Hybrid Model of Singular Spectrum Analysis and ARIMA for Seasonal Time Series Data

Gumgum Darmawan[1,2,*], Dedi Rosadi[1], Budi N Ruchjana[2]

[1]Gadjah Mada University, Yogyakarta, Indonesia
[2]Padjadjaran University, Bandung, Indonesia

*Corresponding Author
Email: gumgum.darmawan@gmail.ugm.ac.id*,
dedirosadi@gadjahmada.edu, budi.nurani@unpad.ac.id

ABSTRACT

Hybrid models between Singular Spectrum Analysis (SSA) and Autoregressive Integrated Moving Average (ARIMA) have been developed by several researchers. In the SSA-ARIMA hybrid model, SSA is used in the decomposition and reconstruction process, while forecasting is done through the ARIMA model. In this paper, hybrid SSA-ARIMA uses two auto grouping models. The purpose of this paper is to analyze seasonal data using the SSA-ARIMA hybrid by auto grouping. The first model namely the Alexandrov method and the second method is alternative auto grouping with long memory approach. The two hybrid models were tested for two types of seasonal pattern, multiplicative and additive seasonal time series data. The analysis results using both methods give accurate result; as seen from the MAPE generated the 12 observations for future, the value is below 5%. For additive seasonal pattern, The hybrid SSA-ARIMA method with Alexandrov auto grouping is more accurate (MAPE= 0.13%) than the hybrid SSA-ARIMA method with Alternative method but for multiplicative seasonal pattern the hybrid SSA-ARIMA with alternative auto grouping is more accurate (MAPE = 3.63%) than the hybrid SSA-ARIMA method with Alexandrov method.

Keywords: ARIMA; Automatic Grouping; Long Memory Effect; Seasonal Pattern, Singular Spectrum Analysis

INTRODUCTION

Singular Spectrum Analysis (SSA) is a relatively new non-parametric method that has proved its capability in various time series types. Solving all these problems correspond to the so-called basic capabilities of SSA. Besides, the method has several extensions. First, the multivariate version of the method permits the simultaneous expansion of several time series data; see, for example, [1]. Second, the SSA ideas lead to several forecasting procedures for time series; see [2]. Third, SSA has been utilized for change-point detection in time series. The SSA technique has been used as a filtering method in [3]. Fifth, a family of the causality test based on the multivariate SSA technique has been introduced in [4]. Sixth, SSA can be applied for missing value imputation [5].

SSA can be applied in various disciplines, from mathematics and physics to economics and financial mathematics, meteorology and oceanography, to social sciences.

For instance, in climatology ([6], [7], [8]) and biomedical data time series analysis [9]. Hybrid modeling of SSA in time series data has been carried out by many researchers. The hybrid model is carried out so that the advantages of two or more models make a positive contribution to the forecasting results.

SSA hybrid model with other time series models includes ARIMA, Neural network, ARIMAX, PAR, VARIMAX, and others. [10], performed hybrid SSA with Neural Network.[11] perform the hybrid SSA-Algorithm Firefly-BP Neural Network process. [12] carried out a hybrid SSA model with ARMAX. [13] Combining the SSA model with PAR(p), this model was applied to wind speed data.[14], built the SSA-VARIMAX hybrid model and used it for climate data. The ARIMA model is often used as a comparison for the SSA model, such as [15], comparing SSA, ARIMA, and other time series models for tourism cases in various countries in Europe. The result has indicated that there is no good time series model for all tourism data. [16] compared SSA and ARIMA for predicting ambulance demand.

The SSA-ARIMA hybrid model studied by [17] was applied to the annual Runoff data. [18], the SSA-ARIMA hybrid model was compared with the basic SSA and ARIMA models. The result showed that the SSA-ARIMA hybrid model was the most accurate. However, many of these papers do not discuss specific data forms (e.g., seasonal patterns), so we consider it necessary to examine this hybrid model for seasonal data. In this study, the SSA and the ARIMA were employed collectively to forecast two types of time series data. Both models run to get fast and accurate computation. In SSA, there are two methods of automatic grouping (Alexandrov and Alternative). The forecasting performance of the hybrid SSA-ARIMA model was compared between the two methods (alternative vs. Alexandrov). This paper contributes to the analysis of the seasonal patterns (additive and multiplicative) by the SSA-ARIMA hybrid. The purpose of this paper was to analyze seasonal data using the SSA-ARIMA hybrid by auto grouping for two types of seasonal patterns.

This paper was organized as follows: The current section was an introduction where we briefly outlined the use of SSA and introduced our study. In the next section, the methods section, we described the detailed methodology of SSA and ARIMA, briefly outlined forecasting using a linear recurrent formula, identification of fractional differencing parameter, identification of hidden periodicities based on Periodogram and Automatic grouping on Alexandrov Method ([19], [20]) also alternative Automatic grouping [21]. This section also included a proposed algorithm for automatic hybrid SSA-ARIMA. In the results and discussion section, we demonstrated the abilities of hybrid SSA-ARIMA in real-time series data. In this part, we also investigated three types of time series data: Seasonal with no trend, multiplicative seasonal with the trend, and Additive seasonal with the trend. This section also discussed the comparison result between hybrid SSA-ARIMA with the Alexandrov method and hybrid SSA-ARIMA with an alternative method for real data analysis.

