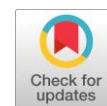


Hybrid SSA-TSR-ARIMA for water demand forecasting

Suhartono ^{a,1,*}, Salafiyah Isnawati ^{a,2}, Novi Ajeng Salehah ^{a,3}, Dedy Dwi Prastyo ^{a,4},
Heri Kuswanto ^{a,5}, Muhammad Hisyam Lee ^{b,6}



^a Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

^b Department of Mathematical Sciences, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

¹ suhartono@statistika.its.ac.id; ² bundo.salaf@gmail.com; ³ noviajeng.s@gmail.com; ⁴ dedy-dp@statistika.its.ac.id;

⁵ kuswanto.its@gmail.com; ⁶ mhl@utm.my

* corresponding author

ARTICLE INFO

Selected paper from The 2018 4th International Conference on Science in Information Technology (ICSITech) (Melaka-Malaysia, 30-31 October 2018) (<http://icsitech.org/>). Peer-reviewed by ICSITech Scientific Committee and Editorial Team of IJAIN journal.

Article history

Received June 13, 2018

Revised August 29, 2018

Accepted September 20, 2018

Available online November 11, 2018

Keywords

Singular spectrum analysis

Time series regression

Automatic ARIMA

Hybrid method

Water demand forecasting

ABSTRACT

Water supply management effectively becomes challenging due to the human population and their needs have been growing rapidly. The aim of this research is to propose hybrid methods based on Singular Spectrum Analysis (SSA) decomposition, Time Series Regression (TSR), and Automatic Autoregressive Integrated Moving Average (ARIMA), known as hybrid SSA-TSR-ARIMA, for water demand forecasting. Monthly water demand data frequently contain trend and seasonal patterns. In this research, two groups of different hybrid methods were developed and proposed, i.e. hybrid methods for individual SSA components and for aggregate SSA components. TSR was used for modeling aggregate trend component and Automatic ARIMA for modeling aggregate seasonal and noise components separately. Firstly, simulation study was conducted for evaluating the performance of the proposed methods. Then, the best hybrid method was applied to real data sample. The simulation showed that hybrid SSA-TSR-ARIMA for aggregate components yielded more accurate forecast than other hybrid methods. Moreover, the comparison of forecast accuracy in real data also showed that hybrid SSA-TSR-ARIMA for aggregate components could improve the forecast accuracy of ARIMA model and yielded better forecast than other hybrid methods. In general, it could be concluded that the hybrid model tends to give more accurate forecast than the individual methods. Thus, this research in line with the third result of the M3 competition that stated the accuracy of hybrid method outperformed, on average, the individual methods being combined and did very well in comparison to other methods.



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



1. Introduction

Due to the rapid expansion of human population and their needs, water supply management becomes challenging to be done effectively [1]. In real application, one of the most important terms for the government water company is satisfying the consumer demand, even though it might cause energy wastage issue and financial problem for the company. Therefore, a precise, efficient and accurate forecasting method is required for cost-effective and sustainable management planning. Moreover, water consumption has significantly increased in the last decades. It is also known that season affects water consumption in the community. Hence, it indicates trend and seasonal pattern in the water demand data.

Previous researches showed that water demand forecasting can use either linear approaches e.g. linear regression or ARIMA [2]-[3] or nonlinear approach [4]-[8]. Otherwise, decomposing patterns in time

series data into smaller subparts can simplify the forecasting process in time series analysis [9]. Hence, a forecasting method that can capture and reconstruct every pattern of components in the data was needed. One of the approaches which can decompose data into trend, seasonal, and oscillatory components is Singular Spectrum Analysis or SSA [10]–[11]. Some researchers have proven that forecasting accuracy can be improved by extracting the pattern in data using SSA. Zhang *et al.* [12] used SSA-ARIMA for forecasting annual runoff data, and they concluded that the SSA-ARIMA model produced more accurate forecast than other individual methods. Moreover, Liu *et al.* [13] also applied SSA-ARIMA for software reliability forecasting and obtained the same conclusion. In addition, another hybrid models such as SSA-NN [14]–[16] and SSA-ANFIS [17]–[18] were also widely developed for water demand forecasting.

This research focused on propose hybrid models based on SSA, Time Series Regression (TSR), and Automatic ARIMA, known as SSA-TSR-ARIMA, for water demand forecasting. Due to automatic forecasting issue in big data analysis, Automatic ARIMA was used because it simplified model, generated models quickly and more robust to unusual time series patterns without user intervention [19]. Two group of different hybrid methods were developed and proposed, i.e. hybrid methods for individual SSA components and for aggregate SSA components. Finally, a comparative study about forecast accuracy was done by applying other forecasting methods, such as ARIMA for actual data, and hybrid SSA-ARIMA for individual and aggregate SSA components.

2. Method

This section discusses about the literature review of the methods that be used in this study.

2.1. Autoregressive Integrated Moving Average (ARIMA)

The Box-Jenkins procedure is one of the most popular procedures for time series analysis and forecasting application with ARIMA model [20]. The general form of $ARIMA(p, d, q)(P, D, Q)^S$ is [21].

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\Theta_Q(B^S)a_t \quad (1)$$

where

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 \dots - \phi_p B^p),$$

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q),$$

$$\Phi_P(B^S) = (1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS}),$$

$$\Theta_Q(B^S) = (1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS}).$$

In general, the Box-Jenkins procedure consists of model identification, estimation and testing of parameter, diagnostic checking and selecting best model, and forecasting step [21].

2.2. Automatic ARIMA

Recently, automatic forecasting becomes one of the main topics in time series analysis particularly forecasting in big data analysis. One of the most popular automatic forecasting algorithms was developed based on ARIMA models, such as **auto.arima** function in R software that be proposed by Hyndman & Khandakar [19]. This research used this **auto.arima** function for applying automatic forecasting. Classical ARIMA methods sometimes tend to be subjective and complicated, particularly in model identification step using Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) plot.

