

An Application of Hybrid Forecasting Singular Spectrum Analysis – Extreme Learning Machine Method in Foreign Tourists Forecasting

Muhammad Fajar

BPS-Statistics Indonesia, mfajar@bps.go.id

doi: <https://doi.org/10.15642/mantik.2019.5.2.60-68>

Abstrak: Wisman adalah salah satu indikator untuk melihat perkembangan pariwisata. Perkembangan pariwisata mempunyai andil penting bagi perekonomian karena pariwisata merupakan booster peningkatan devisa, menciptakan peluang usaha, dan membuka kesempatan kerja. Sebagai bahan input untuk strategi dan program pariwisata adalah prediksi terhadap jumlah wisman di masa depan yang diperoleh dari peramalan. Dalam paper ini, penulis menggunakan metode *Hybrid* singular spectrum analysis – extreme learning machine untuk meramalkan jumlah wisman. Data yang digunakan dalam penelitian adalah jumlah wisman yang bersumber dari Badan Pusat Statistik. Hasil penelitian ini bahwa kemampuan metode *Hybrid* SSA-ELM sangat baik dalam meramalkan jumlah wisman. Hal tersebut ditunjukkan oleh nilai MAPE sebesar 4.91 persen, dengan *out sample* sebanyak delapan observasi.

Kata kunci: wisata mancanegara, singular spectrum analysis, extreme learning machine

Abstract: International tourism is one indicator of measuring tourism development. Tourism development is important for the national economy since tourism could boost foreign exchange, create business opportunities, and provide employment opportunities. The prediction of foreign tourist numbers in the future obtained from forecasting is used as an input parameter for strategy and tourism programs planning. In this paper, the Hybrid Singular Spectrum Analysis – Extreme Learning Machine (SSA-ELM) is used to forecast the number of foreign tourists. Data used is the number of foreign tourists January 1980 - December 2017 taken from Badan Pusat Statistik (Statistics Indonesia). The result of this research concludes that Hybrid SSA-ELM performance is very good at forecasting the number of foreign tourists. It is shown by the MAPE value of 4.91 percent with eight observations out a sample.

Keywords: foreign tourist, singular spectrum analysis, extreme learning machine

1. Introduction

Indonesia is a country that has a lot of mesmerizing landscapes, natural resources, and diverse cultures. These are tourist attractions that could be optimally utilized to advance the national economy. Tourism development is important for the national economy since tourism could boost foreign exchange, create business opportunities, and provide employment opportunities.

One indicator of measuring tourism development is the number of foreign tourists to visit. The number of foreign tourists visiting Indonesia from January 1980 to December 2017 is visually presented in Fig. 1.1. It could be seen that in this period, the number of foreign tourists visiting Indonesia shows an increasing trend each year. The implication for Indonesia as the host is that the strategy of infrastructural development is needed to avoid the decreasing number of foreign tourists and to lower the negative impacts to the environment caused by the increasing number of foreign tourists visit. Such as environment based transportation vehicles, hotels, recreation facilities, etc.

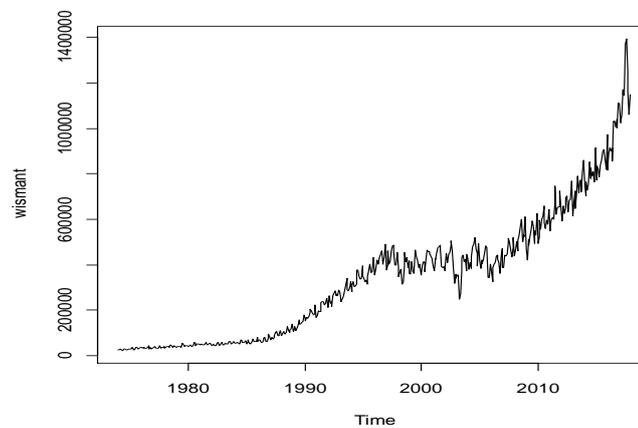


Figure 1.1. The Number of Foreign Tourists January 1980 – December 2017

Therefore, one input for the strategy is the prediction number of foreign tourists obtained from forecasting, specifically, time series forecasting. Time series forecasting is a quantitative method used to analyze a series of data collected in time order using the right technique. The result could be used as a reference to forecast the value of the series in the future [9]. The development of forecasting methods is increasingly rapid and complex as advances in the development of computing technology. The interesting thing from the time series method development is the reconstruction of hybrid time series forecasting method, a time series constructed from two different types of forecasting method [1] – [5].

