International Journal of Advances in Intelligent Informatics
Vol. 5, No. 1, March 2019, pp. 11-23


# Estimating the function of oscillatory components in SSA-based forecasting model

Winita Sulandari [a,b,1], Subanar [a,2,*], Suhartono [c,3], Herni Utami [a,4] , Muhammad Hisyam Lee [d,5]

[a] Department of Mathematics, Universitas Gadjah Mada, Yogyakarta, Indonesia
[b] Study Program of Statistics, Unviersitas Sebelas Maret, Surakarta, Indonesia
[c] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[d] Department of Mathematical Sciences, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[1] winita@mipa.uns.ac.id; [2] subanar@ugm.ac.id; [3] suhartono@statistika.its.ac.id; [4] herni_utami@ugm.ac.id; [5] mhl@utm.my
* corresponding author

ARTICLE INFO

**Article history**
Received December 31, 2018
Revised February 4, 2019
Accepted February 5, 2019
Available online March 22, 2019

**Keywords**
Oscillatory
Time-varying
SSA
OLS
NLM

ABSTRACT

The study of SSA-based forecasting model is always interesting due to its capability in modeling trend and multiple seasonal time series. The aim of this study is to propose an iterative ordinary least square (OLS) for estimating the oscillatory with time-varying amplitude model that usually found in SSA decomposition. We compare the results with those obtained by nonlinear least square based on Levenberg Marquardt (NLM) method. A simulation study based on the time series data which has a linear amplitude modulated sinusoid component is conducted to investigate the error of estimated parameters of the model obtained by the proposed method. A real data series was also considered for the application example. The results show that in terms of forecasting accuracy, the SSA-based model where the oscillatory components are obtained by iterative OLS is nearly the same with that is obtained by the NLM method.



## 1. Introduction

In the last decade, researchers began to be interested in SSA (Singular Spectrum Analysis) development which was first discussed by Broomhead & King [1] and Fraedrich [2]. Vautard and Ghil [3] were also introduced and contributed to this field so that much of the subsequent work refer to the results of their study. Later, some researchers [4]–[13] continued their work on the development of SSA as a tool of time series forecasting.

In time series forecasting, SSA-LRF (linear recurrent formula) proposed by Golyandina *et al.* [4] decomposes time series into two separable components (signal and noise component), estimates the linear recurrent relations (LRR) and applies it to the last points of the signal series to obtain the forecast values. However, by this method, we cannot easily represent the model as a function of time. Meanwhile, Golyandina & Zhigljavsky [14] have proven that the signal governed by LRRs is a linear combination of products of polynomial, exponential and oscillatory series.

In the meantime, researchers [15]–[17] considered SSA as a method to define some separable component series and modeled those each components using the Box-Jenkins method. Each component series is approximated by ARIMA (Autoregressive Integrated Moving Average) model. In fact, not all the components of SSA decomposition can be modeled by ARIMA since the series do not satisfy the assumptions needed in ARIMA modeling. Therefore, the approach may not be implemented in some series.

Recently, Suhartono *et al.* [18] combined the time series regression (TSR) and ARIMA to model components determined by SSA method. In this case, TSR is used to model the trend component while the ARIMA is implemented to model the seasonal component and the noise. Moreover, as mentioned earlier, the signal series described by Golyandina *et al.* [4] and Golyandina & Zhigljavsky [14] consists of the combination of products of polynomial, exponential and harmonic series. While this is true, the performance of SSA-based forecasting model discussed in Sulandari *et al.* [19] is greatly influenced by the determination of each component model. When the choice of function for each component series is not fit to the true line then the performance of the SSA-based model will become worse.

In this work, we highlight the oscillatory components of the SSA decomposition results and discuss the best fit models for those components. There were many studies in the oscillatory or sinusoidal models, especially in analysis and modeling speech signal and audio [20]-[22]. Chen [23] discussed the estimation of parameters of the stationary sinusoidal model by deriving the nonlinear equation as a function of frequency and solving it using Newton method. At the same time, Pantazis *et al.* [21] discussed a time-varying quasi-harmonic model (QHM) by presenting the sinusoidal signal as a sum of products of polynomials and exponentials and estimating the parameters by iterative procedure with considering the frequency mismatch. This algorithm outperformed the FFT-based approach. In the following year, Liu *et al.* [24] also discussed the sinusoidal signal with time-varying amplitude and estimated the parameters using algebraic parametric techniques and modulating functions method, where the estimates are determined by integrals. Later, Valin *et al.* [25] presented the sinusoidal model in the linearization form and estimated the parameters by an iterative method based on the linearization of the model.

This research focuses on estimating the parameters of the oscillatory with time-varying amplitude as a result of SSA decomposition. Based on Golyandina & Zhigljavsky [14], a complex pattern series can be represented as the sum of trend or smoothing and oscillatory components. The oscillatory series may show a stationary or nonstationary sinusoid. In this case, the nonstationary sinusoid may either be the product of a polynomial and stationary sinusoid or product of exponential and stationary sinusoid. Sulandari *et al.* [26] show that the three stages of nonlinear least square based on Levenberg Marquardt (NLM) method produces consistent estimators for the parameters of linear amplitude modulated sinusoids, quadratic amplitude modulated sinusoids and exponential amplitude modulated sinusoids. In this work, we present an iterative OLS method and compare it to the NLM method in modeling the time-varying amplitude oscillatory components. The iterative OLS algorithm has been discussed and been implemented in estimating the stationary sinusoid component of SSA decomposition for the monthly atmospheric concentration of $CO_2$ [27].

The iterative OLS discussed in Sulandari *et al.* [27] cannot be implemented to the time-varying amplitude sinusoid model directly. In this work, we estimate the model by employing the iterative OLS to the three steps in Sulandari *et al.* [26]. We represent the time-varying amplitude sinusoid model in a different way from the references mentioned earlier. We express the model in the trigonometric function of the sum of sine and cosine with time-varying amplitude. Meanwhile, Pantazis *et al.* [21] considered QHM in the form of the exponential function with linear time-varying amplitude and Valin *et al.* [25] presented the model in the form of the cosine function with frequency correction over a finite window. The aim of this study is to present the simple iterative OLS method for estimating parameters of time-varying amplitude sinusoid model, i.e. oscillation with linear amplitude. As stated in Gujarati [28], the OLS method is known simpler than others. Hence, the method presented in this paper is also simple. However, its simplicity does not diminish its ability in estimating parameters of the time-varying sinusoid model. In order to check the estimation performance of the proposed method, we consider nonlinear least square based on Levenberg Marquardt (NLM) method. We compare the results obtained by iterative OLS method with those obtained by NLM.

