

PERAMALAN PADA RUNTUN WAKTU DENGAN POLA TREND MENGUNAKAN SSA-LRF

Diah Safitri^{1,2}, Gunardi¹, Nanang Susyanto¹, Winita Sulandari³

¹Departemen Matematika, Universitas Gadjah Mada, Yogyakarta, Indonesia

²Departemen Statistika, Universitas Diponegoro, Semarang, Indonesia

³Program Studi Statistika, Universitas Sebelas Maret, Surakarta, Indonesia

*e-mail : diahsafitri@mail.ugm.ac.id

DOI: 10.14710/j.gauss.12.2.296-303

Article Info:

Received: 2023-05-02

Accepted: 2023-07-19

Available Online: 2023-07-28

Keywords:

SSA; SSA-LRF; Trend; Time Series

Abstract: *Singular Spectrum Analysis-Linear Recurrent Formulae* (SSA-LRF) is a forecasting method that starts by decomposing time series data into several independent and interpretable components. SSA-LRF does not have any assumptions that must be fulfilled thus it is more flexible to use. In this research, an empirical study of time series forecasting that has a trend data pattern will be carried out using SSA-LRF without difference transformation and with difference transformation. A difference transformation is performed because the data has a trend pattern. Although there are no assumptions that must be met in forecasting using SSA-LRF, it is expected that difference transformation will produce better forecasting accuracy than without difference transformation process. There are three data used in this research. The first is data from Wei's book (2006), this data is called series W8 and is a simulation data. The second data is the number of railway passengers in the Java region. The third data is Mauna Loa atmospheric CO₂ concentration data obtained from R software. Forecasting using SSA-LRF without difference transformation and with difference transformation on all three data resulted in accurate forecasting values, and difference transformation improved the accuracy values.

1. PENDAHULUAN

Pola dalam runtun waktu seringkali dapat didekomposisi menjadi sub pola, dimana dekomposisi tersebut dapat membantu meningkatkan akurasi peramalan. Hassani (2007) mengemukakan bahwa SSA merupakan metode dalam analisis runtun waktu yang mempunyai tujuan membuat dekomposisi dari runtun yang asli menjadi komponen yang independen dan dapat diinterpretasi seperti komponen tren, osilasi, dan *noise*. Golyandina dkk. (2001) mengembangkan *Singular Spectrum Analysis-Linear Recurrent Formulae* (SSA-LRF). SSA-LRF merupakan metode peramalan yang diawali dengan membuat dekomposisi dari runtun waktu menjadi beberapa komponen yang independen dan dapat diinterpretasi. SSA-LRF lebih fleksibel karena tidak terdapat asumsi yang harus dipenuhi.

Beberapa peneliti telah menulis paper yang berkaitan dengan SSA dan runtun waktu, diantaranya adalah Hassani dkk.(2009) melakukan peramalan produksi bulanan industri Eropa khususnya Jerman, Prancis, dan UK, menggunakan SSA-LRF, Holt-Winters dan model ARIMA. Hasil peramalan menggunakan tiga metode tersebut tersebut dibandingkan menggunakan nilai akurasi *Root Mean Square Error* (RMSE) dan diperoleh hasil bahwa ketiga metode tersebut mempunyai performa yang sama untuk peramalan jangka pendek dan SSA-LRF memiliki akurasi yang lebih baik dibandingkan dengan Holt-Winters dan ARIMA untuk peramalan jangka panjang, Hassani dkk.(2010) menggunakan SSA untuk meramalkan nilai tukar, Mahmoudvand dkk. (2019) membandingkan SSA dan *multivariate* SSA (MSSA) dalam meramalkan nilai tukar harian.

