Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.5, October 2025, pp. 2443~2449
e-ISSN: 2548-6861

Forecasting the Number of Passengers for the Jakarta-Bandung High-Speed Rail using SARIMA and SSA Models

Indra Mahib Zuhair Riyanto[1], Elke Frida Rahmawati[2], Muhammad Fahrezi Maulana[3], Keyzha Mutiara Ahdiat[4], Achmad Raihan Nurdin[5], Laily Nissa Atul Mualifah[6*], and Adelia Putri Pangestika[7]
[1,2,3,4,5,6,7]Statistics and Data Science, IPB University
indramahibzuhair@apps.ipb.ac.id [1], 26fridaelke@apps.ipb.ac.id [2], fahrezimuhammad@apps.ipb.ac.id [3], keyzhamutiara@apps.ipb.ac.id [4], aburaihan@apps.ipb.ac.id [5], lailyatul@apps.ipb.ac.id [6], adelia_pangestika@apps.ipb.ac.id [7]

Article Info

Article history:

Received 2025-08-09
Revised 2025-08-31
Accepted 2025-09-20

Keyword:

High Speed Rail Jakarta-Bandung,
SARIMA,
Singular Spectrum Analysis.

ABSTRACT

Time series forecasting is essential for analyzing past data to predict future trends, supporting planning, and decision-making. The SARIMA model is widely used for seasonal data but may be less effective for highly fluctuating or non-stationary data, which can impact forecast accuracy. As an alternative, Singular Spectrum Analysis (SSA) offers a flexible approach, decomposing time series into trend, seasonal, and noise components without strict parametric assumptions, making it effective for complex data patterns. This study compares SARIMA and SSA models in forecasting daily passenger counts on the Jakarta-Bandung high-speed rail, using data from November 1, 2023, to September 30, 2024. The results show that the performance of SSA is more stable compared to SARIMA in the term of MAPE, where SSA provides lower MAPE then SARIMA in all three scenarios of data splits. These results are expected due to the non-linear pattern that appears in the data. Moreover, the predictions on both methods show that slight increment of passengers in the end of 2024 to the beginning of 2025. This finding suggests that the government needs to consider implementing interventions if they wish to change the current trend, such as offering discounts or year-end holiday promotions.

I. INTRODUCTION

Time series forecasting is one of the methods frequently used in analyzing past data to predict future events, particularly in the context of planning and decision-making [1]. One commonly applied model for analyzing seasonal data is the Seasonal Autoregressive Integrated Moving Average (SARIMA). This model is an extension of the Autoregressive Integrated Moving Average (ARIMA) model, designed to handle data with seasonal patterns [2]. However, SARIMA has limitations in addressing data with high fluctuations or seasonal patterns that doesn't occur on fixed seasonal periods, which can result in less optimal forecasting accuracy under specific conditions [3].

With advancements in forecasting methods, Singular Spectrum Analysis (SSA) has emerged as a flexible and effective technique for identifying trend components, seasonality, and random fluctuations in time series data without requiring strict parametric assumptions [4]. The non-parametric approach on modelling time series of SSA can be very useful, especially when domain knowledge expert is included. The ability to decompse noise and choose different part of the decompsition result will bring more understanding to the model.

Previous studies have demonstrated that SSA performs well in time series forecasting, particularly for data with complex patterns. SSA is unique in its ability to handle both stationary and non-stationary data by decomposing the time series into interpretable components through singular value decomposition and diagonal averaging. Furthermore, SSA does not require parametric assumption, making it a robust choice for various applications [5][6]. For instance, SSA has been applied in forecasting monthly mortality rates in North America [7] and predicting carbon price [8] with good result. According to the studies above, we expected that the SSA outperforms the SARIMA, due to its capability to catch non-seasonal pattern in data.