This paper is organized as follows. In section 2, we describe the methods for determining the oscillatory with time-varying amplitude model. The comparison between the oscillatory model obtained by the iterative OLS and NLM method is discussed in Section 3. A Simulation study that was conducted to the linear amplitude modulated sinusoidal model is presented in this section in order to investigate the performance of both iterative OLS and NLM in estimating the parameters of the model. We also discuss the implementation of the algorithms to the real data series in Section 3 and finally, Section 4 delivers the conclusions.

## 2. Method

SSA is a powerful method in decomposing time series with a complex pattern. We do not discuss SSA further in this section, but the reader can find this matter in [4], [7], [29], [30]. Consider that the SSA method decomposes the series into two main separable components, signal and noise component so that the model can be represented as (1).

$$Y_t = S_t + N_t \tag{1}$$

where $Y_t$ is the observation value at time $t$, $S_t$ is the signal value at time $t$, and $N_t$ is the noise value at time $t$. Inspiring by Soares & Medeiros [31], the idea of this work is to approximate the signal component series of (1) by the deterministic function of time and to finish the rest with a stochastic model. The signal $\{S_t, t = 1, 2, \dots N\}$ itself can be decomposed into several components, i.e. trend, smoothing, and oscillatory components and therefore the deterministic function of time can be represented as (2).

$$S_t = \sum_{i=1}^{n_s} f_i(t) g_i(t) + \varepsilon_t \tag{2}$$

where $n_s$ is the number of separable components obtained from the signal, $f_i(t)$ is defined as a polynomial (3).

$$f_i(t) = \sum_{j=0}^{n_{pi}} a_{ij}\, t^j \tag{3}$$

or may other possible functions, i.e. exponential function. Select which one is the most appropriate function for the envelope of the oscillatory series, but generally, the polynomial function is the best one. $n_{pi}$ denotes the order of the $i$th polynomial function and $a_{ij}$ is the $j$-th parameter of $i$-th component of the signal. The function $g_i(t)$ represents the estimate function for the $i$-th stationary sinusoidal series.

$$g_i(t) = \alpha_i + \beta_i \cos(\omega_i t) + \gamma_i \sin(\omega_i t) \tag{4}$$

where $\alpha_i, \beta_i, \gamma_i$ and $\omega_i$ are the parameters of sinusoids. The notation $\omega_i$ denotes the frequency of the model where $\omega_i = 2\pi/T_i$ and $T_i$ is the period of the $i$th component series.

In this case, SSA decomposition results make the identification and determination of the proper deterministic function for each component in (2) easier. When the $i$th component series shows the trend or smoothing pattern, (4) is equal to one (unit) and when it shows stationary sinusoid, (3) is the unit.

Based on Equation (1), it can be seen that $\omega$ is the only nonlinear parameter in the model. This parameter can be estimated simultaneously or separately in the model estimation. This research presents an iterative OLS method for modeling nonstationary periodicity series. The nonstationary periodicity series that discussed in this study is oscillatory with a linear time-varying amplitude. The sinusoidal model obtained by iterative OLS method is then compared with the model determined by NLM method.

In the simulation study, we investigate the performance of the model based on the mean square error of estimators ($MSE_\theta$), root mean square error (RMSE), and mean absolute percentage error (MAPE) of the forecasting values. Meanwhile, in the empirical study, the performance of each oscillatory component of the signal is evaluated based on the RMSE and the $R^2$ of the training data. For further evaluation, the deterministic model that consists of trend and one or more oscillatory components is then combined with other stochastic model to obtain more accurate forecasting values. We use RMSE and MAPE to measure the accuracy performance of the forecasting model.

All evaluation criteria which are employed to evaluate the performance of the model [32]–[34] are defined as follows

$$\mathrm{MSE}_\theta = \sum_{r=1}^{R}\sum_{p=1}^{n_p}(\hat{\theta}_p - \theta_p)^2 / R, \tag{5}$$

$$\mathrm{R}^2 = \sum_{t=1}^{m}(\hat{Y}_t - \bar{Y})^2 / \sum_{t=1}^{m}(Y_t - \bar{Y})^2, \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{t=h}^{h+m-1}(Y_t - \hat{Y}_t)^2}, \tag{7}$$

$$\mathrm{MAPE} = \frac{1}{m}\sum_{t=h}^{h+m-1}\left|(Y_t - \hat{Y}_t)/Y_t\right| \times 100\%, \tag{8}$$

where $\mathrm{MSE}_\theta$ presented in (5) is MSE of estimators, calculated from $R$ data sets, i.e. the number of replication used in simulation study). Notation $n_p$ denotes the number of parameters considered in the model, $\hat{\theta}_p$and $\theta_p$ present the $p$-th estimated parameter and the $p$-th parameter value, respectively. $\mathrm{R}^2$ expressed in (6) measures the relationship between the actual values and the forecast values. $Y_t$ is the actual value at time $t$ and $\hat{Y}_t$ is the forecast value at time $t$. Notation $h$ in (7) and (8) shows the time index when the calculation starts and $m$ denotes the number of the observations used in calculation.

### 2.1. Stationary periodicity series

Consider the stationary sinusoidal model as in (9).

$$S_t = \alpha + \beta\cos(\omega t) + \gamma\sin(\omega t) + \varepsilon_t. \tag{9}$$

The parameters of the model (9) are obtained by iterative OLS method and NLM method as presented as follows:

a. Iterative OLS method

Step 1: Define $j$ where $\omega_j = 2\pi j/N$ is the maximizer of periodogram and $N$ is the sample size.

