

## SSA-based hybrid forecasting models and applications

Winita Sulandari<sup>1</sup>, Subanar<sup>2</sup>, Suhartono<sup>3</sup>, Herni Utami<sup>4</sup>, Muhammad Hisyam Lee<sup>5</sup>  
Paulo Canas Rodrigues<sup>6</sup>

<sup>1,2,4</sup>Department of Mathematics, Universitas Gadjah Mada, Indonesia

<sup>1</sup>Study Program of Statistics, Universitas Sebelas Maret, Indonesia

<sup>3</sup>Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia

<sup>5</sup>Department of Mathematical Sciences, Universiti Teknologi Malaysia, Malaysia

<sup>6</sup>Department of Statistics, Federal University of Bahia, Brazil

### Article Info

#### Article history:

Received Sep 24, 2019

Revised Dec 11, 2019

Accepted May 5, 2020

#### Keywords:

Hybrid

SSA

TLCSSN

TLSAR

TLSNN

### ABSTRACT

This study attempted to combine SSA (Singular Spectrum Analysis) with other methods to improve the performance of forecasting model for time series with a complex pattern. This work discussed two modifications of TLSAR (Two-Level Seasonal Autoregressive) modeling by considering the SSA decomposition results, namely TLSNN (Two-Level Seasonal Neural Network) and TLCSSN (Two-Level Complex Seasonal Neural Network). TLSAR consisted of a linear trend, harmonic, and autoregressive component. In contrast, the two proposed hybrid approaches consisted of flexible trend function, harmonic, and neural networks. Trend and harmonic function were considered as the deterministic part identified based on SSA decomposition. Meanwhile, NN was intended to handle the nonlinearity relationship in the stochastic part. These two SSA-based hybrid models were contemplated to be more flexible than TLSAR and more applicable to the series with an intricate pattern. The experimental studies to the monthly accidental deaths in USA and daily electricity load Jawa-Bali showed that the proposed SSA-based hybrid model reduced RMSE for the testing data from that obtained by TLSAR model up to 95%.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Winita Sulandari,

Study Program of Statistics,

Universitas Sebelas Maret,

Jl. Ir. Sutami 36 A Kampus Kentingan, Surakarta, Central of Java, Indonesia 57126.

Email: winita@mipa.uns.ac.id

## 1. INTRODUCTION

Decomposition is the basis for modeling complex seasonal time series. A time series is said to have a complex seasonal pattern when it has trend and multiple seasonal patterns, perhaps with non-integer period. Recently, this kind of series has been intriguing researchers to study and develop methodologies to improve the forecast accuracy. Soares and Medeiros [1] have discussed TLSAR (two level seasonal autoregressive) model that consists of two parts, deterministic and stochastic component. This method is a simple statistical model that combines linear trend, trigonometry and the autoregressive model. Meanwhile, [2-7] developed an innovation state space exponential smoothing model. Taylor [2] added the second seasonal component to the Holt-Winter exponential smoothing, namely HWT, and in the following year, [6] presented BATS (Box-Cox transform, ARMA errors, trend, and seasonal component) and TBATS (trigonometric BATS) to reduce the parameters of HWT model. Different from other previous methods, TBATS has accommodated the fractional seasonality in the data. Still, it unable to take into account the time varying component.

Meanwhile, other researchers [8-11] considered SSA method in complex time series forecasting. The development of the SSA method was inspired by [12-15]. Much of subsequent works refer to their works. Later, [16-23] continued to develop SSA as a tool of time series forecasting. Vautard *et al.* [16] reviewed the capability of SSA in extracting information from short and noisy time series without any prior knowledge about the data. Yiou *et al.* [17] regarded SSA as a method to identifying and extracting oscillatory component from the original series. Vahabie *et al.* [18] combined SSA with autoregressive (AR) model to predict the electricity demand. In the following year, [19] and [20] reviewed and demonstrated the success of SSA in economic and financial time series forecasting. Golyandina and Korobeynikov [21] described the RSSA package and showed how the methodology of SSA can be implemented for analyzing, estimating parameter and forecasting. Hassani *et al.* [22] studied the effect of outliers and showed that the vector SSA forecasts were more robust than the recurrent SSA forecast. Khan and Poskitt [23] discussed the properties of forecast obtained by SSA in theory and practice. Recently, [24] and [25] implemented SSA method to decompose a series into several components and modeled the components by the Box-Jenkins method. However, some data exhibit a deterministic behavior rather than stochastic. Thus, there will be parts that have not been accommodated by the Box-Jenkins model. Recently, [26] to combine SSA-LRF with NN (neural network) and fuzzy time series to forecast hourly electricity load in Indonesia. The success of NN in dealing the nonlinearity can also be seen in [27-31]. Moreover, the method proposed by [24-26] did not take into account the seasonal with non-integer period, time varying amplitude, and calendar variation. Hence, some modifications are needed to accommodate the features of the series instead.

This work proposes two SSA-based hybrid model, namely TLSNN (two level seasonal neural networks) and TLCSNN (two level complex seasonal neural networks) model, modifications of TLSAR model. TLSAR consists of deterministic and stochastic component. The deterministic is expressed as a combination of linear, harmonic, and dummy variables to take into account the linear trend, seasonal, and holiday influences. Meanwhile, the two proposed models provide flexibility in determining deterministic component, especially in describing nonlinear trends. Moreover, one of the proposed models can capture the oscillation pattern with time varying amplitude. In this study, the two models are constructed by considering the results of SSA decomposition. We take advantage of the capability of SSA in decomposing time series so that we get several separable and interpretable components including the information on how much each component contributes to the series. Further, SSA method makes the identifying process of the component patterns easier. Our methodology is as simple as that in [1]. The main difference is that we take into account the time-varying components of the series that are not included in TLSAR model or other methods mentioned in the literature. Further, we can show that our methodology can improve the multi-step forecasting accuracy.

