

## PERAMALAN VOLUME AIR TERJUAL DI PDAM GIRI TIRTA SARI MENGUNAKAN MODEL HIBRIDA *SINGULAR SPECTRUM ANALYSIS* – *SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE*

Nada Zidan Al Ghifari<sup>1\*</sup>, Etik Zukhronah<sup>2</sup>, Respatiwan<sup>3</sup>

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Sebelas Maret, Surakarta

\*Penulis korespondensi: nadazidanal@student.uns.ac.id

### ABSTRAK

Air bersih merupakan komponen utama dalam kehidupan manusia. Penyedia dan pengelola air bersih di Indonesia dilakukan oleh PDAM. Jumlah air bersih yang digunakan masyarakat Wonogiri berfluktuasi dari waktu ke waktu, seperti penjualan volume air di PDAM Giri Tirta Sari Kabupaten Wonogiri. Data volume air terjual dapat dimodelkan dengan melihat pola datanya, kemudian diramalkan nilai yang akan datang. *Singular Spectrum Analysis* (SSA) merupakan metode nonparametrik dalam peramalan data runtun waktu yang penggunaannya fleksibel karena tidak harus memenuhi berbagai asumsi. *Seasonal Autoregressive Integrated Moving Average* (SARIMA) merupakan model yang digunakan untuk meramalkan data runtun waktu yang mengandung pola musiman. Hibrida merupakan model yang menggabungkan dua atau lebih model peramalan untuk meningkatkan akurasi hasil peramalan. Tujuan dari penelitian ini untuk melakukan peramalan volume air terjual di PDAM Giri Tirta Sari menggunakan model hibrida SSA-SARIMA. Data yang digunakan adalah data periode Januari 2014 hingga Juni 2023. Data *training* dimulai dari Januari 2014 hingga Desember 2021 dan data *testing* dimulai dari Januari 2022 hingga Juni 2023. Data *training* dimodelkan dengan SSA. Selanjutnya, residu SSA dimodelkan dengan SARIMA. Peramalan model hibrida SSA-SARIMA diperoleh dengan menjumlahkan hasil peramalan SSA dan SARIMA. Hasil penelitian menunjukkan bahwa model hibrida SSA-SARIMA(1,0,1)(0,1,0)<sup>12</sup> menghasilkan nilai MAPE sebesar 5,978%. Oleh karena itu, model hibrida SSA-ARIMA dapat meramalkan volume air terjual di PDAM Giri Tirta Sari dengan baik.

**Kata kunci:** volume air terjual, SSA, SARIMA, hibrida

### 1 PENDAHULUAN

Air memiliki peranan yang sangat penting dalam kehidupan manusia. Air dibutuhkan manusia tidak hanya untuk diminum, tetapi juga dibutuhkan hampir disetiap aktivitas manusia sehari-hari. Air di bumi terdiri dari 97,5% air laut yang tersebar di samudra dan lautan serta 2,5% air tawar yang tersebar di uap air, danau, sungai, gletser dan es kutub. 69% dari air tawar di bumi berbentuk gletser dan es di kutub yang berfungsi menstabilkan iklim global. Hanya 31% air tawar yang dapat dimanfaatkan oleh milyaran jiwa manusia di dunia. Ketersediaan air yang dapat digunakan manusia terbatas, yaitu hanya kurang dari 1% air tawar. Hal ini berbanding terbalik dengan besarnya kebutuhan air bersih yang dibutuhkan oleh milyaran manusia di dunia. Keperluan air bersih di Indonesia dikelola oleh beberapa layanan pengelolaan sumber daya air, satu diantaranya adalah PDAM. Data volume air terjual di PDAM Giri Tirta Sari Kabupaten Wonogiri termasuk data runtun waktu yang dapat dimodelkan dengan melihat pola datanya. Pengaruh permintaan dan kebutuhan masyarakat yang tinggi mengakibatkan perubahan kebutuhan air dari waktu ke waktu, pola data tersebut tidak stasioner karena cenderung memuat faktor tren atau musiman.

