

# SSA and ARIMA for Forecasting Number of Foreign Visitor Arrivals to Indonesia

Agustinus Angelus Ete<sup>1\*</sup>, Suhartono<sup>2</sup>, R. Mohammad Atok<sup>3</sup>

<sup>1</sup>IKIP Muhammadiyah Maumere, <sup>2</sup>Institut Teknologi Sepuluh Nopember (ITS) Surabaya

<sup>1</sup>Jl. Jenderal Sudirman, Sikka 86111, <sup>2</sup>Jl. Arief Rahman Hakim, Surabaya 60111

e-mail: <sup>1\*</sup>agustinus\_angelaus@ikipmumaumere.ac.id, <sup>2</sup>suhartono@statistika.its.ac.id,

<sup>3</sup>moh\_atok@statistika.its.ac.id

**Abstract**—Singular Spectrum Analysis (SSA) is the technique of non-parametric analysis of time series used for forecasting. SSA aims to decompose the original time series into a summation of a small number of components that can be interpreted as the trend, oscillatory components, and noise. The purpose of this research is to understand how the SSA model in forecasting the number of foreign tourist arrivals to Indonesia through four entrances. The result of forecasting obtained by using SSA will be compared with ARIMA method to assess its superiority. The data used in this study are the data of the number of foreign tourist arrivals to Indonesia through four entrances in the period January 1996 to August 2016. Four entrances used in this study are Ngurah Rai Airport, Kualanamu Airport, Soekarno-Hatta Airport, and Juanda Airport. The level of forecasting accuracy generated by each forecasting method is measured using the Mean Absolute Percentage Error (MAPE) criterion. The results showed that SSA method is the best forecasting method for forecasting the number of foreign tourist arrivals through Ngurah Rai Airport with an average MAPE value of 9.6%. Forecasting the number of foreign tourist arrivals through Kualanamu Airport, ARIMA method is the best forecasting method with an average MAPE value of 22.4%. In forecasting the number of foreign tourist arrivals through Soekarno-Hatta Airport, ARIMA method is the best forecasting method with an average MAPE value of 10.5%. In forecasting the number of foreign tourist arrivals through Juanda Airport, ARIMA method is the best forecasting method with an average MAPE value of 9.9%.

**Keywords**-- SSA, ARIMA, Foreign Visitor Arrivals.

## I. INTRODUCTION

Forecasting is a process of predicting events that will occur in the future by looking at past data and current data. One way to generate valid and accurate predictions is to choose the correct forecasting method. An important step in choosing an appropriate time series method is to consider the type of data pattern. Pattern data can be divided into four, namely the pattern of stationary, trend, seasonal, and cyclic. One method that can be used to separate the three components of the basic pattern is the Singular Spectrum Analysis (SSA). SSA decomposes the original time series into a summation of a small number of independent components and can be interpreted as a trend that varies slowly, oscillatory components, and noise [1]. The algorithm of SSA consists of two complementary phases, namely the decomposition and reconstruction phase. At the decomposition stage, the first step to do is embedding. In the embedding step, a one-dimensional series is described as a multidimensional series whose dimensions is called window length [1]. Multidimensional time series form the trajectory matrix. Window length is the

main parameter of SSA, in the sense that its improper choice would imply that no grouping activities will help to obtain a good SSA decomposition [1]. The next step is the singular value decomposition (SVD), is the singular value decomposition of the trajectory matrix into a sum of rank-one bi-orthogonal matrices [1]. At the reconstruction stage, the first step to do is grouping. Effect grouping ( $r$ ) is the parameter used in the grouping stage [1]. Effect grouping ( $r$ ) is used to limit the number of eigentriples to be used to identify trend and seasonality components. The grouping step corresponds to splitting the matrices, computed at the SVD step, into several groups and summing the matrices within each group. The last step is diagonal averaging. The last step transfers each resultant matrix into a time series, which is an additive component of the initial series [1]. SSA was first applied to extract tendencies and harmonic components in meteorological and geophysical time series [2]. In recent years SSA has been developed and applied to many practical problems. Hassani et al. [2] forecast European industrial production using SSA. Darmawan et al. [3] used Auto SSA to predict the incidence of flooding in Bandung and surrounding areas. Hassani et al. [4] forecast U.S. tourist arrivals using optimal singular spectrum analysis, ARIMA, exponential smoothing, and neural networks. Hassani et al. [4] use SSA vector forecasting algorithms to get predicted values for U.S. tourist arrivals. In the SSA method, there are two forecasting algorithms that can be used to obtain prediction values, namely recurrent forecasting algorithms and vector forecasting algorithms. Vector forecasting has a larger computational cost than recurrent forecasting [5]. In this study, SSA method will be used to predict the number of foreign tourist arrivals per month to Indonesia according to the entrance. The entrance used in this study consists of Ngurah Rai Airport, Kualanamu Airport, Batam, Soekarno Hatta Airport, and Juanda Airport. In the forecasting stage, SSA recurrent forecasting algorithm is used to get the forecast value. Forecasting the number of foreign tourist arrivals is also beneficial for the tourism office both central and regional to prepare tourism programs associated with the new tourism program that is Wonderful Indonesia. The SSA method will be compared with other methods to assess its superiority. The method used as a comparison in this research is ARIMA. SSA and ARIMA are useful for forecasting series containing seasonal components. The ARIMA method consists of four models, namely Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA).

## II. SINGULAR SPECTRUM ANALYSIS

The main purpose of SSA is to decompose the original time series into a summation of a small number of components that can be interpreted as the trend, oscillatory components, and noise [1]. The algorithm of SSA consists of two complementary stages, namely decomposition and reconstruction.

### A. Decomposition

There are two steps are used at the stage of decomposition, namely embedding and Singular Value Decomposition (SVD). The parameters which have an important role in decomposition is window length ( $L$ ) [1].

