ISSN 2087-3336 (Print) | 2721-4729 (Online)

TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika
Vol. 13, No. 1, 2026, page. 23-33
http://jurnal.sttmcileungsi.ac.id/index.php/tekno
https://doi.org/10.37373/tekno.v13i1

Analysis of the accuracy of the sarimax model in forecasting cocoa production in Central Sulawesi

Nursuci Rafailah Arsya, Wellie Sulistijanti*

*Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, Jl. Prof. Dr. Hamka No.KM. 1, Tambakaji, Kec. Ngaliyan, Kota Semarang, Jawa Tengah, Indonesia 50158

* Corresponding Author: wellie.sulistijanti@itesa.ac.id

Submitted: 6/5/2025 Revised: 24/6/2025 Accepted: 13/7/2025

ABSTRACT

Cocoa production in Central Sulawesi is among the highest in Indonesia and plays a crucial role in supporting national export needs. Cocoa production trends have shown a significant decline over the past two years. This decline is thought to be caused by various factors, including shrinking cultivated land due to land conversion and climate uncertainty, reflected in erratic rainfall patterns, resulting in unstable cocoa supply. Therefore, a scientific and data-driven approach to cocoa production forecasting is crucial for proper planning and monitoring of production and for anticipating imbalances between demand and supply. This study utilized the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) method, considered superior for its ability to capture seasonal patterns while simultaneously accommodating the influence of exogenous variables such as rainfall and land area, resulting in more accurate forecasting. The data used are cocoa production data as endogenous variables and rainfall and land area as exogenous variables for the period January 2020 to December 2023. The analysis stages include SARIMA model identification, pre-whitening, transfer function analysis, and evaluation of model accuracy using Mean Absolute Percentage Error (MAPE). The results of the study show that the best model is SARIMAX (1,1,1)(0,1,0), with land area variables at lag-5 being significant to cocoa production, producing a MAPE value of 3.29%, so this model can be used to predict future cocoa production.

Keywords: Cocoa; forecasting; SARIMAX; land area; production.

1. Introduction

Indonesia, as an archipelagic nation with a strategic geographical location, possesses abundant natural resources. One sector that capitalizes on this wealth is plantations, which contribute significantly to economic development and serve as a primary source of foreign exchange. Among the various plantation commodities available, cocoa is a leading one. Indonesia is even known as one of the world's largest cocoa producers, with this sector playing a vital role in driving national economic growth [1]. This commodity is spread across various regions, especially in Bali, Central Sulawesi, South Sulawesi, West Sulawesi, Lampung, and Southeast Sulawesi [2]. Sulawesi Island is the center of national cocoa production, contributing approximately 75% of Indonesia's total production. In 2018, national cocoa production reached 767,400 tons, with the majority of production coming from smallholder plantations [3].

Among the cocoa-producing regions on Sulawesi Island, Central Sulawesi Province is recorded as one of the highest producing regions. Production trends in recent years have shown a decline. According to data from the Central Statistics Agency (BPS), cocoa production in Central Sulawesi was recorded at 131.5 thousand kilograms in 2021. This figure decreased to 125.9 thousand kilograms in 2023, indicating a two-year decline in production.

The decline in cocoa is caused by various factors, such as the reduction in the area of plantation land which is being converted to develop other crops [4]. This certainly impacts the cocoa production process and harvest. This land conversion reduces the planting area and impacts yields. In addition to land limitations, weather variability, particularly rainfall, also plays a significant role in cocoa production. Dynamically changing rainfall and increases or decreases in the intensity of extreme climate events can impact cocoa plant growth by affecting water availability during the growing and harvesting periods [5].

The impact of this decline in cocoa production is not only felt by farmers but can also impact the economic sector, particularly in meeting export demand. Unexpected production instability can lead to an imbalance between cocoa demand and supply, which in turn leads to price fluctuations. Therefore, strategic steps are needed to anticipate this problem, one of which is forecasting cocoa production for the coming period. Forecasting cocoa production is crucial for proper production planning and monitoring [6].

One method that is considered effective in forecasting seasonal time series data is the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX). This method is not only able to capture seasonal patterns in the data, but also allows the inclusion of exogenous variables such as rainfall and land area, which can affect the forecasting results [7].

Several previous studies have used forecasting approaches using various methods. For example, a study titled "Forecasting Cocoa Production Levels in 2021 in North Sumatra Province Using the Double Exponential Smoothing Brown Method" obtained a forecasting accuracy of 8.96%. However, this method did not consider the influence of exogenous variables that influence harvest yields. Therefore, the SARIMAX method is considered superior because it considers exogenous variables in developing the forecasting model [6].