*Singular Spectrum Analysis* (SSA) merupakan model analisis nonparametrik yang dapat digunakan untuk mengidentifikasi pola tren, musiman dan *noise* pada data runtun waktu (Golyandina dkk., 2001). *Seasonal Autoregressive Integrated Moving Average* (SARIMA) merupakan model yang cocok digunakan untuk data yang memiliki pola musiman (Zhang, 2003). Penggabungan dua atau lebih model peramalan berguna untuk meningkatkan akurasi peramalan dengan memanfaatkan kelebihan dari masing-masing model yang digunakan, dan meminimalisasi kekurangan pada masing-masing model. Salah satu model hibrida yang dapat digunakan adalah hibrida SSA-SARIMA. SSA digunakan untuk mendekomposisi data volume air terjual ke dalam pola tren, musiman, dan *noise*. Selanjutnya, residu dari SSA dimodelkan menggunakan SARIMA. Peramalan model hibrida SSA-SARIMA diperoleh dengan menjumlahkan hasil peramalan SSA dan hasil peramalan residu SSA menggunakan SARIMA. Penelitian sebelumnya mengenai peramalan aliran waduk Biliuhe dan waduk Dahuofang di China dengan model hibrida SSA-ARIMA, ARIMA dan SSA-LRF, diperoleh model terbaik yaitu hibrida SSA-ARIMA (Zhang dkk., 2011). Selain itu, (Darmawan dkk., 2022) melakukan peramalan jumlah penumpang pesawat dan konsentrasi  $CO_2$  pada atmosfer yang terdapat di R *package datasets* dengan metode hibrida SSA dan ARIMA, hasil yang diperoleh dari penelitian ini adalah nilai MAPE sebesar 3,63% dan 0,13%. (Arumsari dkk., 2021) melakukan peramalan inflasi Provinsi Kalimantan Timur tahun 2021 dengan model hibrida SSA-ARIMA dan SSA, hasilnya model hibrida SSA-ARIMA memiliki akurasi yang lebih tinggi dibandingkan dengan model SSA. Irmawati dkk. (2018) melakukan penelitian tentang peramalan Nilai Tukar Petani (NTP) di Jawa Timur dengan metode hibrida SSA-ARIMA dan metode ARIMA, hasilnya model hibrida SSA-ARIMA memiliki akurasi yang lebih tinggi dibandingkan dengan model ARIMA. Hal ini menunjukkan bahwa model hibrida baik digunakan untuk peramalan data runtun waktu. Tujuan dari penelitian ini adalah memodelkan peramalan dan membandingkan hasil peramalan volume air terjual PDAM Giri Tirta Sari Kabupaten Wonogiri menggunakan SSA dan hibrida SSA-SARIMA berdasarkan nilai MAPE. Hasil penelitian ini diharapkan sebagai pendukung keputusan dalam pengambilan kebijakan mengenai volume air terjual.

## 2 METODE

Penelitian ini menggunakan data sekunder yang diperoleh dari PDAM Giri Tirta kabupaten Wonogiri provinsi Jawa Tengah pada tahun 2023. Data yang diperoleh terdiri dari data bulanan volume air terjual di PDAM Giri Tirta kabupaten Wonogiri pada rentang tahun 2014-2023. Data dibagi menjadi dua bagian, yaitu data *training* sebesar 96 data yaitu dari bulan Januari 2014 hingga Desember 2021 dan data *testing* sebesar 18 data yaitu dari bulan Januari 2022 hingga Juni 2023. Model hibrida SSA-SARIMA digunakan dalam meramalkan data runtun waktu.

### 2.1 *Singular Spectrum Analysis* (SSA)

*Singular Spectrum Analysis* (SSA) adalah teknik analisis runtun waktu non parametrik yang digunakan untuk peramalan berdasarkan pada prinsip statistika multivariat dengan tujuan menguraikan komponen runtun waktu asli menjadi sejumlah kecil komponen yang dapat diidentifikasi seperti tren, musiman, dan *noise* (Fajar, 2020). Tahapan model SSA yaitu sebagai berikut.

#### 1) *Embedding*

Prosedur *embedding* memetakan runtun waktu asli ke serangkaian *lagged vector* multidimensi. Runtun waktu  $\mathbb{X} = (x_1, \dots, x_N)$  dengan panjang  $N$ . Data runtun waktu tersebut ditransformasi ke dalam bentuk matriks berukuran  $L \times K$  dengan  $2 < L < \frac{N}{2}$  dan  $K = N - L + 1$ . Prosedur *embedding* membentuk  $K = N - L + 1$  *lagged vectors*.

Didefinisikan matriks lintasan dengan dimensi  $L \times K$ :

$$\mathbf{X} = [X_1: \dots : X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}$$

Matriks lintasan  $\mathbf{X}$  merupakan sebuah matriks Hankel, yang berarti semua elemen pada anti diagonal utama bernilai sama (Febrianti dkk., 2021).

2) *Singular Value Decomposition (SVD)*

Pada tahap ini, matriks lintasan akan dilakukan proses SVD untuk mendapatkan *eigen triple* yang terdiri dari *singular value*  $\sqrt{\lambda_i}$ , *eigenvector*  $\mathbf{U}_i$ , dan *principal component*  $\mathbf{V}_i$ . Tujuan SVD untuk memecah komponen menjadi tiga, yaitu tren, musiman, dan noise.

3) *Pengelompokan atau Grouping*

Proses pengelompokan berfungsi untuk memisahkan komponen aditif dari runtun waktu (Ischak dkk., 2018). Matriks  $\mathbf{X}$  dikelompokkan menjadi beberapa komponen berdasarkan pola pembentukan runtun waktu, yaitu tren, musiman, dan *noise* maka matriks  $\mathbf{X}_I$  berkorespondensi pada kelompok  $I = \{i_1, \dots, i_p\}$  didefinisikan :

$$\mathbf{X}_I = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_p}$$

matriks ini dihitung untuk  $I = I_1, \dots, I_m$  dan ekspansi persamaan sebelumnya menyebabkan dekomposisi yang didefinisikan :

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}$$

Prosedur pemilihan himpunan  $I_1, \dots, I_m$  disebut sebagai pengelompokan *eigen triple* (Golyandina & Zhigljavsky, 2020). Setelah dikelompokkan, plot matriks *w-correlation* antar kelompok dibuat sehingga dapat dilihat adanya *strong separability* yang berarti tidak ada korelasi antar kelompok.