## 2. THE PROPOSED METHOD

### 2.1. Mathematical definition for the proposed SSA-based hybrid model

The two proposed models are called the two-level seasonal neural networks (TLSNN) and the two-level complex seasonal neural networks (TLCSNN). Both models are modifications of the TLSAR model. The higher order of polynomial trend and the time-varying sinusoidal function are included in the deterministic component to capture a more complex pattern in the time series. In this case, the results from the SSA decomposition make the identification and determination of the proper deterministic function for each component of TLSNN and TLCSNN easier. Generally, TLSNN and TLCSNN are expressed in

$$Y_t = S_t + Z_t \quad (1)$$

Where  $Y_t$  is observation at time  $t$ ,  $S_t$  is the deterministic component and  $Z_t$  is the stochastic component. The difference between the two is in the form of  $S_t$ . Meanwhile,  $Z_t$ , is approximated by neural network (NN). These two models are explained further in Definition 1 and Definition 2.

**Definition 1.** A time series  $\{Y_t, t = 1, \dots, N\}$  with the sample size  $N$  and which has a complex pattern follows TLSNN if the deterministic component in (1) can be presented as

$$S_t = \sum_{i=0}^{n_{pol}} a_i t^i + \sum_{h=1}^{n_h} [\alpha_{0h} \cos(2\pi ht/T) + \beta_{0h} \sin(2\pi ht/T)] + \sum_k^{n_D} \delta_k D_{kt} \quad (2)$$

and the stochastic component is

$$Z_t = \sum_l^{n_{hid}} b_l f_l \left( \sum_j^{n_{inp}} c_j Z_{t-j} \right) + e_t \quad (3)$$

where

$$f_l \left( \sum_j^{n_{inp}} c_j Z_{t-j} \right) = \frac{2}{1 - \exp(-2 \sum_j^{n_{inp}} c_j Z_{t-j})} - 1 \quad (4)$$

The notation  $a_i, \alpha_{0h}, \beta_{0h}, \delta_k$  for  $i = 1, \dots, n_{pol}, h = 1, \dots, n_h$  and  $k = 1, 2, \dots, n_D$  are the unknown parameters of the model,  $T$  is the seasonal period,  $n_h$  is the number of harmonics, and  $n_D$  is the number of dummy variables included in the model.  $D_{kt}$  notates the effect of the  $k$ th category at time  $t$ .

**Definition 2.** A time series  $\{Y_t, t = 1, \dots, N\}$  with a complex pattern follows TLCSNN if the deterministic component in (1) can be presented as

$$S_t = \sum_{i=0}^{n_{pol}} a_i t^i + \sum_{v=1}^{n_{os}} \sum_{q=0}^{n_{pol}_v} [\alpha_{vq} t^q \cos(\omega_v t) + \beta_{vq} t^q \sin(\omega_v t)] \\ + \sum_{h=1}^{n_h} [\alpha_{0h} \cos(2\pi h t / T) + \beta_{0h} \sin(2\pi h t / T)] + \sum_k^{n_D} \delta_k D_{kt} \quad (5)$$

and the stochastic component is as in (3) and (4).

The notation  $a_i, \alpha_{vq}, \beta_{vq}, \alpha_{0h}, \beta_{0h}, \delta_k$  for  $i = 1, \dots, n_{pol}, v = 1, \dots, n_{os}, h = 1, \dots, n_h, q = 1, \dots, n_{pol}_v$ , and  $k = 1, 2, \dots, n_D$  are the unknown parameters of the model,  $T$  is the seasonal period,  $n_h$  is the number of harmonics, and  $n_D$  is the number of dummy variables included in the model.  $D_{kt}$  notates the effect of the  $k$ th category at time  $t$ .

The TLSNN and TLCSNN models accommodate, not only trend linear, but also quadratic or another polynomial with higher order to capture other trend behaviors. The differences between TLSNN and TLCSNN are in their harmonic terms. We add the time-varying amplitude sinusoid component, do estimate its parameters separately using iterative ordinary least square to estimate the parameter of the amplitude modulated sinusoid model (see [32, 33]) and then combine its best function with the deterministic component of TLSAR model. Meanwhile, the parameters of the deterministic model in TLSNN are estimated simultaneously using householder transformation. This estimation method was discussed detail in [34-36].

## 2.2. Modeling Strategy for the two proposed SSA-based hybrid model

### 2.2.1. TLSNN model

The steps of TLSNN modeling procedure are explained below.

**Level 0:** the series is decomposed by SSA method into several separable components

**Level 1:** obtaining the deterministic model (2) based on the results of SSA decomposition

- a. identifying the trend function and the dummy variables to accommodate calendar effects.
- b. estimating the parameters of the deterministic model by using least square based on householder method [34-36]
- c. defining the fitted value based on the deterministic model,  $\hat{S}_t$

**Level 2:** obtaining the stochastic model

- a. determining the stochastic component,  $Z_t = Y_t - \hat{S}_t$
- b. testing the linearity in the series by using Terasvirta method [37]
- c. modeling by NN when the linearity test detecting the nonlinearity relationship in the data evaluating the accuracy of the forecast values

### 2.2.2 TLCSNN model

The procedure of TLCSNN modeling are explained in the following steps.

**Level 0:** the series is decomposed by SSA method into several separable components

**Level 1:** obtaining the deterministic model (5) based on the results of SSA decomposition

- a. identifying and estimating each component function
  - estimating parameters of the trend function by ordinary least square (OLS) method
  - estimating parameters of the oscillatory function by iterative OLS [33]
  - defining the appropriate function for representing the irregular component. This function may involve Fourier function with the dummy variables. The number of harmonics can be obtained based on AIC value.
- b. defining the fitted value based on the deterministic model,  $\hat{S}_t$

**Level 2:** obtaining stochastic model

- a. determining the stochastic component,  $Z_t = Y_t - \hat{S}_t$

- b. testing the linearity in the series by using Terasvirta method [37]
- c. modeling by NN when the linearity test detecting the nonlinearity relationship in the data evaluating the accuracy of the forecast values

### 3. RESEARCH METHOD

The SSA-based hybrid model is defined by the method that combines deterministic and stochastic component, where the deterministic component is obtained based on the SSA decomposition result.

