

Contents lists available at DJM DESIMAL: JURNAL MATEMATIKA **[p-ISSN: 2613-9073](http://issn.pdii.lipi.go.id/issn.cgi?daftar&1510197406&1&&) (print)[, e-ISSN: 2613-9081](http://issn.pdii.lipi.go.id/issn.cgi?daftar&1510199675&1&&) (online)[, DOI 10.24042/djm](https://search.crossref.org/?q=Desimal%3A+Jurnal+Matematika&publication=Desimal%3A+Jurnal+Matematika) http://ejournal.radenintan.ac.id/index.php/desimal/index**

Implementation of singular spectrum analysis in the forecasting of seawater wave height

Wika Dianita Utami* , Ade Candra Agustina

UIN Sunan Ampel, Indonesia

ARTICLE INFO ABSTRACT

Article History

Received : 14-08-2023 Revised : 18-09-2023 Accepted : 26-09-2023 Published : 30-12-2023

Keywords:

Forecasting; Seawater Wave Height; Singular Spectrum Analysis.

*Correspondence: E-mail: wikadianita@uinsby.ac.id

Doi: 10.24042/djm.v6i3.18382

Indonesia is renowned as a maritime nation, positioned amidst the Pacific Ocean and the Indian Ocean. This strategic location grants Indonesia the distinct advantage of serving as a global crossroads for maritime traffic, particularly with regards to trade and waterborne transportation. Among Indonesia's bustling ports, Tanjung Priok Port stands out as one of the busiest. In this context, the measurement of seawater wave height assumes a pivotal role in shaping the dynamics of transportation and commercial activities at Tanjung Priok Port. Hence, the availability of predictive insights into forthcoming seawater wave height assumes paramount significance in proactively addressing potential calamities and orchestrating maritime endeavors more efficaciously. This study aims to apply the Singular Spectrum Analysis (SSA) technique to forecast the wave height of seawater at Tanjung Priok Port. The dataset employed encompasses the daily seawater wave height observations recorded at Tanjung Priok Harbor during the timeframe from January 2022 to May 2023. The findings of this research unveil a parameter value of L = 98, a Grouping Effect (r) of 13, and a Mean Absolute Percentage Error (MAPE) value of 10.01%. This MAPE value signifies that the forecasting yielded by the Singular Spectrum Analysis (SSA) methodology exhibits a satisfactory level of accuracy in prognosticating future seawater wave heights at Tanjung Priok Port.

<http://ejournal.radenintan.ac.id/index.php/desimal/index>

INTRODUCTION

Indonesia, often referred to as an archipelago, boasts a larger expanse of sea (75%) compared to its land area (25%) (Rahmadani, Setiawan, & Adinugroho, 2019). From a geographical standpoint, Indonesia holds a significant maritime

position, situated strategically between the Asian and Australian Continents, as well as between the Indian Ocean and the Pacific Ocean. This unique positioning grants Indonesia numerous advantages, serving as a crucial global intersection for maritime traffic and international economic activities (Dewi, 2021). Given these factors, the demand for efficient transportation infrastructure, particularly maritime transport, is evident to enhance connectivity between islands and consequently foster positive growth in regions equipped with ports (Dewi, 2021).

Tanjung Priok Port, situated in North Jakarta, holds the distinction of being Indonesia's most expansive and bustling seaport, serving as a pivotal gateway. Approximately half of Indonesia's importexport commodities and inter-island goods traverse through this port. This prominence establishes Tanjung Priok Port as the primary maritime hub on the island of Java, equipped with contemporary technology and facilities capable of facilitating direct voyages for ships to numerous international trade hubs.

In maritime operations, various elements influence the trajectory of shipping endeavors, with seawater wave height being a significant factor (Rahmadani et al., 2019). Understanding the fluctuations in wave amplitude stands as a critical facet in the realm of transportation and maritime operations, serving as a crucial preemptive measure against potential aquatic disasters. Consequently, early access to seawater wave height data holds paramount importance prior to engaging in shipping activities, playing a pivotal role in decision-making regarding the feasibility of sailing (Rahmadani et al., 2019).