This research is expected to provide more accurate cocoa production forecasting results in Central Sulawesi Province. Furthermore, the results are expected to contribute to policymaking and strategic planning in the cocoa sector, by local governments, farmers, and industry players.

2. Method

This study uses the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) method to forecast cocoa production in Central Sulawesi by considering the influence of exogenous variables such as rainfall and land area. The data used are monthly data from January 2020 to December 2023. The analysis stage begins with a data stationarity test, which includes variance transformation using the Box-Cox method (if necessary), as well as a mean stationarity test through differencing. After the data is declared stationary, an initial SARIMA model is identified based on the ACF and PACF patterns. The best model is selected based on the residual diagnostic test (Ljung-Box) and the smallest Mean Squared Error (MSE) value. Next, pre-whitening is performed on the exogenous data with each SARIMA model, followed by cross-correlation function analysis to evaluate the strength and lag of the relationship between variables, which becomes the basis for determining the transfer function parameters in the form of (b, r, s), where b indicates the lag of the input, r the degree of the input distribution function (δ), and s the degree of the impulse response function (ω). After the initial parameters are determined, the final SARIMAX model estimation is performed. The final stage is forecasting using the SARIMAX model, which is carried out to predict future cocoa production values. Accuracy is evaluated using the Mean Absolute Percentage Error (MAPE) value, which indicates the percentage error of the prediction relative to the actual value. All stages are analyzed systematically and visualized through the following flowchart.

Figure 1 illustrates the research process, which begins with the input of data on cocoa production, rainfall, and land area. The data is then analyzed using ACF and PACF plots to check for stationarity. If the data is not yet stationary, a differencing transformation is performed. Once the data is stationary, a temporary SARIMA model is estimated and diagnostic checking is performed. If the model does not meet the requirements, modifications are necessary to obtain the best model. This best model is developed into SARIMAX through a transfer function that takes into account exogenous variables. The final stage is to forecast cocoa production based on the model.

SARIMA method

Time series forecasting methods aim to identify patterns in historical data series and estimate the value of these patterns for future periods.

a. Model Identification

Identifying models is useful for seeing the type of data pattern and data stationarity.

- Time Series Plot is the first step in choosing the right periodic series or time series method by looking at and considering the types of data patterns which include horizontal patterns, trend patterns, cyclical patterns, and seasonal patterns [8].
- Stationarity refers to the condition where a time series data must meet a constant average (mean) and variance over time, without any growth or decline. To make time series data stationary, it can be done through the Box-Cox transformation and also the average stationary through differencing to see the ACF and PACF plots as determinants of the temporary model [9].

b. Parameter Estimation

Once the model is obtained in the identification stage, the next step is to estimate the coefficients in the model. To assess the model's feasibility, a parameter test is performed to ensure that each estimated coefficient is significant. A parameter is considered significant if its p-value is less than α = 5% (or 0.05). This significance test is important to determine whether the parameters used truly influence the model. If all parameters in the model are proven significant, then the model can be used as a temporary model that will then be further tested for feasibility [10].

c. Diagnostic Checking

Diagnostic tests are designed to evaluate the suitability of the model used and to ensure that it meets important underlying assumptions. In the context of time series analysis, the diagnostic stage typically involves testing for residual independence (to determine whether the residuals exhibit white noise characteristics) and residual normality. A model is considered valid if its residuals are white noise and normally distributed [11]. Among the various models that meet the coefficient significance and white noise criteria, the optimal model will be selected based on the lowest Mean Squared Error (MSE) value. The formula for calculating MSE is as follows [12].