In this paper, singular spectrum analysis and extreme learning machine techniques are combined to forecast the number of foreign tourists visiting Indonesia. Extreme learning machine is an exclusive example of feed-forward neural network in the form of FFNN with only one hidden layer, where the singular spectrum analysis – FFNN method has been developed before [5].

2. Data

Data used in this research is the number of foreign tourists visiting Indonesia from January 1980 until August 2018. Train data (in the sample) covers the number of foreign tourists in January 1980 until August 2018. Meanwhile, test data (out sample) covers the number of foreign tourists from January 2018 until August 2018.

3. Method

3.1 Singular Spectrum Analysis (SSA)

SSA is a non-parametric time series method based on multivariate statistics principle. SSA decomposes time series additively into several independent components. These components could be identified as trend, periodic, quasi-periodic, and noise component. SSA procedure consists of four steps, they are [6]-[10]:

Step 1. Embedding

Given an x_1, x_2, \dots, x_T time series, choose an even number L , where L parameter is the window length defined as $2 < L < T/2$, and $K = T - L + 1$. The cross matrix is:

$$\mathbf{X} = (X_1, \dots, X_T) = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_T \end{pmatrix}$$

The cross matrix proves to be a Hankel matrix, which means every element in the main anti diagonal has the same value. Thus, \mathbf{X} could be assumed as multivariate data with L characteristic and K observations so that the covariance matrix is $\mathbf{S} = \mathbf{X}\mathbf{X}'$ with the dimension of $L \times L$.

Step 2. Singular Value Decomposition (SVD)

Suppose that \mathbf{S} has eigenvalue and eigenvector λ_i and U_i , respectively. Where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$ and U_1, \dots, U_L . Thus, obtained SVD from \mathbf{X} as follows:

$$\mathbf{X} = E_1 + E_2 + \dots + E_d \tag{1}$$

where $E_i = \sqrt{\lambda_i} U_i V_i'$, $i = 1, 2, \dots, d$, E_i is the main component, d is the number of eigenvalue λ_i , and $V_i = \mathbf{X}' U_i / \sqrt{\lambda_i}$.

Step 3. Grouping

In this step, \mathbf{X} is additively grouped into subgroups based on patterns that form a time series. They are trend, periodic, quasi-periodic, and noise component. Partition the index set $\{1, 2, \dots, d\}$ into several groups I_1, I_2, \dots, I_n , then correspond matrix \mathbf{X}_I into group $I = \{i_1, i_2, \dots, i_b\}$ which is defined as:

$$\mathbf{X}_I = E_{i_1} + E_{i_2} + \dots + E_{i_b} \quad (2)$$

Thus, the decomposition represents as:

$$\mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_n} \quad (3)$$

with $\mathbf{X}_{I_j} (j = 1, 2, \dots, n)$ is reconstructed component (RC). \mathbf{X}_I component contribution measured with corresponding eigenvalue contribution: $\sum_{i \in I} \lambda_i / \sum_{i=1}^d \lambda_i$. Using the close frequency range from the main components is based on the study of the grouping process using auto grouping [11]. Main components with relatively close frequency ranges are grouped into one reconstructed component. Soon, until several reconstructed components are formed.