Seeing the importance of information about seawater wave heights, we must be able to find out data about seawater wave heights in the future by making forecasts. The method used is Singular Spectrum Analysis (SSA) which is formed from two stages of the process that enhance each other, namely decomposition (embedding and SVD) and reconstruction (grouping and diagonal averaging) (Ete, 2017). The SSA method is independent of various statistical assumptions, such as the assumption of

residual normality and stationarity, so it can be used on stationary and nonstationary data, making it a more flexible forecasting method than other methods (Lubis, Johra, & Darmawan, 2017).

The implementation of the SSA method is widely carried out in forecasting problems Krishnannair & Aldrich (2019), Sulandari, Subanar, Lee, & Rodrigues (2020), Koad & Jaroensutasinee (2021), Loizou & Christmas (2022), Sodiqin, Sulandari, & Respatiwulan (2021), and Swart, Otter, & Lamoth (2022). Several investigations have been conducted concerning seawater wave height forecasting and the application of the SSA methodology. Vicho et al. employed the ID3 algorithm to predict extreme waves in seawater at Tanjung Priok Port, attaining a substantial accuracy rate of 88% (Reonaldho, Saepudin, & Adytia, 2020). The reliability of the SSA method in forecasting rice prices at the wholesale level in Indonesia, using values of $L = 30$ and group $(r) = 11$, resulting in an impressively low MAPE value of 1.01%. Fitri, Gamayanti, & Gunawan (2017) captured fisheries production data in West Java Province and revealed that the optimal SSA model was achieved with $L = 20$ and a group count (r) of 4, yielding a MAPE value of 6.19%.

Based on this, this research will apply the Singular Spectrum Analysis (SSA) method to predict the seawater wave height at Tanjung Priok Port, which will then determine the level of accuracy using the Mean Absolute Percent Error (MAPE).

METHOD

This study falls under the category of quantitative research, with the primary research variable being the daily seawater wave height data at Tanjung Priok Port, collected from January 2022 to May 2023, comprising a total of 511 data points. The dataset will be divided into two distinct segments: the first segment encompasses data from January 1, 2022, to April 10, 2023, comprising 460 data points for training, while the second segment spans from April 11, 2023, to May 31, 2023, with 51 data points for testing.

Singular Spectrum Analysis (SSA)

SSA serves as a non-parametric technique within the realm of time series analysis, employed for the purpose of forecasting. Its primary objective is to engage in a decomposition procedure on a given dataset, generating multiple smaller components that can be meaningfully interpreted (Irwan, Adnan Sauddin, & Anita Kaimuddin, 2022). The fundamental algorithm of SSA entails breaking down the original time series data into new time series data, encompassing elements such as trend, seasonal, and noise components (Broomhead & King, 1986).

The SSA method is done in two stages: the first is the decomposition, and the second is the reconstruction. The two processes each consist of two separate stages. In the decomposition stage, time series data undergoes a transformation into matrix format, followed by matrix decomposition. During the reconstruction stage, the matrix format is converted back into the time series data format, which is then employed for the forecasting process. 1. Decomposition

• Embeddding

Create a trajectory matrix, denoted as X (Hankel Matrix), with dimensions $L \times K$, using the provided time series data. The choice of the parameter L (window length) is determined through iterative experimentation, ensuring it meets the condition $2 \leq L \leq N/2$, where $N = 460$, and the value of K is calculated as $N - L + 1$.

• Singular Value Decomposition (SVD)

Determine the singular values $(\lambda_1, \lambda_2, ..., \lambda_L)$ of the symmetric matrix $S = X X^T$ based on:

$$
det(S - \lambda I) = 0.
$$

Determine the corresponding eigenvector of the matrix S i.e. (U_1, U_2, \dots, U_L) . Determine the principal component $(V_1, V_2, ..., V_L)$ of matrix S based on equation $V_i = \frac{x^T U_i}{\sqrt{\lambda_i}}$ $\frac{\sqrt{a_i}}{\sqrt{\lambda_i}}$. Once the three elements of the eigentriple have been acquired, specifically the singular value $(\sqrt{\lambda_i})$, eigen vector (U_i) and principal component $(V_i^T$). Subsequently, the Singular Value Decomposition (SVD) of the trajectory matrix X is computed using the following formula:

$$
= U_i \sqrt{\lambda_i} V_i^T \tag{1}
$$

 X where (U_i) is an orthonormal matrix with size $L \times L$, and $(\sqrt{\lambda_i})$ is a diagonal matrix with size $L \times K$. (V_i^T) is an orthonormal matrix with size $K \times K$. Therefore, the Singular Value Decomposition (SVD) for matrix X_i can be expressed as follows (Jatmiko, Rahayu, & Darmawan, 2017):

 $X_i = X_1 + X_2 + \cdots + X_d$.