METHODS

Singular Spectrum Analysis

The (non-parametric) SSA method has received a fair amount of attention in the literature. The first phase of SSA is the decomposition, where the time series are broken down into four components: trend, seasonal, cyclical, and noise. This phase consists of the Embedding and Singular Value Decomposition steps. The second phase, namely the Reconstruction phase, consists of Grouping and Diagonal Average process. The

forecasting process can be done once the four stages have been completed. For the completeness of presentation of our method, we presented the complete phase of the SSA algorithm in the following section.

Embedding

The embedding step will transform one-dimensional time series $X = (x_1, x_2, ..., x_T)$ into multi-dimensional series $X_1, X_2, ..., X_K$ with vectors $X = (X_i, X_{i+1}, X_{i+2}, .., X_{i+L-1})^T \in R^L$, where $i = 1, 2, ..., K$, $K = T - L + 1$. The parameter window length L defines the embedding process, where $2 \leq L \leq T - 1$ [22]. If we need to emphasize the size (dimension) of the vectors X_i, then we shall call them L-lagged vectors. The L-trajectory matrix (or simply the trajectory matrix) of the series X is defined as

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

The lagged vectors X_i are the columns of the trajectory matrix **X**. Both the rows and column of **X** are sub-series of the original series. The (i,j) element of matrix **X** is $x_{ij} = x_{i+j-1}$ which yields that X has equal elements on the 'antidiagonals' $i+j$=const. Hence the trajectory matrix is a *Hankel matrix*.

Singular Value Decomposition

The second step, the SVD step, makes the singular value decomposition of the trajectory matrix **X** and represents it as a sum of rank-one bi-orthogonal elementary matrices. Set $S = XX^T$ and denoted by $\lambda_1, \lambda_2, ..., \lambda_L$ the eigenvalues of S taken in the decreasing order of magnitude $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0)$ and by $U_1, U_2, ..., U_L$ the orthonormal system of the eigenvectors of the matrix S corresponding to these eigenvalues.

$$d = max\{i, such\ that\ \lambda_i > 0\} = rank\ X$$

If we denote $V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}$, then the SVD of the trajectory matrix can be written as

$$X = X_1 + X_2 + \cdots + X_d,$$

where eigenvector U_i, Eigenvalues $\sqrt{\lambda_i}$ form matrix $V_i^T X$. The three elements of SVD forming are called eigen triple.

Grouping

The purpose of this step is to appropriately identify the trend, the oscillatory components with different periods and noise. This step can be skipped if one does not want to extract hidden information by regrouping and filtering components precisely. The grouping procedure partitions the set of indices *1,2,...., L* into m disjoint subsets $I = I_1, I_2, ..., I_m$, so the elementary matrix in equation (2) is regrouped into m groups. Let $I = \{i_1, i_2, ..., i_p\}$. Then the resultant matrix X_i corresponding to the group i is defined as $X_i = X_{i_1} + X_{i_2} + ... + X_{i_p}$. The matrices are computed for $I_1, I_2, ...I_m$, and substituted into equation (2) to obtain the new expansion.

The grouping process is the phase when *the LxK matrix* is grouped into several sub-groups, namely trend patterns, seasonal or periodic, and noise patterns. Here, in this paper, the patterns are identified by Fourier series analysis and long-memory analysis. Fourier series analysis is

used to identify a seasonal pattern, and long memory series analysis is used to identify the differencing parameter of data. We use the GPH method [23] to identify the differencing parameter of time series.

Diagonal Averaging

The next step in Basic SSA transforms each resultant matrix of the grouped decomposition (3) into a new one-dimensional series of length *N* and is called *diagonal averaging.* Let **Y** denote a matrix with orde (*L*x*K*), with the elements $y_{ij}, 1 \le i \le L, 1 \le j \le K$, and define $L^* = min(L, K)$, $K^* = max(L, K)$, and *T=L+K-1*. Let $y_{ij}^* = y_{ij}$ if L<K and $y_{ij}^* = y_{ji}$ otherwise. Diagonal averaging transfers matrix **Y** into a series $y_1, y_2, ..., y_N$ by the formula.

$$
y_k = \begin{cases}
\dfrac{1}{k} \sum_{m=1}^{k} y_{m,k-m+1}^* & for\ 1 \le k < L \\[2em]
\dfrac{1}{L} \sum_{m=1}^{L^*} y_{m,k-m+1}^* & for\ L^* \le k < K^* \\[2em]
\dfrac{1}{T-k+1} \sum_{m=k-K^*+1}^{T-K^*+1} y_{m,k-m+1}^* & for\ K^* < k \le T
\end{cases}
\tag{2}
$$

Using equation (3), the resultant matrix X_{I_i}, $i = 1,2,...,m$ will form the series $\tilde{Y}^{(k)} = \left(\tilde{y}_1^{(k)}, \tilde{y}_2^{(k)}, ..., \tilde{y}_t^{(k)} \right)$; therefore, the original sequence will be decomposed into the sum of the m series.

Forecasting Using Linear Recurrent Formula

The forecasting method used in this paper is the Linear Recurrent Formula (LRF). In the above notation, the *recurrent forecasting* algorithm (briefly, R-Forecasting) can be formulated as follows.

