be effectively used to forecast Khao Dawk Mali 105 (KDML 105) rice production, although its forecasting performance was slightly lower than that of the other regression models evaluated in this study.

Objective 2 To develop time series forecasting models for predicting the production of Khao Dawk Mali 105 (KDML 105) rice using the Box–Jenkins method and Holt's Exponential Smoothing.

2.1 Development of the Forecasting Model Using the Box–Jenkins Method

Based on the Box–Jenkins analysis, the most appropriate model for forecasting the production of Khao Dawk Mali 105 (KDML 105) rice was identified as the ARIMA(2,1,0) model with drift. The forecasting equation can be expressed as follows:

$$(1-0.2503B+0.5627B^{2})(1-B)Y_{t}=28375.913+\varepsilon_{t}$$

The above equation represents the forecasting model used by the program to generate predicted values for the time series data. The forecasted values were then compared with the actual observations to calculate the Mean Absolute Percentage Error (MAPE), which was used as a criterion for evaluating the performance of the model. The developed model was applied to forecast the production of Khao Dawk Mali 105 (KDML 105) rice for the period 2024–2026, as presented in the following table.

Table 6. Forecasted Production of Khao Dawk Mali 105 (KDML 105) Rice Using the Box–Jenkins Method

ปี	ค่าพยากรณ์ (ตัน)	ช่วงความเชื่อมั่น 80%	ช่วงความเชื่อมั่น 95%
2567	914,544.6	848,479.8 – 980,609.4	813,507.1 – 1,015,582.0
2568	940,747.2	847,317.4 – 1,034,177.0	797,858.6 – 1,083,636.0
2569	966,949.8	852,522.1 – 1,081,377.4	791,947.8 – 1,141,952.0

According to the forecasting results presented in the table, the predicted rice production showed a continuous increasing trend throughout the forecasting period. The developed model achieved a Mean Absolute Percentage Error (MAPE) of 4.2956%, indicating a high level of forecasting accuracy. Therefore, the Box–Jenkins model (ARIMA(2,1,0) with drift) is considered an appropriate and reliable model for forecasting the production of Khao Dawk Mali 105 (KDML 105) rice based on the dataset used in this study.

2.2 Development of the Forecasting Model Using Holt's Method

Based on the analysis using Holt's Method, the model was found to appropriately describe the trend of Khao Dawk Mali 105 (KDML 105) rice production. The estimated smoothing parameters obtained from the model were α = 0.0083 and β = 0.0001. The forecasting equation of the model can be expressed as follows:

$$\hat{Y}_{t+h}=0.0083Y_{t}+0.9917(L_{t-1}+T_{t-1})+h\left[0.0001(L_{t}-L_{t-1})+0.9999T_{t-1}\right]$$

Based on the developed model, the production of Khao Dawk Mali 105 (KDML 105) rice was forecasted for the period 2024–2026, as presented in the following table.

Table 7. Forecasted Production of Khao Dawk Mali 105 (KDML 105) Rice Using Holt's Method

ปี	ค่าพยากรณ์ (ตัน)	ช่วงความเชื่อมั่น 80%	ช่วงความเชื่อมั่น 95%
2567	941,149.8	873,663.5 – 1,008,636.0	837,938.5 – 1,044,361.0
2568	967,959.2	900,470.5 – 1,035,448.0	864,744.2 – 1,071,174.0
2569	994,768.6	927,277.5 – 1,062,260.0	891,549.9 – 1,097,987.0

Results of Holt's Method Forecasting Model According to the forecasting results obtained from Holt's Method, the production of Khao Dawk Mali 105 (KDML 105) rice showed a continuous increasing trend throughout the forecasting period. The model achieved a Mean Absolute Percentage Error (MAPE) of 5.30%, indicating that Holt's Method provided a high level of forecasting accuracy and was an appropriate approach for predicting the production of Khao Dawk Mali 105 (KDML 105) rice.