#### 3.1. The benchmark model

The TLSAR is used as the benchmark model in this research. TLSAR model is presented as a combination of a trend linear and trigonometric function of time and the autoregressive model. The TLSAR model is expressed as the combination of deterministic and stochastic component as in (1), where the deterministic component is written as

$$S_t = a_0 + a_1 t + \sum_{h=1}^{n_h} [\alpha_{0h} \cos(2\pi ht/P) + \beta_{0h} \sin(2\pi ht/P)] + \sum_{k=1}^{n_D} \delta_k D_{kt}. \quad (6)$$

The notation  $a_0, a_1, \alpha_{0h}, \beta_{0h}, \delta_k$  for  $h = 1, \dots, n_h$  and  $k = 1, 2, \dots, n_D$  are the unknown parameters of the model,  $P$  is the seasonal period,  $n_h$  is the number of harmonics, and  $n_D$  is the number of dummy variables included in the model. The number of harmonics  $n_h$  in (6) is determined by minimizing the Akaike Information Criteria (AIC),

$$AIC = n \ln(\hat{\sigma}^2) + 2M, \quad (7)$$

Where  $\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$  is the estimate of the residual variance.  $M$  is the number of parameters fitted in the model, and  $n$  is the number of residuals that can be calculated from the series (see [34, 38, 39]). The dummy variable  $D_{kt}$  identifies the effect of the  $k$ th category at time  $t$ , for example, the  $k$ th hours of the day to  $Y_t$  or the  $k$ th days of the weeks to  $Y_t$ . This value of  $D_{kt}$  is one if the  $k$ th category is observed at time  $t$  and zero for otherwise. In the original TLSAR model,  $Z_t$  is the stochastic component that follows the AR model while in this study,  $Z_t$  follows the SAR model. The autoregressive model of order  $p$ , denoted by  $AR(p)$  is given by

$$\phi_p(B)Z_t = e_t, \quad (8)$$

Where  $\phi_p(B) = 1 - \phi_p B - \dots - \phi_p B^p$  and  $e_t$  is zero mean white noise process. This model can be applied to the series in which the preceding values and the random shock influence its present value. The process needs to be stationary so that the roots of  $\phi_p(B)=0$  must lie outside the unit circle. When the seasonal part is stochastic and cannot be handled by the deterministic model, (8) can be extended as

$$\phi_p(B)\varphi_P(B^S)Z_t = e_t, \quad (9)$$

Where  $\phi_p(B) = 1 - \phi_p B - \dots - \phi_p B^p$ ,  $\varphi_P(B^S) = 1 - \varphi_1(B^S) - \dots - \varphi_P(B^{PS})$ ,  $s$  is the seasonal period and  $P$  is the order of the seasonal component. The model presented in (9) denoted by  $SAR(p, 0, 0)(P, 0, 0)_s$

#### 3.2. SSA method in decomposing a time series

SSA method decomposes an original time series into trend, several oscillatory components, and a noise. The major references related to this method can be found in [11, 40-43]. This method consists of four steps, embedding, singular value decomposition (SVD), grouping and diagonal averaging. The time series  $\{Y_t, t = 1, 2, \dots, N\}$  with the sample size  $N$  is transformed into a Hankel matrix  $\mathbf{Y}$  with  $L$ - dimension vectors, that is

$$\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_K \\ Y_2 & Y_3 & \dots & Y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_L & Y_{L+1} & \dots & Y_N \end{bmatrix} \quad (10)$$

Where  $L$  is a window length and  $K = N - L + 1$ . This transformation process is called by the embedding step. The embedding step is then followed by calculating  $\mathbf{Y}\mathbf{Y}'$  and applying SVD to determine its eigentriples  $(\sqrt{\lambda_i}, U_i, V_i)$  for  $i = 1, 2, \dots, L$ . The  $\lambda_j$  represent the  $i$ -th eigenvalue of matrix  $\mathbf{Y}\mathbf{Y}'$  and  $\sqrt{\lambda_i}$  denotes the  $i$ -th singular value of matrix  $\mathbf{Y}$ . The  $U_i$  and  $V_i$  are the left and the right singular values of matrix  $\mathbf{Y}$ . SVD of the Hankel matrix in (10) can be written as

$$\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2 + \dots + \mathbf{Y}_d \quad (11)$$

where  $d$  is the rank of matrix  $\mathbf{Y}$  and  $\mathbf{Y}_j = \sqrt{\lambda_j} U_j V_j'$  for  $j = 1, 2, \dots, d$ . The contribution of the matrix  $\mathbf{Y}_j$  in (11) can be determined by calculating the ratio  $\lambda_j / \sum_{j=1}^d \lambda_j$ . The matrices  $(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_d)$  are arranged in  $m$  groups, where  $m \leq d$  such that (11) can be written as

$$\mathbf{Y} = \mathbf{Y}_{I_1} + \mathbf{Y}_{I_2} + \dots + \mathbf{Y}_{I_m}. \quad (12)$$

In this case, each matrix  $\mathbf{Y}_{I_i}$  in (12) may be reconstructed from one or more eigentriples  $(\sqrt{\lambda_j}, U_j, V_j)$  and the contribution of this matrix can be measured by the ratio  $\sum_{j \in I_i} \lambda_j / \sum_{j=1}^d \lambda_j$ .

Each matrix  $\mathbf{Y}_{I_i}$  is then retransformed into a time series by the diagonal averaging procedure. The diagonal averaging is a Hankelization process that transforms the elements of matrix  $\mathbf{Y}_{I_i}$ , say  $y_{ij}$  into the  $t$ -th observation of the  $i$ -th ( $i = 1, 2, \dots, m$ ) component series by averaging  $y_{ij}$  for all  $i, j$  such that  $i + j = t + 2$ . Thus the original series  $\{Y_t, t = 1, 2, \dots, N\}$  can be represented as the sum of  $m$  series,

$$Y_t = \sum_{i=1}^m Y_t^{(i)} \quad (13)$$

where  $Y_t^{(i)}$  is the series reconstructed from the matrix  $\mathbf{Y}_{I_i}$ .