$$MSE = \sum_{t=1}^{n} \frac{(Z_t - F_t)^2}{n} \qquad (1)$$

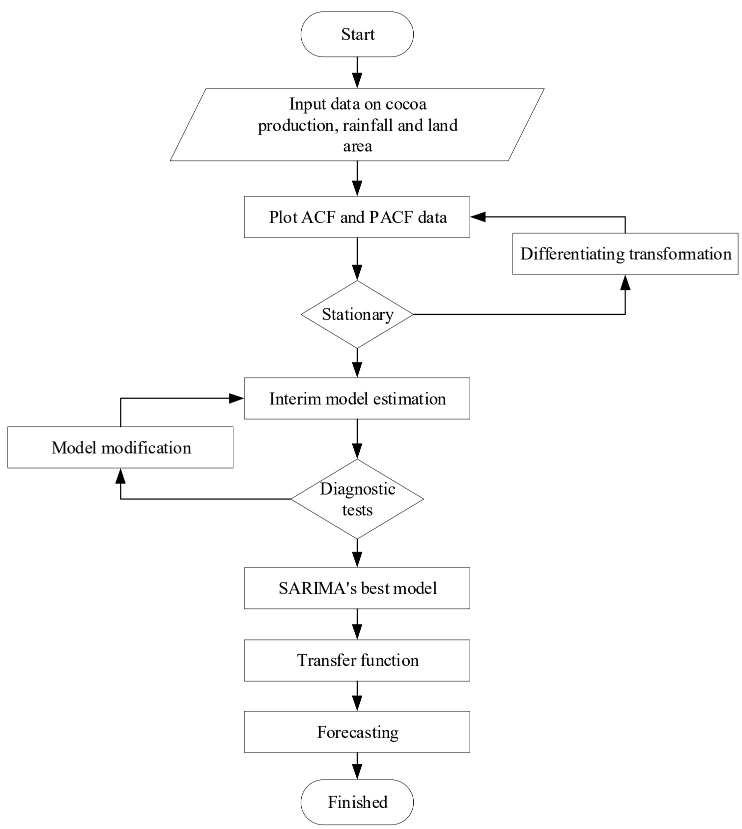

Figure 1. Flowchart SARIMAX

SARIMAX Method

Starting with forming a SARIMA model, then continuing to determine the transfer function model. The steps for determining SARIMAX include [13].

a. Pre-whitening of input and output series
The pre-whitening stage aims to ensure the model meets the SARIMA white noise assumptions by utilizing significant input and output series. This process helps us assess whether the model exhibits white noise characteristics. If the p-value exceeds 0.05, it indicates that the model meets the white noise criteria.
b. Calculation of the cross-correlation function between the input and output series
The cross correlation function is used to assess the strength of the relationship between two variables.
c. Determination of (b,r,s) for the transfer function model
The next step is to develop an initial transfer function model. This involves determining the order of the parameters b, r, and s based on the observed cross-correlation. In this transfer function model, three key parameters have special significance: b represents the delay indicated by the subscript $X_{(t-b)}$, r denotes the autoregressive operator $\delta(B)$, and s and s denotes the moving average operator $\omega(B)$.
d. Final estimation of the SARIMAX model
The final SARIMAX model estimation is performed through two main tests on models that have been proven significant. The first test is the residual autocorrelation test, and the second is the cross-correlation test between the residuals and the input series.

Forecasting

Once the optimal model has been identified, it is ready to be used in forecasting and assessing its effectiveness in predicting cocoa production in Central Sulawesi. Forecasting serves as a benchmark for anticipating future events, relying on historical data for its calculations. When conducting forecasts, it is crucial to focus on accurate and relevant data to ensure that the results can be used as references in various fields, including industry, trade, economics, social sciences, and more [14]. Untuk mengevaluasi keakuratan perkiraan ini, dilakukan dengan menghitung *Mean Absolute* Percentage Error (MAPE). Techniques with MAPE values below 10% demonstrate strong performance because, in general, lower MAPE values reflect better effectiveness. MAPE can be determined using the following equation [15].

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{Z_t - Z_t'}{Z_t} \right| \quad (2)$$

Where:
Z_t: Actual data period t
Z_t': Prediction value in the period t

3. Results and Discussion

This study begins by examining the characteristics of each variable through data patterns, including historical data visualizations, such as trends, seasonality, and fluctuations. This step aims to understand the behavior of cocoa production, rainfall, and land area data before transforming and testing for stationarity.