1. The time series $Y_{T+M} = (y_1, y_{2,...} y_{T+M})$ is defined by

$$
y_i = \begin{cases}
\tilde{x}_i & for\ i = 1,2,....,T \\
\sum_{j=1}^{L-1} a_j y_{i-j} & for\ i = T+1,...,T+M
\end{cases}
\tag{3}
$$

2. The numbers $y_{T+1}, y_{T+2,...} y_{T+M}$ form the M terms of the recurrent forecast

Thus, the R-forecasting is formed by the direct use of the LRF with coefficients $\{a_j, j = 1, ..., L-1\}$.

Autoregressive Integrating Moving Average

A time series $\{x_t\}$ is said to follow the Autoregressive Integrated Moving Average model if the differenc*d* that is $W_t = \nabla^d x_t$ then this process is ARMA processes. If $W_{t\ w_t}$ is ARMA(*p,q*), then $\{X_t\}$ is ARIMA(*p,d,q*) process. The mathematic model of ARIMA(*p,d,q*) is as follow;

$$\emptyset(B)(1-B)^d(X_t - \mu) = \theta(B)a_t \tag{4}$$

Where

t = index of observation

d = differencing parameter (1,2,3,..)

μ = mean of observation

$a_t \sim IID(0, \sigma_a^2)$

$\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$ is a polynomial of AR(p)

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p$ is polynomial of MA(q)

Identification of Fraction differencing parameter

The fractional differencing operator is described as a limitless binomial series growth in powers of the backward-shift operator [24]. One of the methods to estimate the value of the fractional differencing parameter is the Geweke and Porter-Hudak method (GPH). According to the GPH method, the differencing parameter of time series data can be estimated by the least square method of the ARFIMA (autoregressive fractionally integrated moving average) model. The GPH procedure can be described as follows. For the ARFIMA model given in (4), let $W_t = (1-B)^\delta Z_t$ and let $f_W(\omega)$ and $f_Z(\omega)$ be the spectral density function of $\{W_t\}$ and $\{Z_t\}$, respectively. Then,

$$f_z(\omega_j) = f_W(\omega_j)\left\{2sin\left(\omega_j/2\right)\right\}^{-2\delta}, \omega_j \in (-\pi, \pi) \tag{5}$$

where;

$f_z(\omega_j) = \dfrac{\sigma_a^2}{2\pi}\left|\dfrac{\theta_q(e^{-i\omega_j})}{\phi_p(e^{-i\omega_j})}\right|^2$, $\omega_j \in (-\pi, \pi)$ is the spectral density of a regular ARMA(p,q) model.

$\ln[f_z(\omega_j)] = \delta \ln\left\{2sin\left(\omega_j/2\right)\right\}^{-2} + ln[f_W(\omega_j)],$

$\ln[f_z(\omega_j)] = ln[f_W(0)] + \delta \ln\left\{2sin\left(\omega_j/2\right)\right\}^{-2} + ln\left[\dfrac{f_W(\omega_j)}{f_W(0)}\right].$

Replacing ω_j by the Fourier frequencies $\omega_j = \dfrac{2\pi j}{T}$ and adding $ln[I_Z(\omega_j)]$ to both sides, the periodogram$\{Z_t\}$ gives

$\ln[I_Z(\omega_j)] = ln[f_W(0)] + \delta \ln\left\{2sin\left(\omega_j/2\right)\right\}^{-2} + ln\left[\dfrac{I_Z(\omega_j)}{f_Z(\omega_j)}\right] + ln\left[\dfrac{f_W(\omega_j)}{f_W(0)}\right].$

For ω_j closes to zero, i.e. for $j = 1,2,\ldots,T/2$ such that $m/T \to \infty$ we have $ln\left[\dfrac{f_W(\omega_j)}{f_W(0)}\right] \approx 0,$ thus,

$$Y_j = \beta_0 + \beta_1 X_j + e_j, \quad j = 1,2,\ldots,m \tag{6}$$

$Y_j = ln[I_Z(\omega_j)], \quad X_j = ln\left(\dfrac{1}{4sin^2\left(\omega_j/2\right)}\right)$

$$\hat{\beta}_0 = ln[f_W(0)], I_Z = \frac{1}{2\pi}\left[\gamma_0 + \sum_{t=0}^{T-1}\gamma_t cos(\omega_j.t)\right]$$

The least-square estimator of long memory coefficient is given by

$$\hat{\beta}_1 = \hat{\delta} = \frac{\sum_{j=1}^{m}(X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{j=1}^{m}(X_j - \bar{X})^2}; m = g(T) = T^\alpha, 0 < \alpha < 1 \tag{7}$$

Identification of hidden Periodicities based on Periodogram

Identification of hidden periodicities based on Periodogram analysis can be described as follows;

1. Fit the data to equation (7):

$$X_t = \sum_{k=0}^{[T/2]}(a_k cos\omega_k t + b_k sin\omega_k t), \tag{8}$$

where k = 0,1,…,[T/2] and ω_k is Fourier frequency determined by $\omega_k = \frac{2\pi k}{T}$.