Step 4. Reconstruction

In this last step, \mathbf{X}_{I_j} is transformed into a new time series with T observations obtained from diagonal averaging or Hankelization. Suppose that \mathbf{Y} is a matrix with $L \times K$ dimensions and has $y_{ij}, 1 \leq i \leq L, 1 \leq j \leq K$ elements. Then, $L^* = \min(L, K), K^* = \max(L, K)$, and $T = L + K - 1$. Then, $y_{ij}^* = y_{ij}$ if $L < K$ and $y_{ij}^* = y_{ji}$ if $L > K$. Matrix \mathbf{Y} transferred into y_1, y_2, \dots, y_T series with using the following formula:

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

Diagonal averaging on equation (4) is applied to every matrix component \mathbf{X}_{I_j} on equation (3) resulting a $\tilde{\mathbf{X}}^{(k)} = (\tilde{x}_1^{(k)}, \tilde{x}_2^{(k)}, \dots, \tilde{x}_T^{(k)})$ series. Thus, x_1, x_2, \dots, x_T series is decomposed into an addition of reconstructed m series:

$$x_t = \sum_{k=1}^m \tilde{x}_t^{(k)}, t = 1, 2, \dots, T \quad (5)$$

SSA forecasting used in this research is SSA recurrent, with estimating min-norm LRR (Linear Recurrence Relationship) coefficient. The LRR coefficient is calculated with the following algorithm:

1. Input: Matrix $\mathbf{P} = [P_1; \dots; P_r]$, \mathbf{P} is a matrix composed of P_i eigenvector from SVD step, where Select a group of r ($1 \leq r \leq L$) eigenvectors P_1, \dots, P_r . Suppose that $\underline{\mathbf{P}}$ is a \mathbf{P} that the last row is removed, and $\overline{\mathbf{P}}$ is a \mathbf{P} that the first row is removed.

2. For every P_i vector column from \mathbf{P} , calculate π_i , where π_i is the last component from P_i , and \underline{P}_i is a P_i that the last component is removed.
3. Calculate: $v^2 = \sum_{i=1}^r \pi_i^2$. If $v^2 = 1$, then STOP with a warning message "Verticality coefficient equals 1."
4. Calculate the min-norm LRR coefficient (\mathcal{R}):

$$\mathcal{R} = \frac{1}{1 - v^2} \sum_{i=1}^r \pi_i \underline{P}_i$$

5. From point (4) obtained: $\mathcal{R} = (\alpha_{L-1} \dots \alpha_1)'$.
6. Then, calculate the forecasting value with:

$$\hat{x}_n = \sum_{i=1}^{L-1} \alpha_i \tilde{x}_{n-1}, \quad n = T + 1, \dots, T + h$$

3.2 Extreme Learning Machine (ELM)

Extreme learning machine is a learning scheme of feedforward neural network for single-hidden layer feedforward neural networks (SLFN). ELM could adaptively set the number of nodes and randomly chooses the input weight \mathbf{W} and bias \mathbf{b}_i on a hidden layer. Hidden layer weight is obtained by using the least square method [12].

Suppose that a SLFN training process with K hidden nodes and an activation vector function $\mathbf{g}(\tilde{x}) = (g_1(\tilde{x}), g_2(\tilde{x}), \dots, g_K(\tilde{x}))$ for N samples $(\tilde{x}_i, \mathbf{p}_i)$ learning process, with $\tilde{x}_i = [\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in}]'$ and $\mathbf{p}_i = [p_{i1}, p_{i2}, \dots, p_{in}]'$ is performed. If SLFN could approx N samples without any error (zero error), thus:

$$\sum_{j=1}^N \|\mathbf{y}_j - \mathbf{p}_j\| = 0, \tag{6}$$

with \mathbf{y}_j is actual output value of SLFN. There are also parameter $\boldsymbol{\beta}_i = [\beta_{i1}, \dots, \beta_{im}]'$, $\mathbf{w}_i = [w_{i1}, \dots, w_{im}]'$ and b_i which are interconnected in:

$$\sum_{i=1}^K \beta_i g_i(\mathbf{w}_i \tilde{x}_j + b_i) = \mathbf{p}_j, j = 1, \dots, N, i = 1, \dots, K \tag{7}$$