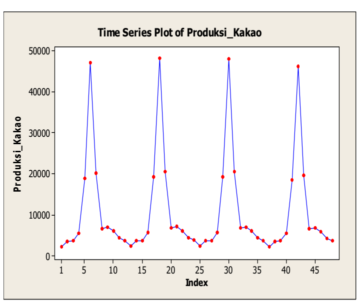

Figure 2. Cocoa production data pattern

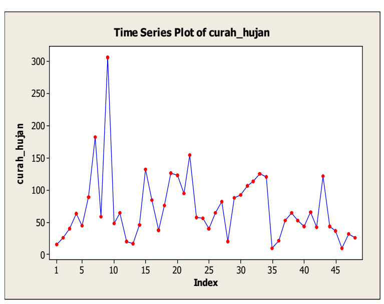

Figure 3. Rainfall data pattern

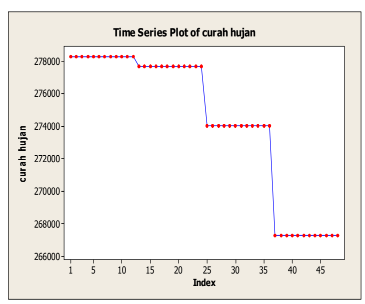

Figure 4. Land area data pattern

The graphs shown above demonstrate the different characteristics of the data patterns for each variable analyzed. Figure 2 shows a strong seasonal component to the cocoa production pattern, with a recurring peak in June each year. Figure 3 shows a highly volatile rainfall pattern with sharp spikes at certain periods, indicating a pattern that is not entirely regular and may be influenced by seasonal factors or extreme weather phenomena. In contrast to these two variables, Figure 4 shows a more stable land area pattern with a stepwise decline, possibly driven by agricultural policies or land-use changes.

Overall, cocoa production and rainfall exhibited more fluctuating patterns with seasonal influences, while land area tended to experience gradual changes without a clear seasonal pattern. This characteristic is an important aspect in transfer function analysis to better understand the relationship between these variables. Furthermore, considering the stationarity of the data, the results of the stationary variance analysis are shown in Table 1.

Table 1. Lambda Value

Variables	Lambda	Information
Cocoa Production	-0,5	Transformation $1/\sqrt{Y}$
	1	Stationary
Rainfall	0,5	Transformation ln (Y)
	1	Stationary
Land area	1	Stationary

Cocoa production and rainfall data exhibited non-stationary variance, so a Box-Cox transformation of the data was performed. Table 1 illustrates the results of this transformation for both variables, showing that the rounded values achieved a stationary variance of 1.00. Meanwhile, the land area variable was already stationary in variance, so no transformation was required.

In the next step, we examine the ACF and PACF plots of cocoa production to illustrate the results of parameter estimation. A differencing process is required for both non-seasonal and seasonal data to ensure the data remains evenly stationary. The temporary model obtained is SARIMA $(1,1,1)(1,1,1)^{12}$ which shows the ACF and PACF cut-off graphs after lag 1 (already stationary). Through this temporary model, several model estimates were also obtained, as shown in Table 2.

Table 2. Parameter Estimation of the SARIMA Cocoa Production model

Model $(p,d,q)(P,D,Q)^s$	type	P-value	Decision
$(1,1,1)(1,1,1)^{12}$	AR 1	0,347	No Sig
	SAR 12	0,001	Sig
	MA 1	0,190	No Sig
	SMA 12	0,001	Sig
$(0,1,1)(1,1,0)^{12}$	SAR12	0,000	Sig
	MA 1	0,615	No Sig
$(1,1,0)(1,1,1)^{12}$	AR 1	0,366	No Sig
	SAR12	0,000	Sig
	SMA 12	0,001	Sig
$(1,1,0)(0,1,1)^{12}$	AR 1	0,671	No Sig
	SMA 12	0,009	Sig
$(1,1,0)(1,1,0)^{12}$	AR 1	0,660	No Sig
	SAR 12	0,000	Sig
$(0,1,0)(1,1,1)^{12}$	SAR	0,001	Sig
	SMA 12	0,001	Sig
$(0,1,1)(1,1,0)^{12}$	SAR 12	0,000	Sig
	MA 1	0,990	No Sig
$(1,1,1)(0,1,1)^{12}$	AR 1	0,000	Sig
	MA 1	0,000	Sig
	SMA 12	0,109	No Sig

Model (p,d,q)(P,D,Q)^s	type	P-value	Decision
$(1,1,1)(1,1,0)^{12}$	AR 1	0,000	Sig
	SAR 12	0,745	No Sig
	MA	0,000	Sig

Table 3 shows several SARIMA models that have been tested with parameter estimation. The decision regarding whether a parameter is significant or not is determined based on its p-value; a parameter is considered significant if its p-value is <0.05. From these models, the model with the significant value is $(0,1,0)(1,1,1)^{12}$.

Meanwhile, the stationarity test for rainfall variables in non-seasonal and seasonal patterns is examined using the ACF and PACF of the data. A differencing process is required for both non-seasonal and seasonal data to ensure the data is evenly stationary. The temporary model obtained is SARIMA $(1,1,1)(1,1,1)^{12}$ which shows the ACF and PACF cut-off graphs after lag 1 (already stationary). Through this temporary model, several model estimates were also obtained, which are presented in Table 3.