- 2. Reconstruction
	- Grouping

During this phase, the matrix X_i undergoes a grouping procedure with the intention of segregating the eigentriple components acquired during the SVD phase into distinct categories, specifically trends, seasonality, and noise. This grouping process involves the aggregation of sets of indices $i = \{1, 2, ..., d\}$ into m subsets that disjoin $I_1, I_2, ..., I_m$ by $m = d$. Afterwards, X_i is adapted to the corresponding group $I = \{I_1, I_2, ..., I_m\}.$ Such that $X_i = X_1 + X_2 + \cdots + X_d$ can be expressed as follows:

 $X_I = X_{I1} + \cdots + X_{Im}$ (2) The steps involved in selecting a collection $I = \{I_1, I_2, \dots, I_m\}$ are called eigentriples which are done by trial and error.

• Diagonal Averaging

The final step in Singular Spectrum Analysis (SSA) involves converting each X_I matrix in Equation (2) into a new time series of length N using diagonal averaging, as outlined below:

Desimal, 6 (3), 2023 - 288 Wika Dianita Utami, Ade Candra Agustina

$$
g_k = \frac{\sum_{(l,k)\in A_s} y_{l,k}}{|A_s|} \tag{3}
$$

where $|A_{\mathcal{S}}|$ denotes the number of set members A_s , and defined $A_s = \{(l, k): l +$ $k = s, 1 \leq l \leq L, 1 \leq k \leq K$ } and $i + j =$ (Golyandina, Nekrutkin, & Zhigljavsky, 2001). Based on Equation (3), for $k = 1$, then $g_1 = y_{1,1}$; for $k = 2$, then $g_2 =$ $y_{1,2}+y_{2,1}$ $\frac{y_{2,1}}{2}$; for $k = 3$, then $g_3 = \frac{y_{1,3} + y_{2,2} + y_{3,1}}{2}$ $\frac{2,2+93,1}{2}$ and the process continues (Asrof, 2017). 3. Prediction with R-forecasting

R-forecasting is related to the Linear Recurrent Formula (LRF) estimation, which is a_1, \ldots, a_d , using the eigenvector obtained from the SVD stage. Let $U =$ $(u_1, u_2, \dots, u_{L-1}, u_L)^T$, $U^{\bar{v}} =$

 $(u_1, u_2, \dots, u_{L-1})^T$, and π_q is the last component of the eigenvector; it can be expressed as $\pi_a = u_L$. . Then, the calculation of the LRF (Local Regression Filter) coefficient can be calculated by:

$$
R = (a_{L-1}, a_{L-2}, ..., a_1)
$$

=
$$
\frac{1}{1 - v^2} \sum_{q=1}^r \pi_q U^{\bar{v}}
$$
 (4)

where $v^2 = \sum_{q=1}^r \pi_q^2$ (Jatmiko et al., 2017).

Then, M new data points will be determined to be predicted. The prediction model is as follows:

$$
g_i = \begin{cases} \tilde{y}_i & \text{for } i = 0, ..., N \\ \sum_{j=1}^{L-1} a_j, g_{i-1} & \text{for } i = N+1, ..., N+M \end{cases}
$$
 (5)

where $g_{N+1}, ..., g_{N+M}$ are the forecast results of SSA (Asrof, 2017).

Accuracy of Forecasting Results

The method used in measuring the level of forecasting accuracy is Mean Absolute Percent Error (MAPE) which is written in the following equation:

$$
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{F_t} \right| \times 100\% \quad (6)
$$

where A_t is the actual value, F_t is the
forecasting value, and *n* is the amount of data.