2. Compute a_k and b_k by (8) and (9) :

$$a_k = \begin{cases} \frac{1}{T}\sum_{t=1}^{T} x_t cos\omega_k t & for\ k = 0\ and\ k = \frac{T}{2}\ if\ T\ even \\ \frac{2}{T}\sum_{t=1}^{T} x_t cos\omega_k t & for\ k = 1,2,…,\frac{(T-1)}{2} \end{cases} \tag{9}$$

$$b_t = \frac{2}{T}\sum_{t=1}^{T} x_t sin\omega_k t, \qquad k = 1,2,…\frac{(T-1)}{2} \tag{10}$$

3. Compute the values of ordinate $I(\omega_k)$ by formula (10) ;

$$I(\omega_k) = \begin{cases} Ta_0^2 & k = 0 \\ \frac{T}{2}(a_k^2 + b_k^2) & k = 1,2,…,\left[\frac{T-1}{2}\right] \\ Ta_{T/2}^2 & k = \left[\frac{T}{2}\right] \end{cases} \tag{11}$$

4. Test for significance of every Fourier frequency as follow;

H_0: $\alpha = \beta = 0$ (data do not have a seasonal pattern)

H_1: $\alpha \neq 0$ or $\beta \neq 0$ (data have a seasonal pattern)

Statistic test:

$$T = \frac{I^{(1)}(\omega_{(1)})}{\sum_{k=1}^{[T/2]} I(\omega_k)} \tag{12}$$

where,

$I^{(1)}(\omega_{(1)})$: maximum ordinate of periodogram of Fourier frequency

$I(\omega_k)$: Value of periodogram ordinate at k-th' Fourier frequency.

Test criteria:

Reject H_0 if T >g_α with α = significant level. The value of g_α can be seen in table Fisher [25].

Automatic Grouping in SSA

In this section, we explain two methods SSA automatic grouping based on trend extraction based namely Alexandrov, and SSA automatic grouping based on long memory approach, namely Alternative method. First, we introduce the periodogram of a time series. Let us consider the Fourier representation of the elements of a time series $X = (x_1, x_2, \ldots, x_T)$ of length T.

$$X_t = C_0 + \sum_{1 \le k \le \frac{N-1}{2}} \left(c_k cos\left(\frac{2\pi t k}{T}\right) \right) + s_k cos\left(\frac{2\pi t k}{T}\right) + (-1)^t C_{T/2}, \qquad (13)$$

Where $k \in \square$, $0 \le t \le T-1$ and $c_{T/2} = 0$ if T is an odd number. The periodogram of X at the frequencies $\omega \in \{k/N\}_{k=0}^{|N/2|}$ is defined as

$$I_X^N\left(\frac{k}{T}\right) = \frac{T}{2} \begin{cases} 2c_0^2 & k = 0 \\ (c_k^2 + s_k^2) & 0 < k < \frac{T}{2} \\ 2c_{T/2}^2 & k = \frac{T}{2}, T = even \end{cases} \qquad (14)$$

Define the kth element of the discrete Fourier transform of X as

$F_k(X) = \sum_{t=0}^{T-1} e^{-i2\pi t k/T} x_t.$

The periodogram $I_X^T(\omega)$ at the frequencies $\omega \in \{k/T\}_{k=0}^{\lfloor T/2 \rfloor}$ can be simplified as

$$I_X^T\left(\frac{k}{T}\right) = \frac{1}{T} \begin{cases} 2|F_k(X)|^2 & for\ 0 < k < T/2 \\ |F_k(X)|^2 & for\ k = 0\ or\ T\ is\ even\ and\ k = T/2 \end{cases}$$

All frequency in the interval (0,0.5) is multiplied by two. This is done to ensure the following property;

$$\|X\|_2^2 = \sum_{t=0}^{T-1} x_t^2 = \sum_{k=0}^{\lfloor T/2 \rfloor} I_X^N\left(\frac{k}{T}\right)$$

Where $\pi_X^T(\omega) = \sum_{k:0 \le k \le \omega} I_T^X\left(\frac{k}{T}\right), \omega \in [0,0.5]$ is the cumulative contribution of the frequencies $[0, \omega]$? Then, for given $\omega_0 \in (0,0.5)$, [19] define the contribution of low frequencies from the interval $[0, \omega_0]$ to $X \in R^T$ as

$$C(X, \omega_0) = \pi_X^T(\omega_0)/\pi_X^T(0.5) \qquad (15)$$

Then, given parameters $\omega_0 \in (0,0.5)$ and $\omega_0 \in [0,1]$, [19] proposed to select those SVD components whose eigenvectors satisfy the following criteria:

$$C(U_j, \omega_0) \le C_0 \qquad (16)$$

where U_j is a corresponding jth eigenvector.