4) *Diagonal Averaging*

Tahapan terakhir dalam SSA adalah mengubah setiap matriks  $\mathbf{X}_{I_j}$  dari dekomposisi yang dikelompokkan diubah menjadi suatu runtun baru dengan panjang  $N$  (Putri dkk., 2023). Misalkan  $\mathbf{Y}$  adalah sebuah matriks berukuran  $L \times K$  dengan unsur-unsur  $y_{ij}$ , dengan  $1 \leq i \leq L, 1 \leq j \leq K$ , dan  $L \leq K$ . Diberikan persamaan  $L^* = \min(L, K)$ ,  $K^* = \max(L, K)$ ,  $n = L + K - 1$ . Misal  $y_{ij}^* = y_{ij}$  jika  $L < K$  dan  $y_{ij}^* = y_{ji}$  jika  $L > K$ . *Diagonal averaging* matriks  $\mathbf{Y}$  ditransformasikan ke dalam bentuk runtun  $y_1, \dots, y_N$  dengan persamaan diagonal averaging yang didefinisikan :

$$y_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k y_{m, k-m+1}^* & ; 1 \leq k < L^* \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+1}^* & ; L^* \leq k < K^* \\ \frac{1}{N-k+1} \sum_{m=k-K^*+1}^{N-K^*+1} y_{m, k-m+1}^* & ; K^* \leq k < N \end{cases}$$

Jika persamaan diagonal averaging diterapkan pada matriks  $\mathbf{X}_{I_k}$ , maka akan diperoleh suatu runtun rekonstruksi  $\tilde{X}^{(k)} = (\tilde{x}_1^{(k)}, \dots, \tilde{x}_N^{(k)})$ . Selanjutnya, runtun awal  $x_1, \dots, x_N$  didekomposisi menjadi jumlahan dari  $m$  hasil rekonstruksi yang didefinisikan :

$$x_n = \sum_{k=1}^m \tilde{x}_n^{(k)}, n = 1, \dots, N$$

Peramalan untuk periode yang akan datang dapat dilakukan jika model telah terbentuk. Terdapat dua metode peramalan dalam SSA, yaitu metode *recurrent (R-Forecasting)* dan *vector (V-Forecasting)*. Peramalan SSA yang digunakan dalam penelitian adalah *SSA recurrent forecasting* (Fajar, 2020). Peramalan model SSA, dibangun dengan bantuan *Linear Recurrence Forecasting (LRF)* (Hassani, 2021). *Eigenvector* yang diperoleh dari proses SVD digunakan untuk menaksir koefisien LRF. Peramalan SSA dilakukan dengan menggunakan

$$\hat{y}_i = \begin{cases} \hat{y}_i & ; i = 1, \dots, N \\ \sum_{j=1}^{L-1} a_j \hat{y}_{i-j} & ; i = N + 1, \dots, N + M \end{cases}$$

Hasil peramalan  $\hat{Y} = (\hat{y}_1, \dots, \hat{y}_{N+M})$  dengan  $\hat{y}_{N+1}, \dots, \hat{y}_{N+M}$  adalah hasil ramalan untuk M periode ke depan.

## 2.2 Seasonal Autoregressive Integrated Moving Average (SARIMA)

Model *Seasonal Autoregressive Integrated Moving Average* (SARIMA) merupakan teknik analisis runtun waktu satu variabel dengan pemodelan dan peramalan dilihat dengan melihat data yang ada tanpa perlu mempertimbangkan variabel lain yang mempengaruhinya. Model SARIMA digunakan untuk data runtun waktu yang berpola musiman (Tokan & Hermawan, 2023). Secara matematis, bentuk umum dari model SARIMA  $(p, d, q)(P, D, Q)^D$  untuk pola data musiman dirumuskan sebagai berikut.

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

dengan

- $\phi_p$  : Koefisien parameter *autoregressive*.
- $\Phi_p$  : Koefisien parameter *autoregressive* periode musiman.
- $B$  : Operator *backshift*.
- $Z_t$  : Nilai runtun waktu pada waktu ke- $t$ .
- $\theta_q$  : Koefisien parameter *moving average*.
- $\Theta_Q$  : Koefisien parameter *moving average* periode musiman.
- $a_t$  : *Error* pada waktu ke- $t$ .
- $d$  : Orde *differencing*.
- $D$  : Orde *differencing* musiman.
- $S$  : Jumlah periode waktu per-musim

Data harus stasioner terhadap rata-rata dan variansi. Apabila data belum stasioner dilakukan *differencing* nonmusiman dan musiman. Orde dari model SARIMA ditentukan berdasarkan plot ACF dan PACF. Model SARIMA terbaik diperoleh dengan tahapan identifikasi model, estimasi parameter, diagnosis model, dan peramalan dengan model terbaik (Christie dkk, 2022).

