

Cocoa price prediction in North Sumatra using Singular Spectrum Analysis (SSA) algorithm

Sintia Fransiska^{1)*}, Ismail Husein²⁾

¹⁾²⁾ Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

¹⁾sintia0703191069@uinsu.ac.id, ²⁾ husein_ismail@uinsu.ac.id

Submitted: Jun 15, 2023 | **Accepted:** Jul 2, 2023 | **Published:** Jul 2, 2023

Abstract: The implementation of agricultural development is basically aimed at increasing the welfare of the people, especially farmers, providing a source of foreign exchange through exports, supplying food and industrial raw materials, alleviating poverty, providing employment and improving people's income. Cocoa is a leading commodity which is a source of income for farmers in North Sumatra. Price fluctuations sometimes make farmers suffer losses, so it is necessary to make a cocoa price prediction to anticipate future losses. This study aims to determine the prediction results of cocoa prices in North Sumatra in 2023 and the accuracy of the method used. The results of the study obtained the prediction of cocoa prices in North Sumatra Province in 2023 using the Singular Spectrum Analysis (SSA) method from January to December, respectively, Rp. 34876 in January, February prediction of Rp. 33967, March prediction of Rp. 33446, in April Rp 33725, prediction in May of Rp. 33986, prediction in June of Rp 33916, in July Rp. 34196, prediction in August of Rp. 34841, prediction in September of Rp. 35228, in October of Rp. 3479, in November Rp 344517, the December prediction is Rp 34770 with a prediction accuracy level based on the standard MAPE value of 0.96%. The MAPE value obtained indicates that the SSA approach with Windows length 18 and 14 groups is very accurate for prediction cocoa price in North Sumatra Province because it is less than 10% and close to 0%.

Keywords: Singular Spectrum Analysis, Prediction, MAPE

INTRODUCTION

The agricultural sector has a strategic role in economic development. The agricultural sector is one of the mainstays of the economy of North Sumatra Province which plays an important role as a contributor to the formation of Gross Regional Product (GRDP), providing a source of foreign exchange through exports, obtaining food and industrial raw materials, alleviating poverty, absorbing labor and increasing community income. In addition, the agricultural sector helps in driving local economic growth and regional development. The main objective of agricultural development is to improve the welfare of the people, especially farmers. Therefore, farmers' welfare is a top priority in development goals at every level of agricultural development efforts. The government seeks to increase agricultural production, ensure food supply security, and increase farmers' income and welfare through various agricultural development policies and programs. (Rusono, Sunari, Candradijaya, Martino, & Tejaningsih, 2013)

The agricultural sector has an important role in Indonesia's economic activities, as evidenced by its contribution to the gross domestic product (GDP) which is quite large, which is around 13.28 percent in 2021, second only to the industrial sector at 19.25 percent. One of the subsectors that has a lot of potential is the agriculture sector. The contribution of the plantation sector in GDP is 3.94 percent in 2021 and 29.67 percent of the agriculture, forestry and fisheries sector and ranks first in this sector. Cocoa is one of the plantation commodities that plays an important role in Indonesia's economic activities. Cocoa is also one of Indonesia's most important export commodities as a foreign exchange earner besides oil and gas. Indonesia is the 6th largest cocoa producer in the world. (BPS, 2021).

North Sumatra Province produces commodities such as rubber, cocoa, tea, oil palm, coffee, cloves, coconut, cinnamon, sugarcane, and tobacco. However, of the several commodities listed as plantation crops, palm oil, rubber cocoa and coffee are the most important commodities in North Sumatra Province. The determination of the four commodities as superior is based on the ability to compete with the same commodities from other regions and even from abroad, both on their sustainable marketing and their ability to offer benefits to their managers. (Hasnudi & Iskandar, 2005)

*name of corresponding author



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The fruit of the cacao plant can also be consumed, but the part that is utilized is the seed. Theobromine, a mild stimulant, is an alkaloid found in cocoa beans. The main component of chocolate is the cocoa bean which is roasted and ground into chocolate powder before being used to make various foods and beverages. Only a small portion of the cocoa produced in Indonesia is used for internal purposes, most of it is exported. The exported products are mostly (78.5%) in the form of primary products, i.e. in the form of dry beans and a small portion (21.5%) in the form of processed products. The marketing of Indonesian cocoa beans has reached international markets and most of Indonesia's cocoa beans are exported abroad, although some industries have processed cocoa beans into semi-finished products. The development of cocoa bean exports from Indonesia shows relative growth from year to year, so it is an opportunity for Indonesia to earn foreign exchange through this commodity. What determines the price level in the international market is the quality of cocoa beans. Improving the quality of cocoa beans must start from the cultivation process to post-harvest handling, especially in the cocoa fermentation process, because without the fermentation process cocoa beans will produce a low chocolate flavor that is more dominantly bitter and astringent, so it is not liked by consumers. (Misnani, 2008) Based on data from the Department of Plantation and Livestock of North Sumatra Province, the selling price of fermented cocoa is more expensive than non-fermented cocoa.