#### Embedding

The embedding procedure maps the original time series  $F = (f_0, \dots, f_{N-1})$  to a multidimensional series of  $L$ -lagged vectors  $\mathbf{X}_1, \dots, \mathbf{X}_K$ . Here  $\mathbf{X}_i = (f_{i-1}, \dots, f_{i+L-2})^T$  for  $i = 1, \dots, K$ , where  $K = N - L + 1$ . The parameter  $L$  is known as the window length and is an integer such that  $1 < L < N$ . The value of the parameter  $L$  lies between  $\frac{N}{2}$  and  $\frac{N}{3}$  [6].

The trajectory matrix of the series  $F$  [1]:

$$\mathbf{X} = [\mathbf{X}_1 : \dots : \mathbf{X}_K] = \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{K-1} \\ f_1 & f_2 & f_3 & \dots & f_K \\ f_2 & f_3 & f_4 & \dots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \dots & f_{N-1} \end{pmatrix} \quad (1)$$

#### Singular Value Decomposition

Define the matrix  $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ . Suppose that  $\lambda_1, \dots, \lambda_L$  are the eigenvalues of  $\mathbf{S}$ , where  $\lambda_1 \geq \dots \geq \lambda_L \geq 0$ .  $\mathbf{U}_1, \dots, \mathbf{U}_L$  are the eigenvector of the matrix  $\mathbf{S}$  corresponding to  $\lambda_1, \dots, \lambda_L$ . The SVD of the trajectory matrix  $\mathbf{X}$  can be written as [6]:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d, \quad (2)$$

where,  $\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T$  for  $i = 1, \dots, d$ .  $d = \max\{i: \lambda_i > 0\}$  is rank of the matrix  $\mathbf{X}$ .  $\mathbf{V}_i = \mathbf{X}^T \mathbf{U}_i / \sqrt{\lambda_i}$  are the principal components. The collection  $(\sqrt{\lambda_i}, \mathbf{U}_i, \mathbf{V}_i)$  will be called  $i$ th eigentriple of the SVD [6].

### B. Reconstruction

There are two steps are used at the stage of reconstruction, namely grouping and diagonal averaging. The parameters which have an important role in reconstruction is grouping effect ( $r$ ).

#### Grouping

The grouping procedure partitions the set of indices  $\{1, \dots, d\}$  into  $m$  disjoint subsets  $I_1, \dots, I_m$ . The resultant matrix  $\mathbf{X}_I$  is

defined as  $\mathbf{X}_I = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}$ . These is computed for  $I = I_1, \dots, I_m$  and leads to the decomposition  $\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}$  [6].

#### Diagonal Averaging

Define  $\mathbf{Y}$  be an  $L \times K$  matrix with elements  $y_{ij}$ ,  $1 \leq i \leq L$ ,  $1 \leq j \leq K$ . Set  $L^* = \min(L, K)$ ,  $K^* = \max(L, K)$ ,  $y_{ij}^* = y_{ij}$  if  $L < K$ , and  $y_{ij}^* = y_{ji}$  if  $L > K$ . The element  $g_k$  as a result of the diagonal averaging of  $\mathbf{Y}$  is given by [1]:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m, k-m+2}^* & \text{for } 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+2}^* & \text{for } L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m, k-m+2}^* & \text{for } K^* \leq k < N. \end{cases} \quad (3)$$

Diagonal averaging applied to a resultant matrix  $\mathbf{X}_{I_k}$  produces a reconstructed series  $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$ ,

where  $f_n = \sum_{k=1}^m \tilde{f}_n^{(k)}$ , for  $n = 0, 1, \dots, N-1$  [1].

## III. SSA FORECASTING

The SSA forecasting method can be applied to time series that approximately satisfy linear recurrent formula. The recurrent forecasting algorithm can be formulated as follows [1]:

a. The time series  $F_{N+M} = (f_0, \dots, f_{N+M-1})$  is defined by

$$f_i = \begin{cases} \tilde{f}_i & \text{for } i = 0, \dots, N-1, \\ \sum_{j=1}^{L-1} a_j f_{i-j} & \text{for } i = N, \dots, N+M-1, \end{cases} \quad (4)$$

where the vector of coefficients  $A = (a_1, \dots, a_{L-1})$  can be expressed as  $A = \frac{1}{1-v^2} \sum_{i=1}^r \pi_i U_i^\nabla$ . For an eigenvector

$U \in \mathbb{R}^L$ , the vector of the first  $L-1$  components of the vector  $U$  can be expressed as  $U^\nabla \in \mathbb{R}^{L-1}$ . Define  $v^2 = \pi_1^2 + \dots + \pi_r^2 < 1$ , where  $\pi_i$  is the last component of the eigenvector  $U_i$ , for  $i = 1, \dots, r$ .

b. The numbers  $f_N, \dots, f_{N+M-1}$  form the  $M$  terms of the recurrent forecast.

## IV. ARIMA

In the univariate model, classical methods that are often used include Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing [7]. The ARIMA model can be used on data that has non-seasonal and seasonal

patterns. If the data used contains seasonal patterns, the ARIMA model used is the seasonal ARIMA model. The ARIMA model  $(P, D, Q)^s$  for the seasonal data pattern is formulated as follows [8]:

$$\phi_p(B)\Phi_P(B^s)Y_t^* = \theta_0 + \theta_q(B)\Theta_Q(B^s)a_t. \quad (5)$$

### V. MEASURES OF ACCURACY

If a good ARIMA model is obtained for more than one model, the best model can be determined using in-sample criteria and out-sampled criteria. If the best model selection is based on in-sample criteria, then the criterion that can be used is Akaike's Information Criterion (AIC) [9]:

$$AIC = n \ln \hat{\sigma}_a^2 + 2M, \quad (6)$$

where  $M$  is the number of parameters that are estimated,  $n$  is the number of observations, and  $\hat{\sigma}_a^2$  represents the residual variance value with maximum likelihood estimation. If the best model selection is based on out-sample criteria, then the criterion that can be used is *Mean Absolute Percentage Error* (MAPE) [10]:

$$MAPE = \left( \frac{1}{M} \sum_{l=1}^M (Z_{n+l} - \hat{Z}_n(l))^2 \right) 100\%. \quad (7)$$

The best model has the smallest MAPE value.