The separability of all component series,  $Y_t^{(i)}, i = 1, \dots, m$ , in (13) can be measured by calculating w-correlation values, which are in the range of  $[-1, 1]$  and visualize them in the form of a w-correlation matrix diagram with different level of color from the brightest (white means no correlation or separable) to the darkest (black means perfect correlation). The strongest separability is represented by the white color and the weakest separability is shown by black color (see [40]). Other alternatives to achieve separability to obtain a more parsimonious forecasting model were proposed by [44] and [45].

### 3.3. Diagnostic checking and evaluating the performance of forecast accuracy

Model diagnostic checking is needed to check whether the model assumption that is zero mean white noise residual series is satisfied. To see whether the residuals are uncorrelated random shock with zero mean and constant variance, we can examine both the plot of residuals and its sample ACF. We can also verify the results by using the Ljung-Box test [46, 47]. In this work, we use RMSE and MAPE to measure the performance of the forecast accuracy.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}, \quad (14)$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|. \quad (15)$$

where  $n$  is the number of observations involved in the calculation.  $Y_t$  and  $\hat{Y}_t$  are the real value at time  $t$  and the forecast value at time  $t$ , respectively.  $\hat{Y}$  is the mean of the in sample data. Equation (14) and (15) are used for evaluating the training data and the testing data.

## 4. RESULTS AND DISCUSSION

The well-known monthly accidental deaths in the USA and the daily load series of Jawa-Bali for hour 1.00 am to hour 4.00 am were used in this empirical study. In this section, we show how the two proposed method implemented to the complex series and compare the results with TLSAR as the benchmark.

### 4.1. Monthly accidental deaths in USA

The monthly accidental deaths series depicted in Figure 1(a) is quite popular and has been discussed in several studies. Brockwell and Davis [48] have used this data as an example of the application of

the ARIMA, Holt-Winter seasonal, ARAR model, while [11] discussed the application of the SSA-LRF model. The data are divided into two parts, the training and the testing data set. We used the data from January 1973 to December 1978 as the training data and January 1979 to June 1979 as the testing data. As in [11], window length  $L$  used in this study is 24, proportional to the period. Based on the w-matrix correlation, the time series can be decomposed into 4 components (see, Figure 1(b)), namely trend, two oscillation components, and noise.

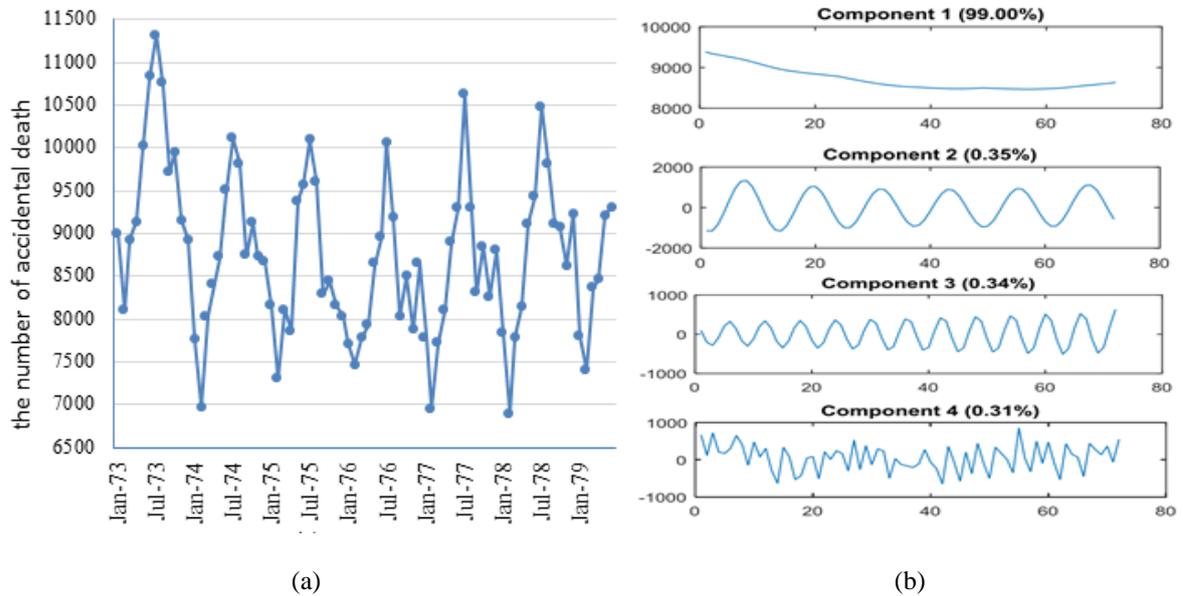


Figure 1. The monthly accidental deaths series from January 1973 to June 1979, (a) The original series, (b) Its decomposition

In TLCSNN model, the quadratic function is the most suitable for modeling trend components, while the first and the second oscillation components can be approximated by quadratic amplitude modulated sinusoidal function and linear amplitude modulated sinusoidal function, respectively. Since Terasvirta test indicates that there is a nonlinearity relationship in the stochastic component, NN must be more appropriate to model the stochastic component than the AR or other linear models. The best NN is determined based on the RMSE value. The number of input variables are chosen between 6 or 12 regarding the seasonal periods while the number of hidden units are chosen between 1 to 10 so that the network produces the smallest RMSE value with random residuals. Based on the experimental results, NN with 6 input units and 8 hidden units, denoted NN (6-8-1) is the best fit NN model for the stochastic component.

Further, in TLSNN model, we also choose the quadratic trend in the deterministic component and NN(6-8-1) for the stochastic component. In this case, the number of harmonic is 5 with period 11.98. The comparison results of RMSE and MAPE for the three models are summarized in Table 1. It can be seen that TLSNN and TLCSNN produce smaller RMSE and MAPE than TLSAR both for the training and testing data. TLSNN and TLCSNN respectively reduce the RMSE of TLSAR by 85.26% and 82.44% for the training data and by 82.73% and 95.00% for the testing data. Meanwhile, sequentially TLSNN and TLCSNN reduce the MAPE of TLSAR by 86.69% and 87.15% for the training data and by 82.52% and 95.58% for the testing data.