Cocoa is one of Indonesia's most popular agricultural products and is traded on the global market. Cocoa prices are set based on prevailing prices on commodity exchanges in New York and London. However, cocoa prices at the farm level may vary due to prices set by traders, as farmers do not have direct access to factories or industries. Local governments also do not have the capacity to set cocoa prices. The price formation mechanism of cocoa beans is adjusted by the market mechanism, considering that cocoa is a global product. (Karim, Fatmawaty, Anas, & Wulandari, 2020)

Cocoa cultivation is always at risk of destruction due to pest and disease attacks, as well as seasonal conditions that do not favor production. Fluctuations in bean prices sometimes also cause huge losses for cocoa farmers. The slow but steady increase in inputs can never be matched by an increase in the selling price of the output. Cocoa prices often fluctuate, based on data obtained from the Plantation and Livestock Service Office of North Sumatra Province. The lowest price was 26172 in the second quarter of 2017, while the highest price was reached in the second quarter of 2016. Over the past two years, cocoa prices have continued to experience significant increases and decreases.

To avoid the negative impacts of cocoa price fluctuations, it is important to forecast cocoa prices, especially in North Sumatra Province. So that it can be used as decision support related to cocoa price issues and the results of the analysis can be used in setting the right policy. Forecasting is an art and science used to project events that will occur in the future. This process is done by using past data and compiling it in a mathematical model to estimate events that will occur in the future. (Prasetya & Lukiasuti, 2009). There are several forecasting methods in statistics that are suitable for forecasting time series data. Exponential smoothing models are recommended as a simple and economical technique with reasonably good prediction results in many applications. The seasonal element is influenced by the global market which will be high in certain months which will affect the price. Based on this, prediction can be done using the Singular Spectrum Analysis (SSA) method. SSA is a more flexible forecasting method compared to other forecasting methods because it uses a nonparametric approach which means that assumption tests such as independence and normality of residuals are not required and are suitable for stationary and non-stationary data. (Irwan, Adnan Sauddin, & Anita Kaimuddin, 2022)

Some previous research was conducted by GilangBimasaktiAndhika et al (2020). about using the SSA method to forecast the Farmer Exchange Rate in Bali Province with a good level of accuracy with a MAPE value of 0.49%. KukuH WahyuHidayat et al (2020) on modeling the number of hot spots in East Kalimantan province using the singular spectrum analysis method. Then further research was conducted by DesyTresnowatiHardi (2019), to forecast GRDP with data containing elements of seasonal patterned trends also showing a low error value based on the standard MAPE value of 1.59% each.

LITERATURE REVIEW

Previous research that discusses the Singular Spectrum analysis (SSA) method in 2022 with the title "Forecasting rainfall in Gorontalo using the Singular Spectrum Analysis method" Research journal article conducted by EkaPurnama. In this study, forecasting was carried out by combining the Linear Recurrent Formula (LRF) method with the Singular Spectrum Analysis (SSA) method. As a result of using SSA with windows length of $L = 36$, the rainfall forecast in Gorontalo resulted in a MAPE forecast on out-of-sample data of 0.029 or 2.9% according to the results. As a result, the rainfall forecast of Gorontalo Province using Singular Spectrum Analysis (SSA) method for 2022-2023 is quite accurate.

Another research that also uses the Singular Spectrum analysis method is in 2022 with the title "Pinrang Regency Rice Production Projection with Singular Spectrum Analysis Method". Research conducted by Irwan, Adnan Sauddin and Anita Kaimuddin. The test results on the projection of rice production in Pinrang Regency with the singular spectrum analysis method obtained the level of forecasting accuracy based on the standard

*name of corresponding author



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

MAPE value obtained of 4.97%. The MAPE value obtained is less than 10% and close to 0%, meaning that the SSA method with windows length 9 and 7 groups is very accurately used to forecast rice production in Pinrang Regency.

Based on the two previous studies, it can be seen that the Singular Spectrum Analysis method is very good in forecasting or prediction, so researchers are interested in using the singular spectrum analysis method to predict cocoa prices.

METHOD

This research uses quantitative research type and the research method used is secondary data analysis method. The data analysis conducted in this study is an analysis using the Singular Spectrum Analysis (SSA) method to predict cocoa prices in North Sumatra province. This research uses the help of *softwareR* studiowith *package rssa*. The research procedures that will be carried out in this study are as follows:

1. Perform descriptive statistics on the data.
2. Data analysis using Singular Spectrum Analysis (SSA).
3. Perform decomposition with Embedding and Singular Value Decomposition (SVD).
4. Reconstruction with grouping and Diagonal Averaging.
5. Calculating the result of modeling
6. Calculating forecaster accuracy