The determination of the best ARIMA model by **auto.arima** function is based on the minimum of AIC value. Stationarity condition in this automatic ARIMA method is tested by unit root test, i.e. Augmented Dickey-Fuller (ADF) test. Automatic ARIMA does not use ACF and PACF plots in determining the best model. This function tries one by one for each possible model, where order p and q start from 0 to 5, P and Q start from 0 to 2, d is 0 to 2, and D is 0 or 1 [19].

2.3. Singular Spectrum Analysis

Singular Spectrum Analysis (SSA) is known as a powerful method for time series analysis. SSA combines elements of classical time series analysis, multivariate statistics, multivariate geometric, dynamical system, and signal processing [22]. The main purpose of SSA is to decipher the original series into a small number of identifiable components such as trend, seasonal, and oscillatory, followed by the reconstruction of the original series [23]. There are two main stages in SSA as follows:

2.3.1. Decomposition (embedding and singular value decomposition)

Given a real-valued time series (Y_1, Y_2, \dots, Y_n) and L is an integer number denoted for window length, $1 < L < n$. In embedding step, the original time series will be mapped into a trajectory matrix which illustrated as follows,

$$\mathbf{X} = [X_1 : \dots : X_K] = \begin{pmatrix} f_1 & f_2 & f_3 & \dots & f_K \\ f_2 & f_3 & f_4 & \dots & f_{K+1} \\ f_3 & f_4 & f_5 & \dots & f_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_L & f_{L+1} & f_{L+2} & \dots & f_n \end{pmatrix} \tag{2}$$

where $K = n - L + 1$ and $X_i = (f_i, f_{i+1}, \dots, f_{i+L-1})^T$, $1 \leq i \leq K$. Let $\lambda_1, \lambda_2, \dots, \lambda_L$ be the eigenvalues of the covariance matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ and U_1, U_2, \dots, U_L are the corresponding eigenvectors. That eigenvalues are arranged in a decreasing order, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$. In second step, the SVD of matrix \mathbf{X} can be stated as follows,

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d \tag{3}$$

where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$. The set $(\sqrt{\lambda_i}, U_i, V_i)$ is called i -th eigentriple to SVD.

2.3.2. Reconstruction (grouping and diagonal averaging)

In grouping step, the set of indices $\{1, 2, \dots, d\}$ will be partitioned into m disjoint subsets I_1, I_2, \dots, I_m and let $I = \{i_1, i_2, \dots, i_p\}$. Then the resultant matrix \mathbf{X}_I corresponds to group I defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \mathbf{X}_{i_2} + \dots + \mathbf{X}_{i_p}$. Computing these matrixes for groups $I = I_1, I_2, \dots, I_m$ and lead to decomposition form $\mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_m}$. The set I_1, I_2, \dots, I_m are called eigentriple groupings.

In the last step, each elementary matrix in the grouped decomposition is transformed into a new component series. Let y_{ij} be the element of matrix \mathbf{Y} ($L \times K$ matrix), $1 \leq i \leq L, 1 \leq j \leq K$, for $L \leq K$. Given the values of $L^* = \min(L, K)$, $K^* = \max(L, K)$ and $n = L + K - 1$. Let $y_{ij}^* = y_{ij}$. If $L < K$ and let $y_{ij}^* = y_{ji}$ if $L > K$ otherwise. Diagonal averaging transforms matrix \mathbf{Y} into the series g_1, g_2, \dots, g_n by the formula,

$$g_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k y_{m, k-m+1}^* & \text{for } 1 \leq k < L^*, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+1}^* & \text{for } L^* \leq k < K^* \\ \frac{1}{n-k+1} \sum_{m=k-K^*+1}^{n-K^*+1} y_{m, k-m+1}^* & \text{for } K^* \leq k \leq n \end{cases} \tag{4}$$

The equation corresponds to the average matrix element over the ‘antidiagonals’ $i + j = k + 1$. If the averaging diagonal is applied to the matrix \mathbf{X}_{jk} then a reconstructed series $F^{(k)} = (f_1^{(k)}, f_2^{(k)}, \dots, f_n^{(k)})$ with length n will be obtained. Therefore, the initial series can be reconstructed by summation as follows,

$$f_j = \sum_{k=1}^m f_j^{(k)} \quad , \quad j = 1, 2, \dots, n \tag{5}$$

2.4. Time Series Regression

Basically, TSR is the same as regression, particularly with regression with dummy variables. In this research, the TSR model is a model for handling trend and seasonal components separately. In general, the trend is defined as the long-term direction that is continuously up or down, and seasonality is a repeating pattern with the same period, for example, 12 months per year [24].

The TSR model for trend pattern can be described as polynomial regression as follows:

$$\hat{T}_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_m t^m \tag{6}$$

whereas, the TSR model for seasonal pattern in general is written as follows:

$$\hat{S}_t = \beta_1 D_1 + \beta_2 D_2 + \dots + \beta_s D_s \tag{7}$$

where D_j (for $j = 1, 2, \dots, s$) are dummy variables for seasonal component.

2.5. Hybrid SSA-ARIMA

The SSA method decomposes the data (Z_t) into several subpart series (PC_t). The hybrid SSA-ARIMA model applies the idea of individual and aggregate component modeling. In this research, the framework of individual and aggregate SSA-ARIMA modeling are shown in Fig. 1 and Fig. 2, respectively [12]-[13].