The maximizer of periodogram, $\omega_j$, can be defined by first considering a Fourier representation of the series [35] expressed by (10).

$$S_t = \sum_{k=0}^{[N/2]}\theta_k\cos(\omega_k t) + \vartheta_k\sin(\omega_k t). \tag{10}$$

$\omega_k = 2\pi k/N, k = 0,1,2,\dots,[N/2]$ are the Fourier frequencies, $[N/2]$ is $N/2$ if $N$ is even, and $[N/2]$ = $(N-1)/2$ if $N$ is odd. The Fourier coefficients $\theta_k$ and $\vartheta_k$ are defined by (11)-(12).

$$\theta_k = \begin{cases}\frac{1}{N}\sum_{t=1}^{N}S_t\cos(\omega_k t), & k = 0 \text{ and } k = N/2 \text{ if } N \text{ is even,}\\ \frac{2}{N}\sum_{t=1}^{N}S_t\cos(\omega_k t), & k = 1,2,\dots,[(N-1)/2],\end{cases} \tag{11}$$

and

$$\vartheta_k = \frac{2}{N}\sum_{t=1}^{N}S_t\sin(\omega_k t), \qquad k = 1,2,\dots,[(N-1)/2]. \tag{12}$$

The periodogram $I(\omega_k)$ is then defined by (13).

$$I(\omega_k) = \begin{cases}N\theta_0^2, & k = 0\\ \frac{N}{2}(\theta_k^2 + \vartheta_k^2), & k = 1,2,\dots,[(N-1)/2],\\ N\theta_{N/2}^2, & k = N/2 \text{ if } N \text{ is even.}\end{cases} \tag{13}$$

Based on (13), we can obtain a certain $j$ where $I(\omega_j)$ is the maximum value of the periodogram, $I(\omega_j) = \max\{I(\omega_k),\ k = 0,1,2,\dots,[N/2]\}$. In this case, $\omega_j$ is called the maximizer of the periodogram.

Step 2: Set $\delta$, a value between 0 and 1 and assume that the true frequency $\omega$ is the value between $2\pi(j-\delta)/N$ and $2\pi(j+\delta)/N$

Step 3: Assign the length of step, $\Delta$, where $\Delta < \delta$.

Step 4: Define $\omega_l = 2\pi(j-\delta)/N$ and $\omega_u = 2\pi(j+\delta)/N$.

Step 5: Estimate $(\alpha_l, \beta_l, \gamma_l)$ and $(\alpha_u, \beta_u, \gamma_u)$ using OLS method for the given $\omega_l$ and $\omega_u$, respectively.

Step 6: Calculate $RMSE$ for the estimated parameters $(\omega_l, \alpha_l, \beta_l, \gamma_l)$, say $RMSE_l$ and for the estimated parameters $(\omega_u, \alpha_u, \beta_u, \gamma_u)$, say $RMSE_u$ .

Step 7: Compare between $\text{RMSE}_l$ and $\text{RMSE}_u$.

i. If $\text{RMSE}_l < \text{RMSE}_u$ then $\omega_l^{new} = \omega_l^{old}$ and adjust $\omega_u^{new} = \omega_u^{old} - 2\pi\Delta/N$, where $\omega_u^{new}$ must be greater than $\omega_l^{new}$. If $\omega_u^{new} < \omega_l^{new}$ then $\widehat{\omega} = \omega_l^{old}$, and continue to Step 8, else back to Step 6.

ii. If $\text{RMSE}_l > \text{RMSE}_u$ then $\omega_u^{new} = \omega_u^{old}$ and adjust $\omega_l^{new} = \omega_l^{old} + 2\pi\Delta/N$, where $\omega_l^{new}$ must be less than $\omega_l^{new}$. If $\omega_l^{new} > \omega_u^{new}$ then $\widehat{\omega} = \omega_u^{old}$ and continue to the Step 8, else back to Step 6

iii. If $\text{RMSE}_l = \text{RMSE}_u$ or $\omega_l^{new} = \omega_u^{new}$ then $\widehat{\omega} = \omega_l^{new} = \omega_u^{new}$ and continue to Step 8.

Step 8: Find $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ using OLS method for the value $\widehat{\omega}$ obtained from Step 7.

b. NLM method

The NLM algorithm was discussed in many literature, such as [36]–[38] and [36]. In this study, we use periodogram to find the initial value of frequency [39], i.e. the maximizer of the periodogram as described in the first step of iterative OLS method while the initial values of other parameters are set to be zeros.

### 2.2. Oscillatory series with time-varying amplitude

The amplitude modulated sinusoidal modeling using NLM method was discussed detail in the previous research [26]. There are three stages in estimating the sinusoidal function with time-varying amplitude, both for the iterative OLS method and the NLM method. The algorithm is presented below.

Stage 1: establish the envelope series and define the best fit function (3) for the series

Stage 2: obtain the stationary sinusoidal series and determine the fit function (4) by

a. Iterative OLS method as presented in Section **2.1**.a.

b. NLM method as presented in Section **2.1**.b.

Stage 3: estimating the sinusoidal model with time-varying as a product of $f(t)$ obtained from Stage 1 and $g(t)$ determined from Stage 2.

## 3. Results and Discussion

This section discusses the simulation study of linear amplitude modulated sinusoidal model and its application. The simulation study is conducted to investigate the properties of estimators that are obtained by the iterative OLS and the NLM method. The comparison results between the two methods are presented.

### 3.1. Simulation

The data used in simulation study were generated based on (14).

$$S_t = (a + bt)(\alpha + \beta\cos(\omega t) + \gamma\sin(\omega t)) + \varepsilon_t, \tag{14}$$

that may also be written as (15).

$$S_t = A + Bt + C\cos(\omega t) + D\sin(\omega t) + Et\cos(\omega t) + Ft\sin(\omega t) + \varepsilon_t, \tag{15}$$

where $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. The parameters of the model (15) are $A = a\alpha, B = b\alpha,\ C = a\beta,\ D = a\gamma,\ E = b\beta,\ F = b\gamma$ and $\omega = 2\pi/T$.

We have considered several data generated processes (DGPs) to investigate the stability of the iterative OLS method. Since the results were almost similar one another so we just present two DGPs with the four levels sample of size ($N$ = 500, 1000, 1500, and 2000) and 50 independent series each. We used

a. $S_t = 200 + 0.02t - 50\cos\left(\frac{2\pi}{48}t\right) + 75\sin\left(\frac{2\pi}{48}t\right) - 0.005t\cos\left(\frac{2\pi}{48}t\right) + 0.0075t\sin\left(\frac{2\pi}{48}t\right) + \varepsilon_t$

b. $S_t = 125 + 0.0375t - 37.5\cos\left(\frac{2\pi}{48}t\right) - 25\sin\left(\frac{2\pi}{48}t\right) + 0.0113t\cos\left(\frac{2\pi}{48}t\right) - 0.0075t\sin\left(\frac{2\pi}{48}t\right) + \varepsilon_t$

where the frequency was $\omega = 2\pi/48$ = 0.1309 and $\sigma^2$ = 0.25. Each series was divided into two parts, the training and the testing dataset. For all series, the last forty eight data were considered as the testing data. The results of the simulation study were summarized in Table 1 and Table 2.