SSA menguraikan (dekomposisi) runtun waktu ke beberapa komponen yang dapat diinterpretasi berdasarkan *Singular Value Decomposition* (SVD) dari matriks yang dibangun oleh runtun waktu. Asumsi yang harus dipenuhi pada model parametrik tidak diasumsikan pada metode SSA, hal ini membuat SSA adalah suatu metode yang bebas model dan memungkinkan mempunyai jangkauan luas dalam penerapannya. Tujuan dari penelitian ini adalah studi empiris untuk menunjukkan bahwa peramalan runtun waktu berpola trend menggunakan SSA-LRF menghasilkan nilai akurasi peramalan yang sangat baik dan dengan digunakan tranformasi diferens dapat meningkatkan akurasi peramalan. Pada penelitian ini menggunakan *Mean Absolute Percentage Error* (MAPE) sebagai nilai akurasi peramalan. Terdapat tiga data yang digunakan dalam penelitian ini, yang pertama adalah data dari buku Wei (2006) yang berjudul “*Time Series Analysis Univariate and Multivariate Methods*”, data yang kedua adalah jumlah penumpang kereta api wilayah Jawa (Jabodetabek dan Non Jabodetabek) yang diperoleh dari <https://www.bps.go.id>, data ketiga adalah data konsentrasi CO2 atmosfer Mauna Loa yang diperoleh dari *software R*.

2. TINJAUAN PUSTAKA

Sebagai metode non parametrik dan bebas model, SSA dapat diterapkan untuk sebarang runtun. Keberhasilan dalam SSA bergantung pada pemisahan yang tepat dari komponen-komponen runtun waktu. Algoritma SSA terdapat empat tahap yaitu *embedding*, *Singular Value Decomposition* (SVD), pengelompokan dan rekonstruksi. Algoritma dijelaskan di dalam Elsner dan Tsonis (1996), Golyandina dkk. (2001), Golyandina dan Zhigljavsky (2013), dan Golyandina dkk. (2018).

Misalkan runtun waktu bernilai real $\mathbf{X}_N = (x_1, \dots, x_N)$ dengan panjang N dan misal L *window length* dengan $1 < L < N - 1$. $K = N - L + 1$ merupakan banyaknya kolom dari matriks \mathbf{X} . Pada tahap *embedding*, runtun waktu dipetakan ke suatu matriks trajektori \mathbf{X} .

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}$$

Tahap kedua adalah *Singular Value Decomposition* (SVD). Hasil dari tahap ini adalah SVD dari matriks trajektori \mathbf{X} , yaitu:

$$\begin{aligned} \mathbf{X} &= \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d \quad (2) \\ &= \sqrt{\lambda_1} U_1 V_1^T + \sqrt{\lambda_2} U_2 V_2^T + \dots + \sqrt{\lambda_d} U_d V_d^T \end{aligned}$$

dengan $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ dan $\lambda_1, \lambda_2, \dots, \lambda_d$ nilai *eigen* positif dari \mathbf{S} serta U_1, \dots, U_d sistem orthonormal dari vektor *eigen* matriks \mathbf{S} yang sesuai dengan nilai *eigen*. $(\sqrt{\lambda_i}, U_i, V_i)$ disebut *eigen triple* ke- i .

Tahap ketiga adalah pengelompokan, dimana pada tahap ini mempartisi himpunan indeks $\{i = 1, 2, \dots, d\}$ menjadi m himpunan yang saling asing sehingga Persamaan (2) menjadi

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m} \quad (3)$$

Tahap terakhir adalah rekonstruksi. Pada tahap ini, dilakukan transformasi setiap matriks X_{I_j} pada Persamaan (3) menjadi runtun waktu berukuran N . Misalkan Y suatu matriks $L \times K$ dengan elemen y_{ij} , $1 \leq i \leq L, 1 \leq j \leq K$. $L^* = \min(L, K)$, $K^* = \max(L, K)$ dan $N = L + K - 1$. Misal $y_{ij}^* = y_{ij}$ jika $L < K$ and $y_{ij}^* = y_{ji}$ dalam keadaan selain itu. Matrix Y ditransfer ke runtun y_1, \dots, y_N menggunakan formula

$$y_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k y_{m,k-m+1}^* & \text{untuk } 1 \leq k < L^* \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+1}^* & \text{untuk } L^* \leq k \leq K^* \\ \frac{1}{N-k+1} \sum_{m=k-K^*+1}^{N-K^*+1} y_{m,k-m+1}^* & \text{untuk } K^* < k \leq N \end{cases} \quad (4)$$