Table 3**.** Parameter Estimation of SARIMA Rainfall Model

Model (p,d,q)(P,D,Q)^s	type	p-value	Decision
$(1,1,1)(1,1,1)^{12}$	AR 1	0,001	Sig
	SAR 12	0,002	Sig
	MA 1	0,000	Sig
	SMA 12	0,001	Sig
$(0,1,1)(1,1,1)^{12}$	SAR12	0,494	No Sig
	MA 1	0,000	Sig
	SMA 12	0,753	No Sig
$(1,1,1)(1,1,0)^{12}$	AR 1	0,020	Sig
	SAR12	0,000	Sig
	MA 1	0,000	Sign
$(1,1,1)(0,1,1)^{12}$	AR 1	0,160	No Sig
	MA 1	0,000	Sig
	SMA 12	0,002	Sig
$(0,1,1)(0,1,1)^{12}$	MA 1	0,000	Sig
	SMA 12	0,000	Sig
$(0,1,1)(1,1,0)^{12}$	SAR	0,000	Sig
	MA 1	0,000	Sig
$(1,1,0)(1,1,1)^{12}$	AR 1	0,000	Sig
	SAR 12	0,000	Sig
	SMA 12	0,000	Sig
$(1,1,0)(0,1,1)^{12}$	AR 1	0,008	Sig
	SMA 12	0,010	Sig
$(1,1,0)(1,1,0)^{12}$	AR 1	0,000	Sig
	SAR 12	0,000	Sig

The several models in Table 3, several SARIMA models have been tested using parameter estimation. The decision regarding whether a parameter is significant or not is based on its p-value; a parameter is considered significant if its p-value is <0.05. These models, the model with the significant value is obtained, namely: $(1,1,1)(1,1,1)^{12}$, $(1,1,1)(1,1,0)^{12}$, $(0,1,1)(0,1,1)^{12}$, $(0,1,1)(1,1,0)^{12}$, $(1,1,0)(1,1,1)^{12}$, $(1,1,0)(0,1,1)^{12}$, $(0,1,0)(1,1,0)^{12}$.

Furthermore, the land area variable is stationary in variance but not yet stationary in mean. Therefore, a differencing process is required for both non-seasonal and seasonal data to ensure the data is evenly stationary. The temporary model obtained is SARIMA $(1,1,1)(0,0,0)^{12}$ which shows the ACF and PACF cut-off plots after lag 1 (already stationary). Through this temporary model, several model estimates were also obtained, which are presented in Table 4.

Table 4. Parameter Estimation of SARIMA Model for Land Area

Model (p,d,q)	type	p-value	Decision
(1,1,1)	AR 1	0,000	Sig
	MA 1	0,000	Sig
(0,1,1)	MA 1	0,728	No Sig
(1,1,2)	AR 1	0,000	Sig
	MA 1	0,000	Sig
	MA 2	0,594	No Sig
(2,1,1)	MA 1	0,730	No Sig
	MA 2	0,034	Sig
(1,1,0)	AR 1	0,745	No Sig
(2,1,1)	AR 1	0,935	No Sig
	AR 2	0,000	Sig
	SAR 12	0,000	Sig
(1,1,2)	AR 1	0,000	Sig
	MA 1	0,000	Sig
	MA 2	0,594	No Sig
(1,1,0)	AR 1	0,000	Sig

Several models in Table 4, the model obtained has a significant value, namely (1,1,1), (1,1,0). After determining several SARIMA estimation models for each variable, the next step is to perform diagnostic checking to ensure the model is appropriate. Diagnostic checking is performed by testing the assumption that residuals must be white noise and normally distributed. The white noise test output from the SARIMA model is presented in Table 5.

Table 5. Sarima model checking diagnostic test

Diagnostic Check			
Data	Model	P-value	Conclusion
Cocoa Production	$(0,1,0)(1,1,1)^{12}$	0,056 0,274	Fulfilled
Rainfall	$(0,1,0)(1,1,0)^{12}$	0,124 0,9626	Fulfilled
Land area	$(1,1,0)(0,0,0)^{12}$	0,221 0,883 0,998	Fulfilled

Based on Table 5, the p-value for each SARIMA model is above 0.05, indicating no significant autocorrelation in the model residuals, indicating that the residuals are white noise. After confirming that the residuals from the SARIMA model are white noise, the next step is to test the residuals for normality, which is important because this assumption supports the model's reliability in making predictions and further analysis. The following plots of the normality test results for cocoa production, rainfall, and land area are presented in Table 6.