\mathbf{w}_i is a weight vector connecting the i -th hidden node and the input node, $\boldsymbol{\beta}_i$ is a weight vector connecting the i -th hidden node and the output node, and b_i is the threshold of the i -th hidden node. Equation (7) could be expressed as:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \tag{8}$$

With $\mathbf{H} = \{h_{ij}\}$ is the output matrix of the hidden layer, $h_{ij} = g(w_j \tilde{x}_i + b_j)$ represents the output of j -th hidden neuron corresponding with \tilde{x}_i , $\boldsymbol{\beta} =$

$[\beta_1, \dots, \beta_K]$ is the weight output matrix, and $\mathbf{T} = [\mathbf{p}_1, \dots, \mathbf{p}_N]'$ is the target matrix. The output weight (weight connecting the hidden layer and the output) is obtained from finding the solution of least square from the linear system given. The solution of the linear system (8) is:

$$\hat{\beta} = \mathbf{H}^+ \mathbf{T} \quad (9)$$

With \mathbf{H}^+ is Moore-Penrose generalized inverse matrix of \mathbf{H} . The solution of equation (8) is unique and has the shortest distance compared to other solutions. As mentioned in the reference [12], ELM tends to give a generally good performance along with the increasing of the learning speed using the Moore-Penrose Generalized Inverse method.

ELM algorithm could be summarized into four steps [12], they are:

1. Define the number of hidden nodes (K), then randomly choose the initial value of β_i and b_i .
2. Calculate matrix \mathbf{H} .
3. Based on equation (9), calculate the weight output $\hat{\beta}$.
4. Then, the forecasting result $\hat{\mathbf{T}}$ is calculated with:

$$\hat{\mathbf{T}} = \mathbf{H} \hat{\beta} \quad (10)$$

3.3 Hybrid Singular Spectrum Analysis – Extreme Learning Machine

In this section, this paper proposes the hybrid SSA – ELM forecasting method as follows:

1. Time series is decomposed into main components by using SSA method.
2. The main components obtained from (1) could be defined as trend, periodic, quasi-periodic, and noise component.
3. The reconstructed component is formed by adding up several main components based on the frequency range closeness.
4. ELM is applied to every reconstructed component thus, ELM architecture is different for every reconstructed component.
5. The final result of hybrid SSA-ELM forecasting is an addition of forecasting results from several ELM architectures using equation (10). Steps above are visually presented in Picture 3.1.

3.4 Forecasting Accuracy

Forecasting accuracy of the test data (out sample) in this research uses MAPE (Mean Absolute Percentage Error) formulated as follows:

$$\text{MAPE} = \frac{1}{v - T} \sum_{t=T+1}^v \left| \frac{F_t - A_t}{A_t} \times 100\% \right| \quad (11)$$

With F_t is the t -th forecasting result value, and A_t is the t -th actual value. MAPE characteristics: (1) If the MAPE < 10%, then the Hybrid SSA – ELM method performance is very good, (2) if the MAPE value is in the range of 10% - 20%, then the Hybrid SSA – ELM method forecasting performance is good, (3) if the

MAPE in the range of 20% - 50%, then the Hybrid SSA – ELM method forecasting performance is adequate, and (4) if the MAPE > 50%, then the Hybrid SSA – ELM method forecasting performance is bad.