The second procedure (alternative method) of the automatic Grouping Algorithm can be described using the following steps:

1. Denote the original series as $X = (x_1, x_2, \dots, x_T)$, the univariate time series data with length T
2. Define Window Length (L) $2 \le L \le {T}/{2}$..
3. Determine the shape of the trajectory matrix based on L as in equation (3).
4. Find the Eigen Triple $\left(\sqrt{\lambda_i}, V_i, U_i\right)$, i = 1,2,....,d (d is the number of positive eigenvalues) of the **XX^T** matrix, where **X** is the Trajectory matrix.
5. Identify the $\hat\delta$ (Order difference) of each series U_i by equation (6) in section 2.3.
6. Make a differencing for all U_i based on the value of $\hat\delta$ (step 5) to obtain the UD_i series. Using this step, then we obtain data with the stationary pattern.
7. Specify the maximum value of seasonal order of P_i of UD_i; Here, the P_i *values are between 0 and 24.*
8. Group UD_i series based on P_i values obtained from the previous step 7, a grouping of UD$_i$ is based on periods, 3,6,12, and 24, UD$_i$ with P$_i$=0 is grouped as non-seasonal data, where when P$_i$>24 is considered as cycle pattern.
9. Create diagonal averaging for grouping components from step 8; see equation (3).
10. Forecasting (Linear Recurrent Method) with equation (4).
11. Determine MAPE values based on the number of our sample data.
    The MAPE of the h- step ahead forecast is defined as

$$MAPE = \left(\frac{1}{h}\sum_{i=1}^{h}\left|\frac{y_{n+h} - \hat{y}_{n+h}}{\hat{y}_{n+h}}\right|\right) \times 100\% \tag{17}$$

where n denotes the last data used in the in-sample data and $\hat{y}_{n+h}$ denotes the forecast values.

12. Go to step 2 until all of the L choices have been calculated. The best model will give the smallest MAPE value.

Algorithm for hybrid SSA-ARIMA

In this section, we proposed our new automatic algorithm on Singular Spectrum Analysis. The procedure of our automatic Grouping Algorithm can be described using the following steps:

1. Denote the original series as $X = (x_1, x_2, \dots, x_T)$, the univariate time series data with length T.
2. Define Window Length (L) $2 \le L \le {T}/{2}$.
3. Determine the shape of the trajectory matrix based on L as in equation (3).
4. Find the Eigen Triple $\left(\sqrt{\lambda_i}, V_i, U_i\right)$, i = 1,2,....,d (d is the number of positive eigenvalues) of the **XX^T** matrix, where **X** is the Trajectory matrix.

5. Use one method of automatic grouping procedures (section 2.4). This procedure takes the best parameter (window length (L)).
6. Build Model ARIMA of all grouping (RC's) by Auto-ARIMA procedures.
7. Forecast a new series of all groups by model ARIMA.
8. Sum all results of forecast to get predicted of new time series.
9. Determine MAPE values based on the number of our sample data (17).
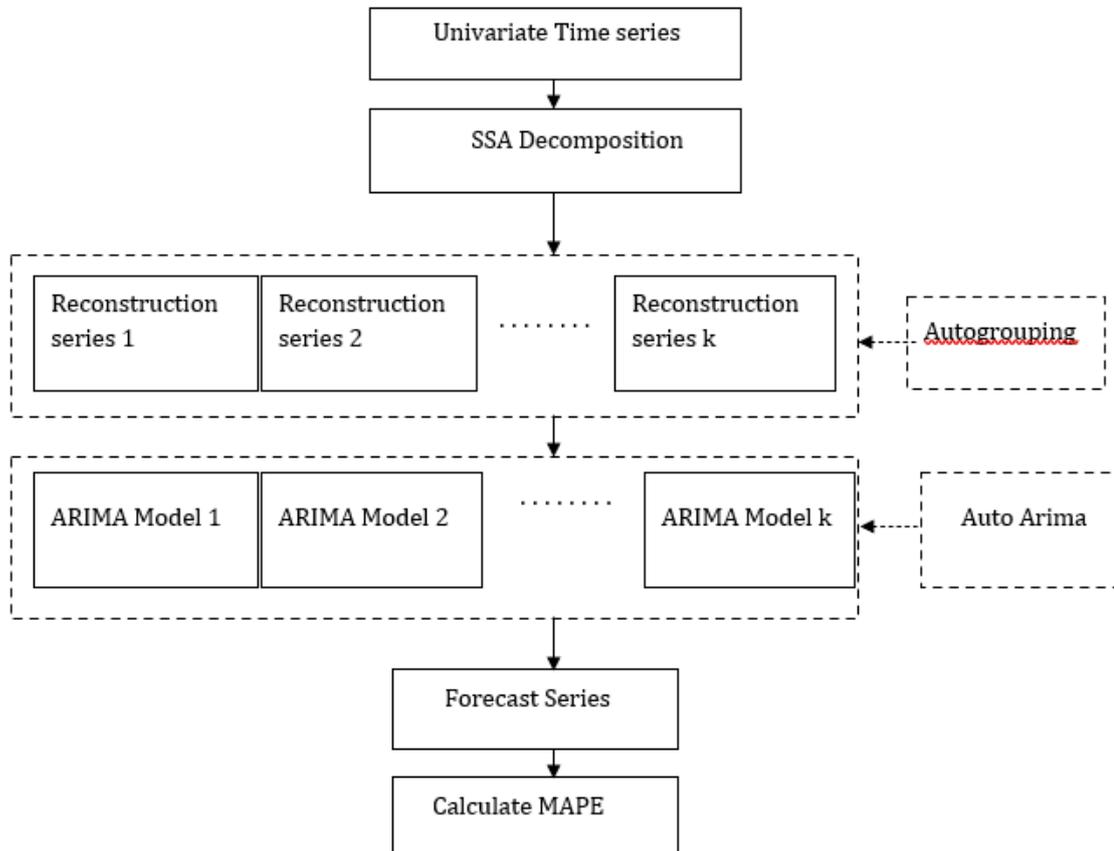   This procedure is described in figure 1.