Linear Recurrent Formulae (LRF) terdapat pada Golyandina dkk. (2001), Golyandina dan Zhigljavsky (2013), dan Golyandina dan Korobeynikov (2014). Misalkan I merupakan *eigen triple* yang dipilih, $P_i \in R^L$, dan $i \in I$ \underline{P}_i merupakan koordinat $L-1$ pertama P_i , π_i merupakan koordinat terakhir P_i , $v^2 = \sum_{i \in I} \pi_i^2$
 $R = (a_{L-1}, \dots, a_1)^T = \frac{1}{1-v^2} \sum_{i \in I} \pi_i \underline{P}_i$. Runtun waktu $Y_{N+M} = (y_1, \dots, y_{N+M})$ yang merupakan hasil peramalan menggunakan SSA-LRF didefinisikan oleh

$$y_i = \begin{cases} \bar{x}_i & \text{untuk } i=1, \dots, N \\ \sum_{j=1}^{L-1} a_j y_{i-j} & \text{untuk } i = N+1, \dots, N+M \end{cases}$$

Pada penelitian ini menggunakan *Mean Absolute Percentage Error* (MAPE) sebagai nilai akurasi peramalan. Jika Y_t merupakan data aktual untuk periode t dan F_t adalah ramalan untuk periode yang sama, maka $e_t = Y_t - F_t$ dan $PE_t = \left(\frac{Y_t - F_t}{Y_t} \right) (100)$, sehingga diperoleh

$$MAPE = \frac{1}{n} \sum_{t=1}^n |PE_t|$$

3. METODE PENELITIAN

Pada penelitian ini digunakan tiga data, yang pertama adalah data dari buku Wei (2006) yang berjudul "*Time Series Analysis Univariate and Multivariate Methods*". Data dari buku Wei (2006) ini dinamakan runtun W8, mempunyai 150 pengamatan dan merupakan data simulasi, kemudian data yang kedua adalah jumlah penumpang kereta api wilayah Jawa (Jabodetabek dan Non Jabodetabek) yang merupakan data bulanan dengan 72 pengamatan dari bulan Januari 2014 sampai dengan Desember 2019 yang diperoleh dari <https://www.bps.go.id>, data ketiga adalah data konsentrasi CO2 atmosfer Mauna Loa yang diperoleh dari *software* R yang merupakan data bulanan dengan 468 pengamatan dari tahun 1959 sampai dengan 1997.

Metode analisis data yang digunakan dalam penelitian ini adalah

Langkah 1 : Menentukan data yang digunakan dalam penelitian ini.

Langkah 2 : Peramalan menggunakan metode SSA-LRF tanpa transformasi data.

Langkah 3: Menghitung MAPE pada hasil peramalan tanpa transformasi data.

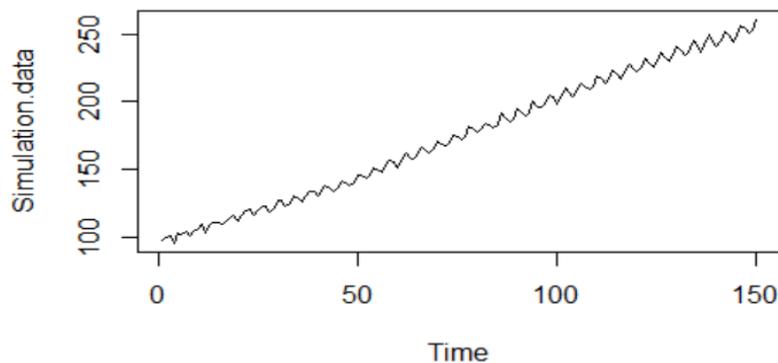
Langkah 4: Peramalan menggunakan metode SSA-LRF dengan transformasi diferens.