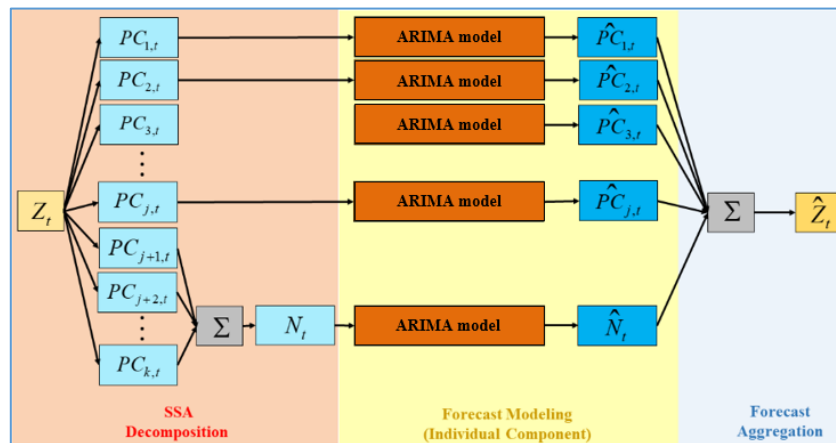


Fig. 1. Hybrid SSA-ARIMA of individual component for time series forecasting

The individual forecasting is done by modeling each series using the ARIMA method and calculate the fitted value for each eigentriple. However, specifically for eigentriple which has noise pattern, the modeling is done in aggregate, i.e. noise eigentriple is combined into noise component (N_t). This component will be fitted using auto ARIMA, so the forecast result of noise component (\hat{N}_t) is formed. Finally, the forecast was made by each SSA-component fitting have been summed up to make final forecast (\hat{Z}_t).

Fig. 2 represents the framework of Aggregate SSA-ARIMA modeling. In contrast to individual SSA-ARIMA modeling, this model is performed by incorporating similar eigentriple. Eigentriples that have

a similar pattern will be summed into one component. Eigentriples that have a trend pattern, summed into trend component (T_t), and so do for seasonal (S_t) and noise (N_t) component. Automatic ARIMA modeling is done on each SSA-component to produce the fitted values of trend (\hat{T}_t), seasonal (\hat{S}_t), and noise (\hat{N}_t) component. The summation of each SSA-component fitting will lead to final forecast of the series (\hat{Z}_t). Both in individual and aggregate modeling, the trend, seasonal, and noise series will be approximated by Automatic ARIMA.

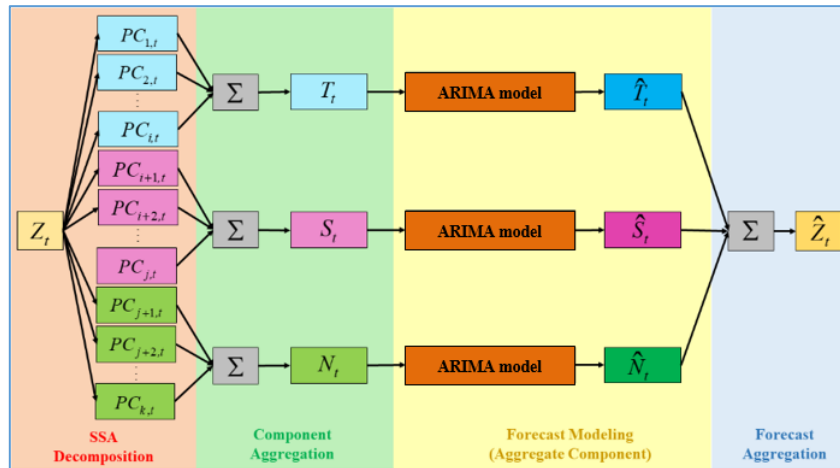


Fig. 2. Hybrid SSA-ARIMA of aggregate component for time series forecasting

2.6. The Proposed Hybrid SSA-TSR-ARIMA

The proposed hybrid method is mainly based on TSR for fitting trend component in aggregate modeling scenario. The idea is motivated by the trend component of SSA-decomposition tend to follow polynomial trend pattern. Thus, TSR as equation (6) will capture well this pattern. In general, the proposed hybrid SSA-TSR-ARIMA method is illustrated as Fig. 3.

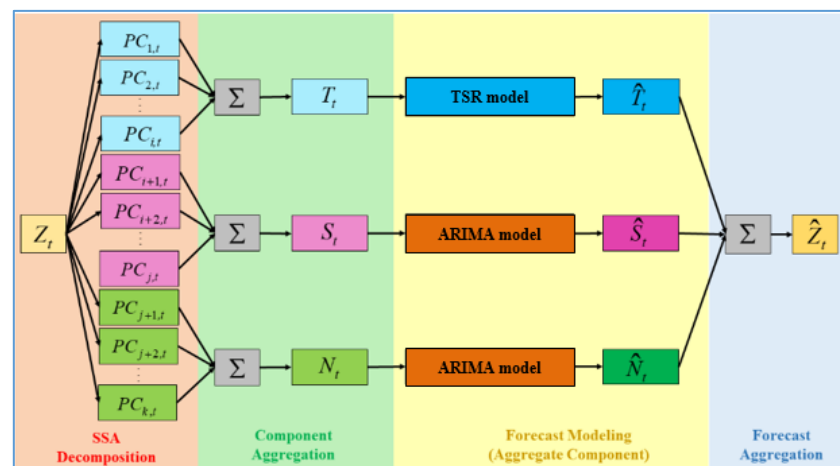


Fig. 3. The proposed hybrid SSA-TSR-ARIMA for time series forecasting

2.7. Model Evaluation

The model evaluation is done based on both in-sample (training) and out-sample (testing) criteria. Automatic ARIMA uses in-sample criteria for selecting the best model based on the smallest AIC or Akaike's Information Criterion as follows:

$$AIC = n \ln \hat{\sigma}_a^2 + 2C \tag{8}$$

where $\hat{\sigma}_a^2$ is maximum likelihood estimation of σ_a^2 , and C is number of parameters.

Furthermore, the best model in this research was selected based on out-sample (testing) criteria as known as cross-validation principle. Two criteria in testing data for determining the best model are Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) that be calculated as follows:

$$RMSE = \sqrt{\frac{1}{R} \sum_{r=1}^R (Y_{n+r} - \hat{Y}_n(r))^2} \quad (9)$$

$$MAPE = \frac{1}{R} \sum_{r=1}^R \left| \frac{Y_{n+r} - \hat{Y}_n(r)}{Y_{n+r}} \right| \times 100\% \quad (10)$$

where R is the forecast periods [25].