**Table 1.** Estimated parameters of the 1st DGP and the forecast accuracy of the model obtained by iterative OLS and NLM method

| | | Estimated parameters by | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Method* | | *Iterative OLS* | | | | *NLM* | | | |
| *Size (N)* | | *500* | *1000* | *1500* | *2000* | *500* | *1000* | *1500* | *2000* |
| *Parameter* | *A = 200* | 199.9791 | 200.0632 | 200.0203 | 200.0132 | 199.9793 | 200.0636 | 200.0203 | 200.0140 |
| | *B = 0.02* | 0.0201 | 0.0199 | 0.0200 | 0.0200 | 0.0201 | 0.0199 | 0.0200 | 0.0200 |
| | *C = -50* | -50.0132 | -49.9966 | -50.0083 | -49.9772 | -49.9888 | -50.0204 | -50.0124 | -50.0006 |
| | *D = 75* | 74.9832 | 75.0359 | 75.0098 | 75.0212 | 74.9995 | 75.0199 | 75.0071 | 75.0060 |
| | *E = -0.005* | -0.0050 | -0.0050 | -0.0050 | -0.0050 | -0.0050 | -0.0050 | -0.0050 | -0.0050 |
| | *F = 0.0075* | 0.0075 | 0.0074 | 0.0075 | 0.0075 | 0.0075 | 0.0074 | 0.0075 | 0.0075 |
| | *ω = 0.1309* | 0.1309 | 0.1309 | 0.1309 | 0.1309 | 0.1309 | 0.1309 | 0.1309 | 0.1309 |
| | *σ= 0.25* | 0.2932 | 0.2707 | 0.2630 | 0.2606 | 0.2908 | 0.2706 | 0.2628 | 0.2604 |
| *MSE of estimators* | | 0.0477 | 0.0241 | 0.0121 | 0.0125 | 0.0344 | 0.0236 | 0.0132 | 0.0115 |
| *RMSE* | *Train* | 0.5076 | 0.5046 | 0.5030 | 0.5031 | 0.5059 | 0.0545 | 0.5027 | 0.5030 |
| | *Test* | 0.5366 | 0.5226 | 0.5116 | 0.5091 | 0.5310 | 0.5236 | 0.5123 | 0.5078 |
| *MAPE(%)* | *Train* | 0.2187 | 0.2178 | 0.2176 | 0.2191 | 0.2181 | 0.2177 | 0.2175 | 0.2190 |
| | *Test* | 0.2279 | 0.2111 | 0.2011 | 0.1924 | 0.2261 | 0.2114 | 0.2016 | 0.1920 |

Table 1 and Table 2 show that the MSE of estimators for both the iterative OLS and the NLM tend to become smaller as the sample sizes increase. In addition, values of RMSEs and MAPEs for the two methods are almost similar. It means that the two methods have nearly the same performance, in term of forecasting values.

Based on the experimental results, it is noted that we need to pay attention in selecting $\delta$ and $\Delta$ in using the iterative OLS method. In case of the results obtained by the iterative OLS, the values $\delta$ = 0.5 and $\Delta$ = 0.001 produce the best estimators for the sample size N = 500 and N = 1500 while $\delta$ = 0.1 and $\Delta$ = 0.00005 is better for the sample size N =1000 and N = 2000.

Comparing to the NLM, the iterative OLS is simpler in the calculation since it does not need to find the differentiation of the objective function as in the NLM. However, it might be time consuming when we determine $\delta$ and $\Delta$ incorrectly. Though it can also occur in the NLM when we do not set the right initial values. Moreover, we can say that the complex method does not always produce better results than the simpler one as stated in Makridakis & Hibon [40].

**Table 2.** Estimated parameters of the 2nd DGP and the forecast accuracy of the model obtained by iterative OLS and NLM method

| | | Estimated parameters by | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Method | | Iterative OLS | | | | NLM | | | |
| Size (N) | | *500* | *1000* | *1500* | *2000* | *500* | *1000* | *1500* | *2000* |
| *parameter* | *A = 125* | 124.9914 | 125.0027 | 125.0573 | 125.0249 | 124.9913 | 125.0025 | 125.0573 | 125.0249 |
| | *B = 0.0375* | 0.0375 | 0.0375 | 0.0374 | 0.0375 | 0.0375 | 0.0375 | 0.0374 | 0.0375 |
| | *C = 37.5* | 37.5056 | 37.4899 | 37.5132 | 37.5022 | 37.4997 | 37.4976 | 37.5108 | 37.5101 |
| | *D = -25* | -24.9832 | -25.0157 | -25.0126 | -25.0140 | -24.9923 | -25.0041 | -25.0163 | -25.0023 |
| | *E = 0.0113* | 0.0113 | 0.0112 | 0.0112 | 0.0112 | 0.0113 | 0.0112 | 0.0112 | 0.0112 |
| | *F = - 0.0075* | -0.0075 | -0.0075 | -0.0075 | -0.0075 | -0.0075 | -0.0075 | -0.0075 | -0.0075 |
| | $\omega = 0.1309$ | 0.1309 | 0.1309 | 0.1309 | 0.1309 | 0.1309 | 0.1309 | 0.1309 | 0.1309 |
| | $\sigma^2 = 0.25$ | 0.2950 | 0.2702 | 0.2633 | 0.2597 | 0.2946 | 0.2701 | 0.2631 | 0.2597 |
| *MSE of estimators* | | 0.0460 | 0.0195 | 0.0123 | 0.0095 | 0.0436 | 0.0191 | 0.0127 | 0.0094 |
| *RMSE* | *Train* | 0.5084 | 0.5042 | 0.5032 | 0.5023 | 0.5091 | 0.5042 | 0.5030 | 0.5023 |
| | *Test* | 0.5424 | 0.5223 | 0.5132 | 0.5079 | 0.5419 | 0.5212 | 0.5130 | 0.5081 |
| *MAPE(%)* | *Train* | 0.3279 | 0.3249 | 0.3240 | 0.3271 | 0.3277 | 0.3249 | 0.3239 | 0.3270 |
| | *Test* | 0.3236 | 0.2761 | 0.2418 | 0.2179 | 0.3235 | 0.2756 | 0.2418 | 0.2181 |

### 3.2. Application

A half hourly Java-Bali electricity load data for period 3 January to 13 February 2010 was considered for the experiment due to its complex pattern. The data were collected from PT. PLN P2B Jawa Bali, Gandul, Cinere, West Java, Indonesia. In this empirical study, the series that consists of 2016 observations splits into two parts. We carry the first 1968 observations as the fitting sample (training data) and the rest are to be the post-sample (testing data).