In the context of modern transportation, the advent of high-speed rail presents a significant innovation for mobility and the economy. In Indonesia, the Jakarta-Bandung High-Speed Rail (KJCB) project has become a topic of public interest [9], especially regarding its financial projections and the long-term benefits it could offer to society and the economy [10]. The Jakarta-Bandung high-speed rail project aims to shorten travel time between Jakarta and Bandung, two cities known for their dynamic and ever-increasing human mobility each year [11]. Given the high investment costs, the success of this project depends heavily on the accuracy of future passenger projections. A time series model capable of producing low forecasting errors is crucial for predicting future passenger numbers on KJCB, as accurate demand estimates enable stakeholders to make informed decisions regarding capacity planning, scheduling, and financial strategies. Minimizing prediction errors reduces the risk of underestimating or overestimating demand, which is essential for ensuring operational efficiency, optimizing resource allocation, and maintaining financial sustainability of the service.

Research has been conducted on the total number of train passengers in Java, Jabodetabek (Jakarta, Bogor, Depok, Tangerang, and Bekasi) and Sumatra using time series regression with integrated calendar variation, yielding a Root Mean Square Error (RMSE) of 2453.827, 7657.821, and 275.901 respectively on the test data [12]. Similarly, a study on forecasting the number of executive train passengers in Java using the SARIMA method produced favourable results, with a Mean Squared Error (MSE) of 119.17 on the training data [13]. However, to date, no studies have addressed passenger forecasting for the KJCB project.

Therefore, this study aims to compare the performance of the SARIMA and SSA models in forecasting daily passenger numbers for the Jakarta-Bandung High-Speed Rail. The choice of SSA and SARIMA reflects their complementary paradigms—non-parametric versus parametric—offering robust and interpretable insights into seasonal and trend dynamics. While more advanced approaches such as Prophet or LSTM exist, employing these well-established models ensures methodological transparency and establishes a reliable baseline for analyzing emerging datasets. The findings of this study will provide novel evidence on the comparative performance of SSA and SARIMA in the Indonesian high-speed rail context. Additionally, for practice purposes, the results of this study will provide recommendations on the most suitable forecasting model to support the development and operational strategies of the high-speed rail, thereby enhancing efficiency and effectively responding to changes in passenger patterns.

II. METODHS

The data used in this study consists of secondary data in the form of daily passenger numbers at the Halim station of the Jakarta-Bandung High-Speed Rail (KJCB). The data was obtained from the Sistem Informasi Angkutan dan Sarana Transportasi Indonesia (SIASATI) website (https://dashboard-siasati.dephub.go.id/).

The KJCB first operate on October 2, 2023, and a lot of promos given during the first month of the opening that can alter the normal condition. Therefore, the period of analysis choosen spans from November 1, 2023, to September 30, 2024, with the hope to reflect a normal operating condition. There are 333 complete daily observation during the period of analysis, without any missing observation. Data analysis was conducted using R software with steps undertaken in this study are as follows:

A. Data Preprocessing

The data division scenario follows the expanding window cross-validation concept. This scenario aims to compare and validate the ability of the developed models in forecasting time series data with different patterns. The data division scenarios are summarized in Table 1.

TABLE I
DATA DIVISIONS SCENARIO

Scenario	Train-test split ratio	Train data period	Test data period
1	70:30	1 November 2023 – 24 June 2024	25 June 2024 – 30 September 2024
2	80:20	1 November 2023 – 29 July 2024	30 July 2024 – 30 September 2024
3	90:10	1 November 2023 – 1 September 2024	2 September 2024 – 30 September 2024

B. SARIMA Modelling

SARIMA is a time series method developed from the Box-Jenkins ARIMA model to accommodate seasonal patterns. Seasonal patterns refer to recurring trends over specific periods. The SARIMA model can be expressed as ARIMA(p,d,q)(P,D,Q)s where the orders (P,D,Q) represent the seasonal model orders linked to the non-seasonal orders (p,d,q). The SARIMA modelling process for each training dataset involves the following steps:

- Exploring the data through time series plots.
- Examining the stationary of the data by performing Augmented Dicky-Fuller (ADF) test and Box-Cox. The time series is stationary in mean if it failed rejected the null hypothesis in ADF test. The time series stationary in variance if the 95% confidence interval of Box-Cox transformation include the value 1 [14].
- Identifying tentative models using ACF plot and Partial Autocorrelation Function (PACF) plots. Seasonal component and non seasonal component can be identified by observing the potential cut-off and tails-off patterns in ACF dan PACF plots at the seasonal periods lag and each period lag [15].
- Estimating parameters for the identified tentative models.