### VI. RESEARCH METHODOLOGY

The data used in the study is data on the number of foreign tourist arrivals per month to Indonesia according to the entrance in the period January 1996 until August 2016. The type of data used in this study is secondary data. Data obtained from the Central Bureau of Statistics. Data is divided into two parts, namely the in-sample data and out-sample data. The number of in-sample data is 236, starting from January 1996 until August 2015. The number of out-sample data is 12, starting from September 2015 until August 2016. The variables used in this study as follows:

$Y_{1,t}$  : Number of foreign tourist arrivals per month through Ngurah Rai Airport,

$Y_{2,t}$  : Number of foreign tourist arrivals per month through Kualanamu Airport,

$Y_{3,t}$  : Number of foreign tourist arrivals per month through Soekarno-Hatta Airport,

$Y_{4,t}$  : Number of foreign tourist arrivals per month through Juanda Airport.

The steps that must be done in the forecasting stage using the SSA method are:

1. Embedding
2. SVD
3. Grouping
4. Diagonal Averaging
5. Forecasting data testing using recurrent forecasting algorithms
6. Calculate the MAPE value of data testing.

The steps that must be done in the forecasting stage using the ARIMA method are:

1. Create a time series chart to detect stationarity of data. If data are not stationary on its variance, then the data need to be transformed. The type of transformation used is the Box-Cox transformation. If data are not stationary on mean, then the data need to be done differencing operations.
2. If the data has been stationary, then ARIMA model can be identified based on ACF and PACF pattern.
3. Test the significance of the parameters of the ARIMA model.
4. Test the assumption of normality and white noise on the residual ARIMA model.
5. If a good ARIMA model is obtained for more than one model, the best model can be determined using AIC criteria.
6. Forecasting the number of foreign tourist arrivals.
7. Calculate the MAPE value of data testing.

### VII. RESULTS

General description of foreign tourist arrivals to Indonesia is explained by using descriptive statistics and time series diagrams in Table 1 and Figure 1.

Based on Table 1, it is known that Ngurah Rai Airport Bali has the highest average tourist arrival rate of foreign tourists compared to the average value of foreign tourist arrivals through other entrances. Bali has a variety of beauty and uniqueness of nature so that Bali became one of the favorite tourist destinations for foreign tourists for a vacation. Bali is also often a venue for international events, such as Miss World, APEC, WTO, and the World Hindu Forum [12]. Some important events affecting the number of tourist arrivals to Indonesia from January 1996 to August 2015 are shown graphically through a time series chart.

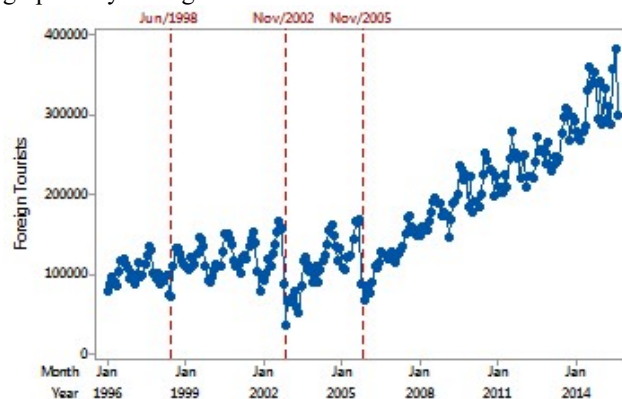


Figure 1. Time Series Chart of the Number of Foreign Visitor Arrivals through Ngurah Rai Airport

Based on Figure 1, it is known that in 1997-1998 the number of foreign tourist arrivals to Ngurah Rai Airport decreased due to the monetary and economic crisis that occurred in Indonesia. The number of foreign tourist arrivals also declined at three other entrances due to the monetary and economic crisis. In November 2002 and November 2005, the

number of foreign tourist arrivals to Ngurah Rai Airport decreased due to the Bali I bombings and the Bali II bombings. The Bali II bombing event in 2005 also affected the number of foreign tourists arriving through Juanda Airport. In October 2006 the number of foreign tourist arrivals to Juanda Airport decreased due to the Bali II bombing. In 2003, the number of foreign tourist arrivals to Indonesia through Kualanamu Airport, Soekarno-Hatta Airport, and Juanda Airport decreased simultaneously due to the SARS outbreak that hit the world. In April 2015, the number of foreign tourist arrivals to Soekarno-Hatta Airport decreased due to the closure of two international flight routes, the Jakarta-Taipei and Jakarta-Haneda routes since January 2015.

**Table 1.** Descriptive Statistics of Foreign Visitor Arrivals to Indonesia through Four Entrances in the Period January 1996 Until August 2015

Entrance	Mean	Standard Deviation	Minimum	Maximum
Ngurah Rai	164030	73532	35107	381890
Kualanamu	11712	5111	3496	29619
Soekarno-Hatta	118120	41964	50370	252914
Juanda	11012	4478	2960	22986

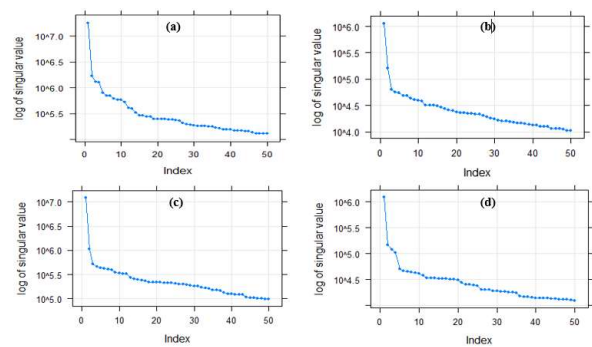
#### A. Forecasting Using SSA Method

In the SSA method, there are two stages used to separate trend components, seasonality components, and noise components from the initial time series, namely the decomposition and reconstruction phase. Window length ( $L$ ) is a single parameter at the decomposition stage. Table 2 presents the number of eigentriples generated based on the value of the window length parameters taken at the decomposition stage at four entrances.