Langkah 5: Menghitung MAPE pada hasil peramalan dengan transformasi diferens.

Langkah 6: Melakukan analisa

4. HASIL DAN PEMBAHASAN

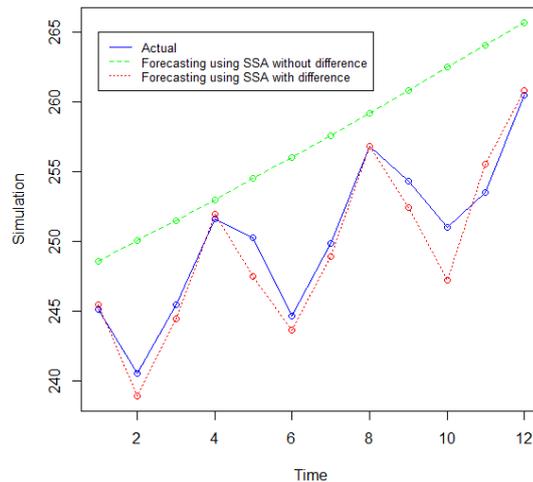
Simulasi diperlukan untuk menyelidiki apakah SSA-LRF tanpa transformasi diferens dan dengan transformasi diferens menghasilkan peramalan dengan nilai akurasi yang bagus, dan apakah tranformasi diferens dapat meningkatkan akurasi peramalan pada data runtun waktu dengan pola trend. Penelitian ini menggunakan runtun W8 dari Wei (2006) yang merupakan data simulasi dengan 150 pengamatan, dengan 138 pengamatan sebagai data training dan 12 pengamatan sebagai data testing. Plot runtun waktu dari data simulasi dapat dilihat pada Gambar 1, yang mana merupakan data runtun waktu dengan pola trend.



Gambar 1. Plot runtun waktu pada data W8 (data simulasi).

Pada data simulasi tersebut, dilakukan peramalan menggunakan SSA-LRF tanpa transformasi diferens dan dengan transformasi diferens, dan keduanya menghasilkan nilai MAPE sebagai nilai akurasi peramalan yang sangat baik yaitu nilai MAPE tanpa transformasi diferens sebesar 2,67% dan dengan transformasi diferens sebesar 0,54%. Dengan menurunnya nilai MAPE dapat diartikan bahwa tranformasi diferens pada data simulasi ini dapat meningkatkan akurasi peramalan. Grafik peramalan 12 langkah ke depan dapat dilihat pada Gambar 2, dimana peramalan menggunakan SSA-LRF dengan transformasi diferens dekat dengan data aktual.

Selanjutnya, pada penelitian ini, peramalan menggunakan SSA-LRF diterapkan pada data dalam kehidupan nyata (bukan data simulasi), yaitu jumlah penumpang kereta api wilayah Jawa (Jabodetabek dan Non Jabodetabek) dan data konsentrasi CO2 atmosfer Mauna Loa.

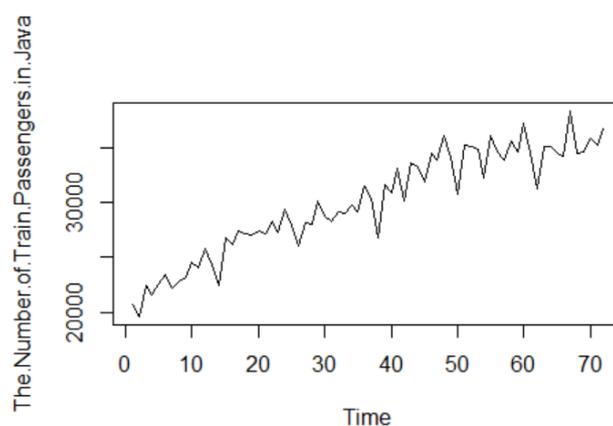


Gambar 2. Grafik peramalan 12 langkah ke depan dari data simulasi.