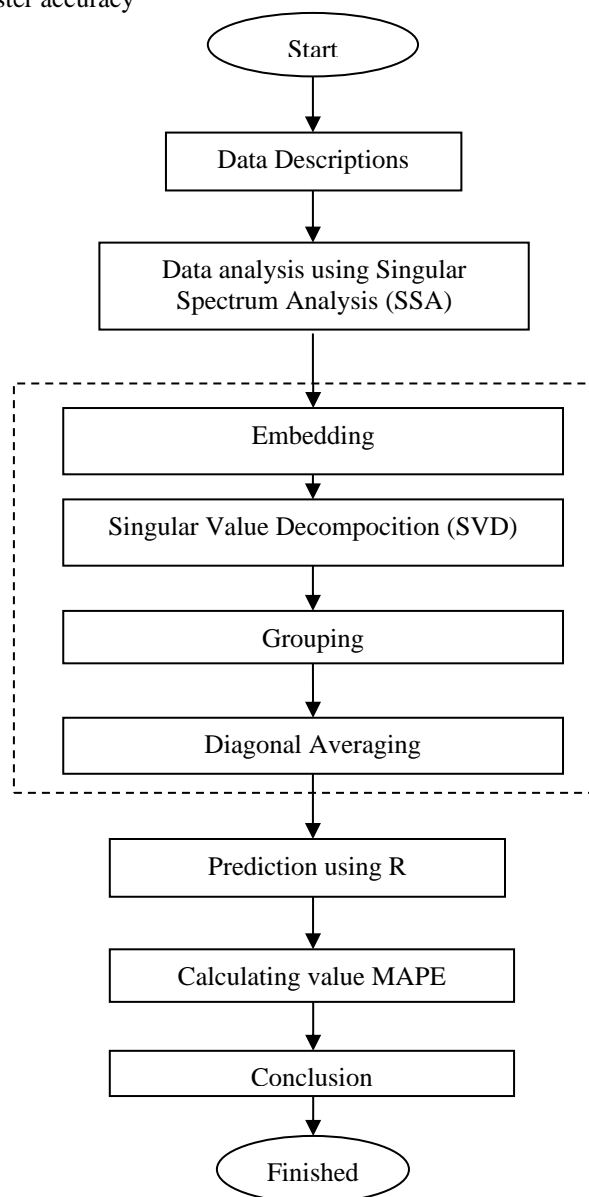


Figure 1. Research *flowchart* diagram

*name of corresponding author



The research data used is secondary data based on existing data in North Sumatra Province, all data used is average price data fermented dry cocoa beans in kilograms in 2015-2022. The data needed comes from the department of livestock and plantations of North Sumatra province, the data used consists of 1 variable which is divided into two parts, namely in sample and out sample data. In Sample data starts from January 2015 to December 2021 which is used to create a trajectory matrix X. while Out Sample data starts from January 2022 to December 2022 which is used to validate the accuracy of forecasting.

RESULT

Before analyzing the data, descriptive statistics of the research variables used are displayed first. These descriptive statistics are useful for displaying an overview of the data used, there are a total of 96 data consisting of 1 variable. The following are descriptive statistics of the data carried out with the help of Microsoft Excel software.

Table 1. Descriptive statistics of data

Month	Description			
	Minimum	Maximum	Mean	Standard deviation
January	28940	36400	32494	1002,084
February	30250	35500	33009,38	2317,593
March	30375	33770	33668,75	2495,451
May	27830	37400	33781,88	3306,55
June	25333	36600	32405,13	3769,244
July	25125	37000	32581,25	4121,1724
August	26400	36100	31971,88	4092,926
September	28125	37500	32910,63	3522,926
October	25212	37800	33004,63	3905,461
November	27400	38125	33647,88	3664,5
December	29000	37875	33389,5	29943,592

In North Sumatra Province, cocoa prices ranged from IDR 25,125 in July to IDR 38,125 in November. The average value ranged from IDR 31,971.88 in August to IDR 33,714.13 in April, the month with the highest average value. There are no outliers or extreme values in the tabulation, as evidenced by the standard deviation of the data being smaller than the mean.

After displaying Table 1, it continues to display data patterns using rstudio software which is displayed in the following image

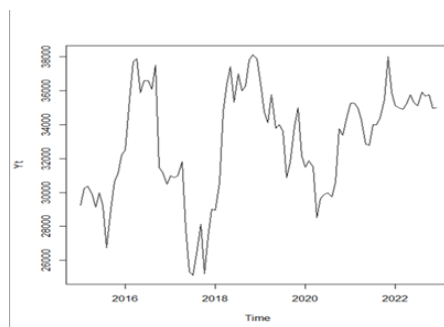


Figure 2 Time series plot of cocoa price data in North Sumatra Province

Figure 2 shows that the pattern of Cocoa Price data in North Sumatra has repeated increases and decreases so that data fluctuations are identified as containing seasonal elements.

The initial stage of cocoa price prediction using the SSA method is done using decomposition and reconstruction. When decomposing data there are two steps, namely embedding and Singular Value Decomposition.

At the embedding stage, the in-sample data is converted into one-dimensional data with a total of 84 data from January 2015 to December 2022 with the following arrangement:

*name of corresponding author



$$\begin{bmatrix} \text{Januari 2015} \\ \text{Februari 2015} \\ \text{Maret 2015} \\ \text{April 2015} \\ \text{Mei 2015} \\ \text{Juni 2015} \\ \text{Juli 2015} \\ \vdots \\ \text{Desember 2021} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ \vdots \\ x_{84} \end{bmatrix} = \begin{bmatrix} 29237 \\ 30250 \\ 30375 \\ 29900 \\ 29125 \\ 30000 \\ 29200 \\ \vdots \\ 35875 \end{bmatrix}$$

Furthermore, the data is converted into multidimensional data which is referred to as the trajectory matrix X with dimensions. $L \times K$. Determination of the L value is based on the minimum MAPE value where the L value analyzed is in the range is $2 < L < 42$. The L value used after the trial and error process is shown in Table 2 below:

Table 2 MAPE results based on window length (L) value

L	MAPE
10	0.009718013
14	0.009771591
17	0.009659174
18	0.009603477
19	0.009694527
20	0.009718662
22	0.009746927
30	0.01009546
38	0.00121105
39	0.01218637
41	0.01256813

In Table 4.6, the L value that produces the minimum MAPE value is in Table 2. $L = 18$. From the value of L , the value of $K = 67$ based on the equation $K = N - L + 1$ where the value of $N = 84$ and $L = 18$. After obtaining the value of $L = 18$, the value of $K = 67$, then the trajectory matrix X can be formed. The formation process is the value x_1 to x_{18} becomes the first column component, the value of x_2 until x_{19} becomes the second column component, the value of x_3 until x_{20} becomes the third column component and so on until x_{49} until x_{67} becomes the 14th column component in the trajectory matrix X which is written as follows:

$$X = \begin{bmatrix} 29237 & 30250 & 30375 & \dots & 30000 \\ 30250 & 30375 & 29900 & \dots & 29750 \\ 30375 & 29900 & 29125 & \dots & 30600 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 36600 & 36600 & 36100 & \dots & 35875 \end{bmatrix}$$

After embedding, the next step is to find the value of Singular Value Decomposition (SDV) or singular value decomposition which aims to obtain the separation of components in the decomposition of time series data. SVD starts by determining the *eigenvalue* ($\lambda_1, \lambda_2, \dots, \lambda_L$) of the symmetric matrix $S = XX^T$ where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L > 0$, as well as the *eigenvector* (u_1, u_2, \dots, u_L) of the S matrix. At this stage, calculations are carried out to find the *eigenvalue* based on the trajectory matrix $X_{(18 \times 67)}$. The first step taken is to form a symmetrical matrix $S = XX^T$ as follows:

$$S = X_{(18 \times 67)} \cdot X_{(67 \times 18)}^T$$

$$S = \begin{bmatrix} 70583353666 & 70499093055 & 70413310830 & \dots & 70617425925 \\ 70499093055 & 70613613997 & 70518986004 & \dots & 70602985275 \\ 70402997379 & 70525023805 & 70638930055 & \dots & 70618855600 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 70617425925 & 706029985275 & 70758858104 & \dots & 73225953997 \end{bmatrix}$$

After obtaining the symmetrical matrix, the eigenvalue values consisting of:

- 1) Singular Value ($\sqrt{\lambda_i}$)

The calculation of the eigenvalue and eigenvector of the square matrix S can be done with the equation $(S - \lambda I) = 0$ (1)

then the eigenvalue λ_i will be obtained for $i = 1, 2, \dots, L$ and $U_i = (u_1, u_2, \dots, u_L)$ is the eigenvector of the corresponding eigenvalue. From the symmetric matrix $S_{18 \times 18}$ From the symmetric matrix, the eigenvalue is calculated and then from the eigenvalue, the singular value is calculated, the results of which are shown in Table 4.7 below:

Table 3 Eigenvalues and Singular Values

No.	Eigenvalues	Singular values
1	1.289779e+12	1135684.348
2	6.544205e+09	80896.262
3	2.408966e+09	49081.223
4	5.768783e+08	24018.292
⋮	⋮	⋮
18	2.739022e+07	5233.567

2) Eigenvector (U_i)

After obtaining the singular value, the eigenvector value is then calculated, the results of which are shown in the following table:

Table 4 Eigenvectors

No.	U_1	U_2	...	U_{18}
1	-0.2321849	-0.29331908	...	0.09747742
2	-0.2324414	-0.30794999	...	-0.18878879
3	-0.2326742	-0.30826778	...	0.20702479
⋮	⋮	⋮	⋮	⋮
18	-0.2367240	0.27139824	...	-0.09991382

3) Principal Component (V_i)

After the eigenvalue and eigenvector are obtained, the principal component value is then calculated using the following equation

$$V_i = \frac{x^T U_i}{\sqrt{\lambda_i}} \tag{2}$$

$i^T V$ is the transpose of the matrix V_i

$$V_i^T = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1K} \\ v_{21} & v_{22} & \dots & v_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ v_{K1} & v_{K2} & \dots & v_{KK} \end{bmatrix} \tag{3}$$

After obtaining the *singular value* (λ_i), *eigenvector* (U_i) and *principal component* (V_i), the SVD of the trajectory matrix X is:

$$X = U_i \sqrt{\lambda_i} V_i^T \tag{4}$$

The three components in matrix X_i namely *singular value* (λ_i), *eigenvector* (U_i) and *principal component* (V_i) are called the *itheigtriple* of SVD. Therefore, the SVD for matrix X_i can be written as follows:

$$X = X_1 + X_2 + \dots + X_d \tag{5}$$

whose results are presented as follows:

Table 5 Principal component values

No.	V_1	V_2	...	V_{18}
1	-0.1191465	-0.126772323	...	-0.053670729
2	-0.1206872	-0.142107018	...	0.044859010
3	-0.1219146	-0.153815784	...	-0.075545540
⋮	⋮	⋮	⋮	⋮
67	-0.1264599	0.050284487	...	0.013202883