2.8. Real Data of Water Demand in Wonogiri Regency, Indonesia

PDAM or Regional Company for Water Utility is one of the regional owned business units which is engaged in distributing and providing fresh water for the public. PDAMs exist in every province, district, and municipality throughout Indonesia. Wonogiri regency is one of the districts in Central Java province, Indonesia. The water needs are managed by PDAM in Wonogiri Regency based on the instruction of the Minister of Home Affairs. Water demand could be detected by monitoring the amount of fresh water consumed by customers. The calculation of water demand in Wonogiri is done every 20th date of the corresponding month. Water demand data in February is calculated from 21 January to 20 February. The water demand in March is calculated from February 21 to March 20, and so on. Then, the collected time series data will be used as a representative data on this work.

3. Results and Discussion

This study conducted simulation and empirical study, and this section discussed about the results and analysis of both studies.

3.1. Simulation study

The purpose of the simulation study is to know the performance of the proposed hybrid method to forecast the time series with trend and seasonal pattern. As discussed in the previous section, the hybrid SSA-ARIMA method will be applied individually (written as first model or M-1) or aggregately (written as second model or M-2), and the proposed hybrid SSA-TSR-ARIMA method in aggregate scheme (written as third model or M-3). This study simulation was conducted by assuming for monthly data with periods from January 2006 to December 2017 or generated for 144 observations where 120 initial data as training dataset and last 24 data as testing dataset. The simulation data that containing trend, seasonal, and noise components were generated as follows:

$$Y_t = T_t + S_t + N_t^* \quad (11)$$

where $T_t = 0.5 t$ for trend, $S_t = 7.5 \sin\left(\frac{2\pi t}{12}\right)$ for seasonal, and $N_t^* \sim N(0,1)$ for noise components.

The time series plot of the simulation data is shown in Fig. 4.

Each original series was decomposed by using SSA. In this simulation data, 50 eigentriple formed with 4 main eigentriples and 46 noise eigentriples. The main eigentriple consists of two trend eigentriples and two seasonal eigentriples. Similar patterned eigentriples were grouped in to the same component. In general, the trend component consists of two trend-patterned eigentriples, a seasonal component consisting of two seasonal-patterned eigentriples, and a noise component composed of 46 eigentriple noise patterns.

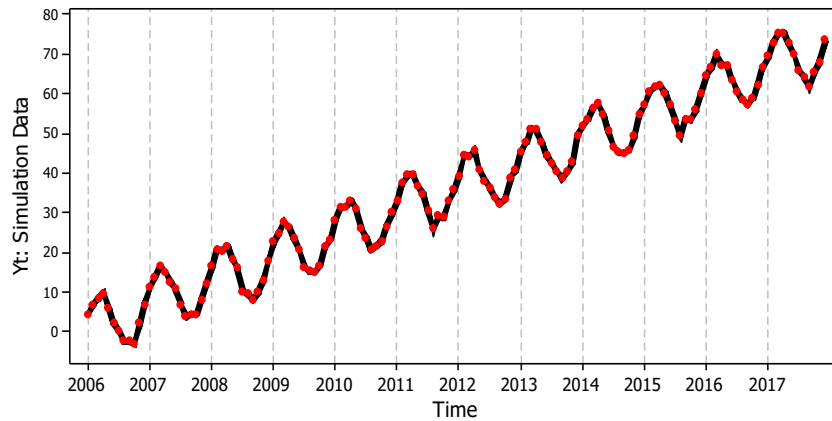


Fig. 4. Time Series Plot of Simulation Data that contain trend and seasonal

Then, these decompositions series were modeled using hybrid SSA-ARIMA in individual and aggregate scheme, and hybrid SSA-TSR-ARIMA in aggregate scheme. The results were showed at Table 1.

Table 1. Forecast Accuracy of Hybrid Methods in Simulation Data

Method	ARIMA Model	AIC	Testing Dataset	
			MAPE	RMSE
<i>M-1: SSA-ARIMA (individually)</i>			12.69%	11.09
PC_1	(1,2,1)	-722.22		
PC_2	(1,2,4)(1,1,0) ¹²	-555.60		
PC_3	(0,0,0)(1,1,0) ¹²	187.17		
PC_4	(3,2,0)	-567.71		
N_t^*	(0,0,1)	309.49		
<i>M-2: SSA-ARIMA (aggregate)</i>			7.59%	5.66
T_t	(1,2,1)	-722.22		
S_t	(0,1,0)(1,1,0) ¹²	-83.14		
N_t^*	(0,0,1)	309.49		
<i>M-3: SSA-TSR-ARIMA (aggregate)</i>			2.83%	2.20
T_t	* \hat{T}_t	-		
S_t	(0,1,0)(1,1,0) ¹²	-83.14		
N_t^*	(0,0,1)	309.49		

$$* \hat{T}_t = 0.824 + 0.467 t + 0.000266 t^2$$

Table 1 showed that the proposed hybrid SSA-TSR-ARIMA in aggregate scheme outperformed other hybrid models, i.e. yielded the lowest MAPE and RMSE. Hence, it indicated that in time series data containing trend and seasonal patterns, the hybrid SSA-TSR-ARIMA under aggregate scheme produced the most accurate forecast at the testing dataset.

3.2. Water Demand Forecasting

Consumer behavior in water demand can be seen from the characteristics of water consumption data. This information can be used to make water resources planning to minimize the waste of water resources and company losses. The monthly of water demand by community in Wonogiri regency, Indonesia during the period January 2006 to August 2017 is shown in Fig. 5.