## 2.3 Hibrida SSA-SARIMA

Model hibrida dalam peramalan runtun waktu merupakan kombinasi beberapa model yang berbeda (Arumsari dkk., 2021). Data yang dimodelkan dengan SSA menghasilkan komponen rekonstruksi dan *noise*. Komponen *noise* dimodelkan dengan SARIMA. Hasil peramalan hibrida SSA-SARIMA diperoleh dengan menjumlahkan hasil peramalan model SSA dan hasil peramalan residu SSA dengan model SARIMA (Ilahi dkk., 2023) . Hasil peramalan model hibrida yaitu:

$$\hat{H}_t = \hat{Y}_t + \hat{Z}_t$$

dengan  $\hat{H}_t$  merupakan hasil peramalan hibrida SSA-SARIMA,  $\hat{Y}_t$  merupakan hasil peramalan SSA,  $\hat{Z}_t$  merupakan hasil peramalan residu SSA menggunakan model SARIMA.

## 2.4 Mean Absolute Percentage Error (MAPE)

Metode peramalan dikatakan baik apabila model yang dihasilkan mampu memberikan hasil peramalan dengan tingkat kesalahan yang rendah (Setyaningrum dkk., 2023). *Mean Absolute Percentage Error* (MAPE) merupakan alat seleksi model berdasarkan rata-rata dari hasil pembagian nilai mutlak *error* dengan data asli.

$$\text{MAPE} = \left( \frac{1}{N} \sum_{t=1}^N \left| \frac{H_t - \hat{H}_t}{H_t} \right| \right) \times 100\%$$

dengan  $H_t$  merupakan data uji *training* atau *testing*, dan  $\hat{H}_t$  merupakan data peramalan.

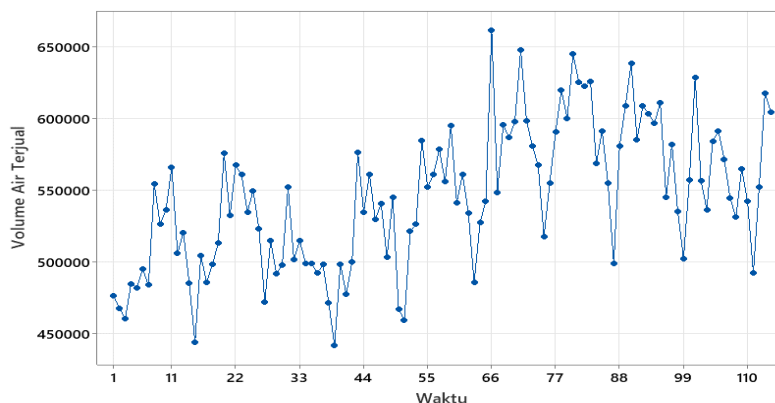
## 2.5 Langkah Analisis

Langkah – langkah analisis data runtun waktu menggunakan model hibrida SSA-SARIMA sebagai berikut :

- 1) Pemodelan SSA.
  - a. Menentukan besarnya parameter *window length* ( $L$ ) dari data *training* untuk membentuk matriks lintasan  $X$  pada proses *embedding*.
  - b. Menentukan *Singular Value Decomposition* (SVD) dari matriks lintasan  $X$  menjadi penjumlahan matriks orthogonal.
  - c. Mengelompokkan (*grouping*) pola hasil dekomposisi ke dalam tiga pola runtun waktu yaitu, tren, musiman, atau *noise*.
  - d. Mengubah matriks hasil dekomposisi yang dihasilkan ke dalam runtun baru pada proses *diagonal averaging*.
  - e. Menghitung nilai residu data *training*.
  - f. Meramalkan data *testing* pada kelompok komponen yang telah terbentuk dengan LRF.
- 2) Pemodelan SARIMA menggunakan data residu dari model SSA yang diperoleh pada poin 1(e).
  - a. Mengidentifikasi pola data residu dari model SSA dengan melihat plot runtun waktu dan plot ACF. Jika data tidak stasioner terhadap variansi maka dilakukan transformasi *Box-Cox*. Jika data tidak stasioner terhadap terhadap rata-rata maka dilakukan *differencing*.
  - b. Mengidentifikasi model SARIMA dugaan berdasarkan plot ACF dan plot PACF dari data yang telah stasioner.
  - c. Mengestimasi parameter model dugaan ARIMA dan uji signifikansi parameter pada setiap model SARIMA dugaan.
  - d. Melakukan pemeriksaan diagnosis, yaitu asumsi *white noise* dan normalitas pada residu setiap model SARIMA dugaan.
  - e. Memilih model SARIMA yang memenuhi semua asumsi.
  - f. Meramalkan model SARIMA terbaik.
- 3) Meramalkan model hibrida SSA-SARIMA dengan menjumlahkan hasil peramalan SSA pada poin 1(f) dan hasil peramalan SARIMA pada poin 2(f).
- 4) Mengevaluasi tingkat kesalahan hasil peramalan SSA dan hibrida SSA-SARIMA dengan melihat nilai MAPE.

## 3 HASIL DAN PEMBAHASAN

Langkah awal dalam peramalan data runtun waktu yaitu menampilkan plot runtun waktu data *training* pada **Gambar 1**.