Table 1. Comparison of RMSE and MAPE obtained by TLSAR, TLSNN, and TLCSNN model for the death series

Model	Training		Testing	
	RMSE	MAPE	RMSE	MAPE
TLSAR ( $h=2, P=12, SARIMA(0,1,1)(0,1,1)_{12}$ )	270.4575	2.4997	334.7116	3.1314
TLSNN, $n_{pol}=2, n_h=5, NN(6-8-1)$	39.8595	0.3327	57.7906	0.5474
TLCSNN, $n_{pol}=2, n_{os}=2, n_h=0, NN(6-8-1)$	47.4913	0.3213	16.7392	0.1384

#### 4.2. Daily electricity load of Jawa-Bali in the specific hours

The electricity load of Jawa-Bali, in megawatt per hour (MWh), obtained from a power utility company in Indonesia, Jawa-Bali control center (JCC), that were observed hourly are considered in this work. However, the data discussed is only the electricity load at 01.00 am, 02.00 am, 03.00 am, and 04.00 am. These data have a unique pattern compared to other hours due to the influence of the habits of Indonesian citizens in the month of Ramadan. These four periods are considered to represent complex seasonal patterns, which have seasonal patterns, trends, and calendar influences. The data pattern of time series 01.00, 02.00, 03.00, and 04.00 for the period of January 1, 2009, to December 31, 2011, is presented in Figure 2. Each series consists of 1095 observations. It is divided into two parts, the first 1088 observations are the estimation sample (the training data) and the last 7 observations (a week) are the test sample.

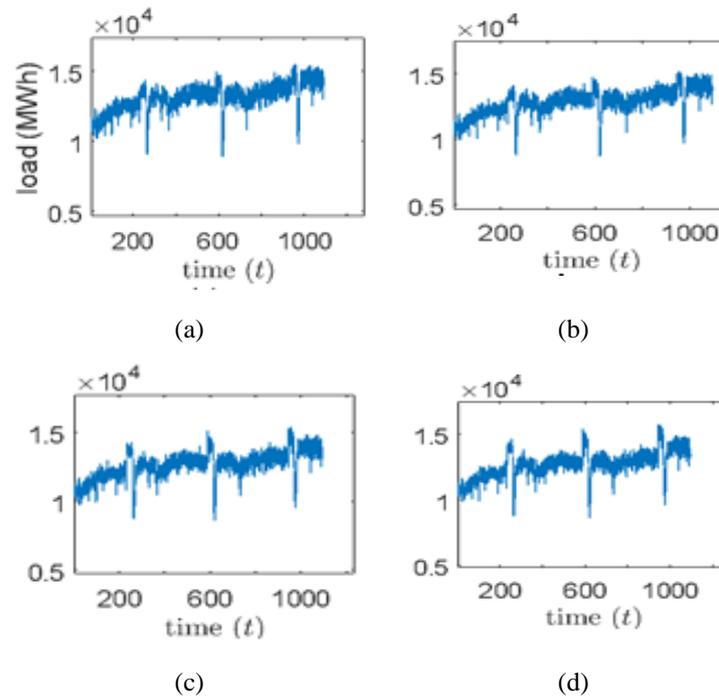


Figure 2. Daily load series from 1 January 2009 to 31 December 2011 for, (a) 01.00, (b) 02.00, (c) 03.00, and (d) 04.00

There was a noticeable surge in several days and repeated from year to year. Higher surges occur at 03.00 and 04.00. This surge occurred during the *sahur* hours in the month of Ramadan and eventually dropped dramatically before and after the Eid al-Fitr. Each series exhibits trend and seasonal. The demand for electricity that increases from time to time is related to population growth and society needs. This causes a positive trend in the series and can be modeled by the function of time. Due to the similarity results in the component patterns of SSA decomposition for 01.00 to 04.00, we only report 01.00 in detail. Nevertheless, the overall results will be discussed. The results of SSA decomposition for time series 04.00 is shown in Figures 3(a) and 3(b).

Based on Figure 3(c), the dummy variables in the TLSAR, TLSNN, and TLCSNN models can be described in Table 2. In this case, the electricity load tends to fall on the next day (the day after a holiday). Especially for the month of Ramadan, the electricity load increases. The  $k$  dummy variable  $\{k = 1, 2, \dots, 10\}$  is one if time  $t$  shows the event as described in Table 2 and has a value of 0 for the other. All the models described in Table 3 have harmonic components with the number of harmonics  $n_h = 3$  and dummy variables, as stated in Table 2. The results of the comparison of RMSE and MAPE are presented in Table 3.