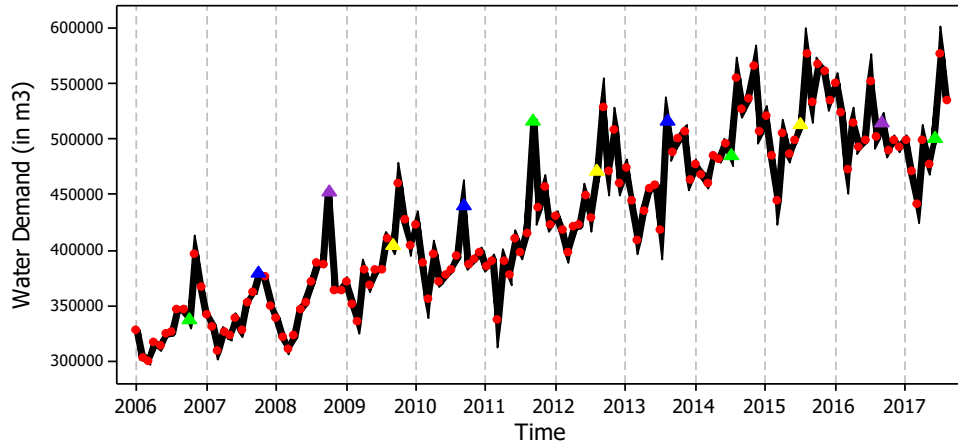


Fig. 5. Time Series Plot of Monthly Water Demand in Wonogiri Regency, Indonesia

The monthly water demand increased over time and has seasonal pattern (Fig. 5). There are 140 observations with average monthly water demand of 428.7 thousand m³. In this period, the smallest demand occurred in March 2006 and the highest is in July 2017, i.e. 299.4 and 576.3 thousand m³, respectively. Moreover, due to dry season in Wonogiri regency, the average water demand in August to November tends to be higher than other months.

3.2.1. Hybrid SSA-ARIMA

SSA decomposition of the monthly water demand data generated 50 eigentriples (as shown in Fig. 6a), which in the SVD stage produces 3 main eigentriples with 1 trend eigentriple and 2 seasonal eigentriples (as shown in Fig. 6b). The noise component consists of 4 to 50 eigentriples. Fig. 6 and Fig. 7 illustrate the reconstruction results of trend, seasonal, and noise components of monthly water demand data.

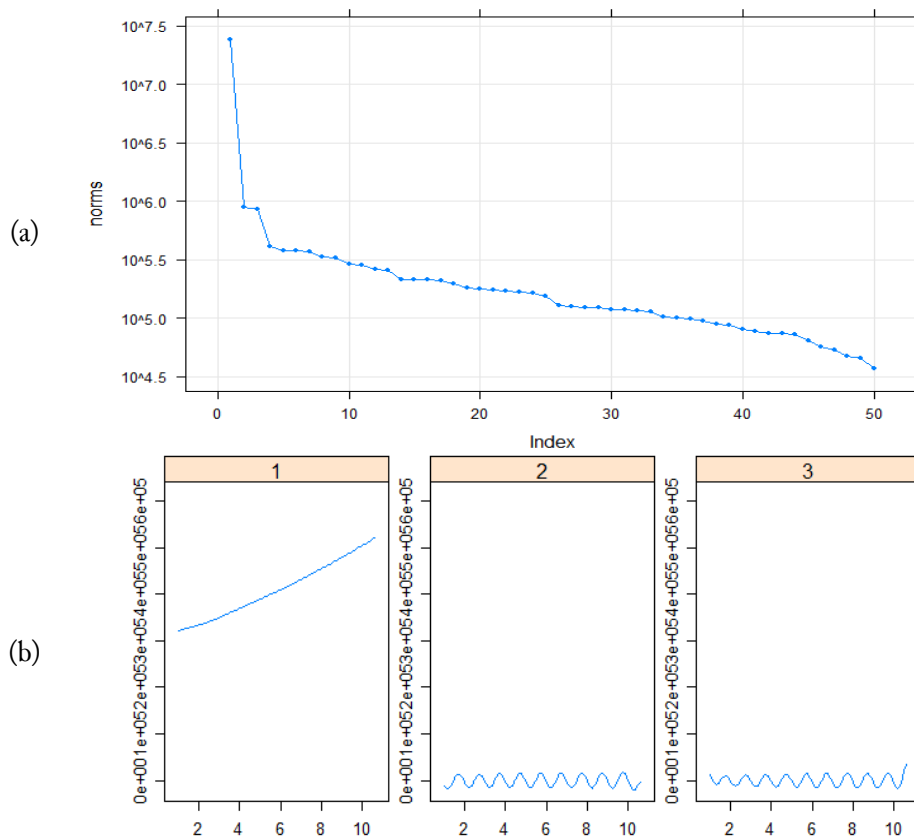


Fig. 6. Eigentriples (a) and Principal Component (b) plot of water demand

Due to the results of simulation study showed that hybrid SSA-ARIMA in aggregate scheme give better forecast than individual scheme, then monthly water demand data were analyzed only by hybrid SSA-ARIMA in aggregate scheme. Each component was modeled using Automatic ARIMA and the best model for trend, seasonal, and noise components were ARIMA(1,2,0), ARIMA(1,0,0)(1,1,1)¹², and ARIMA(1,0,2)(1,0,0)¹², respectively. The MAPE and RMSE of this hybrid model in training dataset were 5.26 and 45.900, respectively, whereas at testing dataset were 13.02 and 79.039, respectively.

