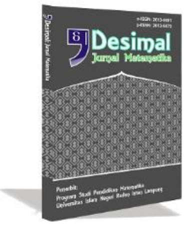
Contents lists available at DJM

DESIMAL: JURNAL MATEMATIKA

p-ISSN: 2613-9073 (print), e-ISSN: 2613-9081 (online), DOI 10.24042/djm

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



Application of singular spectrum analysis (SSA) method on forecasting train passengers data in sumatera

Debi Nur Fitriani, Widiarti*, Aang Nuryaman, Eri Setiawan

Universitas Lampung, Indonesia

ARTICLE INFO

Article History

Received : 03-10-2023

Revised : 10-10-2023

Accepted : 06-11-2023

Published : 30-12-2023

Keywords:

Forecasting; ; MAPE; Singular Spectrum Analysis; Windows Length.

*Correspondence: E-mail: widiarti.1980@fmipa.unila.ac.id

Doi: 10.24042/djm.v6i3.19040

ABSTRACT

A time series is a series of observations of a variable that is collected, recorded, or observed over a period of time in sequence. Singular Spectrum Analysis is a powerful method to analyze time series data by decomposing the original time series data into several small components that can be identified, such as trend, periodic, and noise components. One of the datasets that can be used is data on the number of train passengers in Sumatera in 2013–2022. In this study, the Singular Spectrum Analysis method is used to forecast the number of train passengers in Sumatera in 2013–2022. The best Singular Spectrum Analysis model in this study was obtained at a window length of 22 and a number of groups of 8, with a MAPE value of 19.55%.

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

INTRODUCTION

Forecasting is a technique for estimating a value in the future by looking at past and current data. The thing that must be considered in forecasting is the error value. The smaller the error value, the better the result. The important thing in choosing the appropriate forecasting method is to pay attention to the type of data pattern. One of the methods that can be used for forecasting is the Singular Spectrum Analysis method.

The Singular Spectrum Analysis method is a more flexible forecasting method compared to other forecasting methods because it uses a non-parametric approach that does not require several types of assumptions, such as independence and normality of residuals, and is suitable for stationary and non-stationary data (Hidayat, Wahyuningsih, & Nasution, 2020).

The data used in this study is data on the number of train passengers on Sumatera in 2013–2022 (Badan Pusat Statistik, 2022). Based on data, the number of train passengers always increases by several months each year. This increase usually occurs during Eid al-Fitr as well as during Christmas and the New Year. Therefore, it is important to

Copyright © 2023, Desimal, Print ISSN: 2613-9073, Online ISSN: 2613-9081

forecast the number of train passengers so that PT Kereta Api Indonesia can prepare additional facilities to deal with the surge in passenger numbers in the future.

Some previous research on the Singular Spectrum Analysis method includes forecasting farmer exchange rates (Andhika, Sumarjaya, & Srinadi, 2020), forecasting the number of foreign tourist visits to the Special Capital Region of Jakarta (Sodiqin, Sulandari, & Respatiwulan, 2021), forecasting rainfall in Gorontalo Province (Purnama, 2022), monthly rainfall forecasting at the Jatisrono Rain Post (Utami, Sulandari, & Handajani, 2021), predicting sea level change in Japanese coast (Niu et al., 2020), forecasting the amount of oil palm production in East Kalimantan Province (Siringoringo, Wahyuningsih, Purnamasari, & Arumsari, 2022), forecasting Balikpapan city inflation (Sergio, Wahyuningsih, & Siringoringo, 2023), forecasting Indonesian composite data (Wijayanti & Kartikasari, 2023), forecasting the CPI in South Sulawesi (Satriani & Ibnas, 2020).

METHOD

The Singular Spectrum Analysis (SSA) is a time series analysis technique that combines classical time series analysis, multivariate statistics, multivariate geometrics, dynamical systems, and signal processing (Golyandina & Zhigljavsky, 2013). This method is also quite flexible compared to similar forecasting methods because, in its application, neither the parametric model nor stationarity conditions must be assumed (Golyandina & Zhingljavsky, 2020). This makes SSA an analysis with a non-parametric approach. In this method, there are two stages: decomposition and reconstruction. In the decomposition stage, the steps taken are embedding and singular value decomposition. Meanwhile, in the reconstruction stage, the steps taken are grouping and diagonal averaging.