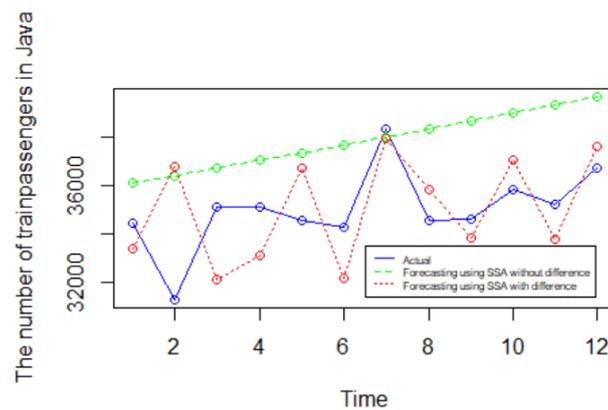
4.1. Penerapan pada data jumlah penumpang kereta api wilayah Jawa (Jabodetabek dan Non Jabodetabek)

Data jumlah penumpang kereta api wilayah Jawa (Jabodetabek dan Non Jabodetabek) merupakan data bulanan dengan 72 pengamatan dari bulan Januari 2014 sampai dengan Desember 2019 diperoleh dari <https://www.bps.go.id>. Dari 72 pengamatan tersebut, 60 pengamatan menjadi data training dan 12 pengamatan menjadi data testing. Plot runtun waktu dapat dilihat pada Gambar 3, dimana dapat dilihat bahwa data tersebut merupakan data runtun waktu dengan pola trend.

Peramalan pada data jumlah penumpang kereta api wilayah Jawa (Jabodetabek dan Non Jabodetabek) menggunakan SSA-LRF tanpa transformasi diferens dan dengan transformasi diferens berturut-turut menghasilkan nilai MAPE sebesar 8,43% dan 5,33%, dapat diartikan keduanya mempunyai nilai akurasi peramalan yang sangat baik dan tranformasi diferens pada data jumlah penumpang kereta api wilayah Jawa ini dapat meningkatkan akurasi peramalan. Peramalan menggunakan SSA-LRF dengan transformasi diferens lebih dekat dengan data aktual, hal ini dapat dilihat pada grafik peramalan 12 langkah ke depan (Gambar 4).



Gambar 3. Plot runtun waktu data jumlah penumpang kereta api wilayah Jawa

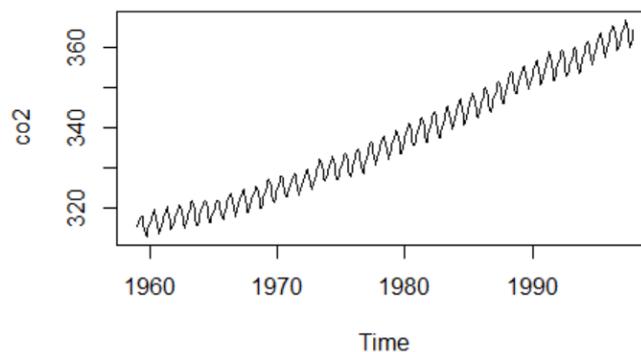


Gambar 4. Grafik peramalan 12 langkah ke depan dari data jumlah penumpang kereta api wilayah Jawa

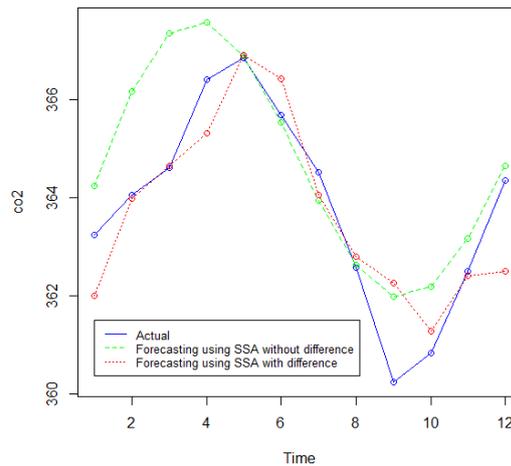
4.2. Penerapan pada data konsentrasi CO2 atmosfer Mauna Loa

Data konsentrasi CO2 atmosfer Mauna Loa diperoleh dari *software* R dan merupakan data bulanan dengan 468 pengamatan dari tahun 1959 sampai dengan 1997, yang mana dari plot runtun waktu (Gambar 5) dapat dilihat bahwa data tersebut merupakan runtun waktu berpola trend. Dari 468 pengamatan, 456 pengamatan sebagai data training dan 12 pengamatan sebagai data testing.