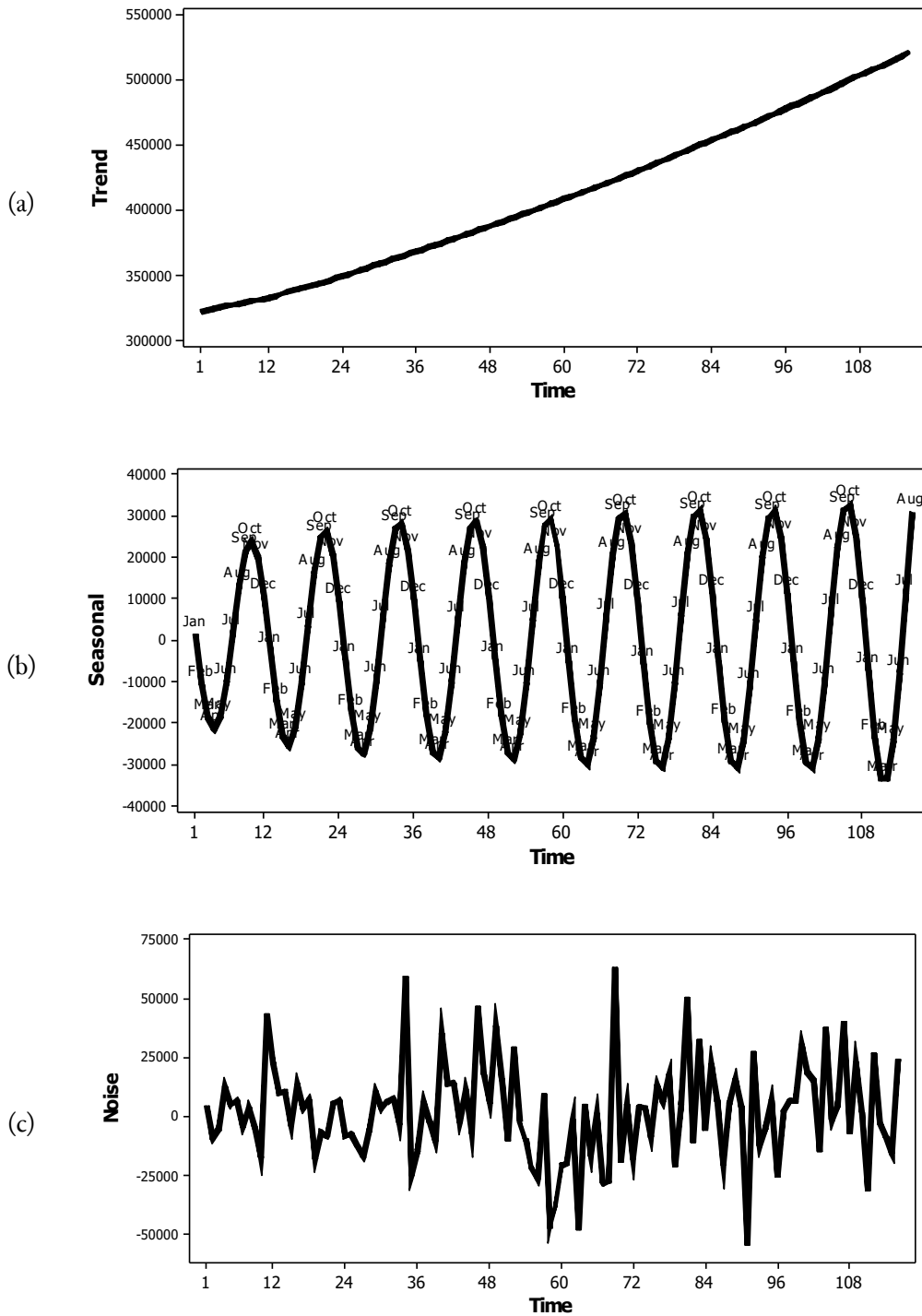


Fig. 7. Grouping of Trend (a), Seasonal (b), and Noise (c) Components of Water Demand Data

3.2.2. Hybrid SSA-TSR-ARIMA

The main difference between hybrid SSA-TSR-ARIMA and SSA-ARIMA in aggregate scheme is the modeling of trend component using polynomial regression instead of ARIMA model. The model for trend component obtained from TSR is,

$$\hat{T}_t = 318079 + 1225,87t + 4,4867t^2 \quad (12)$$

Moreover, the MAPE and RMSE in training dataset of this hybrid SSA-TSR-ARIMA method were 3.82 and 20.398, respectively, whereas at testing dataset were 9.37 and 58.751, respectively.

3.2.3. The Results of ARIMA Model

The best ARIMA models for monthly water demand data was ARIMA(2,1,[35])(1,1,0)¹². This model has fulfilled white noise and normally distributed assumptions for the residual. The equation of this ARIMA model is as follows:

$$Z_t^* = 0.34 Z_{t-1}^* + 0.36 Z_{t-2}^* - 0.29 Z_{t-3}^* + 0.71 Z_{t-12}^* + 0.12 Z_{t-13}^* - 0.25 Z_{t-14}^* - 0.20 Z_{t-15}^* + \\ 0.29 Z_{t-24}^* - 0.10 Z_{t-25}^* - 0.10 Z_{t-26}^* - 0.08 Z_{t-27}^* + a_t + 0.50 a_{t-35} \quad (13)$$

where $Z_t^* = 1/Z_t$. The MAPE and RMSE in training dataset of this ARIMA model were 4.24 and 22.940, respectively, whereas at testing dataset were 9.64 and 60.836, respectively.

3.2.4. Model Selection and Forecast Accuracy Comparison

Three forecasting methods have been applied for forecasting monthly water demand data in Wonogiri regency, i.e. ARIMA, hybrid SSA-ARIMA, and hybrid SSA-TSR-ARIMA methods. The actual and forecasted values of water demand at testing dataset are shown in Fig. 8.

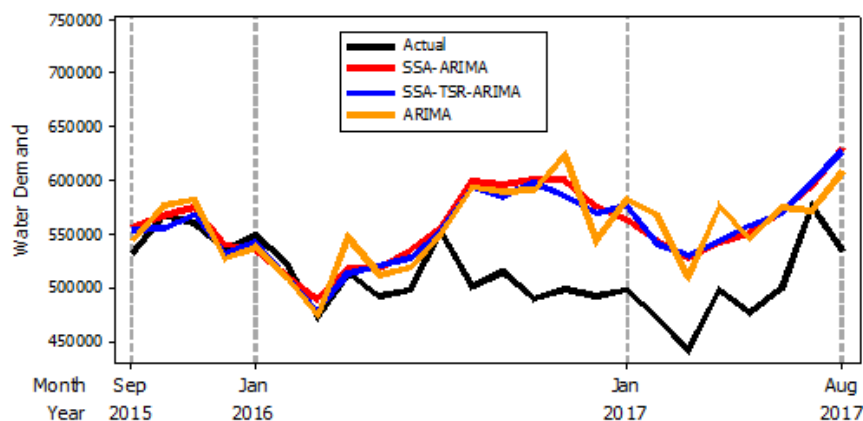


Fig. 8. The Actual and Forecast Values of Water Demand at Testing Dataset

Moreover, a numerical evaluation based on RMSE and MAPE for both training and testing datasets to provide more precise comparison on the performance of this water demand forecasting from each model was done and presented in Table 2. Column ratio of MAPE and RMSE showed the ratio values to ARIMA model.