**Table 2.** Number of Eigentriples Generated Based on Window Length Value ( $L$ ) at the Four Entrance Entrances

Entrance	Window Length ( $L$ )	Number of Eigentriples
Ngurah Rai	96	50
Kualanamu	60	50
Soekarno-Hatta	72	50
Juanda	108	50

In the reconstruction phase, there are two steps used to obtain the trend components, seasonality components, and noise components of the eigentriples produced at the decomposition stage, namely grouping and diagonal averaging. Effect grouping ( $r$ ) is the parameter used in the grouping stage. Effect grouping ( $r$ ) is used to limit the number of eigentriples to be used to identify trend and seasonality components. The value of the effect grouping parameter ( $r$ ) is determined based on the number of eigentriples that do not reflect the noise on the plot of the singular value. In the plot of the singular value, the slowly decreasing sequence of singular values reflects the noise component.



**Figure 2.** Plot of 50 Singular Values on Case Number of Foreign Visitor Arrivals through Ngurah Rai Airport (a), Kualanamu (b), Soekarno-Hatta (c), and Juanda (d)

Based on Figure 2, it is known that the value of the effect grouping parameter ( $r$ ) for Ngurah Rai Airport is 18 because the number of eigentriples that do not reflect the noise on the plot of the singular value is 18. Eigentriple 19 to eigentriple 50 is identified as a noise component because the singular value decreases slowly on eigentriple 19 to eigentriple 50. The value of the effect grouping parameter ( $r$ ) for Kualanamu Airport is 11. The value of the effect grouping parameter ( $r$ ) for Soekarno-Hatta Airport is 12. The value of the effect grouping parameter ( $r$ ) for Juanda Airport is 11. Although eigentriples reflecting noise has been identified, there is a possibility that the number of eigentriples reflecting noise can increase. The remaining eigentriples that are not related to the trend and seasonality of the  $r$  eigentriples will be grouped into the noise group. After the noise component is successfully grouped, the next step is to grouping the eigentriples associated with trend and seasonality. The plot of the reconstructed series can be used to identify eigentriples related to trends and seasonality. Trend is a slowly varying component of a time series which does not contain oscillatory components [2]. All slowly varying components in the plot of the reconstructed series are identified as trend components. The grouping of eigentriples associated with seasonality is based on similarity of singular values of two consecutive eigentriples. In the plot of the reconstructed series, the similarity of the singular value results in a series reconstructed by an eigentriple having the same seasonal pattern and the same seasonal period as the series reconstructed by the other eigentriple. Two eigentriples with singular singular similarities can be grouped into seasonality groups if the seasonal periods of the series reconstructed by the two eigentriples are 12 months, 6 months, 4 months, 3 months, 2.4 months, and 2 months [13]. Figure 3 shows the number of reconstructed series based on the value of the effect grouping parameters obtained at Ngurah Rai entrance.

Based on Figure 3, it is known that the series reconstructed by eigentriple 1, eigentriple 2, and eigentriple 5 contain slowly varying components, so that the three eigentriples are grouped into the trend group. According to Figure 3, it is known that eigentriple 3 has a singular value similarity with eigentriple 4 because the series reconstructed by eigentriple 3 has the same seasonal pattern as the series reconstructed by eigentriple 4. The series reconstructed by

eigentruple 3 has the same seasonal period as the series reconstructed by eigentruple 4, which is 12 months. Thus eigentruple 3 and eigentruple 4 are grouped into seasonality groups. Table 3 provides a complete 50 eigentriples and related components at the four entrances. Figure 4 shows each of the reconstructed components corresponding to the corresponding eigentruple of each entrance.

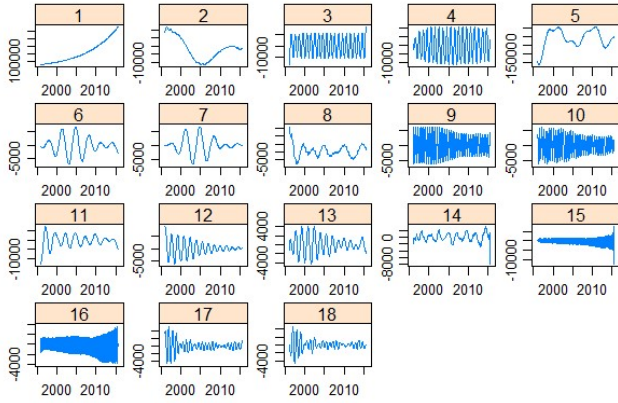


Figure 3. Plot of the Reconstructed Series on the Case of Number of Foreign Visitor Arrivals through Ngurah Rai Airport

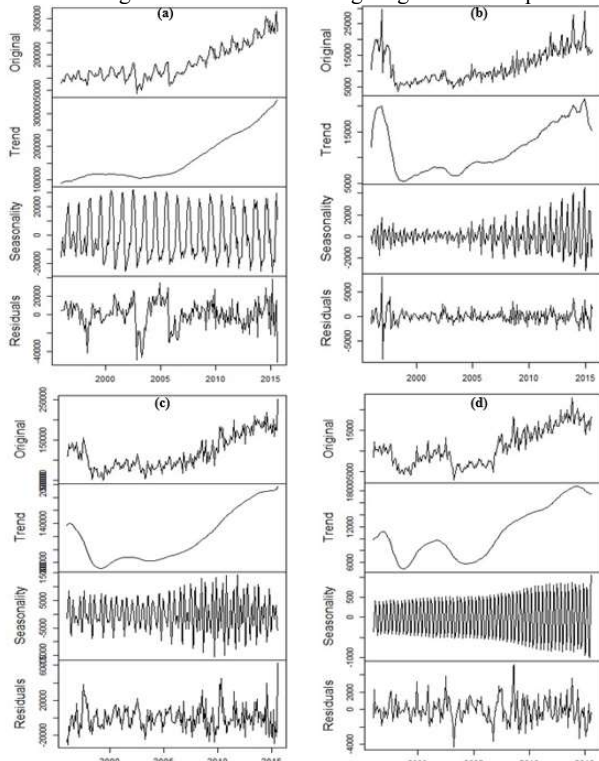


Figure 4. Tourist Series and Reconstructed Components on Cases Number of Foreign Visitor Arrivals Through Ngurah Rai Airport (a), Kualanamu (b), Soekarno-Hatta (c), and Juanda (d).