- Selecting the best tentative model. The best model is chosen based on the hypotesis testing of the parameters, the value of Akaike Information Criterion (AIC), and also the residual diagnostic.
- Performing overfitting on the best tentative model. Overfitting involves increasing the ARMA orders incrementally to confirm that the selected model is indeed the best. The overfitted models are compared with the initially selected model based on smaller AIC values [16], significance of overfitting parameters, and adherence to residual diagnostic assumptions.
- Forecasting for the test data in all three scenarios. The finalized models are used to make forecasts for the test datasets across the three scenarios.

C. SSA Modelling

SSA is a non-parametric method for time series forecasting that combines elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems, and signal processing. The goal of SSA is to decompose the original time series into interpretable components such as trends, seasonal components, and noise. The SSA modelling process involves the following steps [17]:

- Determining the parameter L (window length). L is a critical parameter in the decomposition process. It must satisfy 1<L<N, where N is the total number of data length. For seasonal data, L is typically proportional to the seasonal period. The process of determining L is done by series of trial and error.
- Constructing trajectory matrix (X) of size L×K, where K = N - L +1. This matrix is formed by arranging overlapping segments of the time series into columns, providing a structured representation of the data for decomposition. An illustration of the trajectory matrix formation is shown in equation 3.

$$\boldsymbol{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{bmatrix} \tag{3}$$

- Decomposing trajectory matrix ($\boldsymbol{X}$) using *Singular Value Decomposition* (SVD) as expressed in equation 4. The value of $\lambda_1 > \lambda_2 > \cdots > \lambda_d > 0$ are the eigenvalue of $\boldsymbol{XX}'$ matrix, U_i are the left eigenvectors, and V_i is the right eigenvectors. The triplet $\sqrt{\lambda_i}$, U_i, dan V_i are also referred as *eigentriple.*
- Combining multiple eigentriple based on certain patterns of the data that has been decomposed.
- Transforming each set into a new data through diagonal averaging.
- Forecasting for the test data to evaluate the model. The forecast methods that implemented in this reasearch is using Linear Recurrent Formula (LRF).

D. Model Evaluation

To evaluate the performance of each model, we use Mean Absolute Percentage Error (MAPE) metrix, which compares how far the prediction of the model built with the training data in predicting the test data. The value of MAPE is given in equation 6.

$$MAPE = \sum_{t=1}^{n} \frac{|y_t - \widehat{y_t}|}{y_t} \times 100\% \tag{7}$$

where y_t is the actual data and $\widehat{y_t}$ is the forecast value. In this study, the Mean Absolute Percentage Error (MAPE) was selected as the primary evaluation metric in preference to alternatives such as MAE, MSE, or RMSE. The rationale for this choice lies in MAPE's superior interpretability, as it expresses error in relative percentage terms, thereby offering a scale-independent measure of accuracy. This universality makes MAPE particularly advantageous for comparing model performance across datasets with different magnitudes, where the interpretability of scale-dependent metrics can be limited.

E. Prediction

The best model for each method is picked and use to predict the next 120 days the gain and insight and provide suggestion.

III. RESULTS

The initial step in the analysis involves splitting the data into two parts: training data and test data. The split is conducted based on the three aforementioned scenarios. This division is based on the differing patterns of data changes in each of the three scenarios. The difference in data patterns between the training and test sets significantly influences the modelling process. The results from these three scenarios provide a comprehensive view of the model's ability to handle variations in data patterns. The data split scenarios are presented in Table 1, and the corresponding plots are shown in Figure 1.