In general, the results at Table 2 provided two contradictive conclusions about the hybrid or combination methods (or complex method) compared to individual or simple methods. The first comparison between ARIMA and hybrid SSA-ARIMA showed that ARIMA as a simple method yielded more accurate forecast than hybrid SSA-ARIMA as a complex method. Moreover, this first result was in line with the first result of M3 competition that stated statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones [26]. Otherwise, this first result did not in line with the results from Zhang et al. [12] and Liu et al. [13].

Table 2. Comparison Results of Accuracy Performance for each Model

Model	Training Dataset		Testing Dataset		Ratio	
	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
<i>SSA-ARIMA</i>	5.26	45.900	13.02	79.039	1.35	1.32
<i>SSA-TSR-ARIMA</i>	3.82	20.398	9.37	58.751	0.97	0.97
<i>ARIMA</i>	4.24	22.940	9.64	60.836	1.00	1.00

Furthermore, the second comparison between hybrid SSA-TSR-ARIMA, hybrid SSA-ARIMA, and ARIMA showed that hybrid SSA-TSR-ARIMA method as combination method provided better forecast than hybrid SSA-ARIMA as a simpler method and ARIMA as an individual method. Thus, this second result was in line with the third result of M3 competition that stated the accuracy when various methods were being combined outperforms, on average, the individual methods being combined and did very well in comparison to other methods [26]–[28]. In addition, this second result also showed that forecaster must understand well the advantage of each model before combining, particularly the relation between model and the time series pattern that could be handled it.

In general, the results on Table 2 showed that the best method for water demand forecasting in Wonogiri regency was the proposed hybrid SSA-TSR-ARIMA in aggregate scheme. Moreover, this proposed method provided more accurate forecast both in training and testing dataset by using MAPE and RMSE criteria. The greater values of MAPE and RMSE in testing dataset compared to in training dataset was caused by the fluctuation of trend behavior change lower in testing dataset. If there is such a change due to a special event, the resulting forecast of SSA decomposition tends to be grouped into other components and can produce misleading forecasts [29].

4. Conclusion

This research proposed a hybrid SSA-TSR-ARIMA method for water demand forecasting. The results of simulation study showed that hybrid SSA-ARIMA method in aggregate scheme yielded more accurate forecast than individual scheme. Moreover, the simulation study also showed that the proposed hybrid SSA-TSR-ARIMA method in aggregate scheme provided better forecast than hybrid SSA-ARIMA method both in aggregate and individual scheme. Additionally, the results of simulation study also showed that SSA decomposition was appropriately applied on time series data containing trend, seasonal, and noise components [30].

Furthermore, the results of real data about monthly water demand forecasting in Wonogiri regency showed that the proposed hybrid SSA-TSR-ARIMA in aggregate scheme as a combining some individual methods provided more accurate forecast than hybrid SSA-ARIMA as a simpler method and ARIMA as an individual method. Thus, these results in line with the third result of M3 forecasting competition that be done by Makridakis & Hibon [26]. Moreover, this study also clearly shown that by applying TSR for fitting the trend component of SSA decomposition can increase the accuracy of the prediction compared to only applying ARIMA directly to actual data or to each SSA component. This leads a suggestion for further research, particularly by involving nonlinear methods (such as neural networks, deep learning networks, or ANFIS) when combining to SSA for handling nonlinearity issue in time series forecasting.

Acknowledgment

This study was supported by DRPM-DIKTI under scheme of “Penelitian Berbasis Kompetensi”, project No. 851/PKS/ITS/2018. The authors thank to the General Director of DIKTI for funding and to anonymous referees for their useful suggestions.

References

- [1] W. J. Cosgrove and D. P. Loucks, “Water management: Current and future challenges and research directions,” *Water Resour. Res.*, vol. 51, no. 6, pp. 4823–4839, Jun. 2015, doi: <https://doi.org/10.1002/2014WR016869>.