Figure 1. Flowchart of Hybrid SSA-ARIMA

RESULTS AND DISCUSSION

For our empirical study, we analyzed *AirPassengers*, and CO_2 data, available in R package *datasets*. Here, we consider two seasonal data, namely *AirPassengers*, which has a trend and multiplicative seasonal pattern; and CO_2, which contains an additive seasonal pattern with the trend.

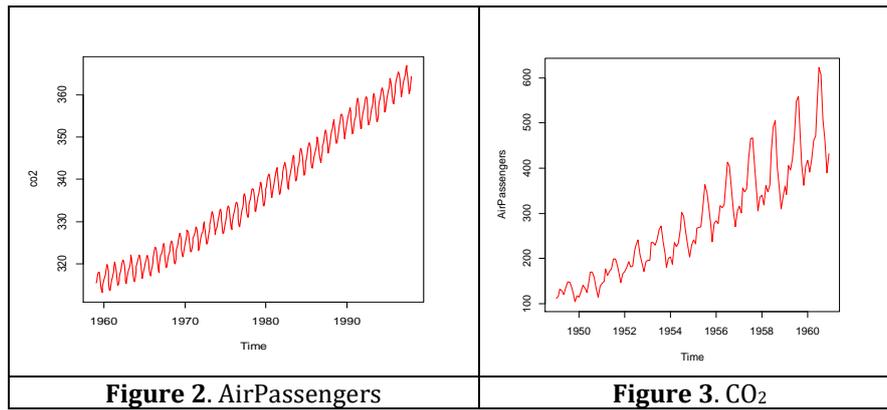| **Figure 2**. AirPassengers | **Figure 3**. CO$_2$ |

Figure 2 reveals that there has been a gradual increase in several passengers with a multiplicative seasonal pattern from 1949 to 1960. Figure 3 shows that there has been a sharp increase in Mauna Loa Atmospheric CO$_2$ Concentration with an additive seasonal pattern.

Table 1. Grouping and Contribution of Data CO$_2$

RC	Hybrid-SSA-ARIMA (Alexandrov Grouping)		Hybrid-SSA-ARIMA (Alternative Grouping)	
	Principle Component (PC)	**Contribution**	**Principle Component (PC)**	**Contribution**
RC1	PC1,PC2,PC3,PC6 PC7,PC8	99.49%	PC1,PC8,PC9,PC16,PC17 PC21,PC23,PC25	98.40%
RC2	PC4,PC5,PC12,PC15,PC16 PC17	0.28%	PC2,PC3,PC10,PC12,PC13 PC14,PC15,PC18,PC19,PC20	0.93%
RC3	PC10,PC11,PC18,PC19 PC22,PC28,PC29	0.07%	PC24	0.01%
RC4	PC13,PC14,PC26,PC30,PC31	0.06%	PC4,PC5,PC6,PC22,PC28 PC31,PC32,PC33,PC34,PC35	0.40%
RC5	PC20,PC21,PC23,PC24,PC25 PC27	0.06%	PC11,PC26,PC27,PC37,PC39 PC40,PC41	0.06%

CO$_2$ data had been successfully decomposed into five groups through SSA, each group consisting of 5 or 6 Principle Components. PC1 to PC31 were PCs that had successfully identified the pattern, while PC32 and so on were PCs with random patterns. RC1 was the largest contribution of 99.49%; this showed that the trend pattern was very dominant in this data.

On the right side of this table, CO$_2$ data had been successfully decomposed into five groups through SSA; the number of PCs for each RC is uneven. RC3 only consists of 1 PC. PC1 to PC41 were PCs that had successfully identified the pattern, while PC42 and so on were PCs with random patterns. RC1 had the largest contribution of 98.49%; this showed that the trend pattern was very dominant in this data.

Table 2. Separability of CO$_2$

RC	Hybrid-SSA-ARIMA (Alexandrov Grouping)					Hybrid-SSA-ARIMA (Alternative Grouping)				
	RC1	**RC2**	**RC3**	**RC4**	**RC5**	**RC1**	**RC2**	**RC3**	**RC4**	**RC5**
RC1	1	0.021	0.023	0.000	0.006	1	0.010	0.006	0.351*	0.000
RC2	0.021	1	0,000	0,001	0,001	0.010	1	0.000	0.000	0.001
RC3	0.023	0,000	1	0,001	0.001	0.006	0.000	1	0.000	0.002
RC4	0.000	0,001	0,001	1	0.001	0.351*	0.000	0.000	1	-0.002
RC5	0.00	0,001	0.001	0.001	1	0.000	0.001	0.002	-0.002	1

This table shows the correlation values between groups from RC1 to RC5. This best model turned out to not have a very good separability which can be seen from the correlation between RC1 and RC4, which was quite large. However, the correlation values for the others were very small.

The right side of this table shows the correlation values between groups from RC1 to RC5. This best model turned out to not have a very good separability. It can be seen from the correlation between RC1 and RC4, which is quite large. However, the correlation values for the others were very small.