**Gambar 1.** Plot runtun waktu volume air terjual

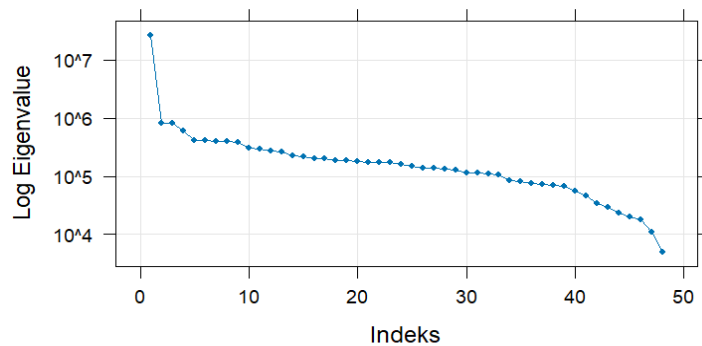
**Gambar 1** menunjukkan plot data bulanan volume air terjual di PDAM Giri Tirta Sari periode Januari 2014 hingga Desember 2021. Terlihat bahwa data mengandung pola tren karena terdapat peningkatan pada beberapa periode waktu dan pola musiman karena terdapat pola berulang di selang waktu tertentu.

### 3.1 Pemodelan *Singular Spectrum Analysis (SSA)*

Analisis data Volume Air Terjual di PDAM Giri Tirta Sari Kabupaten Wonogiri dengan model SSA, perlu dilakukan dua tahap yaitu dekomposisi dan rekonstruksi. Pada tahap dekomposisi dilakukan pemilihan parameter *window length (L)*, parameter *window length* yang digunakan adalah  $L = 48$ . Berdasarkan nilai  $L$  diperoleh nilai  $K = N - L + 1 = 49$ , sehingga matriks lintasan  $X$  dengan dimensi  $48 \times 49$ .

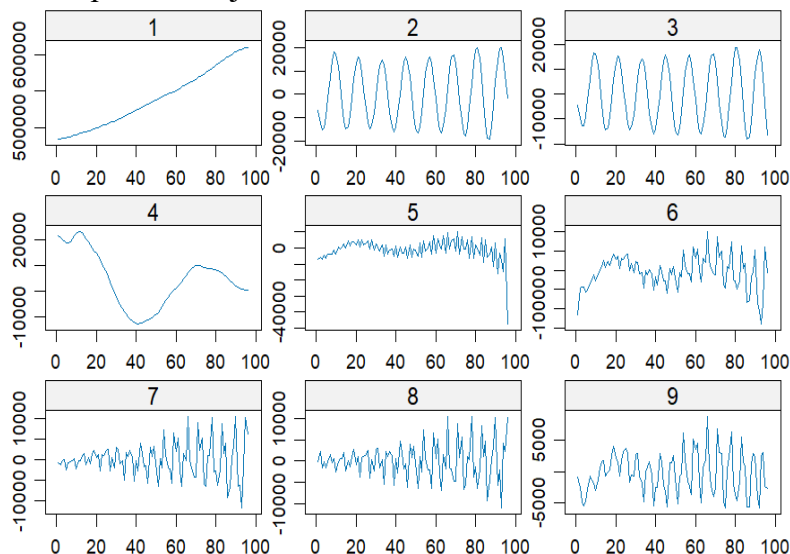
$$X = (x_{i,j})_{i,j=1}^{48,49} = V = \begin{bmatrix} 476045 & \dots & 544658 \\ 467598 & \dots & 466887 \\ \vdots & \ddots & \vdots \\ 503034 & \dots & 544702 \end{bmatrix}$$

Matriks lintasan  $X$  dilakukan proses SVD, sehingga menghasilkan 48 *eigen triple* yang terdiri dari *singular value* ( $\sqrt{\lambda_i}$ ), *eigenvector* ( $U_i$ ), dan *principal component* ( $V_i$ ). Selanjutnya dibentuk grafik *singular value* hasil dari SVD.



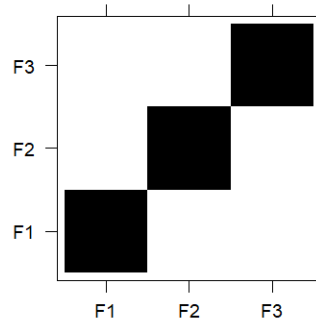
**Gambar 2.** Scree plot singular value

**Gambar 2** menunjukkan *scree plot* dari nilai singular. Terlihat bahwa nilai singular ke-10 sampai 48 menunjukkan penurunan secara lambat sehingga dapat diidentifikasi ke dalam komponen *noise*. Oleh karena itu, nilai dari parameter  $r = 9$  dapat digunakan dalam mengelompokkan komponen menjadi tren dan musiman.