- [2] J. Adamowski, H. Fung Chan, S. O. Prasher, B. Ozga-Zielinski, and A. Sliusarieva, "Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada," *Water Resour. Res.*, vol. 48, no. 1, Jan. 2012, doi: <https://doi.org/10.1029/2010WR009945>.
- [3] A. S. Polebitski and R. N. Palmer, "Seasonal residential water demand forecasting for census tracts," *J. Water Resour. Plan. Manag.*, vol. 136, no. 1, pp. 27–36, 2009, doi: [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000003](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000003).
- [4] S. P. Zhang, H. Watanabe, and R. Yamada, "Prediction of Daily Water Demands by Neural Networks," 1994, pp. 217–227, doi: https://doi.org/10.1007/978-94-017-3083-9_17.
- [5] J. Bougadis, K. Adamowski, and R. Diduch, "Short-term municipal water demand forecasting," *Hydrol. Process. An Int. J.*, vol. 19, no. 1, pp. 137–148, 2005, doi: <https://doi.org/10.1002/hyp.5763>.
- [6] J. F. Adamowski, "Peak daily water demand forecast modeling using artificial neural networks," *J. Water Resour. Plan. Manag.*, vol. 134, no. 2, pp. 119–128, 2008, doi: [https://doi.org/10.1061/\(ASCE\)0733-9496\(2008\)134:2\(119\)](https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(119)).
- [7] M. Ghiassi, D. K. Zimbra, and H. Saidane, "Urban water demand forecasting with a dynamic artificial neural network model," *J. Water Resour. Plan. Manag.*, vol. 134, no. 2, pp. 138–146, 2008, doi: [https://doi.org/10.1061/\(ASCE\)0733-9496\(2008\)134:2\(138\)](https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(138)).
- [8] M. Herrera, L. Torgo, J. Izquierdo, and R. Pérez-García, "Predictive models for forecasting hourly urban water demand," *J. Hydrol.*, vol. 387, no. 1–2, pp. 141–150, Jun. 2010, doi: <https://doi.org/10.1016/j.jhydrol.2010.04.005>.
- [9] B. L. Bowerman and R. T. O'Connell, "Forecasting and time series: An applied approach. 3rd," 1993, available at : <http://ecsocman.hse.ru/text/19151946/>.
- [10] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure*, 2001, vol. 90, doi: <https://doi.org/10.1201/9781420035841>.
- [11] J. Liao, L. Gao, and X. Wang, "Numerical Simulation and Forecasting of Water Level for Qinghai Lake Using Multi-Altimeter Data Between 2002 and 2012," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 2, pp. 609–622, Feb. 2014, doi: <https://doi.org/10.1109/JSTARS.2013.2291516>.
- [12] 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, Sep. 2011, doi: <https://doi.org/10.1007/s11269-011-9833-y>.
- [13] G. Liu, D. Zhang, and T. Zhang, "Software Reliability Forecasting: Singular Spectrum Analysis and ARIMA Hybrid Model," in *2015 International Symposium on Theoretical Aspects of Software Engineering*, 2015, pp. 111–118, doi: <https://doi.org/10.1109/TASE.2015.19>.
- [14] S. L. Zubaidi, J. Dooley, R. M. Alkhaddar, M. Abdellatif, H. Al-Bugharbee, and S. Ortega-Martorell, "A Novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks," *J. Hydrol.*, vol. 561, pp. 136–145, Jun. 2018, doi: <https://doi.org/10.1016/j.jhydrol.2018.03.047>.
- [15] M. Sun, X. Li, and G. Kim, "Precipitation analysis and forecasting using singular spectrum analysis with artificial neural networks," *Cluster Comput.*, Jan. 2018, doi: <https://doi.org/10.1007/s10586-018-1713-2>.
- [16] L. Latifoğlu, Ö. Kişi, and F. Latifoğlu, "Importance of hybrid models for forecasting of hydrological variable," *Neural Comput. Appl.*, vol. 26, no. 7, pp. 1669–1680, Oct. 2015, doi: <https://doi.org/10.1007/s00521-015-1831-1>.
- [17] Y. Xiao, J. J. Liu, Y. Hu, Y. Wang, K. K. Lai, and S. Wang, "A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting," *J. Air Transp. Manag.*, vol. 39, pp. 1–11, Jul. 2014, doi: <https://doi.org/10.1016/j.jairtraman.2014.03.004>.
- [18] M. Abdollahzade, A. Miranian, H. Hassani, and H. Iranmanesh, "A new hybrid enhanced local linear neuro-fuzzy model based on the optimized singular spectrum analysis and its application for nonlinear and chaotic time series forecasting," *Inf. Sci. (Ny)*, vol. 295, pp. 107–125, Feb. 2015, doi: <https://doi.org/10.1016/j.ins.2014.09.002>.

- [19] R. J. Hyndman and Y. Khandakar, *Automatic time series for forecasting: the forecast package for R*, no. 6/07. Monash University, Department of Econometrics and Business Statistics, 2007, available at: http://webdoc.sub.gwdg.de/ebook/serien/e/monash_univ/wp6-07.pdf.
- [20] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed. Prentice Hall, 1994, available at: <https://books.google.com/books?id=sRzvAAAAMAAJ>.
- [21] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Addison Wesley, 2006, available at: <https://books.google.com/books?id=aY0QAQAIAAJ>.
- [22] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*, 2013, doi: <https://doi.org/10.1007/978-3-642-34913-3>.
- [23] H. Hassani, "Singular spectrum analysis: methodology and comparison," *J. Data Sci.*, vol. 5, no. 2, pp. 239–257, 2007, available at : <https://mpra.ub.uni-muenchen.de/4991/>.
- [24] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018, available at: https://books.google.com/books?id=_bBhDwAAQBAJ.
- [25] 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: <https://doi.org/10.1016/j.ijforecast.2006.03.001>.
- [26] S. Makridakis and M. Hibon, "The M3-Competition: results, conclusions and implications," *Int. J. Forecast.*, vol. 16, no. 4, pp. 451–476, Oct. 2000, doi: [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1).
- [27] Suhartono and M. H. Lee, "A Hybrid Approach based on Winter's Model and Weighted Fuzzy Time Series for Forecasting Trend and Seasonal Data," *J. Math. Stat.*, vol. 7, no. 3, pp. 177–183, 2011, doi: <https://doi.org/10.3844/jmssp.2011.177.183>.
- [28] Suhartono, I. Puspitasari, M. S. Akbar, M. H. Lee, "Two-level seasonal model based on hybrid ARIMA-ANFIS for forecasting short-term electricity load in Indonesia," in *Statistics in Science, Business, and Engineering (ICSSBE), 2012 International Conference on*, 2012, pp. 1–5, doi: <https://doi.org/10.1109/ICSSBE.2012.6396642>.
- [29] H. Hassani, A. S. Soofi, and A. A. Zhigljavsky, "Predicting daily exchange rate with singular spectrum analysis," *Nonlinear Anal. Real World Appl.*, vol. 11, no. 3, pp. 2023–2034, Jun. 2010, doi: <https://doi.org/10.1016/j.nonrwa.2009.05.008>.
- [30] W. Sulandari, Suhartono, Subanar, and H. Utami, "Forecasting time series with trend and seasonal patterns based on SSA," in *Science in Information Technology (ICSITech), 2017 3rd International Conference on*, 2017, pp. 648–653, doi: <https://doi.org/10.1109/ICSITech.2017.8257193>.