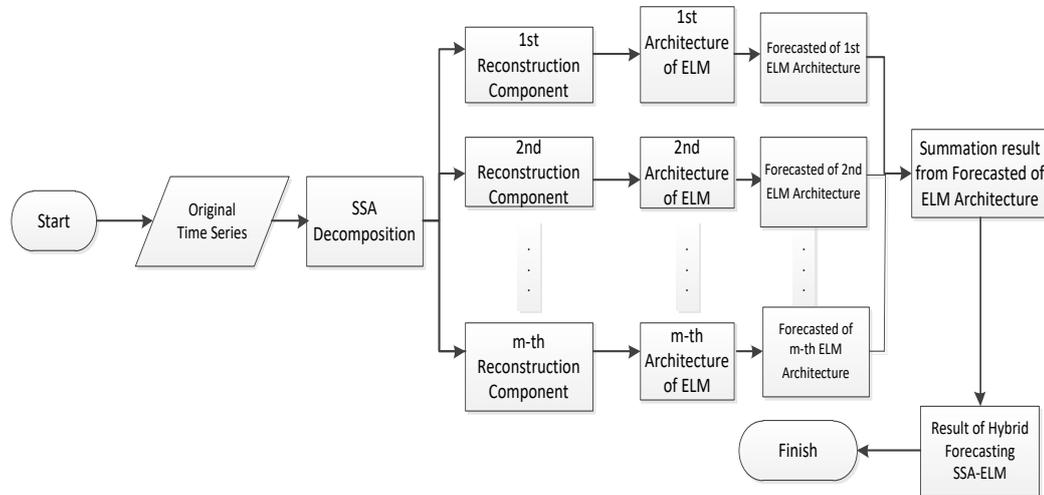


Figure 3.1. Hybrid SSA – ELM Method Flowchart

4. Results and Discussion

In the process of Hybrid SSA – ELM forecasting, the first step is decomposing foreign tourists data with SSA. In SSA, defining the value of L in this paper is the number of train data (528 observations) divided by two. Thus L is 264. Based on SVD (Singular Value Decomposition), 264 components are obtained from SSA process of train data, but only the first ten components are picked because with only these first ten components could explain the variation of foreign tourists 99.38 percent to be analyzed further.

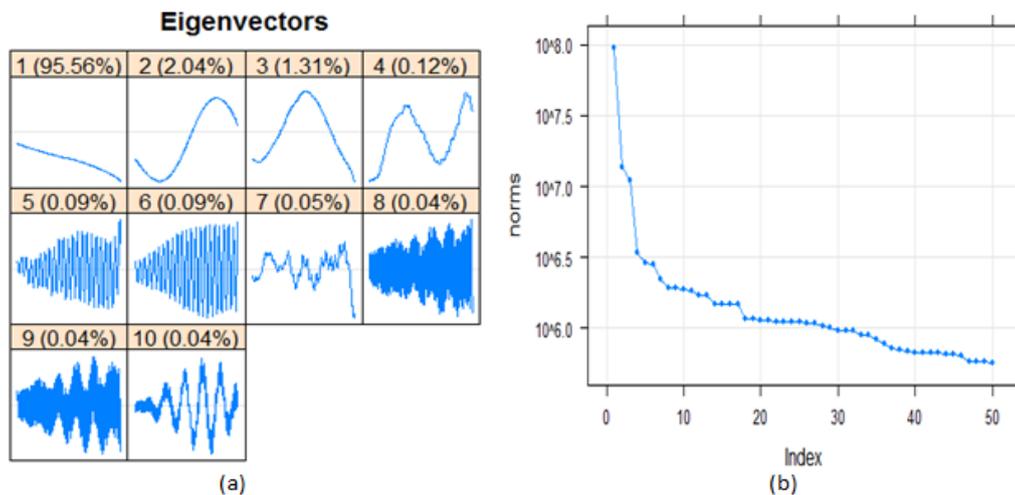


Figure 4.1. (a). Variation of the first ten components, (b). Data scree plot

Grouping is performed by looking at the similarities of 10 components plot patterns that indirectly indicate the similarities of the components. Ten components are grouped into ten groups. Group 1 consists of the 1st components,

group 2 consists of the 2nd components, group 3 consists of the 3rd components, and so on.

Table 4.1 presents the forecasting result of SSA-ELM, SSA dan ELM methods in test data forecasting. SSA-ELM has the lowest MAPE (4.91 percent) compared to SSA (28 percent) and ELM (9.07 percent). Based on MAPE characteristics, SSA-ELM and ELM methods' performance is very good. Meanwhile, SSA performance is adequate.