Objective 3 To compare the performance of regression models and time series forecasting models for predicting the production of Khao Dawk Mali 105 (KDML 105) rice using the Mean Absolute Percentage Error (MAPE) as the evaluation criterion.

To compare the forecasting performance of the regression models and time series models for predicting Khao Dawk Mali 105 (KDML 105) rice production, the researchers used the Mean Absolute Percentage Error (MAPE) as the evaluation criterion. A lower MAPE value indicates a higher forecasting accuracy of the model.

Table 8. Mean Absolute Percentage Error (MAPE) Values of the Forecasting Models

ประเภทตัวแบบ	ตัวแบบพยากรณ์	ค่า MAPE (%)
การถดถอย	Stepwise Regression	13.01
การถดถอย	Ridge Regression	14.85
อนุกรมเวลา	Box–Jenkins (ARIMA)	4.30
อนุกรมเวลา	Holt method	5.30

As shown in Table 8, the ARIMA(2,1,0) with drift model developed using the Box–Jenkins method achieved the lowest Mean Absolute Percentage Error (MAPE) of 4.30%. The second-best performance was obtained from Holt's Method, with a MAPE value of 5.2950%. In comparison, the Multiple Regression and Stepwise Multiple Regression models produced the same MAPE value of 13.01%, while the Ridge Regression model resulted in the highest MAPE value of 14.85%. The comparison results indicate that the time series forecasting models demonstrated higher forecasting performance than the regression models for the dataset used in this study. In particular, the Box–Jenkins model (ARIMA(2,1,0) with drift) provided the lowest forecasting error and was therefore considered the most appropriate model for predicting the production of Khao Dawk Mali 105 (KDML 105) rice based on the dataset from Ubon Ratchathani Province.

Conclusions

This study investigated the relationships between climatic factors and the production of Dok Mali 105 rice in Ubon Ratchathani Province and compared the forecasting performance of regression models and time series models. The results indicated that average temperature and average rainfall were moderately positively correlated with rice production, whereas average wind speed and average relative humidity showed negative correlations.

Among the regression models, the Stepwise Multiple Regression model provided the most suitable forecasting equation, with a MAPE of 13.01%. For the time series models, the

ARIMA(2,1,0) with drift model developed using the Box–Jenkins method achieved the highest forecasting accuracy, with a MAPE of 4.3 0 %, followed by Holt's method, which yielded a MAPE of 5.30%.

Overall, the findings demonstrate that the Box–Jenkins model outperformed the regression models in forecasting Dok Mali 105 rice production. Therefore, it can be considered the most appropriate model for forecasting annual rice production under the conditions of this study. The forecasting models developed in this research may serve as useful tools for agricultural production planning and decision-making by farmers and relevant organizations.

Acknowledgments

The project entitled "A Comparative Study of the Performance of Stepwise Multiple Regression, Ridge Regression, Box–Jenkins, and Holt's Method for Forecasting the Production of Khao Dawk Mali 105 (KDML 105) Rice" was successfully completed with the support and encouragement of many individuals.

The researchers would like to express their sincere gratitude to Mr. Thaweesak Charoentia, the school director, for kindly providing facilities and support for conducting this research. The researchers also would like to thank Mr. Pirapat Phumkhong, a mathematics teacher, for his valuable guidance, advice, and suggestions throughout the development of this project. The researchers would also like to extend their appreciation to all teachers in the Mathematics Department for providing opportunities, encouragement, and continuous support, which greatly contributed to the successful completion of this project. Finally, the researchers would like to express their deepest gratitude to their parents and families for their constant encouragement, support, and motivation throughout the entire process of conducting this research until its completion.

References

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis* (3rd ed.). Wiley.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Lewis, C. D. (1982). *Industrial and Business Forecasting Methods*. Butterworth Scientific.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis* (6th ed.). Wiley.

Office of Agricultural Economics. (2024). Rice production data. https://www.oae.go.th

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia, 126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Thai Meteorological Department. (2024). Climate statistics. https://www.tmd.go.th

Yoshida, S. (1981). *Fundamentals of Rice Crop Science*. International Rice Research Institute.