Since the load series is influenced by the economic and demographic factor, it exhibits trend and multiple seasonal patterns (see Fig. 1). Fig. 2(a) shows a w-correlation matrix of SSA decomposition with L=976 which is enlarged to clarify which eigentriple groups are not correlated. Based on Fig. 2(a), we can define five separable components, denoted by the colors. The first four components (blue, red, green, and purple) reconstructed by the first eleven eigentriples are then considered as the deterministic component while the 973 eigentriples left are considered as a part of the stochastic component. The blue line in Fig.2(b) is the trend component obtained by the first eigentriple. The red (see Fig. 2(c)) and the green (see Fig. 2(d)) charts show the oscillatory with time-varying amplitude while the purple chart (see Fig. 2(e)) displays the stationary oscillatory component. The last component is the irregular component (see Fig. 2(f)).

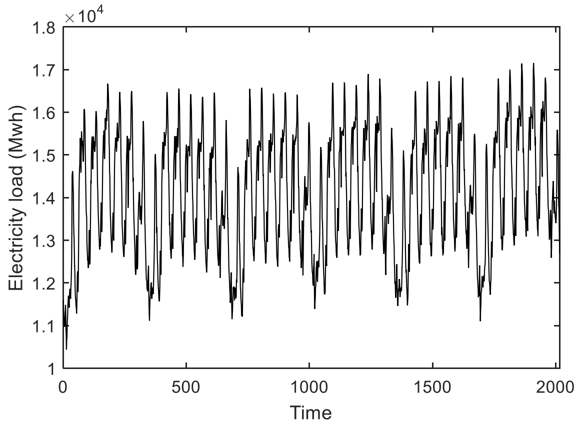


**Fig. 1.** The series of half-hourly Jawa-Bali electricity load for period 3 January to 13 February 2010.

Though we focus on modeling the oscillatory components, we need to find the best approximation function for the trend for further evaluation of the SSA-based forecasting model. There were three possible functions that can be applied to fit the trend line, namely the linear, quadratic and exponential functions. In this point, the exponential function which can be written as

$$f_1(t) = 13902 \exp(0.00001t) + 0.1080 \exp(0.0038t) \tag{16}$$

yields the lowest RMSE and the greatest $R^2$ compared to the two alternative functions.

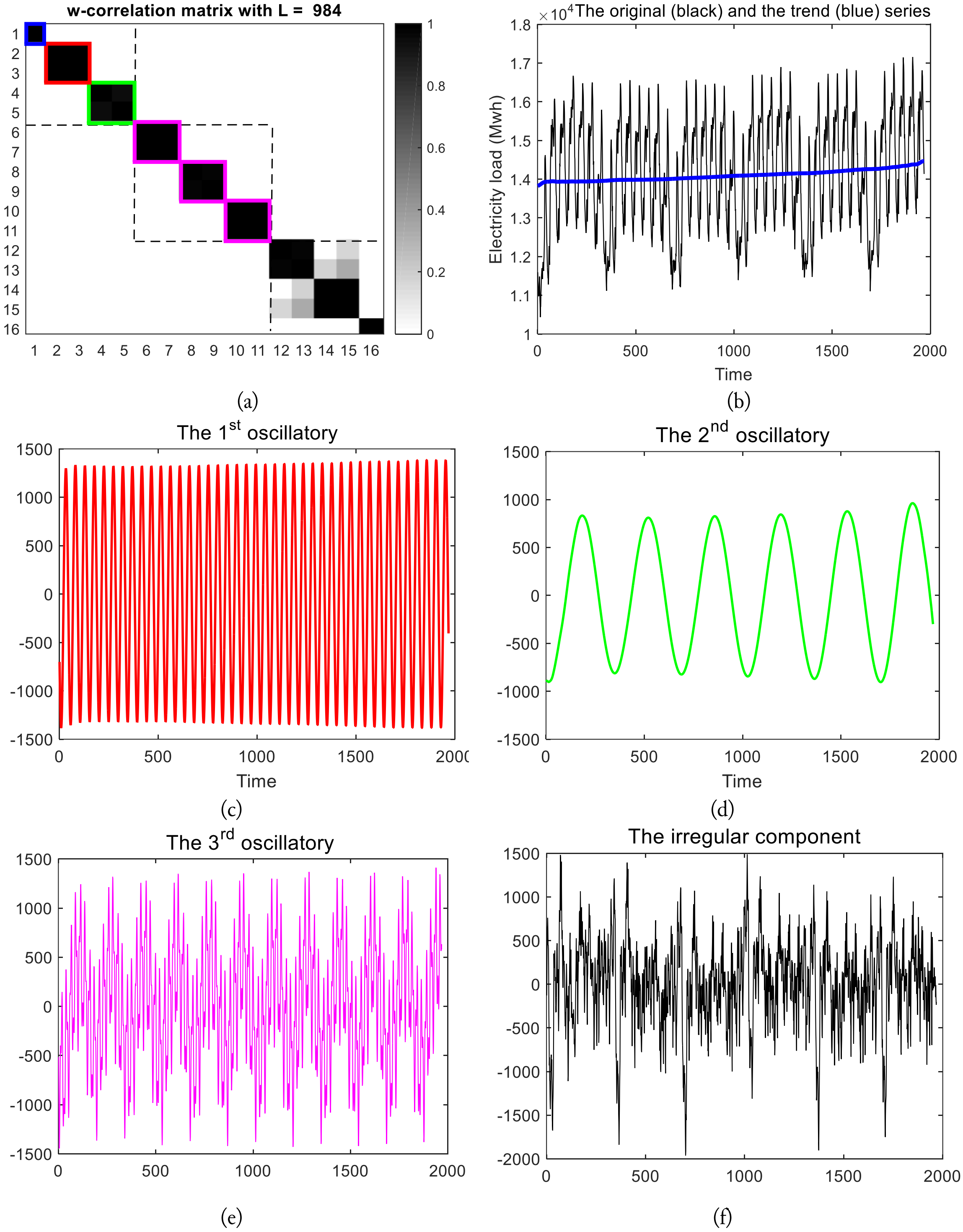


**Fig. 2.** The half-hourly Jawa-Bali electricity load data for period 3 January to 12 February 2010 and its decomposition

Later, the best fit function for each oscillatory component is presented in Table 3. The first and the second oscillatory series have a linear time-varying amplitude as seen in Fig. 2(c) and Fig. 2(d), respectively. Meanwhile, the third oscillatory is reconstructed from several eigentriples and it seems to produce a stationary periodicity function. Thus, the sum of stationary sinusoids based on Fourier series with order 7 is the most appropriate function for this series.