Table 6. Normal distribution

Variables	p-value	Information
Cocoa Production	0,150	Normal
Rainfall	0,150	Normal
Land area	0,150	Normal

Based on Table 6, the results of the residual normality test using a probability plot, it is known that the residuals from the three SARIMA models for cocoa production, rainfall, and land area have a p-value of 0.150. Since the p-value is > 0.05, it can be concluded that the residuals from the three models are normally distributed.

After ensuring that the model residuals are normally distributed, the next step is to evaluate the model's accuracy using the Mean Squared Error (MSE). Table 7 below presents the MSE evaluation results for each selected SARIMA model.

Table 7. Nilai MSE Model Sarima

Variable	Model SARIMA	MSE
Cocoa Production	$(0,1,0)(1,1,1)^{12}$	1,1267
Rainfall	$(0,1,0)(1,1,0)^{12}$	1,2076
Land area	(1,1,0)	1,2080

In Table 7, the evaluation using Mean Squared Error (MSE) shows that the SARIMA model used has a low error rate, so it can be implemented for forecasting. The obtained SARIMA model is suitable for further analysis to understand the relationship between cocoa production and exogenous factors such as rainfall and land area. After obtaining the best SARIMA model, the next step is to pre-whiten the input and output series in the SARIMA model for each variable. The SARIMA model used in the transfer function can be seen in Tabel 8.

Tabel 8. Prewhitening result model

Variable	Model SARIMA After Pre-whiteningg
Cocoa Production	$(0,1,0)(1,1,1)^{12}$
Rainfall	$(0,1,0)(1,1,0)^{12}$
Land area	(1,1,0)

Tabel 8 shows the pre-whitening model for each series in the transfer function analysis, where the estimation results from each SARIMA model indicate that the resulting residuals are white noise, meaning they no longer show significant autocorrelation. The following is the pre-whitening model for the output series:

$$\alpha_{1t} = \frac{(1-B^{12})}{(1-363583B^{12})} X_{1t} \tag{3}$$

$$\alpha_{2t} = \frac{(1-B)}{(1-0{,}974418B)} X_{2t} \tag{4}$$

The pre-whitening stage on the output series is performed using the same technique as that for the input series. However, in this process, the pre-whitening model for the output series is formed based on the model used in the input series, namely rainfall with the model $(0,1,0)(1,1,0)^{12}$ and land area with model (1,1,0). The following is a model of the prewhitening equation for the output series that follows the model of each input series.

$$\beta_{1t} = \frac{(1-B^{12})}{(1-0{,}363583B^{12})} Z_{1t} \tag{5}$$

$$\beta_{2t} = \frac{(1-B)}{(1-0{,}974418B)} Z_{2t} \tag{6}$$

After going through the prewhitening process of the input and output series α_t and β_t then calculate the cross correlation. The results of the cross correlation between α_t and β_t can be seen in Table 9.

Table 9**.** Cross-correlation results of rainfall

Lag	Correlation (Cocoa Production, Rainfall)
2	0,3156
3	0,4577

Lag	Correlation (Cocoa Production, Rainfall)
5	0,0311

Table 9 shows the highest correlation with cocoa production at lag 3, with a value of 0.457. Changes in land area had the greatest impact on cocoa production after five months. Although the correlation at lag 4 was still quite high at 0.3156, after lag 3 this relationship weakened significantly. Based on these results, in the transfer function modeling, lag 3 for rainfall was used.

Table 10**.** Cross correlation results of land area

Lag	Correlation (Cocoa Production, Land Area)
4	0,2438
5	0,5538
17	0,03236

Meanwhile, Table 10 shows the highest correlation with cocoa production at lag 5, with a value of 0.5538. Changes in land area have the greatest impact on cocoa production after five months. Although the correlation at lag 4 is still quite high (0.2438), after lag 4 this relationship weakens significantly, as seen at lag 17, which only has a correlation of 0.03236. Based on these results, in the transfer function modeling, lag 5 is used for rainfall. Therefore, the determination of (b, r, s) is presented in Table 11.

Table 11. Value(r,s,b) Model SARIMAX

Input	r	s	b
Rainfall	1	0	3
Land area	1	0	5

The estimation results in Table 12 show that the cross-correlation pattern of cocoa production as an input variable with land area as an output variable has an order value of (b,r,s), namely (b=3,r=1,s=0), while for the other output variable, namely land area, it has an order of (b,r,s) with (b=5,r=1,s=0).