When the trend, seasonality, and noise components are successfully separated, the next step is to make forecasting on each component except for the noise component. The forecast values generated by the SSA model of each component consists of forecast values for in-sample data and forecast values for out-sample data. The SSA model used in forecasting trend components for each entrance can be written as follows:

Table 3. Eigentriples and Related Components in the Four Entrances

Entrance	Eigentruple		
	Trend	Seasonal	Noise
Ngurah Rai	1-2, 5	3-4, 9-10, 15-18	6-8, 11-14, 19-50
Kualanamu	1-3, 8, 11	4-7, 9-10	12-50
Soekarno-Hatta	1, 2, 3	6-7, 9-10	4-5, 8, 11-50
Juanda	1-5	6-7	8-50

1. Ngurah Rai Airport

a. Model for in-sample data

i. SSA generates reconstructed trend, namely  $\tilde{y}_0, \tilde{y}_1, \dots, \tilde{y}_{235}$ .

ii.  $\hat{y}_i = \tilde{y}_i$ , for  $i = 0, 1, 2, \dots, 94$ .

iii.  $\hat{y}_i = 0,04196\tilde{y}_{i-1} + 0,042331\tilde{y}_{i-2} + \dots$

$+ 0,00364\tilde{y}_{i-95}$ ,  
for  $i = 95, 96, 97, \dots, 235$ .

b. Model for out-sample data

i. SSA generates reconstructed trend, namely  $\tilde{y}_0, \tilde{y}_1, \dots, \tilde{y}_{235}$ .

ii. SSA generates reconstructed trend, namely  $\tilde{y}_0, \tilde{y}_1, \dots, \tilde{y}_{235}$ .

iii.  $\hat{y}_{235+k} = 0,04196\tilde{y}_{235+k-1} + 0,042331\tilde{y}_{235+k-2} + \dots$   
 $+ 0,00364\tilde{y}_{235+k-95}$ ,

where

$$\tilde{y}_{235+k-j} = \begin{cases} \tilde{y}_{235+k-j}^T, & \text{for } j \geq k \\ \hat{y}_{235+k-j}^T, & \text{for } j < k \end{cases}$$

2. Kualanamu Airport

a. Model for in-sample data

i. SSA generates reconstructed trend, namely  $\tilde{y}_0, \tilde{y}_1, \dots, \tilde{y}_{235}$ .

ii.  $\hat{y}_i = \tilde{y}_i$ , for  $i = 0, 1, 2, \dots, 58$ .

iii.  $\hat{y}_i = 0,186104\tilde{y}_{i-1} + 0,183775\tilde{y}_{i-2} + \dots$

$+ 0,033577\tilde{y}_{i-59}$ ,  
for  $i = 59, 60, 61, \dots, 235$ .

b. Model for out-sample data

i. SSA generates reconstructed trend, namely  $\tilde{y}_0, \tilde{y}_1, \dots, \tilde{y}_{235}$ .

ii.  $\hat{y}_{235+k} = 0,186104\tilde{y}_{235+k-1} + 0,183775\tilde{y}_{235+k-2} + \dots$

$+ 0,033577\tilde{y}_{235+k-59}$ ,

where

$$\tilde{y}_{235+k-j}^T = \begin{cases} \tilde{y}_{235+k-j}^T, & \text{for } j \geq k \\ \hat{y}_{235+k-j}^T, & \text{for } j < k \end{cases}$$

### 3. Soekarno-Hatta Airport

#### a. Model for in-sample data

- i. SSA generates reconstructed trend, namely

$$\tilde{y}_0^T, \tilde{y}_1^T, \dots, \tilde{y}_{235}^T.$$

- ii.  $\hat{y}_i^T = \tilde{y}_i^T$ , for  $i = 0, 1, 2, \dots, 70$ .

$$\hat{y}_i^T = 0,065025\tilde{y}_{i-1}^T + 0,063867\tilde{y}_{i-2}^T + \dots$$

$$+ 0,00298\tilde{y}_{i-71}^T,$$

for  $i = 71, 72, 73, \dots, 235$ .

#### b. Model for out-sample data

- i. SSA generates reconstructed trend, namely

$$\tilde{y}_0^T, \tilde{y}_1^T, \dots, \tilde{y}_{235}^T.$$

$$\hat{y}_{235+k}^T = 0,065025\tilde{y}_{235+k-1}^T + 0,063867\tilde{y}_{235+k-2}^T + \dots$$

$$+ 0,00298\tilde{y}_{235+k-71}^T$$

where

$$\tilde{y}_{235+k-j}^T = \begin{cases} \tilde{y}_{235+k-j}^T, & \text{for } j \geq k \\ \hat{y}_{235+k-j}^T, & \text{for } j < k \end{cases}$$

### 4. Juanda Airport

#### a. Model for in-sample data

- i. SSA generates reconstructed trend, namely

$$\tilde{y}_0^T, \tilde{y}_1^T, \dots, \tilde{y}_{235}^T.$$

- ii.  $\hat{y}_i^T = \tilde{y}_i^T$ , for  $i = 0, 1, 2, \dots, 106$ .

$$\hat{y}_i^T = 0,08485\tilde{y}_{i-1}^T + 0,082533\tilde{y}_{i-2}^T + \dots$$

$$+ 0,037337\tilde{y}_{i-107}^T,$$

for  $i = 107, 108, 109, \dots, 235$ .