**Table 3.** Comparisons of RMSEs and $R^2$s of accuracy performances of oscillatory models obtained by OLS and NLM method.

| Oscillatory | Function | RMSE | | $R^2$ | |
|---|---|---|---|---|---|
| | | *Iterative OLS* | *NLM* | *Iterative OLS* | *NLM* |
| 1 | Linear time-varying amplitude | 10.9054 | 10.9049 | 99.99% | 99.99% |
| 2 | Linear time-varying amplitude | 27.8919 | 27.8919 | 99.79% | 99.79% |
| 3 | Stationary[a] | 224.6702 | 224.6460 | 86.88% | 86.88% |

[a.] Sinusoid model based on Fourier series with order 7.

The performances of the deterministic model as the sum of the trend, two linear time-varying amplitude, and the stationary periodicity functions are described in Table 4. The table shows the comparisons between the deterministic functions where their oscillatory functions are obtained by OLS and NLM method. Table 4 also presents the further evaluation for the hybrid model. In this case, neural network (NN) is chosen to model the stochastic component due to its capability in dealing the uncertainty in the noise. What is meant by the noise here is the sum of all residuals of each component models and the irregular component or can also be considered as the residuals of the deterministic model. This noise is not necessarily white noise and by chance, the noise type in the illustrated example is not a Gaussian white noise. In addition, NN is a general and flexible modeling tool that does not need any specific assumptions regarding the relationships in the data [41]. This has even succeeded in eliminating the spiky event for the case discussed in Sulandari *et al.* [19]. Therefore, we consider that NN can handle the noise well.

**Table 4.** Comparisons of RMSEs and MAPEs of accuracy performances of SSA-based models obtained by iterative OLS and NLM method

| No | Model | Training | | | | Testing | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | RMSE | | MAPE(%) | | RMSE | | MAPE(%) | |
| | | *Iterative OLS* | *NLM* | *Iterative OLS* | *NLM* | *Iterative OLS* | *NLM* | *Iterative OLS* | *NLM* |
| 1 | Deterministic SSA-based model: Exponential-oscillatory model | 813.5209 | 813.5211 | 4.6011 | 4.6011 | 790.2083 | 790.2232 | 4.6925 | 4.6924 |
| 2 | Hybrid SSA-based model: Exponential-oscillatory-NN model | 105.5063 | 113.8603 | 0.5818 | 0.6184 | 122.7776 | 143.6725 | 0.6871 | 0.8425 |

Based on the experimental results presented in Table 4, the performance of deterministic SSA-based model where the oscillatory functions are obtained by iterative OLS method is not much different from the results of that obtained by NLM method. Indeed, for the deterministic SSA-based model, NLM method produces smaller MAPE and RMSE than the results obtained by iterative OLS method. However, we get the opposite results when we combine the deterministic SSA-based model with NN (see Fig. 3). Fig. 3 shows that there is a strong decline in MAPE when the deterministic model is combined with NN. In this case, NN model is a viable choice to improve the accuracy performance for further evaluation of SSA-based model.

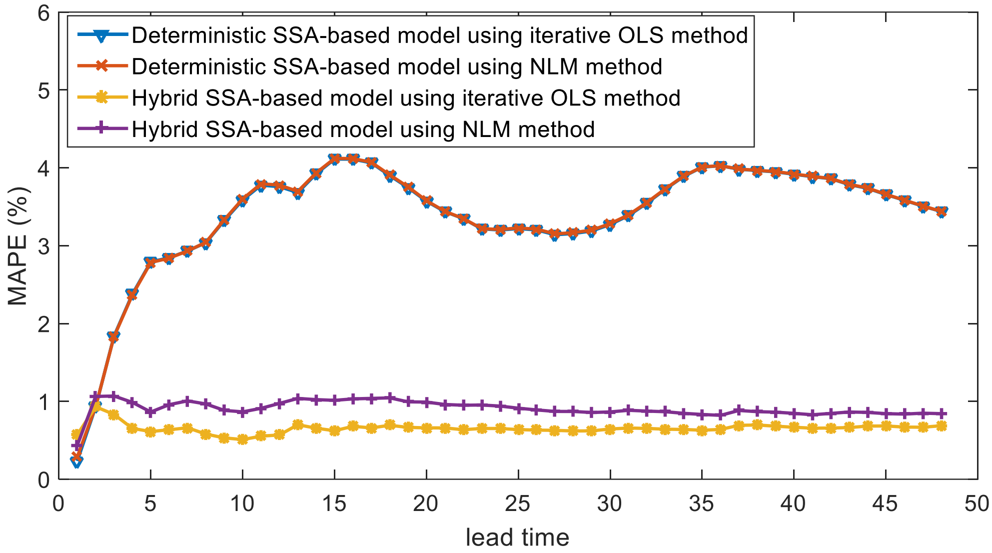


**Fig. 3.** MAPEs for the post-sample (in Frebruary, 13$^{th}$ 2010)

The actual and post-sample results that are 48-steps-ahead forecast values of electricity load using hybrid SSA-based models are shown in Fig. 4. We can see that the results of both models follow the actual values well. With the MAPEs of accuracy performance less than 1%, this hybrid model is acceptable in dealing the trend and multiple seasonal in the data.