After determining the transfer function parameters based on the results of the cross-correlation analysis, the SARIMAX model estimation process was continued. The initial estimation was carried out by entering the rainfall variable at lag 3 and land area at lag 5 as input variables, as well as the seasonal components AR(12) and MA(12) to capture the annual seasonal pattern in the cocoa production data. The initial estimation results showed that the rainfall variable (lag-3) and the MA(12) component were insignificant, so they were not included in the final model specifications. The final model only retained the land area variable at lag 5 and the AR(12) component, which are presented in Table 12.

Table 12. Final sarimax estimation model results

Variables/Parameters	Coefficient	Std. Error	t-count	p-value
Land area (-5)	0.0719	0.0241	2.9793	0.0051
AR(12)	0.9991	0.0006	1801.476	0.0000

Table 12 shows that land area with a five-month lag significantly influences cocoa production, with a p-value of 0.0051, where each increase in land area by one unit is estimated to increase production by 0.0719 units. Meanwhile, the seasonal component of AR(12) is also significant with a p-value of 0.0000, indicating a strong annual seasonal pattern in cocoa production data.

After the final SARIMAX model was established, the next step was to evaluate the model's accuracy based on historical data. Model accuracy was measured using the Mean Absolute Percentage Error (MAPE). The MAPE value of 3.29% reflects the model's average error in representing historical cocoa production data. In other words, the model was able to map past data patterns quite well, and the error rate of 3.29% indicates that the final SARIMAX model has high accuracy for use in future analysis and projections. To demonstrate the model's compatibility with historical data, Figure 5 presents a comparison between actual cocoa production values and the predicted results from the SARIMAX model during the observation period.

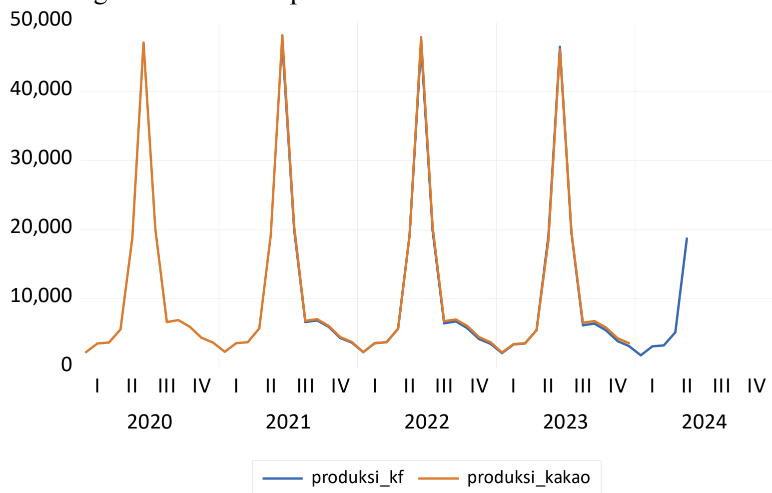

Figure 5. Data graph of Cocoa Production Prediction results

Figure 5 shows how the SARIMAX model is able to track the movement of cocoa production data, where the forecast results appear close to the actual data values, and there are forecasts for the period from 2024 to the second quarter. This indicates that the model can capture annual seasonal patterns and production changes that tend to recur annually, such as mid-year increases. The graph of the model results applied to historical data is used to assess the extent to which the model can accurately represent historical data. The success of this model is further strengthened by the inclusion of land area as an exogenous variable with a five-month lag (lag-5). The combination of the SARIMA model with these exogenous inputs strengthens the performance of the SARIMAX model, allowing it to reflect cocoa production behavior realistically and consistently with historical data.

4. Conclusion

This study demonstrates that the SARIMAX approach can be effectively applied to analyze cocoa production in Central Sulawesi Province by incorporating the influence of exogenous variables. The best model produced is SARIMAX $(1,1,0)(0,1,0)^{12}$, which has a land area variable with a five-month lag and includes a significant AR(12) seasonal aspect. The results of the model accuracy evaluation indicated by the Mean Absolute Percentage Error (MAPE) value of 3.29% reflect a high level of prediction accuracy, so that the SARIMAX model $(1,1,0)(0,1,0)^{12}$ can be used as an analysis tool in strategic cocoa production planning, especially in anticipating production fluctuations influenced by changes in land area.