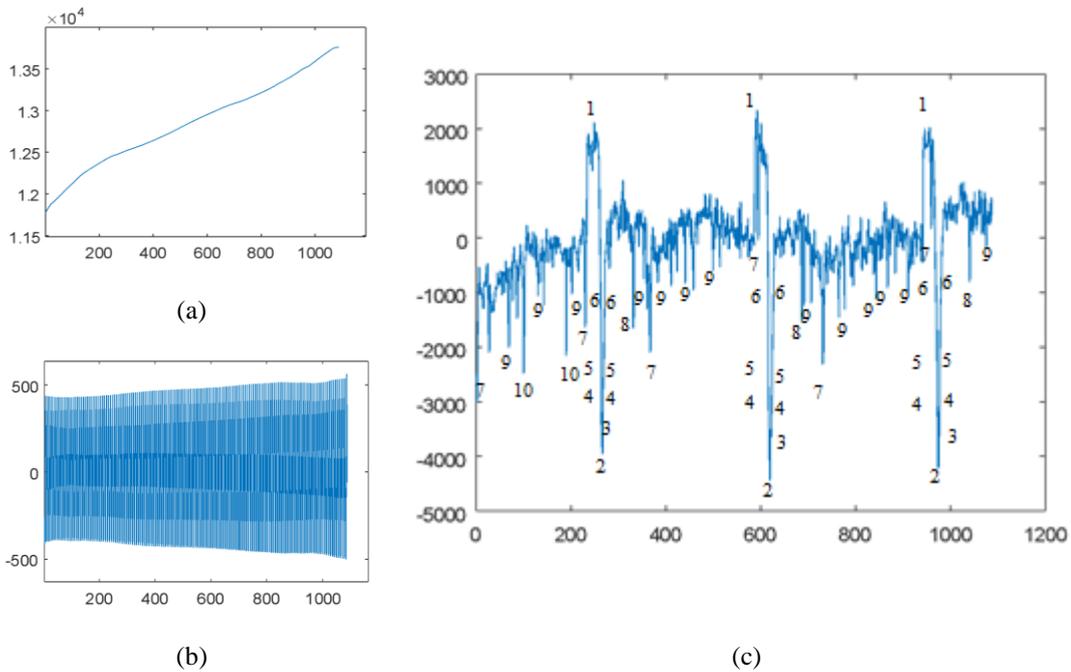


Figure 3. The components of series 04.00 obtained by SSA decomposition with  $L = 490$ , (a) Trend component, (b) Oscillation component, (c) Irregular component

Table 2. Types of days presented as dummy variables in the TLSAR, TLSNN, and TLCSNN model

Code	Description
1	The second day of Ramadan up to five days before the first Eid Al-Fitr
2	The two days of Eid Al-Fitr and the two days after the second Eid Al-Fitr
3	The third day after the second Eid Al-Fitr
4	A day before the first Eid Al-Fitr = and the fourth day after the second Eid Al-Fitr
5	The second day before the first Eid Al-Fitr, and the fifth and sixth day after the second Eid Al-Fitr
6	The third day before the first Eid Al-Fitr, a day before the first Eid Al-Fitr, and the seventh day after the second Eid Al-Fitr
7	A day after the independent day, new year, and holidays related new year.
8	Eid al Adha and a day after Eid Al Adha
9	A day after Christmas, a day after Waisya, a day after Nyepi, a day after Imlek, a day after the good Friday, a day after the ascension day of Jesus, a day after New Hijr. =
10	A day after presidential and legislative elections

Table 3. RMSE and MAPE calculated from the three models for the training and testing data

Hour	Method	RMSE		MAPE							
		Training	Testing	Training	Testing ( 25-31 December 2011)						
					1	2	3	4	5	6	7
1	TLSAR	258.1585	604.5183	1.5269	3.0608	2.2952	2.6805	2.7826	2.7215	3.3189	3.8516
	TLSNN	225.0217	210.5599	1.3292	1.7816	1.1340	1.6498	1.4100	1.4120	1.3194	1.3108
	TLCSNN	219.6763	223.3640	1.2854	1.9569	1.8204	2.0931	1.6147	1.3931	1.2624	1.3573
2	TLSAR	258.4460	598.5510	1.5406	1.9241	1.5728	2.2643	2.5783	2.7039	3.3912	3.8528
	TLSNN	226.9396	291.9397	1.3586	1.5159	1.8073	1.8611	1.9852	1.9391	1.6678	1.9340
	TLCSNN	226.5588	284.2629	1.3412	0.7976	0.8453	1.3011	1.9835	2.0730	1.7419	1.5428
3	TLSAR	258.0119	594.6462	1.5426	2.6900	1.8375	2.4459	3.0090	3.0498	3.6538	3.9846
	TLSNN	228.1649	364.7621	1.3492	3.0535	1.6601	2.0184	2.9127	2.3784	2.2030	1.9297
	TLCSNN	216.0240	282.1139	1.2868	1.1977	1.1545	2.0739	2.3221	2.1597	1.8226	1.5671
4	TLSAR	166.6101	674.3093	1.0052	3.2022	1.6647	2.1011	2.8500	3.1066	3.7152	4.3406
	TLSNN	230.2454	431.8189	1.3081	2.6571	1.6434	2.0005	2.7185	2.4458	2.7858	2.7745
	TLCSNN	218.3380	229.3032	1.2888	1.6704	1.6563	1.3914	1.9303	1.7635	1.4701	1.2627

Based on Table 3, it can be concluded that the TLSNN and TLCSNN models provide a more accurate forecasting value than the TLSAR model. The modulated oscillation component in the TLCSNN

model increases the forecast accuracy of the TLSNN model, especially at 03.00 and 04.00. The TLSNN model is able to reduce the value of the RMSE of TLSAR between 35.96% and 65.17% for the testing data, while the TLCSNN model is able to reduce the RMSE of TLSAR between 52.51% and 65.99% for the testing data. For the case of electricity load forecasting at 01.00, 02.00, 03.00, and at 04.00, the TLSNN and TLCSNN models are more acceptable than the TLSAR model. Furthermore, the results of the TLSNN model forecasting for 02.00 and the TLCSNN model for 04.00 provide MAPE values that meet PLN standards, which are below 2%.

## 5. CONCLUSION

In this study, we proposed two SSA-based hybrid forecasting methods for complex seasonal time series and provided five data series to show the application examples. The proposed methods named TLSNN and TLCSNN were developed from TLSAR by considering components of SSA decomposition results. The proposed hybrid models consist of deterministic and stochastic components. The deterministic components of TLSNN and TLCSNN models constructed based on the SSA decomposition results accommodated not only trend linear but also quadratic or another polynomial with higher order to capture other trend behaviors. The differences between TLSNN and TLCSNN were in their harmonic term. Parameters of the deterministic model in TLSNN were estimated simultaneously using householder transformation. Meanwhile in TLCSNN, we added the time-varying amplitude sinusoid components to the trend and harmonic or stationary sinusoids. The parameters for each functions were then estimated separately. Finally, the stochastic component of TLSNN and TLCSNN were modeled by NN to overcome nonlinearity relationship problem in the data. Being modified, the models tend to produce better performance for the multi-steps forecast values than TLSAR for the case of the monthly accidental death series and Jawa-Bali load series. The findings of this study have several implications for future practice. Another possible future research would be to develop the methods for multivariate problems.