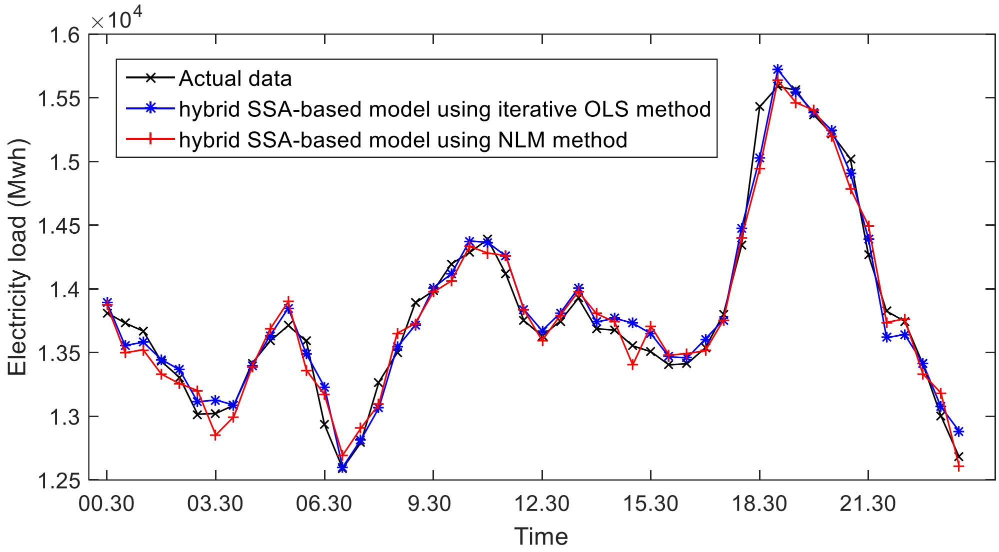


**Fig. 4.** Actual and forecast values of electricity load for the post-sample (in Frebruary, 13$^{th}$ 2010)

Finally, based on the simulation and application results, we can see that the simple iterative OLS method provides nearly the same accuracy performance as the NLM method. As mentioned before, the iterative OLS is simpler than the NLM method. Perhaps, iterative OLS method is time consuming regarding to the determination of step length, lower and upper bound. However, in using NLM method, the problem may arise when we do not establish a good initial estimate for the parameters of the model.

In further evaluation of electricity load forecasting model, we combine the deterministic model with NN, called by the hybrid SSA-based model. The conclusion we get is that the better performance of the deterministic SSA-based model does not necessarily generate a better performance of hybrid SSA-based model.

## 4. Conclusion

This study compares the iterative OLS method with the NLM method in estimating the oscillatory with time-varying amplitude as a result of SSA decomposition. Indicated by nearly the same values of MAPEs and RMSEs between the two methods both for the training and testing data, it can be concluded that the complex NLM method does not necessarily produce superior results than the simple iterative OLS method.

A further evaluation for the hybrid SSA-based model where the oscillatory components were estimated by iterative OLS and NLM method was conducted on electricity load data. The results show that the better performance of the deterministic SSA-based model does not necessarily generate a better performance of the hybrid SSA-based model, either for the oscillatory model obtained by iterative OLS or NLM method. It is noted that results from one case may differ for other case. Therefore, further study on other types of time-varying amplitude model need to be conducted due to the possibility we encounter a more complex time-varying amplitude sinusoid in SSA decomposition results.

### Acknowledgment

We would like to thank PT PLN P2B Jawa Bali - Indonesia that providing the data for this study. We appreciate and thank the reviewers for their careful reading of our manuscript.

### References

[1] D. S. Broomhead and G. P. King, "Extracting qualitative dynamics from experimental data," *Phys. Nonlinear Phenom.*, vol. 20, no. 2–3, pp. 217–236, 1986, doi: 10.1016/0167-2789(86)90031-X.

[2] K. Fraedrich, "Estimating the dimensions of weather and climate attractors," *J. Atmospheric Sci.*, vol. 43, no. 5, pp. 419–432, 1986, doi: 10.1175/1520-0469(1986)043<0419:ETDOWA>2.0.CO;2.

[3] R. Vautard and M. Ghil, "Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series," *Phys. Nonlinear Phenom.*, vol. 35, no. 3, pp. 395–424, 1989, doi: 10.1016/0167-2789(89)90077-8.

[4] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and related techniques*, vol. 90. Chapman & Hall/CRC, Boca Raton, FL, 2001, doi: 10.1201/9781420035841.

[5] N. Golyandina, "On the choice of parameters in singular spectrum analysis and related subspace-based methods," *Stat Interface*, vol. 3, no. 3, pp. 259–279, 2010, doi: 10.4310/SII.2010.v3.n3.a2.

[6] N. Golyandina and A. Korobeynikov, "Basic singular spectrum analysis and forecasting with R," *Comput. Stat. Data Anal.*, vol. 71, pp. 934–954, 2014, doi: 10.1016/j.csda.2013.04.009.

[7] H. Hassani, "Singular Spectrum Analysis: Methodology and Comparison," *J. Data Sci.*, vol. 5, pp. 239–257, 2007, available at: http://www.jds-online.com/file_download/133/JDS-396.pdf.

[8] H. Hassani, S. Heravi, and A. Zhigljavsky, "Forecasting European industrial production with singular spectrum analysis," *Int. J. Forecast.*, vol. 25, no. 1, pp. 103–118, 2009, doi: 10.1016/j.ijforecast.2008.09.007.

[9] H. Hassani, R. Mahmoudvand, H. N. Omer, and E. S. Silva, "A preliminary investigation into the effect of outlier (s) on Singular Spectrum Analysis," *Fluct. Noise Lett.*, vol. 13, no. 04, p. 1450029, 2014, doi: 10.1142/S0219477514500291.

[10] H. Hassani, A. S. Soofi, and A. Zhigljavsky, "Predicting Daily Exchange Rate with Singular Spectrum Analysis Data," *Nonlinear Anal. Real World Appl.*, vol. 11, no. 3, pp. 2023–2034, 2010, doi: 10.1016/j.nonrwa.2009.05.008.

[11] M. A. R. Khan and D. S. Poskitt, "On The Theory and Practice of Singular Spectrum Analysis Forecasting," Monash University, Department of Econometrics and Business Statistics, 2014, avilable at: https://ideas.repec.org/p/msh/ebswps/2014-3.html.

[12] R. Vautard and M. Ghil, "Interdecadal oscillations and the warming trend in global temperature time series," *Nature*, vol. 350, no. 6316, p. 324, 1991, doi: 10.1038/350324a0.

[13] R. Vautard, P. Yiou, and M. Ghil, "Singular-spectrum analysis: A toolkit for short, noisy chaotic signals," *Phys. Nonlinear Phenom.*, vol. 58, no. 1, pp. 95–126, 1992, doi: 10.1016/0167-2789(92)90103-T.

[14] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for time series*. Springer Science & Business Media, 2013, doi: 10.1007/978-3-642-34913-3.