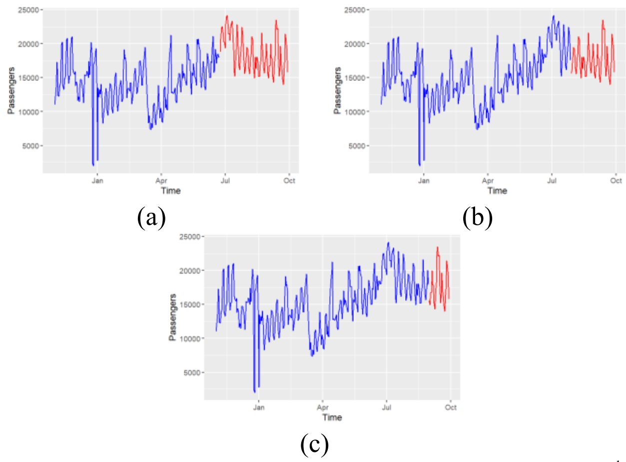


Figure 1 Time series plot with train test split for (a) 1st scenario, (b) 2nd scenario, and (c) 3rd scenario. Blue is for training set and red is for testing set.

Data exploration is conducted for each training dataset across the three scenarios by creating time series plots. These plots provide insights into trends and patterns, especially highlighting non-linear trends in data with seasonal patterns. A significant decline is observed on December 25 and 26, 2023, followed by a rise starting in mid-March 2024, which is attributed to the long religious holiday in Indonesia. The time series plots for the three scenarios illustrate different characteristics of the data split. This variation is intended to assess how well the models perform under different time series conditions.

The SARIMA modelling process begins with exploring the ACF plot. In all three scenarios, the ACF plots reveal a strong seasonal pattern, indicated by significant lag values at multiples of 7. This suggests a weekly seasonal period, providing strong indicator that the model to be developed is a SARIMA model with a weekly seasonal period. This observation helps guide the identification of the appropriate seasonal components (p, d, q) and (P, D, Q) for the SARIMA model.

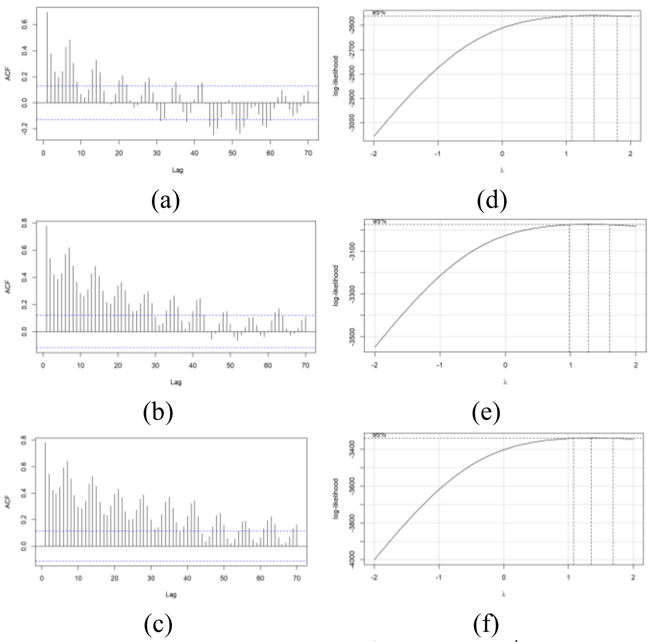


Figure 2. ACF plot for train data in (a) 1st scenario, (b) 2nd scenario, and (c) 3rd scenario and Box-Cox plot for train data in (d) 1st scenario, (e) 2nd scenario, and (f) 3rd scenario.

The ACF plot of the data shows a slow decline in autocorrelation, indicating that the data tends to be non-stationary in mean. This is further supported by the ADF test results, which show p-values greater than 0.05 (0.368, 0.405, and 0.332) for all three training data scenarios, suggesting the data is non-stationary in mean. The Box-Cox plot shows that the 95% confidence interval includes the value of one for scenario 2, while it does not include the value of one for 1st and 3rd scenarios. This indicates the need for data transformation for training data in 1st and 3rd scenarios, and differencing must be applied to both the non-seasonal and seasonal components in the entire test data. As a result, all the training data becomes stationary in both variance and mean.