#### b. Model for out-sample data

- i. SSA generates reconstructed trend, namely

$$\tilde{y}_0^T, \tilde{y}_1^T, \dots, \tilde{y}_{235}^T.$$

$$\hat{y}_{235+k}^T = 0,08485\tilde{y}_{235+k-1}^T + 0,082533\tilde{y}_{235+k-2}^T + \dots$$

$$+ 0,037337\tilde{y}_{235+k-107}^T$$

where

$$\tilde{y}_{235+k-j}^T = \begin{cases} \tilde{y}_{235+k-j}^T, & \text{for } j \geq k \\ \hat{y}_{235+k-j}^T, & \text{for } j < k \end{cases}$$

The SSA model used in forecasting seasonal components for each entrance can be written as follows:

#### 1. Ngurah Rai Airport

##### a. Model for in-sample data

- i. SSA generates reconstructed seasonal, namely

$$\tilde{y}_0^S, \tilde{y}_1^S, \tilde{y}_2^S, \dots, \tilde{y}_{235}^S.$$

- ii.  $\hat{y}_i^S = \tilde{y}_i^S$ , for  $i = 0, 1, 2, \dots, 94$ .

$$\hat{y}_i^S = -0,00667\tilde{y}_{i-1}^S - 0,02489\tilde{y}_{i-2}^S + \dots$$

$$+ 0,022311\tilde{y}_{i-95}^S,$$

for  $95, 96, 97, \dots, 235$ .

##### b. Model for out-sample data

- i. SSA generates reconstructed seasonal, namely

$$\tilde{y}_0^S, \tilde{y}_1^S, \tilde{y}_2^S, \dots, \tilde{y}_{235}^S.$$

$$\hat{y}_{235+k}^S = -0,00667\tilde{y}_{235+k-1}^S - 0,02489\tilde{y}_{235+k-2}^S + \dots$$

$$+ 0,022311\tilde{y}_{235+k-95}^S$$

where

$$\tilde{y}_{235+k-j}^S = \begin{cases} \tilde{y}_{235+k-j}^S, & \text{for } j \geq k \\ \hat{y}_{235+k-j}^S, & \text{for } j < k \end{cases}$$

### 2. Kuala Lumpur Airport

#### a. Model for in-sample data

- i. SSA generates reconstructed seasonal, namely

$$\tilde{y}_0^S, \tilde{y}_1^S, \tilde{y}_2^S, \dots, \tilde{y}_{235}^S.$$

- ii.  $\hat{y}_i^S = \tilde{y}_i^S$ , for  $i = 0, 1, 2, \dots, 58$ .

$$\hat{y}_i^S = 0,011502\tilde{y}_{i-1}^S - 0,08198\tilde{y}_{i-2}^S + \dots$$

$$+ 0,032616\tilde{y}_{i-59}^S,$$

for  $59, 60, 61, \dots, 235$ .

#### b. Model for out-sample data

- i. SSA generates reconstructed seasonal, namely

$$\tilde{y}_0^S, \tilde{y}_1^S, \tilde{y}_2^S, \dots, \tilde{y}_{235}^S.$$

$$\hat{y}_{235+k}^S = 0,011502\tilde{y}_{235+k-1}^S - 0,08198\tilde{y}_{235+k-2}^S + \dots$$

$$+ 0,032616\tilde{y}_{235+k-59}^S$$

,where

$$\tilde{y}_{235+k-j}^S = \begin{cases} \tilde{y}_{235+k-j}^S, & \text{for } j \geq k \\ \hat{y}_{235+k-j}^S, & \text{for } j < k \end{cases}$$

### 3. Soekarno-Hatta Airport

#### a. Model for in-sample data

- i. SSA generates reconstructed seasonal, namely

$$\tilde{y}_0^S, \tilde{y}_1^S, \tilde{y}_2^S, \dots, \tilde{y}_{235}^S.$$

- ii.  $\hat{y}_i^S = \tilde{y}_i^S$ , for  $i = 0, 1, 2, \dots, 70$ .

$$\hat{y}_i^S = 0,005071\tilde{y}_{i-1}^S - 0,03435\tilde{y}_{i-2}^S + \dots$$

$$+ 0,028249\tilde{y}_{i-71}^S,$$

for  $i = 71, 72, 73, \dots, 235$ .

b. Model for out-sample data

- i. SSA generates reconstructed seasonal, namely  $\tilde{y}_0^S, \tilde{y}_1^S, \tilde{y}_2^S, \dots, \tilde{y}_{235}^S$ .
- ii.  $\hat{y}_{235+k}^S = 0,005071\tilde{y}_{235+k-1}^S - 0,03435\tilde{y}_{235+k-2}^S + \dots + 0,028249\tilde{y}_{235+k-71}^S$

where

$$\hat{y}_{235+k-j}^S = \begin{cases} \tilde{y}_{235+k-j}^S, & \text{for } j \geq k \\ \hat{y}_{235+k-j}^S, & \text{for } j < k \end{cases}$$

4. Juanda Airport

a. Model for in-sample data

- i. SSA generates reconstructed seasonal, namely  $\tilde{y}_0^S, \tilde{y}_1^S, \tilde{y}_2^S, \dots, \tilde{y}_{235}^S$ .
- ii.  $\hat{y}_i^S = \tilde{y}_i^S$ , for  $i = 0, 1, 2, \dots, 106$ .
- iii.  $\hat{y}_i^S = 0,008869\tilde{y}_{i-1}^S - 0,02072\tilde{y}_{i-2}^S + \dots - 0,00416\tilde{y}_{i-107}^S$ , for  $i = 107, 108, 109, \dots, 235$ .

b. Model for out-sample data

- i. SSA generates reconstructed seasonal, namely  $\tilde{y}_0^S, \tilde{y}_1^S, \tilde{y}_2^S, \dots, \tilde{y}_{235}^S$ .
- ii.  $\hat{y}_{235+k}^S = 0,008869\tilde{y}_{235+k-1}^S - 0,02072\tilde{y}_{235+k-2}^S + \dots - 0,00416\tilde{y}_{235+k-107}^S$

where

$$\hat{y}_{235+k-j}^S = \begin{cases} \tilde{y}_{235+k-j}^S, & \text{for } j \geq k \\ \hat{y}_{235+k-j}^S, & \text{for } j < k \end{cases}$$

The SSA model used to obtain forecast values for in-sample data and out-sample data as a whole can be written as follows:

1. In-sample data

$$\hat{y}_i = \hat{y}_i^T + \hat{y}_i^S, \text{ for } i = 0, 1, 2, \dots, 235.$$

2. Out-sample data

$$\hat{y}_{235+k} = \hat{y}_{235+k}^T + \hat{y}_{235+k}^S, \text{ for } k = 1, 2, \dots, 12.$$

The next step is the in-sample data and the out-sample data is predicted based on the model obtained. The result of forecasting the number of foreign tourist arrivals using the SSA model in the four entrances is shown in Figure 6.