[15] 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," in *Power Tech, 2007 IEEE Lausanne*, 2007, pp. 1090–1093, doi: 10.1109/PCT.2007.4538467.

[16] Q. Zhang, B.-D. Wang, B. He, Y. Peng, and M.-L. Ren, "Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting," *Water Resour. Manag.*, vol. 25, no. 11, pp. 2683–2703, 2011, doi: 10.1007/s11269-011-9833-y.

[17] H. Li, L. Cui, and S. Guo, "A hybrid short-term power load forecasting model based on the singular spectrum analysis and autoregressive model," *Adv. Electr. Eng.*, vol. 2014, 2014, doi: 10.1155/2014/424781.

[18] S. Suhartono, S. Isnawati, N. A. Salehah, D. D. Prastyo, H. Kuswanto, and M. H. Lee, "Hybrid SSA-TSR-ARIMA for water demand forecasting," *Int. J. Adv. Intell. Inform.*, vol. 4, no. 3, pp. 238–250, Nov. 2018, doi: 10.26555/ijain.v4i3.275.

[19] W. Sulandari, S. Subanar, S. Suhartono, and H. Utami, "Forecasting Time Series with Trend and Seasonal Patterns Based on SSA," in *2017 3rd International Conference on Science in Information Technology (ICSITech) "Theory and Applicattion of IT for Education, Industry and Society in Big Data Era,"* Bandung, Indonesia, 2017, pp. 694–699, doi: 10.1109/ICSITech.2017.8257193.

[20] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, 1 edition. Upper Saddle River, N.J: Prentice Hall, 1999, available at: Google Scholar.

[21] Y. Pantazis, O. Rosec, and Y. Stylianou, "Iterative estimation of sinusoidal signal parameters," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 461–464, 2010, doi: 10.1109/LSP.2010.2043153.

[22] A. Bánhalmi, K. Kovács, A. Kocsor, and L. Tóth, "Fundamental frequency estimation by least-squares harmonic model fitting," in *Ninth European Conference on Speech Communication and Technology*, 2005, available at: http://www.inf.u-szeged.hu/-kocsor/publications/Papers/2005/konferencia-telj/Conf-2005-EuroSpeech-BA/Web/BKK05.pdf.

[23] K. F. Chen, "Estimating parameters of a sine wave by separable nonlinear least squares fitting," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 12, pp. 3214–3217, 2010, doi: 10.1109/TIM.2010.2047298.

[24] D. Liu, O. Gibaru, and W. Perruquetti, "Parameters estimation of a noisy sinusoidal signal with time-varying amplitude," in *Control & Automation (MED), 2011 19th Mediterranean Conference on*, 2011, pp. 570–575, doi: 10.1109/MED.2011.5983186.

[25] J.-M. Valin, D. V. Smith, C. Montgomery, and T. B. Terriberry, "Low-Complexity Iterative Sinusoidal Parameter Estimation," *ArXiv Prepr. ArXiv160301824*, 2016, available at: https://arxiv.org/pdf/1603.01824.pdf.

[26] W. Sulandari, Subanar, Suhartono, and H. Utami, "Amplitude-Modulated Sinusoidal Model for The Periodic Components of SSA Decomposition," in *The 2018 International Symposium on Advanced Intelligent Informatics (SAIN) "Revolutionize Intelligent Informatics Spectrum for Humanity,"* Yogyakarta-Indonesia, 2018, pp. 66–71, available at: https://ieeexplore.ieee.org/Xplore/home.jsp.

[27] W. Sulandari, S. Subanar, S. Suhartono, and H. Utami, "An Empirical Study of Error Evaluation in Trend and Multiple Seasonal Time Series Forecasting Based on SSA," *Pak. J. Stat. Oper. Res.*, vol. 14, no. 4, pp. 945–960, 2018, doi: 10.18187/pjsor.v14i4.2443.

[28] D. N. Gujarati, *Basic econometrics*. Tata McGraw-Hill Education, 2009, available at : Google Scholar.

[29] H. Hassani and D. Thomakos, "A review on singular spectrum analysis for economic and financial time series," *Stat. Interface*, vol. 3, no. 3, pp. 377–397, 2010, doi: 10.4310/SII.2010.v3.n3.a11.

[30] J. B. Elsner and A. A. Tsonis, *Singular spectrum analysis: a new tool in time series analysis*. Springer Science & Business Media, 1996, doi: 10.1007/978-1-4757-2514-8.

[31] L. J. Soares and M. C. Medeiros, "Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data," *Int. J. Forecast.*, vol. 24, no. 4, pp. 630–644, 2008, doi: 10.1016/j.ijforecast.2008.08.003.

[32] M. H. Kutner, C. J. Nachtsheim, and J. N. Dr, *Applied Linear Regression Models- 4th Edition with Student CD*, 4 edition. Boston; Montreal: McGraw-Hill Education, 2004, available at : Google Scholar.

[33] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, Oct. 2006, doi: 10.1016/j.ijforecast.2006.03.001.

[34] G. Lebanon, *Bias, variance, and mse of estimators*. Technical Notes, Georgia: Georgia, 2010, available at: http://theanalysisofdata.com/notes/estimators1.pdf.

[35] W. W.-S. Wei, *Time series analysis: Univariate and Multivariate Methods*, 2nd ed. Pearson Addison-Wesley, 2006, avilable at : Google Scholar.

[36] H. Gavin, "The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems," *Dep. Civ. Environ. Eng. Duke Univ.*, pp. 1–15, 2011, available at: http://people.duke.edu/~hpgavin/ce281/lm.pdf.

[37] D. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, Jun. 1963, doi: 10.1137/0111030.

[38] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical analysis*, Springer, 1978, pp. 105–116, doi: 10.1007/BFb0067700.

[39] B. G. Quinn, "Estimating frequency by interpolation using Fourier coefficients," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1264–1268, 1994, doi: 10.1109/78.295186.

[40] S. Makridakis and M. Hibon, "The M3-Competition: results, conclusions and implications," *Int. J. Forecast.*, vol. 16, pp. 451–476, 2000, doi: 10.1016/S0169-2070(00)00057-1.

[41] G. P. Zhang, B. E. Patuwo, and M. Y. Hu, "A simulation study of artificial neural networks for nonlinear time-series forecasting," *Comput. Oper. Res.*, vol. 28, no. 4, pp. 381–396, Apr. 2001, doi: 10.1016/S0305-0548(99)00123-9.