Table 1. Criteria for MAPE Value

No.	MAPE Value	Criteria
1	$< 10\%$	Very Good
\mathcal{L}	$10 - 20 %$	Good
3	$20 - 50%$	Ouite Good
4	>50%	Bad

The steps conducted when implementing the Singular Spectrum Analysis (SSA) technique for forecasting the seawater wave height are presented in Figure 1.

RESULTS AND DISCUSSION

Figure 2 shows the results of the time series data plot of all daily seawater wave height data at Tanjung Priok Port:

Figure 2. Time Series Plot of Seawater Wave Height Data

Figure 2 illustrates the existence of a recurring pattern, evident from the alternating upward and downward trend. Based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), plots reveal that the partial autocorrelation function (PACF) exhibits a significant correlation at lag 24. This means that the pattern repeats every 24 days, as shown by the high wave height data every 24 days. It can be concluded that the data on seawater wave height falls within the category of seasonal data. Data is characterized as seasonal when it displays repetitive cyclic patterns within a consistent time interval.

The forecasting of seawater wave height using the Singular Spectrum Analysis (SSA) method involves a dualphase procedure: the decomposition process and the reconstruction process.

1. Decomposition

• Embeddding

The value of L (window length) is determined by trial and error. Because the data used is as large as 460, the suitable range for L is determined to be within the bounds of $2 \leq L \leq 230$. To search for the optimum L, a trial of the values of $L =$ 50, 100, 150, 200, 210, 220, and 230 was conducted to determine the L value corresponding to the minimum MAPE. The minimum MAPE value obtained was $L =$ 100. Furthermore, tracking is done around the value of $L = 100$ in the same way to get the L value with the smallest MAPE. The results of trial and error can be seen in Table 2.

Table 2. MAPE Value Results based on

Based on Table 2, the minimum MAPE value is 10.01%, so $L = 98$ and $K = 460 - 98 + 1 = 363$ are obtained. As a result, the trajectory matrix X is obtained as follows:

• Singular Value Decomposition (SVD)

The process of Singular Value Decomposition (SVD) encompasses three components: singular values, eigenvectors, and principal components referred to as eigentriples. The eigentriple value is established from a symmetric matrix $S = XX^T$. Using the symmetric matrix $S_{98\times98}$, the eigentriple values are then calculated. The following are the results of the eigentriple values:

Table 3. Singular Value

i	Eigen Value (λ_i)	Singular Value $(\sqrt{\lambda_i})$
1	48003.9000	219.0978
2	639.7467	25.2932
3	430.5403	20.7494
4	106.7630	10.3326
5	97.7484	9.8867
6	55.5127	7.4506
7	40.9054	6.3957
\vdots		
98	0.2072	0.4552

Desimal, 6 (3), 2023 - 290 Wika Dianita Utami, Ade Candra Agustina

No.	\boldsymbol{U}_1	U,		U_{98}
1	-0.100543	0.070213		-0.087073
2	-0.100535	0.075731		-0.176642
3	-0.100522	0.081074		-0.070963
4	-0.100495	0.086211		-0.124939
5	-0.100444	0.091194		0.1404716
6	-0.100387	0.096000		0.0471743
7	-0.10359	0.100587		-0.142686
\vdots			፡	
98	-0.100216	-0.070180		-0.009815

Table 4. Eigen Vector

Table 5. Principal Component

2. Reconstruction

• Grouping

At this stage, the eigentriple grouping of SVD results is based on the characteristics of each component related to trends, seasonality, and noise. The grouping procedure involves a graphical analysis approach, specifically by examining the graphs of singular values and eigenvectors.

Figure 3. Singular Value Chart

If the singular value decreases slowly, a noise component is detected. Referring to Figure 3, the singular value chart demonstrates a gradual decrease from the 14th to the 98th eigentriple, signifying it as a noise component. Therefore, the Grouping Effect (r) values,
comprised of 13 eigentriples, are comprised of 13 eigentriples, are identified as non-noise components, and they will be employed for detecting trend and seasonal elements. Additionally, a plot of the reconstructed series identifies these 13 eigentriples as trend and seasonal components.