Tentative model identification is carried out using ACF, PACF, and EACF plots for the non-seasonal components, and ACF and PACF plots for the seasonal components. Several combinations of the plots lead to different tentative models. Each model is estimated for its parameters, and diagnostic tests on residuals are performed. The best tentative model is the one where all the parameters are significant, the AIC value is the smallest compared to other tentative models, and the residual diagnostics are satisfied, except for the assumption of normally distributed residuals. The best tentative models obtained from the three scenarios are summarized in Table 2.

TABLE 2
SUMMARY ON BEST TENTATIVE MODELS

Scenario	Model	Parameter Significant	AIC
1	ARIMA(1,1,1) $(0,1,1)_7$	All	6037.968
2	ARIMA(1,1,1) $(0,1,1)_7$	All	4789.202
3	ARIMA(1,1,1) $(0,1,1)_7$	All	7336.472

The selected best tentative model has the same order across all three scenarios and shows significant estimates across all parameters. The model is then tested to overfitting to ensure that the obtained model is indeed the best. Each model generated from the overfitting process is evaluated based on the significance of the overfitting parameters, the AIC value compared to the initial model, and the residual diagnostics of the model. The results of the overfitting process are summarized in Table 3.

TABLE 3
MODEL SELECTED AFTER OVERFITTING

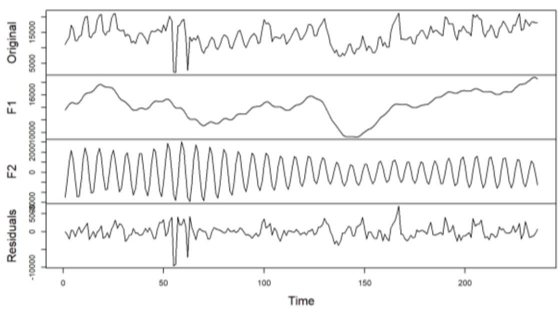
Scenario	Model	Parameter Significant	AIC
1	ARIMA(1,1,1) $(1,1,1)_7$	All	6035.225
2	ARIMA(1,1,1) $(1,1,1)_7$	All	4785.046
3	ARIMA(1,1,1) $(1,1,1)_7$	All	7332.144

The final selected model is the model with and increase AR order of seasonal component, which shows a decrease in the AIC value while maintaining the significance of each estimated parameter. The selected best model is then subjected to residual diagnostics, including normality test of residuals, autocorrelation test of residuals, and homogeneity of variance test of residuals. These tests are necessary because the estimation method used is parametric, which requires certain assumptions for the estimates to be valid. All three models satisfy the required assumptions, except for the normality of residuals. This assumption violation can be disregarded due to the relatively large amount of data. A summary of the residual diagnostics for the selected models is presented in Table 4.

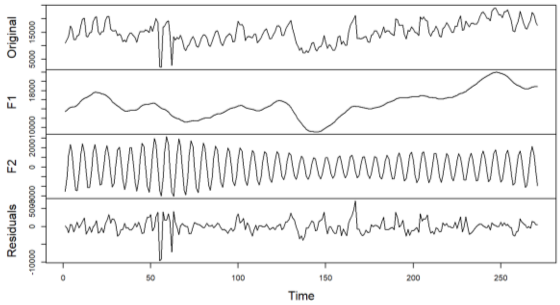
TABLE 4
SUMMARY ON RESIDUAL DIAGNOSTIC TEST

Scenario	Diagnostic test	*p-value*	Description
1	Jarque-Bera Test	0.000	Residuals are normally distributed
	Box-Ljung Test on Residuals	0.376	Residuals are independent
	Box-Ljung Test on Squared Residuals	0.184	Residuals variance is homogenous
2	Jarque-Bera Test	0.000	Residuals are normally distributed
	Box-Ljung Test on Residuals	0.437	Residuals are independent
	Box-Ljung Test on Squared Residuals	0.069	Residuals variance is homogenous
3	Jarque-Bera Test	0.000	Residuals are normally distributed
	Box-Ljung Test on Residuals	0.401	Residuals are independent
	Box-Ljung Test on Squared Residuals	0.147	Residuals variance is homogenous