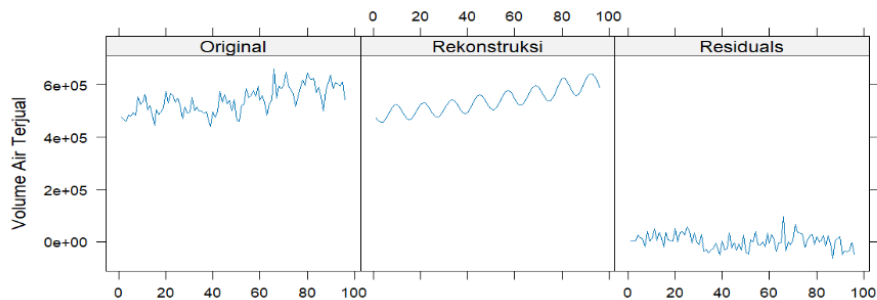
**Gambar 3.** Plot rekonstruksi runtun

**Gambar 3** menunjukkan runtun yang telah direkonstruksi. Terlihat bahwa *eigen triple* 1 memuat komponen yang bervariasi lambat dan tidak mengandung komponen musiman sehingga *eigen triple* 1 diidentifikasi ke dalam komponen tren. Pengelompokan komponen musiman dapat ditunjukkan dari kemiripan plot *eigen triple* yang berurutan. Berdasarkan **Gambar 3** terdapat *eigen triple* yang memiliki pola mirip yaitu *eigen triple* 2 dan 3. Selanjutnya dilakukan proses identifikasi *eigen triple* dapat dilihat melalui plot matriks *W-correlation*.



**Gambar 4.** Plot matriks *W-correlation*

**Gambar 4** menunjukkan bahwa F2 merupakan keterikatan antara *eigen triple* 2 dan 3 yang kuat, F3 merupakan keterikatan antara *eigen triple* 4 hingga 48 yang kuat. Oleh karena itu, kelompok yang dihasilkan dari tahap *grouping* adalah *eigen triple* 1 merupakan komponen tren, *eigen triple* 2 dan 3 merupakan komponen musiman, dan *eigen triple* 4 hingga 48 merupakan komponen *noise*. Selanjutnya dilakukan rekonstruksi pada masing masing kelompok dan hasilnya disajikan pada **Gambar 5**.



**Gambar 5.** Plot hasil rekonstruksi data training

**Gambar 5** menunjukkan perbandingan antara *original* yang merupakan data training dengan hasil rekonstruksi dan *residuals* merupakan selisih antara data training dengan hasil rekonstruksi.

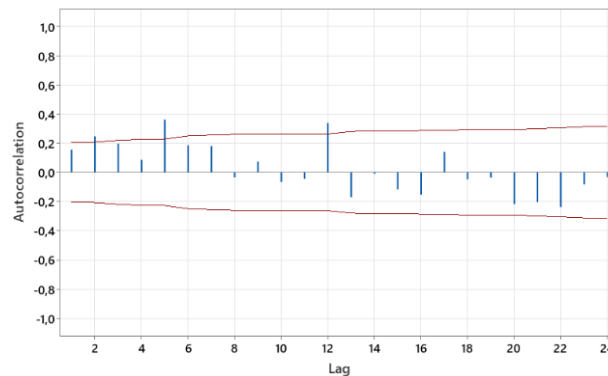
**Tabel 1.** Data residu

$t$	Hasil Rekonstruksi	Data Training	Residu
1	472920,557	476045	3124,443
2	463329,004	467598	4268,996
⋮	⋮	⋮	⋮
96	591656,470	544702	-46954,470

**Tabel 1** menunjukkan data residu dari model SSA. Selanjutnya residu SSA dimodelkan dengan SARIMA.

### 3.2 Pemodelan *Seasonal Autoregressive Integrated Moving Average* (SARIMA)

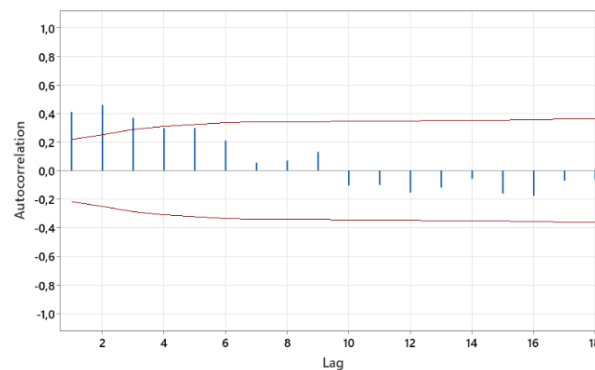
Tahap pertama dalam analisis SARIMA adalah pengecekan stasioneritas. Pengecekan stasioneritas dalam variansi dilakukan melalui transformasi *Box-Cox* yang menghasilkan nilai



**Gambar 6.** Plot ACF residu

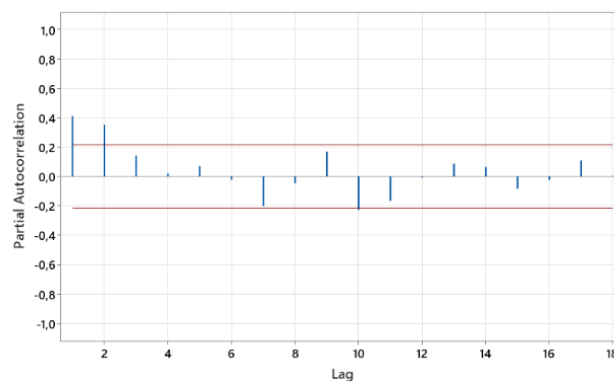
$\lambda = 0.833$ . Nilai  $\lambda$  mendekati nilai  $\lambda = 1$ , sehingga dapat dikatakan bahwa komponen residu stasioner dalam variansi. Stasioneritas data terhadap rata-rata dilihat dari plot ACF yang disajikan pada **Gambar 6**.