Figure 4. Eigen Vector Chart

The slowly changing elements depicted in Figure 4 are recognized as trend components. It can be seen that the series reconstructed by eigenvector 1

depicts a slowly varying component and can therefore be grouped into the trend component. Next is the grouping of seasonal components.

Periodograms of eigenvector pairs and circular patterns identify the presence of seasonal components. Where the pattern increasingly resembles a circle, the component period will be greater. Based on Figure 5, it is known that eigenvector pairs 4-5, 9-10, and 11-12 have a period of 24, which produces seasonality. This means that the eigenvector pair is highly correlated between them and does not correlate with others.

The identification of trend and seasonal components can also be seen in the w-correlation plot. This plot is used to see how much correlation there is between eigenvectors; the darker the color of the plot, the stronger the correlation. Here are the results of the wcorrelation plot:

Figure 6. W-correlation Plot

Referring to Figure 6, it becomes evident that component F1 exhibits selfcorrelation and lacks significant correlation with other components, thus warranting grouping as trend components. On the other hand, components F4 and F5, F9 and F10, as well as F11 and F12, display strong correlation, as indicated by the dark slices between them, justifying their classification as seasonal components. Meanwhile, components F2, F3, F6, F7, F8, F13, and F14 show weaker correlation, with lightcolored slices between them, suggesting their categorization as noise components. The outcomes of the grouping process based on the w-correlation plot results are summarized in Table 6.

Table 6. Grouping Component Results

• Diagonal Averaging

At this stage, the transformation of the grouping results into a new time series is carried out. The purpose of this stage is to get the singular value of the components that have been separated, which is then used in forecasting.

Table 7. Diagonal Averaging Results

Afterward, a forecast is conducted using the R-forecasting method.

3. Prediction with R-forecasting

Moreover, a forecast is applied to the test data using the R-forecasting prediction technique. The following results were obtained.

Graph of Actual Data and Forecasting Results

Figure 7. Comparison Plot of Actual Data and Forecasting Results with SSA

Based on Figure 7, the forecasting data plot has a looping pattern that is almost the same as the actual data, where the forecasting data is close to the actual data. The disparity between the forecasted value and the actual data is referred to as the forecasting error, which is an inherent component of forecasting. The forecasting method with the lowest forecasting error is deemed the most appropriate for use.

Furthermore, computations of forecast errors using the Mean Absolute Percentage Error (MAPE) method are as follows: 5.190182

$$
MAPE = \frac{5.190492}{511} \times 100\%
$$

$$
= 10.01595\%
$$

The MAPE value obtained is 10.01595%. This means that the SSA method is good for predicting seawater wave heights in the future. Furthermore, predictions were made for 30 days, and the following results were obtained:

Table 8. Seawater Wave Height Forecast Result Data (Meters) in 2023

Based on the data in Table 8, the highest recorded seawater wave height at Tanjung Priok Port over a span of 30 days occurred on May 27, 2023, with a value of 1.039 m, while the lowest value was registered on June 25, 2023, reaching 1.020 m. The computed average wave height stands at 1.027833 m, indicating a categorization of the wave height at Tanjung Priok Port as medium during the middle of 2023. This conclusion is reaffirmed by the duplicate mention of the average seawater wave height, signifying the medium classification at Tanjung Priok Port during that time frame.

CONCLUSIONS AND SUGGESTIONS

The Singular Spectrum Analysis (SSA) method has an accuracy MAPE value of 10.01% in predicting the seawater wave height at Tanjung Priok Port for the future. Based on the MAPE value category, it shows that the SSA method is a good method for forecasting. Based on the SSA method, a value of $L = 98$ and a Grouping Effect $(r) = 13$ are obtained. The predicted seawater wave height for the next 30 days has an average wave height of 1.027833 m, meaning that the seawater wave height at Tanjung Priok Port is classified as medium in mid-2023.

This research focuses on one research variable to forecast seawater wave height using Singular Spectrum Analysis (SSA). For future research, it is recommended to use more than one research variable so that it can use the Multivariate Singular Spectrum Analysis (MSSA) method.