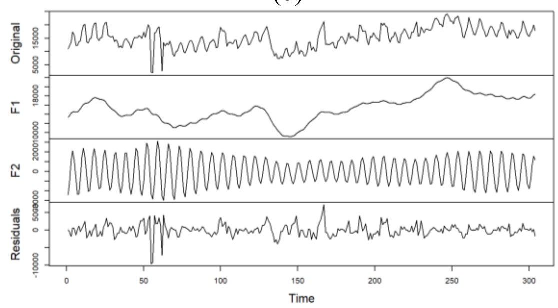
The next step is to model Singular Spectrum Analysis (SSA). The SSA modelling begins with determining the value of L which will be used for signal decomposition. The value of L in time series data with seasonal patterns should ideally be proportional to the seasonal aspect of the time series [18]. In this study, L=35 was found to perform well in separating the time series data into seasonal and trend components. In all three scenarios, eigentriples 1, 2, and 5 represent components that explain the trend in the data, eigentriples 3 and 4 explain the seasonal components, while the remaining eigentriples are considered noise and were not used in the modelling process. The results of the time series reconstruction from the SSA modelling are presented in Figure 2.



(a)



(b)



(c)

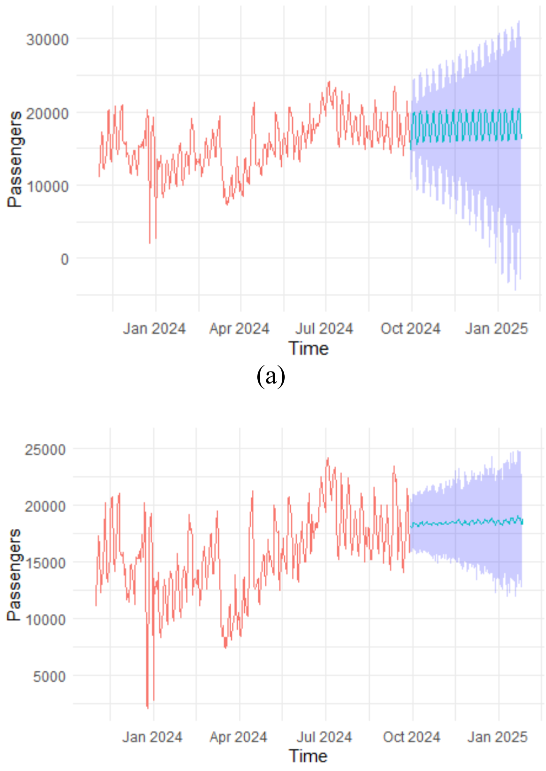
Figure 2. SSA decomposition and reconstruction result for train data in (a) 1st scenario, (b) 2nd scenario, and (c) 3rd scenario.

The results from the modelling between the SARIMA and SSA methods show almost identical outcomes, with SSA slightly outperforming SARIMA in all scenarios except the 3rd scenario. The largest difference in MAPE values occurred in 2nd scenario, where the SSA model showed a better performance, with a 10.03% MAPE difference. The division of data into training and test sets can influence the goodness of fit of the model, as certain models can only accommodate specific patterns. In 2nd scenario, the data was split at a point where the trend of the data underwent a change. The SARIMA model is only capable of accommodating linear trends, meaning that if the previous pattern was increasing, the predicted result would also increase. On the other hand, the SSA model has the ability to forecast nonlinear trend patterns in the data, allowing it to consider the shift in trend, as seen in 2nd scenario. A summary of the SARIMA and SSA modelling and forecasting results is provided in Table 5.