*name of corresponding author



The last stage of SSA analysis is Reconstruction. At this stage, reconstruction of the SVD result data is carried out with 2 steps, namely *grouping* and *diagonal averaging*, the steps of which are as follows:

a. Grouping

At this stage, grouping of the X_i matrix is carried out with the aim of separating the eigentriple components obtained at the SVD stage into several subgroups, namely trends, seasonality, and noise. The grouping process is done by grouping the index sets $i = \{1, 2, \dots, d\}$ into m disjoint subsets I_1, I_2, \dots, I_m with $m = d$. Then X_i is adjusted to the group $I = \{I_1, I_2, \dots, I_m\}$. Then, $X_i = X_1 + X_2 + \dots + X_d$ can be expanded into:

$$XI = XI_1 + XI_2 + \dots + XI_m \tag{6}$$

The step to select the set $I = \{I_1, I_2, \dots, I_m\}$ is called eigentriple grouping which is done by trial and error. The determination of group members is by the scatter diagram of XI , where a group will be formed if the scatter diagram of XI has the same shape, meaning that the components have similar characteristics. At this stage, the eigentriple grouping of SVD results is based on the characteristics of each component. To determine the members of the group is done by looking at the plot of the eigenvector. The following is a plot of the eigenvector with a value of $i = 1, 2, \dots, 18$:

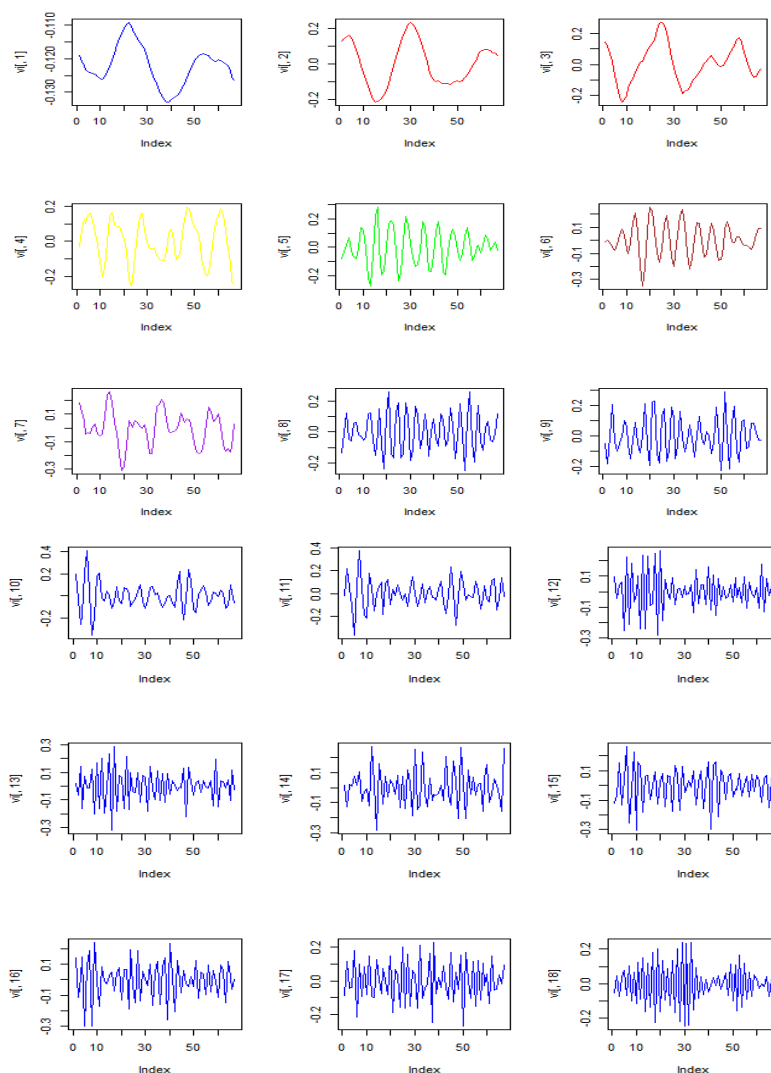


Figure 2 Data Eigenvector Plot

From Figure 2, it can be seen that the first component is a trend and the other components contain seasonal and noise patterns. Grouping using eigenvector patterns is subjective. There are several plots that have the same pattern and it is difficult to distinguish their characteristics. To see the similarity of characteristics between components more clearly, it can be seen from the W-correlation plot. The W-correlation plot is shown as follows

*name of corresponding author



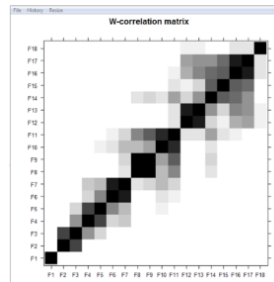


Figure 3 W-correlation plot

From Figure 3, based on the strong and low correlation of each group, it can be seen that the 8 components can be divided into 6 groups. The blackness of the slice between components F1 and F14 shows a strong and low correlation. The blackness of the slices between components F2 and F3, F5 and F7, F9 and F11, and components F2 and F3 all show a high association in Figure 3. The lightness of the slices between components F1, F4, F6, F8, F10, F12, and F14 indicates that there is only a weak association between these components. The groups that can be formed based on the eigenvector plots and W-correlation plots shown above are as follows:

Table 6 Grouping Results

Component	Group
1	1
2,3	2
4	3
6	4
5,7	5
8	6
9	7
10,11	8
12	9
13	10
14,15	11
16	12
17	13
18	14