Pada data konsentrasi CO2 atmosfer Mauna Loa, dilakukan peramalan menggunakan SSA-LRF tanpa transformasi diferens dan dengan transformasi diferens, dan keduanya menghasilkan nilai MAPE sebagai nilai akurasi peramalan yang sangat baik yaitu nilai MAPE tanpa transformasi diferens sebesar 0,27% dan dengan transformasi diferens sebesar 0,15%. Dengan menurunnya nilai MAPE dapat diartikan bahwa tranformasi diferens pada data konsentrasi CO2 atmosfer Mauna Loa ini meningkatkan akurasi peramalan. Grafik peramalan 12 langkah ke depan dapat dilihat pada Gambar 6, terlihat bahwa peramalan menggunakan SSA-LRF dengan transformasi diferens dekat dengan data aktual.



Gambar 5. Plot runtun waktu konsentrasi CO2 atmosfer Mauna Loa



Gambar 6. Grafik peramalan 12 langkah ke depan dari data konsentrasi CO2 atmosfer Mauna Loa

5. KESIMPULAN

SSA-LRF tidak mempunyai asumsi yang harus dipenuhi sehingga lebih fleksibel dalam penggunaannya. Asumsi yang harus dipenuhi pada model parametrik tidak diasumsikan pada metode SSA-LRF, sehingga SSA-LRF adalah suatu metode yang bebas model dan memungkinkan mempunyai jangkauan luas dalam penerapannya.

Peramalan menggunakan SSA-LRF tanpa transformasi diferens dan dengan transformasi diferens pada studi empiris ini menghasilkan peramalan dengan nilai akurasi peramalan yang sangat baik dan transformasi diferens meningkatkan nilai akurasi peramalan.

DAFTAR PUSTAKA

- Elsner, J. B. dan A. A. Tsonis. 1996. *Singular Spectrum Analysis A New Tool in Time Series Analysis*. Springer Science & Business Media.
- Golyandina, N. dan A. Korobeynikov. 2014 Basic singular spectrum analysis and forecasting with R. *Computational Statistics & Data Analysis*. 71:934-954.
- Golyandina, N. dan A. Zhigljavsky. 2013. *Singular Spectrum Analysis for Time Series*. Springer.
- Golyandina, N., V. Nekrutkin, dan A. A. Zhigljavsky. 2001. *Analysis of Time Series Structure : SSA and Related Techniques*, Chapman & Hall/CRC.
- Hassani, H. 2007. Singular spectrum analysis: methodology and comparison. *Journal of Data Science*, 5:239-257.
- Hassani, H., A. S. Soofi, dan A. A. Zhigljavsky. 2010. Predicting daily exchange rate with singular spectrum analysis. *Nonlinear Analysis:Real World Applications*, 11: 2023-2034.
- Hassani, H., S. Heravi, dan A. Zhigljavsky. 2009. Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting*, 25:103-118.
- Mahmoudvand, R., P. C. Rodrigues, dan M. Yarmohammadi. 2019. Forecasting daily exchange rates: A comparison between SSA and MSSA. *Statistical Journal*, 17 (4):599-616.

- Makridakis, S. G., S. C. Wheelwright, dan R. J. Hyndman. 2008. *Forecasting Methods and Applications*. Wiley.
- Wei, W. W. S. 2006. *Time Series Analysis Univariate and Multivariate Methods*, second edition, Pearson Education Inc.