TABLE 5
SUMMARY OF SARIMA AND SSA MODELLING AND FORECASTING PERFORMANCE ON TEST DATA

Data	Scenario	MAPE of SARIMA	MAPE of SSA
Train	1st Scenario	13.81 %	13.61%
	2nd Scenario	12.37 %	12.35%
	3rd Scenario	11.59%	11.40%
Test	1st Scenario	13.33%	12.79%
	2nd Scenario	22.05 %	12.02%
	3rd Scenario	10.11%	11.36%

From Table 5, we can see that 3rd scenario is the best scenario in this study, as it has the best performance based on the MAPE values compared to the other scenarios. The model from this scenario will be used on the entire dataset and then employed to forecast the number of passengers for the KCJB for the next 120 days. This forecast aims to provide an outlook on the number of passengers beyond the observation period and to understand the emerging trends.



(a)

(b)

Figure 3. Forecasting results for the next 120 days using (a) the SARIMA model and (b) the SSA model

The 120-period forecast indicates a seasonal trend with a very gradual increase with pronounced seasonal pattern. The SARIMA model's predictions also show significant instability for longer forecast periods, as evident by the widening confidence intervals as the forecast horizon extends. Similarly, the SSA model's forecast shows the same trend, but the seasonal pattern is less pronounced compared to the SARIMA forecast. The confidence intervals for the SSA model's forecast also widen towards the end of the forecast period, but not as dramatically as with SARIMA.

The forecast results suggest that the KCJB passenger numbers at Halim Station is not showing high increment, rather slow and slightly. This means that the government faced a serious challenged due to the investment return time that will be slow. The government needs to consider implementing interventions if they wish to change the current trend. Business strategies such as offering discounts or year-end holiday promotions can be considered to increase the number of KCJB passengers and gain better momentum in meeting investment target returns.

IV. Conclusion

The SSA method can serve as an alternative method for forecasting time series with nonlinear trends. Its non-parametric approach provides flexibility, allowing researchers to forecast without being constrained by parametric assumptions. In this study, the SSA method demonstrates forecasting capabilities comparable to SARIMA. This is evident in the case of predicting the number of passengers at the KCJB Halim station, where the SSA model outperforms SARIMA. The 120-day forecast shows a clear weekly pattern in the SARIMA predictions, while the SSA model exhibits a more random pattern. The predictions on both methods alert the government and all stakeholders of the KCJB project that investment return will be slow and there should be an intervention to prevent the snowball effect in terms of financial aspect.

This study acknowledges the limitation of not incorporating external factors such as national holidays, promotional campaigns, or other exogenous variables, which may significantly influence passenger demand. Future research should therefore integrate these elements to enhance forecasting accuracy and reliability, while also exploring advanced approaches such as Prophet, LSTM, or hybrid frameworks for broader comparative insights. Although SSA proves valuable for uncovering underlying patterns through data reconstruction, its reliance on trial-and-error in selecting the window length (L) can be time-consuming; hence, future studies are encouraged to adopt systematic strategies, such as cross-validation, to improve both efficiency and predictive performance.

References

[1] E. Choiriyah, U. D. Syafitri, and I. M. Sumertajaya, “Pengembangan Model Peramalan Space-Time* Studi Kasus: Data Produksi Padi di Sulawesi Selatan,” *Indones. J. Stat. Its Appl.*, vol. 4, no. 4, pp. 579–589, 2020.

[2] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forescasting and Control*, 5th ed. Hoboken: John Wiley & Sons, Inc, 2015.

[3] C. Tang, Z. Tong, Y. Wei, X. Wu, X. Tian, and J. Yang, “Time–Frequency Characteristics and SARIMA Forecasting of Atmospheric Water Vapor in East Asia,” *Atmosphere (Basel).*, vol. 14, no. 5, p. 899, May 2023, doi: 10.3390/atmos14050899.

[4] E. Zukhronah, W. Sulandari, and E. P. S. P. Ilahi, “Hybrid Model of Singular Spectrum Analysis with Autoregressive Integrated Moving Average and Fuzzy Time Series for Indonesian Crude Price Forecasting,” *Barekeng J. Ilmu Mat. dan Terap.*, vol. 18, no. 3, pp. 1519–1526, Jul. 2024, doi: 10.30598/barekengvol18iss3pp1519-1526.