REFERENCES

- Asrof, A. (2017). Peramalan produksi cabai merah di jawa barat menggunakan metode singular spectrum analysis (ssa). *STATISTIKA: Journal of Theoretical Statistics and Its Applications*, *17*(2), 77–87. https://doi.org/10.29313/jstat.v17i 2.2839
- Broomhead, D. S., & King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, *20*(2–3), 217– 236. https://doi.org/10.1016/0167- 2789(86)90031-X
- Dewi, Z. Z. R. (2021). *Prediksi tinggi gelombang di pelabuhan ketapang menggunakan model fungsi transfer*. UIN Sunan Ampel Surabaya, Surabaya.
- Ete, A. A. (2017). *Forecasting the amount of foreign tourist travelers to indonesia by entrance guide using singular spectrum analysis and arima*. Institut Teknologi Sepuluh Nopember, Surabaya.
- Fitri, F., Gamayanti, N. F., & Gunawan, G. (2017). Metode ssa pada data produksi perikanan tangkap di provinsi jawa barat. *Jurnal Ilmiah Matematika Dan Pendidikan Matematika*, *9*(2), 95. https://doi.org/10.20884/1.jmp.201 7.9.2.2870
- Golyandina, N., Nekrutkin, V., & Zhigljavsky, A. A. (2001). Analysis of time series structure. In *Analysis of Time Series Structure*. https://doi.org/10.1201/97803678 01687
- 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). https://doi.org/10.24252/msa.v10i1 .29869
- Jatmiko, Y. A., Rahayu, R. L., & Darmawan, G. (2017). Perbandingan keakuratan hasil peramalan produksi bawang merah metode holt-winters dengan singular spectrum analysis (ssa). *Jurnal Matematika "MANTIK," 3*(1), 13.

https://doi.org/10.15642/mantik.20 17.3.1.13-24

- Koad, P., & Jaroensutasinee, K. (2021). Application of Singular Spectrum Analysis in Deep-Ocean Tide Reconstruction and Prediction. *Walailak Journal of Science and Technology (WJST)*, *18*(3), 1–21. https://doi.org/10.48048/wjst.2020 .7115
- Krishnannair, S., & Aldrich, C. (2019). Process monitoring and fault detection using empirical mode decomposition and singular spectrum analysis. *IFAC-PapersOnLine*, *52*(14), 219–224. https://doi.org/10.1016/j.ifacol.201 9.09.190
- Loizou, A., & Christmas, J. (2022). Sea state from ocean video with singular spectrum analysis and extended kalman filter. *Signal, Image and Video Processing*, *16*(6), 1497–1504. https://doi.org/10.1007/s11760- 021-02103-0
- Lubis, D. A., Johra, M. B., & Darmawan, G. (2017). Peramalan indeks harga konsumen dengan metode singular

spectral analysis (ssa) dan seasonal autoregressive integrated moving average (sarima). *Jurnal Matematika "MANTIK," 3*(2), 74–82. https://doi.org/10.15642/mantik.20 17.3.2.74-82

- Rahmadani, N., Setiawan, B. D., & Adinugroho, S. (2019). Prediksi ketinggian gelombang laut menggunakan metode jaringan saraf tiruan backpropagation. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(7), 6517–6525.
- Reonaldho, J. V., Saepudin, D., & Adytia, D. (2020). Prediksi gelombang ekstrim air laut di pelabuhan tanjung priok menggunakan algoritma id3. *EProceedings of Engineering*, *7*(1), 2700–2713.
- Sodiqin, M. A., Sulandari, W., & Respatiwulan. (2021). The application of singular spectrum analysis method in forecasting the number of foreign tourists visit to special capital region of jakarta. *Jurnal Riset Dan Aplikasi Matematika (JRAM)*, *5*(2), 92–102.
- Sulandari, W., Subanar, S., Lee, M. H., & Rodrigues, P. C. (2020). Time series forecasting using singular spectrum analysis, fuzzy systems and neural networks. *MethodsX*, *7*, 101015. https://doi.org/10.1016/j.mex.2020. 101015
- Swart, S. B., Otter, A. R. den, & Lamoth, C. J. C. (2022). Singular spectrum analysis as a data-driven approach to the analysis of motor adaptation time series. *Biomedical Signal Processing and Control*, *71*, 103068. https://doi.org/10.1016/j.bspc.2021 .103068