Table 4.1. The Forecasting Result of Hybrid SSA-ELM, SSA, and ELM

The year 2018 Month:	The actual number of Foreign Tourists	Forecast Result			MAPE		
		SSA-ELM	SSA	ELM	SSA- ELM	SSA	ELM
January	1100677	1215770.91	925266.00	1141727.56	10.46	15.94	3.73
February	1201001	1218719.07	929846.08	1057204.11	1.48	22.58	11.97
March	1363339	1277141.99	935022.03	1093649.66	6.32	31.42	19.78
April	1300277	1322031.55	941383.50	1205315.22	1.67	27.60	7.30
May	1242588	1321390.12	944461.27	1182573.77	6.34	23.99	4.83
June	1318094	1347513.63	948271.13	1178043.33	2.23	28.06	10.63
July	1540549	1424376.79	954890.47	1404689.88	7.54	38.02	8.82
August	1510764	1461883.81	960299.96	1427398.44	3.24	36.44	5.52
Average					4.91	28.00	9.07

Source: author

5. Conclusions

Based on the previous discussion, it could be concluded that Hybrid SSA-ELM performance is very good in forecasting the number of foreign tourists. It is shown by the MAPE value of 4.91 percent with eight observations out the sample.

References

- [1] C.H. Aladag, E. Egrioglu, and C. Kadilar, "Improvement in forecasting accuracy using the hybrid model of arfima and feed forward neural network american," *Journal of Intelligent Systems*, vol.2, no.2, pp. 12-17, 2012.
- [2] D. Rahmani, "A Forecasting algorithm for singular spectrum analysis based on bootstrap linear recurrent formula coefficients," *International Journal of Energy and Statistics*, vol.2, no.4, pp. 287-299, 2014.
- [3] M. Fajar, "Perbandingan kinerja peramalan pertumbuhan ekonomi Indonesia antara ARMA, FFNN dan hybrid ARMA-FFNN," 2016. DOI:10.13140/RG.2.2.34924.36483.
- [4] M. Fajar, "Meningkatkan akurasi peramalan dengan menggunakan metode hybrid singular spectrum analysis-multilayer perceptron neural networks," 2018. DOI: 10.13140/RG.2.2.32839.60320.
- [5] M. Fajar, "Perbandingan kinerja peramalan antara metode hybrid singular spectrum analysis-multilayer perceptrons neural network dan hybrid singular spectrum analysis-feed forward neural network pada data inflasi,"

2018. DOI: 10.13140/RG.2.2.10312.98561.
- [6] N. Golyandina, V. Nekrutkin, and A. Zhiglovsky, *Analysis of Time Series: SSA and Related Techniques*. Chapman and Hall/CRC, 2001.
 - [7] R. Siregar, D. Prariesa, and G. Darmawan, “Aplikasi Metode Singular Spectral Analysis (SSA) dalam Peramalan Pertumbuhan Ekonomi Indonesia Tahun 2017”, *mantik*, vol. 3, no. 1, pp. 5-12, Oct. 2017
 - [8] Y. Jatmiko, R. Rahayu, and G. Darmawan, “Perbandingan Keakuratan Hasil Peramalan Produksi Bawang Merah Metode Holt-Winters dengan Singular Spectrum Analysis (SSA)”, *mantik*, vol. 3, no. 1, pp. 13-22, Oct. 2017.
 - [9] D. Lubis, M. Johra, and G. Darmawan, “Peramalan Indeks Harga Konsumen dengan Metode Singular Spectral Analysis (SSA) dan Seasonal Autoregressive Integrated Moving Average (SARIMA)”, *mantik*, vol. 3, no. 2, pp. 74-82, Oct. 2017.
 - [10] Th. Alexandrv, and N. Golyandina, “automatic extraction and forecast of time series cyclic components within the framework of SSA,” . *In Proceedings of the 5th St.Petersburg*, 2005.
 - [11] W. Makridakis, and MacGee, *Metode dan Aplikasi Peramalan*. Binarupa Aksara, 1999.
 - [12] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, “Extreme learning machine: algorithm, theory and applications,” *Artificial Intelligence Review*, vol.44, no.1 pp. 103-115, 2013.