In the decomposition stage, the parameter that has an important role is the window length (L). The window length value is determined by trial and error, as there is no specific method to determine its value. Embedding is the process of converting one-dimensional time series data into multidimensional time series data and generating the trajectory matrix X. For example, $X = (X_1, X_2, \ldots, X_N)$ is one-dimensional time series data and there is no missing data, then X will be transformed into a trajectory matrix of size $L \times K$ with $2 < L < \frac{N}{2}$ and $K = N - L + 1$. The trajectory matrix X can also be referred to as a Hankel matrix, which is a matrix with all anti-diagonal elements of the same value. The Hankel matrix can be written as:

$$X = (x_i)_{L \times K} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_{LK} \end{bmatrix} \quad (1)$$

Singular Value Decomposition (SVD) aims to obtain component separation in the decomposition of time series data. The determination of the singular matrix in SSA can be defined by $S = XX^T$. If $\lambda_1, \ldots, \lambda_L$ are eigenvalues of the matrix S where $\lambda_1 \geq \ldots \geq \lambda_L \geq 0$ and $U_1, \ldots, U_L$ is eigenvector of each eigenvalue. Then the principal component can be written as $V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}$. So that the SVD of the trajectory matrix X is obtained as:

$$X = \sum_{i=1}^{d} U_i \sqrt{\lambda_i} V_i^T \quad (2)$$

The matrix X is formed from eigenvectors U_i, singular values $\sqrt{\lambda_i}$, and principal components V_i^T. These three elements can be called eigentriples.

The reconstruction stage is the stage where the data is reconstructed into new time series data. The first step in this stage is grouping. Grouping is a step where the matrix X is divided into several groups with the aim of separating the eigentriple components obtained at the SVD stage into several subgroups, trend, seasonal, and noise. This stage is done by grouping the index sets $i = \{1,2,K\ ,d\}$ into m disjoint subsets $I_1, I_2, K\ , I_m$ with $m = d$. Then X is matched with the group $I = \{I_1, I_2, K\ , I_m\}$. Then $X = X_1 + K + X_d$ can be expanded into $X_I = X_{I1} + K + X_{\mathrm{Im}}$.

The transformation of the grouping result X_I into a new time series of length N will be performed. This stage has the objective of obtaining the singular value of the separated components, and these components will be used in the forecasting. Diagonal averaging can be formulated as:

$$g_k = \begin{cases} \frac{1}{k}\sum_{m=1}^{k} f^{*}_{m,k-m+1} \\ \frac{1}{L^*}\sum_{m=1}^{L^*-1} f^{*}_{m,k-m+1} \\ \frac{1}{N-k+1}\sum_{m=k-K^*+1}^{N-K^*+1} f^{*}_{m,k-m+1} \end{cases} \quad (3)$$

where $L^* = \min(L,K)$ and $K^* = \max(L,K)$. So the resultant matrix X_{Im} will be formed into a series $\tilde{Y}^{(k)} = (\tilde{y}_1^{(k)}, K\ , \tilde{y}_N^{(k)})$. Therefore, the original sequence will turn into the sum of m sequences as:

$$y_n = \sum_{k=1}^{m} \tilde{Y}_N^{(k)} \quad (4)$$

Recurrent Forecasting is one of the most commonly used forecasting techniques of the SSA method. Recurrent Forecasting (R-forecasting) is performed directly with the help of the Linear Recurrent Formula (LRF). The time series used is a reconstructed series obtained from the results of diagonal averaging.

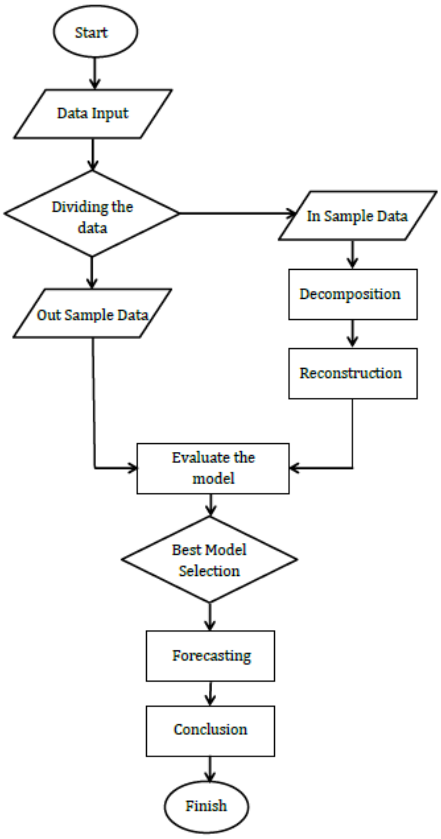


Figure 1. SSA Method

Forecasting error comparison is a simple way to determine whether a forecasting technique is worth choosing to use in calculating data forecasting or not. Forecasting accuracy will be higher if the values of MAD, MSE, and MAPE are smaller (Heizer & Render, 2011). The amount of forecasting error can be calculated using forecasting error measures; Mean Absolute Percentage Error (MAPE) is one of them. MAPE indicates how much the forecasting error is compared to the actual value. A data model will have very good performance if the MAPE value is below 10%. The following is the MAPE formula:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\% \quad (5)$$

RESULTS AND DISCUSSION

In the SSA method, the first stage to be done is the decomposition stage, which has two steps: embedding and singular value decomposition (SVD).