From Table 5, 14 matrices of 18 different components are obtained. The 14 groups are matrices ($X_{I1}, X_{I2}, \dots, X_{I14}$) each of which is 14×67 which is as follows:

$$X_{I1} = \begin{bmatrix} 31417.60 & 31823.87 & 32147.51 & \dots & 33346.06 \\ 31452.31 & 31859.03 & 32183.02 & \dots & 33382.90 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 32031.79 & 32446.00 & 32775.97 & \dots & 33997.95 \end{bmatrix}$$

$$X_{I2} = \begin{bmatrix} -4040.351 & -4042.541 & -3692.069 & \dots & -631.1968 \\ -4040.815 & -4046.917 & -3696.148 & \dots & -631.8942 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -4119.337 & -4121.478 & -3764.246 & \dots & -643.5363 \end{bmatrix}$$

$$X_{I14} = \begin{bmatrix} 65.21827 & -54.51066 & 91.79956 & \dots & -16.04355 \\ 65.29032 & -54.57089 & 91.90098 & \dots & -16.06128 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 66.49324 & -55.57631 & 93.59418 & \dots & -16.36719 \end{bmatrix}$$

b. Diagonal Averaging

The final step in SSA is to transform each matrix X_I in equation into a new time series of length N . Suppose matrix Y is an arbitrary matrix of size $L \times K$ with elements y_{ij} for $1 \leq j \leq L$ and $1 \leq j \leq K$, where $L \leq K$.

*name of corresponding author



$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_K \\ y_{21} & y_{22} & \dots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & y_N \end{bmatrix} \tag{7}$$

The matrix Y is converted into the form of series g_1, g_2, \dots, g_N through *diagonal avergaingin* the following way:

$$g_k = \frac{\sum_{(l,k) \in A_s} y_{lk}}{|A_s|} \tag{8}$$

Where $|A_s|$ denotes the number of members of set. For $A_s = \{(l, k): l + k = s, 1 \leq l \leq L, 1 \leq k \leq K\}$ and $i + j = s$. Based on equation (2.12) suppose that the matrix Y is chosen. For $k = 1$ gives $g_1 = y_{11}$, for $k = 2$, then:

$$g_2 = \frac{y_{1,2} + y_{2,1}}{2}, g_3 = \frac{y_{1,3} + y_{2,2} + y_{3,1}}{3} \text{ and so on.}$$

From the previous step, 14 matrices representing the characteristics of each component are obtained. Each matrix will be converted into new time series data with a length of 14 according to each characteristic using the equation. The results of diagonal averaging of the 14 matrices are as follows:

Table 7 Diagonal averaging results

No.	Group1	Group2	Group3	...	Group14
1	31286.47	-359.89216	-399.641046	...	-31.091148
2	31530.08	-1073.63658	151.125142	...	45.283633
3	31743.80	-1909.48986	515.920295	...	-56.246008
⋮	⋮	⋮	⋮	⋮	⋮
96	35512.19	-429.68833	496.743684	...	-24.319937

Next, calculate the forecasting results from the time series data of the diagonal averaging results with the *R-forecasting* method. *SSA forecasting* using LRF or known as *R-forecasting*, the time series used is the reconstructed series obtained from the results of *diagonal averaging*. Then M new data points will be determined to be forecasted. The forecasting model is as follows:

$$g_t = \begin{cases} \tilde{y}_i & \text{untuk } i = 0, \dots, N \\ \sum_{j=1}^{L-1} a_j g_{i-j} & \text{untuk } i = N + 1, \dots, N + M \end{cases} \tag{9}$$

R-forecasting is concerned with the estimation of *Linear Recurrent Formula (LRF)* i.e. a_1, \dots, a_d using the *eigenvectors* obtained from the SVD stage. Suppose $U = (u_1, u_2, \dots, u_{L-1}, u)$, $U^T U^{\bar{v}} = (u_1, u_2, \dots, u)_{L-1}^T$, and π_q is the last component of the *eigenvector* U or can be written $\pi_q = u_L$ then the LRF coefficient can be calculated by :

$$\mathfrak{R} = (a_{L-1}, a_{L-2}, \dots, a)_1^T = \frac{1}{1-v^2} \sum_{q=1}^r \pi_q U^{\bar{v}} \text{ With } v^2 = \sum_{q=1}^r \pi_q^2 \tag{10}$$

After 14 groups of diagonal averaging results are obtained, the next step is to forecast the out sample data using the *R-forecasting* method. The following shows the results of the out sample data forecast:

Table 8 Forecasting results on out sample data

Year	Month	Actual Data	Forecast Data	Difference
2022	January	35125	36227	1102
	February	35000	35026	26
	March	34900	34873	27
	April	35250	35023	227
	May	35750	35365	385
	June	35300	35401	101
	July	35125	35636	511
	August	35900	35045	855
	September	35660	35704	44
	October	35750	35957	207
	November	34958	35544	586
	December	35000	34996	4

From Table 8, it can be seen that the difference between the actual data and the forecast data is quite small. To see the comparison between the two, it can also be presented in the form of a plot, which is as follows:

*name of corresponding author



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

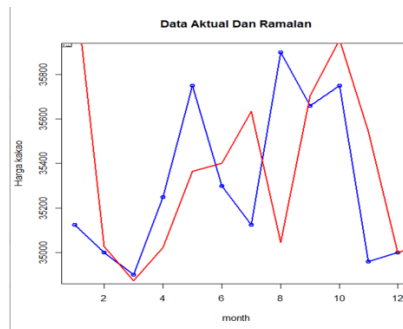


Figure 4 Plot of Actual Data and Forecast Data

In Figure 4, it can be seen that the plot results of actual and forecast data have almost the same pattern, where the forecast data is close to the actual data.