## ACKNOWLEDGEMENTS

We are grateful to M. Reza Pulungan (Universitas Gadjah Mada) for his constructive remarks and to the anonymous reviewers for their valuable comments which helped the authors to improve the quality of the manuscript. The authors would also like to thank to PLN for its support by providing Indonesia electricity load.

## REFERENCES

- [1] L. J. Soares and M. C. Medeiros, "Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data," *International Journal of Forecasting*, vol. 24, no. 4, pp. 630-644, 2008.
- [2] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *Journal of the Operational Research Society*, vol. 54, no. 8, pp. 799-805, 2003.
- [3] J. W. Taylor, "Triple seasonal methods for short-term electricity demand forecasting," *European Journal of Operational Research*, vol. 204, no. 1, pp. 139-152, July 2010.
- [4] J. W. Taylor, *et al.*, "A comparison of univariate methods for forecasting electricity demand up to a day ahead," *International Journal of Forecasting*, vol. 22, no. 1, pp. 1-16, 2006.
- [5] P. G. Gould, *et al.*, "Forecasting time series with multiple seasonal patterns," *European Journal of Operational Research*, vol. 191, no. 1, pp. 207-222, Nov 2008.
- [6] A. M. De Livera, *et al.*, "Forecasting time series with complex seasonal patterns using exponential smoothing," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1513-1527, 2011.
- [7] M. Bernardi and L. Petrella, "Multiple seasonal cycles forecasting model: the Italian electricity demand," *Statistical Methods & Applications*, vol. 24, no. 4, pp. 671-695, 2015.
- [8] A. Miranian, *et al.*, "Day-ahead electricity price analysis and forecasting by singular spectrum analysis," in *IET Generation, Transmission & Distribution*, vol. 7, no. 4, pp. 337-346, 2013.
- [9] K. Afshar and N. Bigdeli, "Data analysis and short term load forecasting in Iran electricity market using singular spectral analysis (SSA)," *Energy*, vol. 36, no. 5, pp. 2620-2627, May 2011.
- [10] Moisés, *et al.*, "Electricity consumption forecasting using singular spectrum analysis," *DYNA*, vol. 82, pp. 138-146, April 2015.
- [11] H. Hassani, "Singular Spectrum Analysis: Methodology and Comparison," *Journal of Data Science*, vol. 5, pp. 239-257, 2007.
- [12] D. S. Broomhead and G. P. King, "Extracting qualitative dynamics from experimental data," *Physica D: Nonlinear Phenomena*, vol. 20, no. 2-3, pp. 217-236, 1986.
- [13] K. Fraedrich, "Estimating the dimensions of weather and climate attractors," *Journal of the atmospheric sciences*, vol. 43, no. 5, pp. 419-432, March 1986.
- [14] R. Vautard and M. Ghil, "Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series," *Physica D: Nonlinear Phenomena*, vol. 35, no. 3, pp. 395-424, May 1989.