B. Forecasting Using ARIMA Method

The Box-Jenkins procedure for obtaining the best ARIMA model in the case of the number of foreign tourist arrivals through four entries consists of four stages: model identification, parameter estimation, diagnostic checking, and application of model for forecasting. The first step in the identification stage is to check the stationarity of data both in

variance and in the mean. If data are not stationary on its variance, then the data need to be transformed. The type of transformation used is the Box-Cox transformation. Based on the Box-Cox plot of data on the number of foreign tourist arrivals at four entrances, it is known that the estimated values obtained for  $\lambda$  at four entrances are 0.57, -0.04, 0.27, and 0.36, respectively. The data has not been stationary in variance because the confidence interval of the estimated value for  $\lambda$  at each entrance does not contain grades one so transformations are necessary. In-sample data is also not stationary in the average because the autocorrelation value is close to one and tends to drop slowly. If data are not stationary in the mean, then differencing operations need to be performed on the transformed data. Because the autocorrelation value in non-seasonal lag tends to fall slowly, then the first non-seasonal differencing operation ( $d=1$ ) needs to be performed on the transformed data. The results of the first non-seasonal differencing operations on the transformed data show that the data has not been stationary in mean for seasonal 12 because the autocorrelation value at seasonal lag 12, 24, 36, and 48 tends to decrease slowly. Since the autocorrelation value in the seasonal lag tends to fall slowly, the first differencing operation of seasonal 12 needs to be performed on the transformed data and the first non-seasonal differencing operations. Figure 5 shows the ACF and PACF diagrams of transformed data, the first non-seasonal differencing, and the first differencing for seasonal 12 on the case of number of foreign visitor arrivals through Ngurah Rai Airport.

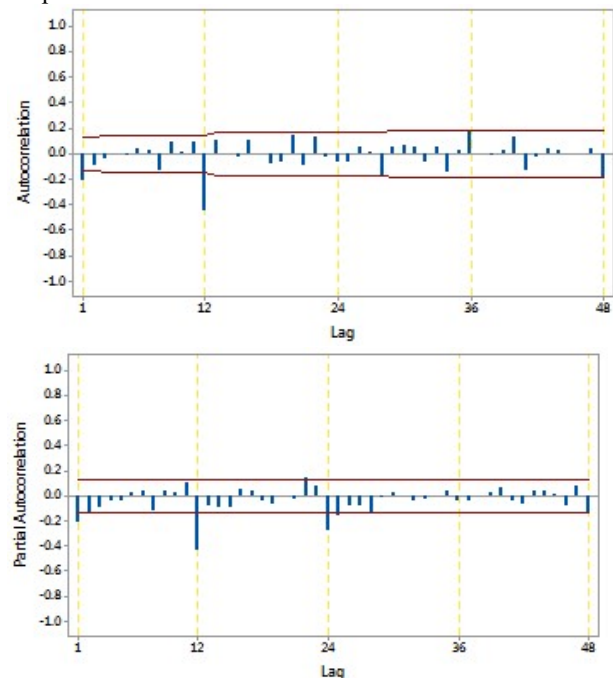


Figure 5. The ACF and PACF Diagrams of Transformed Data, the First Non-Seasonal Differencing, and the First Differencing for Seasonal 12 on Case of Number of Foreign Visitor Arrivals Through Ngurah Rai Airport.

After the transformed data has been stationary in variance and averaging, the next step is to determine the interim ARIMA model based on significant lags on the ACF and PACF diagrams. The best ARIMA model was obtained based on the model's goodness criteria and outlier handling. Outlier handling is necessary because the residuals of the best ARIMA model at two entrances, namely the Ngurah Rai and Kualanamu gates are not normally distributed. Table 4 presents the best ARIMA model obtained in the four main arrival gates.

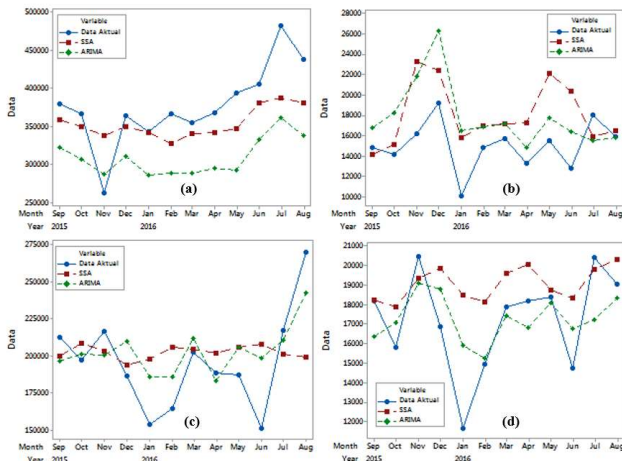
**Table 4.** Best ARIMA Models are Obtained at Each Entrance

Entrance	ARIMA Models
Ngurah Rai	$(0,1,1)(0,1,1)^{12}$
Kualanamu	$(0,1,1)(0,1,1)^{12}$
Soekarno-Hatta	$(0,1,[1,35])(1,1,1)^{12}$
Juanda	$([1,2,35],1,0)(0,1,1)^{12}$

Since the best ARIMA model in the four main arrival gates has met all test both parameter significance tests and model conformity tests, the model can be used for forecasting out-sample data for the next 12 months. The result of forecasting the number of foreign tourist arrivals using the ARIMA model in the f entrances is shown in Figure 6.