**Gambar 6** menunjukkan plot *autocorrelation function* (ACF) residu, terdapat *lag* yang keluar pita konfidensi yaitu *lag* 2, 5, dan 12 dapat disimpulkan bahwa data belum stasioner. Pada *lag* 12 signifikan sehingga diduga ada pola musiman, sehingga dilakukan *differencing lag* 12.



**Gambar 7.** Plot ACF *differencing* musiman *lag* ke-12

**Gambar 7** menunjukkan plot ACF hasil *differencing* musiman pada *lag* ke-12, diketahui bahwa data sudah stasioner. Berdasarkan plot ACF, *lag* 1,2, dan 3 keluar pita konfidensi maka orde  $MA(q)$  yang mungkin adalah 0, 1, 2 dan 3. Selanjutnya plot *partial autocorrelation function* (PACF) *differencing* musiman pada *lag* ke-12 disajikan pada **Gambar 8**.



**Gambar 8.** Plot PACF *differencing* musiman *lag* ke-12



**Gambar 8** menunjukkan plot PACF hasil differencing musiman pada lag ke-12, berdasarkan plot PACF, lag 1 dan 2 keluar pita konfidensi maka orde  $AR(p)$  yang mungkin adalah 0, 1, dan 2. Model dugaan SARIMA dari data residu berdasarkan ACF dan PACF yaitu SARIMA  $(0,0,1)(0,1,0)^{12}$ , SARIMA  $(0,0,2)(0,1,0)^{12}$ , SARIMA  $(0,0,3)(0,1,0)^{12}$ , SARIMA  $(1,0,0)(0,1,0)^{12}$ , SARIMA  $(1,0,1)(0,1,0)^{12}$ , SARIMA  $(1,0,2)(0,1,0)^{12}$ , SARIMA  $(1,0,3)(0,1,0)^{12}$ , SARIMA  $(2,0,0)(0,1,0)^{12}$ , SARIMA  $(2,0,1)(0,1,0)^{12}$ , SARIMA  $(2,0,2)(0,1,0)^{12}$ , SARIMA  $(2,0,3)(0,1,0)^{12}$ .

Semua model dugaan SARIMA dilakukan uji estimasi parameter, sehingga didapatkan enam model yang memenuhi asumsi. Selanjutnya dilakukan uji asumsi *white noise* dan uji asumsi normalitas pada model SARIMA yang memenuhi uji signifikansi parameter. Hasil uji asumsi *white noise* disajikan pada **Tabel 2**.

**Tabel 2.** Uji *white noise*

Model SARIMA	$p - value$	<i>Chi-Square</i>	Keterangan
SARIMA $(0,0,1)(0,1,0)^{12}$	0,000	36,71	Tidak <i>white noise</i>
SARIMA $(0,0,2)(0,1,0)^{12}$	0,005	25,44	Tidak <i>white noise</i>
SARIMA $(1,0,0)(0,1,0)^{12}$	0,020	22,60	Tidak <i>white noise</i>
SARIMA $(1,0,1)(0,1,0)^{12}$	0,155	14,42	<i>white noise</i>
SARIMA $(2,0,0)(0,1,0)^{12}$	0,044	18,69	Tidak <i>white noise</i>
SARIMA $(2,0,2)(0,1,0)^{12}$	0,159	11,82	<i>white noise</i>

**Tabel 2** menunjukkan hasil uji *white noise* dari model yang memenuhi uji signifikansi parameter. Model SARIMA $(1,0,1)(0,1,0)^{12}$  dan SARIMA  $(2,0,2)(0,1,0)^{12}$  memiliki residu yang bersifat *white noise* sampai lag 12 karena  $p - value > \alpha = 0,05$ , sehingga dapat dikatakan kedua model tersebut memenuhi asumsi *white noise*. Selanjutnya model yang memenuhi asumsi *white noise* dilakukan uji normalitas. Hasil uji distribusi normal disajikan pada **Tabel 3**.

**Tabel 3.** Uji distribusi normal

Model SARIMA	$p - value$	$D_{hitung}$	Keterangan
SARIMA $(0,0,1)(0,1,0)^{12}$	0,144	0,084	Berdistribusi normal
SARIMA $(0,0,2)(0,1,0)^{12}$	$> 0,150$	0,078	Berdistribusi normal
SARIMA $(1,0,0)(0,1,0)^{12}$	$> 0,150$	0,067	Berdistribusi normal
SARIMA $(1,0,1)(0,1,0)^{12}$	0,111	0,088	Berdistribusi normal
SARIMA $(2,0,0)(0,1,0)^{12}$	0,150	0,084	Berdistribusi normal
SARIMA $(2,0,2)(0,1,0)^{12}$	$> 0,150$	0,050	Berdistribusi normal