Reference

[1] N. Aisyah, A. I. Sukarni, D. Candra, and R. Novita, "Pengelompokan Hasil Perkebunan di Indonesia Menggunakan Fuzzy Classification Plantation Products in Indonesia Using Fuzzy C-Means," vol. 9, no. 2, pp. 361–372, 2024. https://doi.org/10.24114/cess.v9i2.50682

[2] F. Akhirul--, T. La Ola, A. Bafadal, and B. Budiyanto, "Kajian Produksi Dan Pendapatan Usahatani Kakao Pada Program Gernas (Gerakan Nasional) Kakao Di Sulawesi Tenggara," *J. Sosio Agribisnis*, vol. 3, no. 2, pp. 10–14, 2018, doi: 10.33772/jsa.v3i2.7586. https://doi.org/10.33772/jsa.v3i2.7586

[3] D. Erianto, "Komoditas Kakao: Sejarah, Manfaat, Produsen Dunia, dan Ekspor Indonesia," Kompas pedia.

[4] albaenati Aminah, "Ekspor Kakao Indonesia : Analisis Pengaruh Faktor Produksi , Luas".

[5] pratama harya Reski, "STRATEGI ADAPTASI DAMPAK PERUBAHAN IKLIM PADA TEKNOLOGI BUDIDAYA KAKAO (Theobroma cacao L.)," pp. 1–23, 2023.

[6] T. Handayani, R. S. Lubis, and R. Aprilia, *Peramalan Tingkat Produksi Kakao Tahun 2021 Di Provinsi Sumatera Utara Dengan Metode Double Exponential Smoothing Brown*, vol. 3, no. 1.

2021. doi: 10.15548/map.v3i1.2753. https://doi.org/10.15548/map.v3i1.2753

[7] A. Fauzan, M. E. Dalimunthe, P. Siagian, P. Studi, T. Elektro, and K. Medan, "Sistem Forecasting Penggunaan Daya Harian dengan Metode SARIMAX," no. 3, pp. 316–319, 2024.

[8] M. C. K. Putri and W. Anggraeni, "Penerapan Metode Campuran Autoregressive Integrated Moving Average Dan Quantile Regression (ARIMA-QR) untuk Peramalan Harga Cabai sebagai Komoditas Strategis pertanian Indonesia," *J. Tek. ITS*, vol. 7, no. 1, 2018, doi: 10.12962/j23373539.v7i1.28219. https://doi.org/10.12962/j23373539.v7i1.28219

[9] P. R. Kelembapan, M. Metode, and V. Exogenous, "Prediksi rata-rata kelembapan menggunakan metode sarimax dengan rata-rata temperatur sebagai variabel exogenous," vol. 03, no. 02, pp. 93–106, 2024.

[10] L. Hablinawati and J. Nugraha, "Peramalan Nilai Tukar Petani di Daerah Istimewa Yogyakarta Menggunakan Metode ARIMA," *Emerg. Stat. Data Sci. J.*, vol. 2, no. 1, pp. 85–96, 2024, doi: 10.20885/esds.vol2.iss.1.art9. https://doi.org/10.20885/esds.vol2.iss.1.art9

[11] R. Riung, N. Nainggolan, and Y. A. R. Langi, "Peramalan Jumlah Penumpang Kapal Laut di Pelabuhan Melonguane Kabupaten Kepulauan Talaud Menggunakan Metode SARIMA," *J. MIPA*, vol. 13, no. 2, pp. 62–66, 2024, doi: 10.35799/jm.v13i2.55741. https://doi.org/10.35799/jm.v13i2.55741

[12] A. R. Dimashanti and Sugiman, "Peramalan Indeks Harga Konsumen Kota Semarang Menggunakan SARIMA Berbantuan Software Minitab," *Prisma*, vol. 4, pp. 565–576, 2021, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[13] S. Nurwahdania and W. Sulistijanti, "MUSIMAN AUTOREGRESSIVE TERINTEGRASI MOVING AVERAGE DENGAN METODE EXOGENOUS INPUT (SARIMAX) Prosiding Seminar Edusainstech FMIPA UNIMUS 2020 ISBN : 978-602-5614-35-4," pp. 451–461, 2018.

[14] Poppy Marselina Kristiani and David Andrian, "Peramalan permintaan produksi wafer stick di PT. GarudaFood Putra Putri Jaya Tbk, Gresik," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 228–235, 2023, doi: 10.37373/tekno.v10i2.453. https://doi.org/10.37373/tekno.v10i2.453

[15] M. I. Rizki and T. A. Taqiyyuddin, "Penerapan Model SARIMA untuk Memprediksi Tingkat Inflasi di Indonesia," *J. Sains Mat. dan Stat.*, vol. 7, no. 2, pp. 62–72, 2021, doi: 10.24014/jsms.v7i2.13168. https://doi.org/10.24014/jsms.v7i2.13168