[5] N. Golyandina and A. Korobeynikov, “Basic Singular Spectrum Analysis and Forecasting with R,” *Comput. Stat. Data Anal.*, Mar. 2014, doi: 10.1016/j.csda.2013.04.009.

[6] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 3rd ed. in Springer Texts in Statistics. Springer International Publishing, 2016.

[7] H. Hassani, "Singular Spectrum Analysis: Methodology and Comparison," *J. Data Sci.*, vol. 5, no. 2, pp. 239–257, Jul. 2021, doi: 10.6339/JDS.2007.05(2).396.

[8] C. Zhang and B. Lin, "Carbon Prices Forecasting Based on the Singular Spectrum Analysis, Feature Selection, and Deep Learning: Toward a Unified View," *Process Saf. Environ. Prot.*, vol. 177, pp. 932–946, Sep. 2023, doi: 10.1016/j.psep.2023.07.015.

[9] H. Partawijaya, "Sentiment Analysis of the Public Towards the Jakarta-Bandung High-Speed Train Service Using the Multiclass Support Vector Machine Method," Institut Pertanian Bogor, 2024.

[10] O. T. Giantara, A. Purba, and D. Herianto, "Analisis Ekonomi dan Finansial Kereta Cepat Jakarta - Bandung," *JRSDD*, vol. 6, no. 4, pp. 323–334, 2018.

[11] T. S. Hamzah and T. Kurniawan, "Analisis Manfaat dan Biaya: Kebijakan Transportasi Kereta Cepat Jakarta Bandung dari Kacamata Proyek Strategis Nasional," *J. Darma Agung*, vol. 32, no. 3, pp. 101–105, 2024, doi: https://dx.doi.org.10.46930/ojsuda.v32i3.4288.

[12] M. Silfiani and F. N. Hayati, "Forecasting Number of Train Passengers Using Time Series Regression Integrated Calendar Variation and Covid 19," *J. Mat. Sains, dan Teknol.*, vol. 25, no. 1, pp. 9–22, 2024.

[13] L. L. Ibrahim and E. Kurniati, "Peramalan Jumlah Penumpang Kereta Api Eksekutif di Pulau Jawa Menggunakan Model SARIMA," *J. Ris. Mat.*, vol. 3, no. 1, pp. 73–82, Jul. 2023, doi: 10.29313/jrm.v3i1.1747.

[14] L. Martínez-Acosta, J. P. Medrano-Barboza, Á. López-Ramos, J. F. R. López, and Á. A. López-Lambraño, "SARIMA Approach to Generating Synthetic Monthly Rainfall in the Sinú River Watershed in Colombia," *Atmosphere (Basel).*, vol. 11, no. 6, pp. 1–16, 2020, doi: 10.3390/atmos11060602.

[15] Z. Zeinawaqi and A. A. Dzikrullah, "Factors Affecting The Number Of Domestic Flights In Indonesia During Covid-19 Pandemic Using SARIMAX Method," *J. Mat. Stat. dan Komputasi*, vol. 21, no. 1, pp. 1–9, 2024, doi: 10.20956/j.v21i1.34557.

[16] P. Stoica and Y. Selen, "Model-order selection," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004, doi: 10.1109/MSP.2004.1311138.

[17] N. Golyandina and A. Zhigljavsky, "Basic SSA," in *Singular Spectrum Analysis for Time Series*, Berlin, Heidelberg, 2013, ch. 2, pp. 11–70. doi: 10.1007/978-3-642-34913-3.

[18] W. Sulandari, Subanar, M. H. Lee, and P. C. Rodrigues, "Indonesian Electricity Load Forecasting Using Singular Spectrum Analysis, Fuzzy Systems and Neural Networks," *Energy*, Jan. 2020, doi: 10.1016/j.energy.2019.116408.