Table 3. Grouping and Contribution of AirPassengers

RC	Hybrid-SSA-ARIMA (Alexandrov Grouping)		Hybrid-SSA-ARIMA (Alternative Grouping)	
	Principle Component (PC)	Contribution	Principle Component (PC)	Contribution
RC1	PC1,PC2,PC3,PC6,PC11 PC14,PC15	83.58%	PC15,PC19	0.80%
RC2	PC4,PC5,PC16,PC17,PC23 PC24,PC29	7.20%	PC1,PC2,PC3,PC16,PC17,PC18 PC21,PC22,PC23,PC28,PC29	81.51%
RC3	PC9,PC10,PC19,PC20 PC33,PC35,PC36	3.26%	PC4,PC5,PC6,PC20,PC24,PC26, PC30	8.54%
RC4	PC7,PC8,PC25,PC26,PC30 PC31	3.13%	PC7,PC8,PC9,PC10,PC11 PC33,PC34,PC36,PC37	5.98%
RC5	PC12,PC13,PC18,PC21,PC22 PC27,PC28,PC34	2.91%	PC31,PC32	0.31%

AirPassengers data had been successfully decomposed into five groups through SSA with each group consisted of 6 to 8 Principle Components. PC1 to PC36 were PCs that had successfully identified the pattern, while PC37 and so on were PCs with random patterns. RC1 had the largest contribution of 83.589% which indicated that the trend pattern dominated in this data.

AirPassengers data had been successfully decomposed into five groups via SSA, RC1 and RC5 consisted of 2 PCs, RC2 consists of 11 PCs, RC3 consisted of 7 PCs, and RC4 9PCs. PC1 to PC36 were PCs whose patterns had been identified, while PC38 and so on were PCs with random patterns. RC2 had the most considerable contribution of 81, 51%, which indicated that seasonal patterns dominated in this data.

Table 4. Separability of AirPassengers

RC	Hybrid-SSA-ARIMA (Alexandrov Grouping)					Hybrid-SSA-ARIMA (Alternative Grouping)				
	RC1	RC2	RC3	RC4	RC5	RC1	RC2	RC3	RC4	RC5
RC1	1	-0.053	-0.003	0.008	-0.015	1	-0.017	0.000	0.000	0.000
RC2	-0.053	1	0.005	0.001	0.000	-0.017	1	-0.133	0.033	-0.001
RC3	-0.003	0.005	1	-0.003	-0.002	0.000	-0.133	1	0.000	-0.002
RC4	0.008	0.001	-0.003	1	-0.001	0.000	0.033	0.000	1	0.000
RC5	-0.015	0.000	-0.002	-0.001	1	0.000	-0.001	-0.002	0.000	1

This table showed the correlation values between groups from RC1 to RC5. This best model displayed good separability, which could be seen from RC's very small correlation value. The Hybrid-SSA-ARIMA (Alternative Grouping) model showed the correlation value between groups from RC1 to RC5. This best model showed good separability which could be seen from the very small correlation value between RC.

Table 5. MAPE of Two Methods from AirPassengers dan CO_2

Data	Hybrid-SSA-ARIMA (Alternative Grouping)	Hybrid-SSA-ARIMA (Alexandrov-Grouping)
AirPassengers	3.63%	4.64%
CO_2	0.48%	0.13%

This table showed that the AirPassengers data using the Hybrid-SSA-ARIMA (Alternative Grouping) method was more accurate than the hybrid-SSA-ARIMA (Alexandrov grouping) method. The MAPE value of the Hybrid-SSA-ARIMA (Alternative Grouping) method was 3.63%, while the the value of Hybrid-SSA-ARIMA (Alexandrov grouping) method was 4.64%. On the CO_2 data, on the other hand, the Hybrid SSA-ARIMA (Alexandrov grouping) method was more accurate than the Hybrid SSA-ARIMA (Alternative Grouping) method because the MAPE value for CO_2 data using the hybrid-SSA-ARIMA (Alexandrov grouping) method was 0.13%, while for the hybrid-SSA-ARIMA (Alternative Grouping) method, the value was 0.48%.

CONCLUSIONS

Based on the analysis results, it appeared that the SSA-ARIMA method with Alexandrov auto grouping was better for data with additive seasonal patterns such as CO_2 data. The MAPE value for this method was 0.13% for CO_2 data while the SSA-ARIMA method with alternative auto grouping was 0.48% for CO_2 data. On the other hand, for data with multiplicative seasonal patterns such as AirPassengers data, the SSA-ARIMA hybrid method with alternative auto grouping produced more accurate forecasts with the MAPE value of 3.63%. In comparison, the SSA-ARIMA hybrid method with Alexandrov auto grouping containing the same data gave a MAPE value of 4.64%.

In general, the separability value had good separability as the correlation value was close to zero - only one significant correlation value is 0.351. The correlation value between RC1 and RC4 for CO_2 data generated from the SSA-ARIMA hybrid method with alternative auto grouping. As for the contribution, in general, it is dominated by RC1, except for AirPassengers data which was analyzed through the SSA-ARIMA auto hybrid method with alternative auto grouping; the contribution is dominated by RC2, amounting to 81.51%. In general, both the SSA-ARIMA with Alexadarov auto grouping and the SSA-ARIMA with alternative auto grouping provide satisfactory forecast results. The MAPE value of the two methods was below 5% for the two different data patterns.