### C. Comparison of SSA and ARIMA Methods

Comparison of forecasting results on each main arrival gate using SSA and ARIMA methods is shown in Figure 6. Based on Figure 6, it is known that the result of forecasting the out-sample data at Ngurah Rai Airport using SSA method is closer to the actual data compared with the result of forecasting using ARIMA method.



**Figure 6.** Comparison of Forecasting Results Using SSA and ARIMA Methods with Actual Data on Cases Number of Foreign Visitor Arrivals Through Ngurah Rai Airport (a), Kualanamu (b), Soekarno-Hatta (c), and Juanda (d).

The SSA method produces better prediction accuracy compared to the ARIMA method for each forecasting period. Thus the SSA method is the best method that provides accurate forecasting results at Ngurah Rai Airport. Measuring the accuracy of the model in the four main arrival gates is not sufficiently visual, but it is necessary to consider the MAPE

value of the best model on the data from the four entrances. The accuracy of forecast results in in-sample data and out-sample data is measured by the MAPE value. If the MAPE value in the in-sample data is smaller, then the model is better used for forecasting. If the MAPE value in the out-sample data is smaller, then the result of the prediction obtained from the model is more accurate. The accuracy of forecasting the number of foreign tourist arrivals in four entrances using SSA and ARIMA methods can be seen in Table 5. These values can reinforce the temporary conclusions obtained visually.

**Table 5.** Comparison of MAPE Values from Forecasting Results in Out-Sampled Data Using SSA and ARIMA Methods

Entrance	MAPE (%)			
	SSA		ARIMA	
	In-Sample (%)	Out-Sample (%)	In-Sample (%)	Out-Sample (%)
Ngurah Rai	8,0	9,6	7,1	18,6
Kualanamu	11,0	24,9	11,8	22,4
Soekarno-Hatta	10,2	13,8	9,9	10,5
Juanda	11,5	14,3	11,4	9,9

The best forecasting method is the method that produces the smallest MAPE value compared to the MAPE value generated from other forecasting methods. Based on Table 5, it is known that ARIMA method is the best method that gives more accurate prediction result in three main door compared with SSA method. The three main gates are Kualanamu Airport, Soekarno-Hatta Airport, and Juanda Airport. While the SSA method produces better prediction accuracy compared with ARIMA method at main gate of Ngurah Rai.

## VIII. CONCLUSIONS

The conclusion that can be obtained based on research conducted is the SSA method produces better prediction accuracy compared with ARIMA method for each period of forecasting at the main door of Ngurah Rai. The MAPE value in the out-sampled data obtained from the forecasting result of the number of foreign tourist arrivals at the door of the Ngurah Rai is 9,6%. The result of forecasting on out-sample data using SSA method is closer to actual data when compared with ARIMA method. The results of forecasting on out-sample data in the other three doors using ARIMA method closer to the actual data when compared with SSA method. The three main gates are Kualanamu Airport, Soekarno-Hatta Airport, and Juanda Airport. The MAPE value in the out-sampled data obtained from the forecasting result of the number of foreign tourist arrivals at the door of the Kualanamu is 22,4%. The MAPE value in the out-sampled data obtained from the forecasting result of the number of foreign tourist arrivals at the door of the Soekarno-Hatta is 10,5%. The MAPE value in the out-sampled data obtained from the forecasting result of the number of foreign tourist arrivals at the door of the Juanda is 9,9%.



REFERENCES

- [1] N. Golyandina, V. Nekrutkin and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*, Chapman & Hall/CRC, 2001.
- [2] H. Hassani, S. Heravi and A. Zhigljavsky, "Forecasting European Industrial production with Singular Spectrum Analysis," *International Journal of Forecasting*, vol. 25, pp. 103-118, 2009.
- [3] G. Darmawan, T. Hendrawati and R. Arisanti, *Model Auto Singular Spectrum untuk Meramalkan Kejadian Banjir di Bandung dan Sekitarnya*, Yogyakarta: UNY, 2015.
- [4] H. Hassani, A. Webster, E. S. Silva and S. Heravi, "Forecasting U.S. Tourist Arrivals Using Optimal Singular Spectrum Analysis," *Tourism Management*, vol. 46, pp. 322-335, 2015.
- [5] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series.*, New York: Springer, 2013.
- [6] J. L. Vile, J. W. Gillard, P. R. Harper and V. A. Knight, "Predicting Ambulance Demand Using Singular Spectrum Analysis," *Palgrave Macmillan UK*, vol. 63, pp. 1556-1565, 2012.
- [7] R. J. Hyndman, A. B. Koehler, J. K. Ord and R. D. Snyder, "Monitoring Processes with Changing Variances," *International Journal of Forecasting*, vol. 25, no. 3, pp. 518-525, 2009.
- [8] Aswi and Sukarna, *Analisis Deret Waktu*, Makasar: Andira Publisher, 2006.
- [9] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transaction on Automatic Control*, pp. 716-723, 1974.
- [10] W. W. S. Wei, *Time Series Analysis*, New York: Addison Wesley, 2006.
- [11] Badan-Pusat-Statistik-[BPS], *Jumlah Kedatangan Wisatawan Mancanegara per bulan ke Indonesia Menurut Pintu Masuk*, Jakarta: Badan Pusat Statistik, , 2016.
- [12] E. Simanjuntak, M. Aulia, A. B. Wicaksono, B. A. Anwar, Z. H. Putri and C. Dwiriastya, *Konsep dan Arahan Kebijakan Daerah Provinsi Dalam Pengembangan Infrastruktur Wilayah Pulau Jawa-Bali*, Pusat Perencanaan Infrastruktur Pekerjaan Umum Dan Perumahan Rakyat, 2016.
- [13] N. Golyandina and A. Korobeynikov, *Basic Singular Spectrum Analysis and Forecasting with R.*, St. Petersburg: St. Petersburg State University, 2013.