After doing SSA forecasting, then test the forecasting accuracy. To test the error rate of a forecast, it can be done using the *Mean Absolute Error* (MAPE) method with the formula :

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{g_i - \hat{g}_i}{g_i} \right| \right) \times 100\% \quad (11)$$

Table 9 Mape Value Criteria

Mape Value	Criteria
$x < 10\%$	excellent forecasting
$10\% \leq x < 20\%$	Good forecasting
$20\% \leq x < 50\%$	Good enough forecasting
$x \geq 50\%$	Bad forecasting

The Mean Absolute Percentage Error (MAPE) approach is used in this study to assess the forecasting accuracy of the SSA method. The out-of-sample data is calculated using equation 11 to provide the MAPE value, which is shown in Table 9 below:

Table 10 MAPE value on out sample data forecasting

Year	Quarter	Actual Data	Forecast Data	Error
2022	January	35125	36227.15	0.0313778430
	February	35000	35026.26	0.0007502709
	March	34900	34873.00	0.0007736038
	April	35250	35022.70	0.0064483220
	May	35750	35364.78	0.0107753559
	June	35300	35401.04	0.0028623853
	July	35125	35635.51	0.0145342002
	August	35900	35044.52	0.0238294964
	September	35660	35704.16	0.0012382808
	October	35750	35957.42	0.0058019067
	November	34958	35543.53	0.0167496085
	December	35000	34996.48	0.0001004467
Total Error				0.11524172
Mean Absolute Percentage Error (MAPE)				0.009603477

Table 10 shows that the error value between actual data and forecast data is quite small. Judging from the MAPE value obtained of 0.96%. As stated that predictions are said to be accurate if the MAPE value is less than 10% and close to 0%. Therefore, with windows length 18, forecasting is carried out on cocoa prices in 2023. The forecasting results are as follows:

Table 11 Prediction results

Month	Prediction Result
January	34867
February	33967
March	33446
April	33725
May	33986

*name of corresponding author



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

June	33916
July	34196
August	34841
September	35228
October	34792
November	34517
December	34770

From Table 10 it can be seen that the result of the prediction of the average price of fermented cocoa in 2023 in January is Rp 34876, prediction in February is Rp33967, prediction in March is Rp 33446, in April is Rp 33725, prediction in May is Rp 33986, predictions in June of Rp 33916, in July of Rp 34196, predictions in August of Rp 34841, predictions in September of Rp 35228, in October of Rp34792, predictions in November of Rp 34517, predictions in December of Rp 34770.

DISCUSSIONS

The data analyzed in this study is the average price data of fermented cocoa in North Sumatra Province from 2015 to 2022 using the Singular Spectrum Analysis (SSA) method. Based on the analysis results obtained, the forecast results in January 2022 tend to be higher than the actual data with a difference of 110 and an error of 0.0313778430, while the forecast results in January 2023 tend to be lower than the actual data in January 2022.

The prediction results in February 2022 tend to be higher than the actual data with a difference of 26 and an error of 0.0007502709, while the prediction results in February 2023 tend to be lower than the actual data in February 2022.

The prediction results in March 2022 tend to be lower than the actual data with a difference of 27 and an error of 0.0007736038, while the prediction results in March 2023 tend to be lower than the actual data in March 2022.

The prediction results for April 2022 tend to be lower than the actual data with a difference of 227 and an error of 0.0064483220, while the prediction results for April 2023 tend to be lower than the actual data in April 2022.

The predicted results in May 2022 tend to be lower than the actual data with a difference of 385 and an error of 0.017753559, while the forecast results in May 2023 tend to be lower than the actual data in May 2022.

The prediction results in June 2022 tend to be higher than the actual data with a difference of 27 and an error of 0.0028623853, while the prediction results in June 2023 tend to be lower than the actual data in June 2022.

The prediction results in July 2022 tend to be higher than the actual data with a difference of 511 and an error of 0.0145342002, while the prediction results in July 2023 tend to be lower than the actual data in July 2022.

The prediction results in August 2022 tend to be higher than the actual data with a difference of 855 and an error of 0.0238294964, while the prediction results in August 2023 tend to be lower than the actual data in August 2022.

The predicted results in September 2022 tend to be lower than the actual data with a difference of 44 and an error of 0.0012382808, while the predicted results in September 2023 tend to be lower than the actual data in September 2022.

The predicted results in October 2022 tend to be higher than the actual data with a difference of 207 and an error of 0.0058019067, while the predicted results in October 2023 tend to be lower than the actual data in October 2022.

The predicted results in November 2022 tend to be higher than the actual data with a difference of 586 and an error of 0.0167496085, while the predicted results in November 2023 tend to be lower than the actual data in November 2022.

The predicted results in December 2022 tend to be lower than the actual data and almost the same with a difference of 4 and an error of 0.0001004467, while the predicted results in December 2023 tend to be lower than the actual data in December 2022.