The first thing in the embedding stageis that the data will be divided into two parts: data from January 2013 until December 2021 as in-sample data and data from January 2022 until December 2022 as out-sample data. The length of the in-sample data is 108, and the length of the out-sample data is 12. In this stage, determining the window length (L) value is done through trial and error checking, with $2 < L < \frac{N}{2}$, the selection of the window length value is done by looking at the smallest MAPE value.

$L = 22$ has the smallest MAPE value so that the trajectory matrix X with dimensions $L \times K$ is obtained, with $K = 87$ which is obtained based on $K = N - L + 1$ equation where $N = 108$ and $L = 22$.

$$X = \begin{bmatrix} 327 & 279 & 305 & \Lambda & 476 \\ 279 & 305 & 276 & \Lambda & 85 \\ 305 & 276 & 318 & \Lambda & 8 \\ M & M & M & O & M \\ 420 & 370 & 484 & \Lambda & 220 \end{bmatrix}$$

In this SVD step, calculations will be carried out to find triple eigenvalues, which include singular values, eigenvectors, and principal components based on a symmetric matrix $S = XX^T$.

The next stage is reconstruction, which has two steps: grouping and diagonal averaging. In this grouping stage, the triple eigenvalues resulting from SVD will be grouped based on the characteristics of each component they have. To determine the members of the group, this will be done by looking at the plot of the eigenvector results.

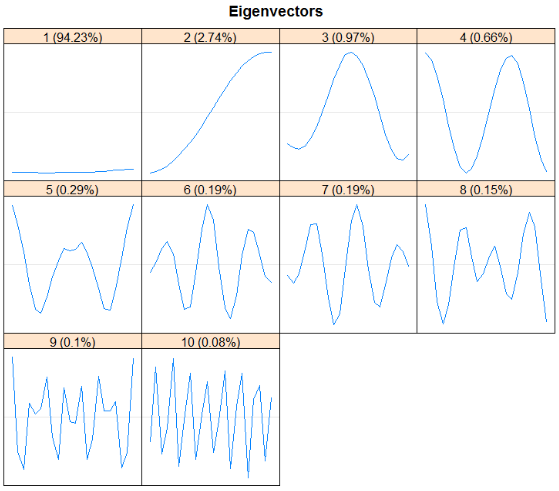


Figure 2. Plot of Eigenvector Values

There are some plots that have the same pattern, and it is difficult to distinguish their characteristics. Therefore, to see the similarity of characteristics between components more clearly, it can be seen in the W-correlation plot.

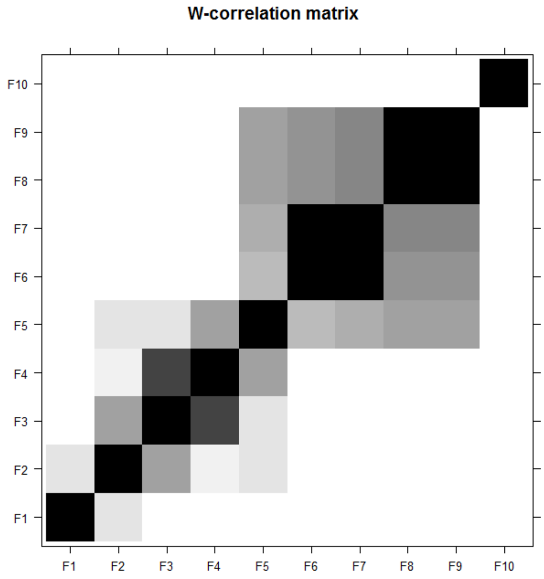


Figure 3. W-correlation Plot

From the results of the W-correlation plot, the 10 components can be grouped into 8 groups based on the high and low correlation of each component. The amount of correlation can be seen from the darkness and lightness of the intersection between components F1 and F10: the darker the color, the higher the correlation.

Copyright © 2023, Desimal, Print ISSN: 2613-9073, Online ISSN: 2613-9081

In the diagonal averaging stage, the groups that have been formed are rearranged into a new time series. Following are the results of diagonal averaging.

Table 1. Diagonal Averaging Results

No.	Group 1	Group 2	Group 3	Λ	Group 8
1	366.28	-63.31	-2.62	Λ	0.78
2	368.03	-60.71	-5.29	Λ	-3.29
3	372.11	-59.16	-5.77	Λ	5.19
M	M	M	M	O	M
108	160.56	9.92	-64.36	Λ	-2.65

Forecasting will be carried out on out-sample data using the R-forecasting method.