ACKNOWLEDGMENTS

REFERENCES

[1] K. Patterson, H. Hassani, S. Heravi and A. Zhigljavsky, "Multivariate Singular Spectrum Analysis for forecasting revisions to real-time series data," Journal.Application.Stat,vol.38,no.10,pp.2183-2211 2011.

[2] H. Hassani, E. S. Silva, N. Antonakakis, G. Filis, and R. Gupta, Forecasting Accuracy evaluation of tourist arrivals, ann. Tour. Res, no.63.pp.112-127, 2017.

[3]   T. J. Harris and H. Yuan, "Filtering and frequency interpretations of Singular Spectrum Analysis," Phys. D Nonlinear Phenom, vol. 239, no. 20–22, pp. 1958–1967, 2010.

[4]   R. Menezes, A. Dionísio, and H. Hassani, "On the globalization of stock markets: An application of Vector Error Correction Model, Mutual Information and Singular Spectrum Analysis to the G7 countries," Q. Rev. Econ. Finance., vol. 52, no. 4, pp. 369–384, 2012.

[5]   P. C. Rodrigues and M. de Carvalho, "Spectral modeling of time series with missing data," Appl. Math. Model., vol. 37, no. 7, pp. 4676–4684, 2013.

[6]   C. L. Wu and K. W. Chau, "Rainfall-runoff modeling using artificial neural network coupled with singular spectrum analysis," J. Hydrology., vol. 399, no. 3–4, pp. 394–409, 2011.

[7]   C. Sivapragasam, S. Liong, and M. Pasha, "Rainfall and runoff forecasting with SSA-SVM approach," J. Hydroinformatics, no. April 2016, pp. 141–152, 2001.

[8]   R. Benzi, R. Deidda, and M. Marrocu, "Characterization of temperature and precipitation fields over Sardinia with principal component analysis and singular spectrum analysis," Int. J. Climatol., vol. 17, no. 11, pp. 1231–1262, 1997.

[9]   S. Hassani Hossein; Sanei, Singular spectrum analysis of biomedical signals. 2016.

[10]   F.Lisi, O.Nicolas, M.Sandri, "Combining Singular Spectrum Analysis and Neural Networks for time series letters, vol.2, No. forecasting", Neural processing letters, vol.2, No.4, 6-10, 1995.

[11]   Y, Gao, C, Qu and K, Zhang, "A Hybrid Method Based on Singular Spectrum Analysis, Firefly Algorithm and BP Neural Network for Short-Term Wind Speed Forecasting," Energies 9, 757.

[12]   K. Laksmi, A.Rao, and N.Gopalakhrisnan, "Singular Spectrum Analysis combined with ARMAX model for structural damage detection," structure Control Health Monit,1-21, 2016.

[13]   M.L, Menezes, R.C, Sousa and Pessanha, "Combining Singular spectrum Analysis and PAR(p) Structures to Model Wind Speed Time Series," J Syst Sci Complex,27,29-46, 2014.

[14]   Y.Feliks, Y.Groth, A.W, Robertson and M, Gill, "Oscillatory climate modes in the Indian monsoon, North Atlantic, tropical Pacific" Journal of Climate,vol26, 9528-9544,2013.

[15]   H.Hassani, E.Sirimal,N. Antonakakis and R.Gupta," Forecasting Accuracy Evaluation of Tourism Arrivals," Annals of Tourism Research,63,112-127, 2017.

[16]   J.Vile, J.W, Gillard, P.R, Harper and V.A.Knight," Predicting ambulance demand using Singular Spectrum Analysis, "Journal of Operation Research Society,63,1556-1565,2012.

[17]   O.Zhang,B.D,Wang,B.He, Y.Peng and M.LRen" Singular Spectrum Analysis and ARIMA hybrid Model for Annual Runoff Forecasting, "Water Resource Manage,25:2683-2703,2011.

[18]   G, Liu, G.Zhang and T.Zhang "Software Reliability Forecasting: Singular Spectrum Analysis and ARIMA Hybrid Model" 2015 International Symposium on Theoretical Aspects of Software Engineering, IEEE Computer Society,111-118, 2015.

[19] T. Alexandrov, "A Method of Trend Extraction Using Singular Spectrum Analysis," pp. 1–23, 2008

[20] T. Alexandrov and N. Golyandina, "Automatic extraction and forecast of time series cyclic components within the framework of SSA," Proc. St. Petersburg. Work. Simul., no. September, pp. 45–50, 2005

[21] G.Darmawan, D.Rosadi, and B.N, Ruchjana, "Daily Forecast for COVID-19 During Ramadhan by Singular Spectrum Analysis in Indonesia", Jour of Adv Research in Dynamical & Control Systems, Vol. 12, Issue-02, 2020.

[22] N. Golyandina, A. Zhigljavsky, G. N., and Z. A., Singular Spectrum Analysis for Time Series. 2013.

[23] S. Geweke J and Porter-Hudak, "The Estimation and Application of Long memory Time Series Models," vol. 4, pp. 221–238, 1983.

[24] J. R.. Hosking, "Fractional differencing," vol. 68, no. 1, pp. 165–176, 1981.

[25] W. Wei, Time Series Analysis Univariate and Multivariate Methods. California: Addison-Wesley Publishing Company, 2006