- [15] R. Vautard and M. Ghil, "Interdecadal oscillations and the warming trend in global temperature time series," *Nature*, vol. 350, no. 6316, pp. 324-327, 1991.
- [16] R. Vautard, *et al.*, "Singular-spectrum analysis: A toolkit for short, noisy chaotic signals," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 95-126, Sep 1992.
- [17] P. Yiou, *et al.*, "Spectral analysis of climate data," *Surveys in Geophysics*, vol. 17, no. 6, pp. 619-663, 1996.
- [18] A. H. Vahabie, M. M. R. Yousefi, B. N. Araabi, C. Lucas and S. Barghinia, "Combination of Singular Spectrum Analysis and Autoregressive Model for Short Term Load Forecasting," *2007 IEEE Lausanne Power Tech*, Lausanne, pp. 1090-1093, 2007.
- [19] H. Hassani and A. Zhigljavsky, "Singular Spectrum Analysis: Methodology and Application to Economics Data," *Journal of Systems Science and Complexity*, vol. 22, no. 3, pp. 372-394, 2009.
- [20] H. Hassani and D. Thomakos, "A review on singular spectrum analysis for economic and financial time series," *Statistics and its Interface*, vol. 3, no. 3, pp. 377-397, 2010.
- [21] N. Golyandina and A. Korobeynikov, "Basic singular spectrum analysis and forecasting with R," *Computational Statistics & Data Analysis*, vol. 71, pp. 934-954, March 2014.
- [22] H. Hassani, *et al.*, "A preliminary investigation into the effect of outlier (s) on Singular Spectrum Analysis," *Fluctuation and Noise Letters*, vol. 13, no. 4, pp. 1450029-1-1450029-23, 2014.
- [23] M. A. R. Khan and D. S. Poskitt, "Forecasting stochastic processes using singular spectrum analysis: Aspects of the theory and application," *International Journal of Forecasting*, vol. 33, no. 1, pp. 199-213, 2017.
- [24] Q. Zhang, *et al.*, "Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting," *Water resources management*, vol. 25, no. 11, pp. 2683-2703, 2011.
- [25] H. Li, *et al.*, "A Hybrid Short-Term Power Load Forecasting Model Based on The Singular Spectrum Analysis and Autoregressive Model," *Advances in Electrical Engineering*, vol. 2014, pp. 1-7, 2014.
- [26] W. Sulandari, *et al.*, "Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks," *Energy*, vol. 190, p. 116408, Jan 2020.
- [27] G. Zhang, *et al.*, "Forecasting with artificial neural networks: The state of the art," *International journal of forecasting*, vol. 14, no. 1, pp. 35-62, March 1998.
- [28] G. P. Zhang, *et al.*, "A simulation study of artificial neural networks for nonlinear time-series forecasting," *Computers & Operations Research*, vol. 28, no. 4, pp. 381-396, April 2001.
- [29] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, Jan 2003.
- [30] E. M. Laadissi, *et al.*, "A Nonlinear TSNN Based Model of a Lead Acid Battery," *Bulletin of Electrical Engineering and Informatics*, vol. 7, no. 2, pp. 169-175, Jun. 2018.
- [31] E. Jensemiriam, *et al.*, "Prediction of Lead-Acid Battery Performance Parameter: An Neural Network Approach," *Bulletin of Electrical Engineering and Informatics*, vol. 2, no. 1, pp. 65-74, Mar. 2013.
- [32] W. Sulandari, Subanar, H. Utami, Suhartono and M. H. Lee, "Amplitude-Modulated Sinusoidal Model for The Sinusoidal Components of SSA Decomposition," *2018 International Symposium on Advanced Intelligent Informatics SAIN*, Yogyakarta, Indonesia, pp. 66-71, 2018.
- [33] W. Sulandari, *et al.*, "Estimating the function of oscillatory components in SSA-based forecasting model," *International Journal of Advances in Intelligent Informatics*, vol. 5, no. 1, pp. 11-23, 2019.
- [34] G. Kitagawa, "Introduction to time series modeling," Boca Raton, FL: Chapman & Hall, *CRC press*, 2010.
- [35] G. Golub, "Numerical methods for solving linear least squares problems," *Numerische Mathematik*, vol. 7, no. 3, pp. 206-216, 1965.
- [36] P. Businger and G. H. Golub, "Linear least squares solutions by Householder transformations," *Numerische Mathematik*, vol. 7, no. 3, pp. 269-276, 1965.
- [37] T. Teräsvirta, *et al.*, "Power of the neural network linearity test," *Journal of time series analysis*, vol. 14, no. 2, pp. 209-220, 1993.
- [38] K. P. Burnham and D. R. Anderson, "Multimodel inference understanding AIC and BIC in model selection," *Sociological methods & research*, vol. 33, no. 2, pp. 261-304, 2004.
- [39] M. B. Priestley, "Spectral analysis and time series: probability and mathematical statistics," New York: *Academic Press*, vol. 1, 1981.
- [40] N. Golyandina, *et al.*, "Analysis of Time Series Structure: SSA and related techniques," vol. 90. Boca Raton, FL: *Chapman & Hall/CRC*, Boca Raton, FL, 2001.
- [41] J. B. Elsner, "Analysis of Time Series Structure: SSA and Related Techniques," *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 1207-1208, 2002.
- [42] N. Golyandina and A. Zhigljavsky, "Singular Spectrum Analysis for time series," *Springer Science & Business Media*, 2013.
- [43] J. B. Elsner and A. A. Tsonis, "Singular spectrum analysis: a new tool in time series analysis," *Springer Science & Business Media*, 2013.
- [44] R. Mahmoudvand and P. C. Rodrigues, "A new parsimonious recurrent forecasting model in singular spectrum analysis," *Journal of Forecasting*, vol. 37, no. 2, pp. 191-200, 2018.
- [45] P. C. Rodrigues and R. Mahmoudvand, "A new approach for the vector forecast algorithm in singular spectrum analysis," *Communications in Statistics-Simulation and Computation*, vol. 49, no. 3, pp. 591-605, 2018.
- [46] W. W.-S. Wei, "Time series analysis: Univariate and Multivariate Methods," *2nd ed. Boston: Pearson Addison-Wesley*, 1989.
- [47] E. Hanke John, *et al.*, "Business Forecasting", 8th ed. New Jersey: Pearson, *Prentice Hall*, 2005.
- [48] P. J. Brockwell and R. A. Davis, "Introduction to time series and forecasting," *2nd ed. New York: Springer-Verlag*, 2002.

## BIOGRAPHIES OF AUTHORS



**Winita Sulandari** currently is an Assistant Professor in the Study Program of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Sebelas Maret, Surakarta. She got Master of Science and Doctorate degree in Mathematics from Universitas Gadjah Mada. She published several scientific papers in reputed journals or conferences. Her research interests include time series forecasting, neural networks, and fuzzy time series.



**Subanar** currently is a professor of Department of Mathematics, Universitas Gadjah Mada. He holds PhD degree from University of Wisconsin-Madison. He has been the Centre Director of SEAMEO Regional Centre for QITEP in Mathematics (SEAQiM) for two periods, from 2011 to 2016, and was instrumental in the progress for mathematics education in Indonesia. He has more than 30 years of teaching and research experience. His research interests include mathematical statistics, time series forecasting, wavelet neural networks, and fuzzy time series



**Suhartono** currently is the chief of forum pendidikan tinggi statistika (statistics higher education forum), Indonesia. He is a faculty at Institut Teknologi Sepuluh Nopember, Surabaya. He got Master of Science in Statistics from University of Manchester Institute of Science and Technology. He holds PhD degree in mathematics from Universitas Gadjah Mada. He has published more than 100 research papers in journals or conferences. His research interests include time series forecasting, neural networks, fuzzy time series, and spatio-temporal model.



**HERNI UTAMI** currently is a researcher and lecturer at Universitas Gadjah Mada, Yogyakarta. She got Master of Science and Doctorate degree in Mathematics from Universitas Gadjah Mada. Her research interests include time series analysis, and regression analysis.



**Muhammad Hisyam Lee** currently is a professor of Department of Mathematical Sciences, Universiti Teknoogi Malaysia (UTM). He got Master of Science (Statistics) from Universiti Kebangsaan Malaysia and PhD (Mathematics) from Universiti Teknologi Malaysia. He has published more than 100 research papers in reputed journals or conferences. His research interests include high dimensional data modeling, time series analysis, neuro fuzzy, and regression model.



**Paulo Canas Rodrigues** is an Assistant Professor in the Department of Statistics, Federal University of Bahia, Brazil. He is the current President of the Brazilian Region of the International Biometric Society, and Member of the Executive Committee of the International Association for Statistical Computing. He published more than 45 scientific papers with more than 70 co-authors, and is member of the editorial board several international scientific journals.