There is a not too significant difference in the forecast results with the actual data in 2022 from January to December as seen from the standard Mean Absolute Percentage Error (MAPE) value of 0.96%. The MAPE value shows that the forecasting results are accurate enough to forecast the Average Price in the cocoa month period in 2023, where the forecast results obtained are slightly lower than in 2022. In January, the prediction was Rp 34876, in February the prediction was Rp 33967, in March the prediction was Rp 33446, in April the prediction was Rp 33725, in May the prediction was Rp 33986, in June the prediction was Rp 33916, in July the prediction was Rp 34196, in August the prediction was Rp 34841, in September the prediction was Rp 35228, in

*name of corresponding author



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

October the prediction was Rp 34792, in November the prediction was Rp 34517, in December the prediction was Rp 34770.

CONCLUSION

Based on the results and discussion above, it is concluded that the results of the prediction of the average price of fermented cocoa in North Sumatra Province in 2023 using the Singular Spectrum Analysis (SSA) method are respectively from January at Rp 34876, prediction in February at Rp33967, prediction in March at Rp 33446, in April by Rp 33725, prediction in May by Rp33986, prediction in June by Rp 33916, in July by Rp 34196, prediction in August by Rp 34841, prediction in September by Rp 35228, in October by Rp34792, prediction in November by Rp 34517, prediction in December by Rp 34770. As well as the level of accuracy of cocoa price forecasting with the SSA method based on the standard MAPE value obtained of 0.96% which means the resulting prediction is very good.

REFERENCES

- BPS. (2021). *Statistik Kakao Indonesia*. Jakarta: Badan Pusat Statistik.
- E.P.Box, G., Jenkins, G. M., C.Reinsel, G., & M.Ljung, G. (2016). *Time Series Analysis Forecasting and Control*. New Jersey: Jhon Wiley & Sons.
- Golyandina, & Zhigljavsky. (n.d.). *Singular Spectrum analysis for Time Series*. STAT ME.
- Hajjah, A., & Nora Marlim, Y. (2021). Analisis Error Terhadap Peramalan Data Penjualan Error Analysis Toward Sales Data Forecasting. *Februari*, 20(1), 1–9.
- Hardi, D. T., Safitri, D., & Rusgiyono, A. (2019). Peramalan Produk Domestik Bruto (Pdb) Sektor Pertanian, Kehutanan, Dan Perikanan Menggunakan Singular Spectrum Analysis (Ssa). *Jurnal Gaussian*, 8(1), 68–80. doi:10.14710/j.gauss.v8i1.26623
- Hasnudi, & Iskandar, S. (2005). *Rencaan Statigis Pembangunan Perkebunan di Provinsi Sumatera Utara Tahun 2005-2012*. Medan: Lecturer Papers Fakultas Pertanian Universitas Sumatera Utara.
- HIDAYAT, K. W., WAHYUNINGSIH, S., & NASUTION, Y. N. (2020). Pemodelan Jumlah Titik Panas Di Provinsi Kalimantan Timur Dengan Metode Singular Spectrum Analysis. *Jambura Journal of Probability and Statistics*, 1(2), 78–88. doi:10.34312/jjps.v1i2.7287
- Idrus, R. A., Ruliana, & Aswi. (2022). Penerapan Metode Singular Spectrum Analysis dalam Peramalan Jumlah Produksi Beras di Kabupaten Gowa. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(2), 49–58. doi:10.35580/variansiunm40
- Irwan, Adnan Sauddin, & Anita Kaimuddin. (2022). Proyeksi Produksi Padi Kabupaten Pinrang Dengan Metode Singular Spectrum Analysis. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, 10(1), 100–109. doi:10.24252/msa.v10i1.29869
- Jenderal, S. (2007). *Gambaran Sekilas Industri Kakao*.
- Karim, I., Fatmawaty, Anas, & Wulandari, E. (2020). *Agribisnis Kakao*. Yogyakarta: Deepublish.
- Kusumasturi, A., Khoiron, A. M., & Achmadi, T. A. (2020). *Metode Penelitian Kuantitatif*. Yogyakarta: Deepublish.
- Misnani. (2008). *Physico-Chemical Changes During Cocoa Fermentation and Key Enzymes Involved*. Review Penelitian Kopi dan Kakao.
- Prasetya, H., & Lukiastuti, F. (2009). *Manajemen Operasi*. Yogyakarta: MedPress.
- Riyanto, S., & Putera, A. R. (n.d.). *Metode Riset Penelitian Kesehatan dan Sains*. Yogyakarta: Deepublish.
- Ruhiat, D., Andiani, D., & Kamilah, W. N. (2020). Forecasting Data Runtun Waktu Musiman Menggunakan Metode Singular Spectrum Analysis (Ssa). *Teorema: Teori Dan Riset Matematika*, 5(1), 47. doi:10.25157/teorema.v5i1.3286
- Rusono, N., Sunari, A., Candradijaya, A., Martino, I., & Tejaningsih. (2013). *Analisis Nilai Tukar Petani (NTP) sebagai bahan penyusunan RPJMN Tahun 2015-2019*. Jakarta: Kementerian Perencanaan Pembangunan Nasional.
- Sahabannur. (2018). *Teknologi Fermentasi Biji Kakao*. Bogor: IPB Press.