Table 2. Forecasting Results of Out Sample Data

Month	Actual Data	Forecast Data	Difference
January	238	309	71
February	210	255	45
March	305	231	74
April	277	242	35
May	480	317	163
June	452	408	44
July	469	421	48
August	368	394	26
September	400	380	20
October	467	361	106
November	451	361	90
December	579	359	220

It can be seen that the difference between actual data and forecast data is not too large. The forecasting accuracy of the SSA method in this study uses Mean Absolute Percentage Error (MAPE). With a window length of 22, the MAPE value is 19.55% obtained, so it can be said that the forecasting accuracy is good because it is between 10 and 20% (Chang, Wang, & Liu, 2007). The forecasting results obtained for 2023 are shown in Table 3.

Table 3. Forecasting Results for 2023

Month	Forecasting Results
January	339
February	299
March	295
April	294
May	300
June	306
July	288
August	271
September	274
October	236
November	172
December	120

CONCLUSIONS AND SUGGESTIONS

The results of forecasting the number of train passengers on Sumatera for 2023 with the Singular Spectrum Analysis method using a window length of 22 can be seen in Table 3. The level of forecasting accuracy based on the MAPE value obtained is 19.55%. Because the MAPE value obtained is between 10 and 20%, it can be concluded that the forecasting accuracy is relatively good.

As a suggestion, because the forecasting results for the number of train passengers on Sumatera with the SSA method tend to produce an upward or downward trend, the number of train passengers data in Indonesia can be analyzed with hybrid methods such as time series regression (ARIMA).

REFERENCES

Andhika, G. B., Sumarjaya, I. W., & Srinadi, I. G. A. M. (2020). Peramalan nilai tukar petani menggunakan metode singular spectrum analysis. *E-Jurnal Matematika*, *9*(3), 171. https://doi.org/10.24843/MTK.2020.v09.i03.p295

Badan Pusat Statistik. (2022). Jumlah penumpang kereta api (ribu orang). Retrieved November 17, 2023, from https://www.bps.go.id/ indicator/17/72/1/jumlah-penumpang-kereta-api.html

Chang, P.-C., Wang, Y.-W., & Liu, C.-H. (2007). The development of a weighted evolving fuzzy neural network for pcb sales forecasting. *Expert Systems with Applications*, *32*(1), 86–96. https://doi.org/10.1016/j.eswa.2005.11.021

Golyandina, N., & Zhigljavsky, A. (2013). *Singular spectrum analysis for time series* (1st ed.). Springer.

Golyandina, N., & Zhingljavsky, A. (2020). *Singular spectrum analysis for time series* (2nd ed.). Springer.

Heizer, J., & Render, B. (2011). *Manajemen operasi*. Salemba Empat.

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. https://doi.org/10.34312/jjps.v1i2.7287

Niu, Y., Guo, J., Yuan, J., Zhu, C., Zhou, M., Liu, X., & Ji, B. (2020). Prediction of sea level change in Japanese coast using singular spectrum analysis and auto regression moving average. *Chinese Journal of Geophysics*, *63*(9), 3263–3274.

Purnama, E. (2022). Aplikasi metode singular spectrum analysis (ssa) pada peramalan curah hujan di provinsi gorontalo. *Jambura Journal of Probability and Statistics*, *3*(2), 161–170.

Satriani, S., & Ibnas, R. (2020). Peramalan indeks harga konsumen (ihk) di sulawesi selatan dengan menggunakan metode singular spectrum analysis (ssa). *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, *8*(1), 82–89. https://doi.org/10.24252/msa.v8i1.17441

Sergio, A., Wahyuningsih, S., & Siringoringo, M. (2023). Peramalan inflasi kota balikpapan menggunakan metode singular spectrum analysis. *Jurnal EKSPONENSIAL*, *14*(1), 21–30.

Siringoringo, M., Wahyuningsih, S., Purnamasari, I., & Arumsari, M. (2022). Peramalan jumlah produksi kelapa sawit kalimantan timur menggunakan metode singular spectrum analysis. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(3), 162–172. https://doi.org/10.35580/variansiunm46

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.

Utami, N. A. G., Sulandari, W., & Handajani, S. S. (2021). Peramalan curah hujan bulanan di pos hujan jatisrono dengan metode singular spectrum analysis (ssa). *Prosiding Seminar Nasional Aplikasi Sains & Teknologi*.

Wijayanti, L. N., & Kartikasari, M. D. (2023). Application of singular spectrum analysis method in forecasting indonesia composite data. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *17*(1), 0513–0526. https://doi.org/10.30598/barekengvol17iss1pp0513-0526