**Tabel 3** menunjukkan hasil uji distribusi normal dari model yang memenuhi uji normalitas SARIMA $(0,0,1)(0,1,0)^{12}$ , SARIMA $(0,0,2)(0,1,0)^{12}$ , SARIMA $(1,0,0)(0,1,0)^{12}$ , SARIMA $(1,0,1)(0,1,0)^{12}$ , SARIMA $(2,0,0)(0,1,0)^{12}$ , dan SARIMA $(2,0,2)(0,1,0)^{12}$  memiliki residu yang berdistribusi normal karena  $p - value > \alpha$ , sehingga dapat dikatakan keenam model tersebut memenuhi asumsi normalitas. Setelah dilakukan pemeriksaan diagnostik, diperoleh model SARIMA $(1,0,1)(0,1,0)^{12}$  dan SARIMA $(2,0,2)(1,0,1)^{12}$  yang memenuhi uji asumsi *white noise* dan uji normalitas residu. Model SARIMA $(1,0,1)(0,1,0)^{12}$  tanpa konstanta ditunjukkan persamaan berikut.

$$Z_t = 0,890Z_{t-1} + Z_{t-12} - 0,890Z_{t-13} + 0,582a_{t-1} + a_t$$

Sementara itu, model SARIMA $(2,0,2)(1,0,1)^{12}$  tanpa konstanta ditunjukkan pada persamaan berikut.

$$Z_t = 1,769Z_{t-1} - 1,769Z_{t-13} - 0,882Z_{t-2} + 0,882Z_{t-14} + Z_{t-12} + 1,740a_{t-1} - 0,987a_{t-2} + a_t$$

### 3.3 Peramalan Hibrida SSA-SARIMA

Langkah selanjutnya yaitu tahap perhitungan hasil peramalan pada data *testing* menggunakan model hibrida SSA-SARIMA. Data hasil peramalan disajikan pada **Tabel 4**.

**Tabel 4.** Data hasil peramalan model hibrida SSA-SARIMA

$t$	$\hat{Y}_t$	$\hat{Z}_{t_1}$	$\hat{Z}_{t_2}$	$\hat{H}_{t_1}$	$\hat{H}_{t_2}$	$H_t$
97	578441,847	-746,461	-2404,534	577695,386	576037,313	581746
98	572841,743	-25610,736	-31200,635	547231,006	541641,108	535026
99	578493,522	-81212,764	-87301,011	497280,758	491192,511	501996
⋮	⋮	⋮	⋮	⋮	⋮	⋮
114	665696,208	4700,064	8176,829	670396,271	673873,036	604361

dengan

$t$  : Waktu ke- $t$ .

$\hat{Y}_t$  : Hasil peramalan SSA.

$\hat{Z}_{t_1}$  : Hasil peramalan SARIMA(1,0,1)(0,1,0)<sup>12</sup>.

$\hat{Z}_{t_2}$  : Hasil peramalan SARIMA(2,0,2)(0,1,0)<sup>12</sup>.

$\hat{H}_{t_1}$  : Hasil peramalan hibrida SSA-SARIMA(1,0,1)(0,1,0)<sup>12</sup>.

$\hat{H}_{t_2}$  : Hasil peramalan hibrida SSA-SARIMA(2,0,2)(0,1,0)<sup>12</sup>.

$x_t$  : Data *testing*.

**Tabel 4** menunjukkan hasil peramalan model SSA dan hibrida SSA-SARIMA. Data peramalan SSA diperoleh dengan model LRF, sedangkan peramalan SARIMA diperoleh dengan model SARIMA terbaik. Berdasarkan hasil peramalan hibrida SSA-SARIMA dapat dihitung tingkat kesalahan model menggunakan nilai MAPE. Tingkat kesalahan hasil peramalan disajikan pada **Tabel 5**.

**Tabel 5.** Tingkat kesalahan hasil peramalan

Model	MAPE	
	<i>Training</i>	<i>Testing</i>
SSA-SARIMA(1,0,1)(0,1,0) <sup>12</sup>	3,646%	5,978%
SSA-SARIMA(2,0,2)(0,1,0) <sup>12</sup>	3,542%	7,467%
SSA	4,215%	11,138%

**Tabel 5** menunjukkan menunjukkan bahwa model SSA-SARIMA(1,0,1)(0,1,0)<sup>12</sup>, SSA-SARIMA(2,0,2)(0,1,0)<sup>12</sup> dan SSA memiliki nilai MAPE *training* yang sama baik, sedangkan nilai MAPE *testing* pada model SSA-SARIMA (1,0,1)(0,1,0)<sup>12</sup> dan SSA-SARIMA(2,0,2)(0,1,0)<sup>12</sup> memiliki nilai MAPE yang baik dan model SSA memiliki nilai yang cukup baik. Model hibrida SSA-SARIMA (1,0,1)(0,1,0)<sup>12</sup> adalah model terbaik dalam meramalkan data runtun waktu volume air terjual di PDAM Giri Tirta Sari Kabupaten Wonogiri, karena memiliki tingkat kesalahan hasil peramalan terkecil.

## 4 KESIMPULAN

Model SSA yang digunakan adalah SSA dengan parameter *window length* sebesar 48 dan *grouping* sebanyak tiga kelompok yaitu komponen tren, musiman, dan *noise*. Residu SSA dimodelkan dengan SARIMA, dan diperoleh model hibrida SSA-SARIMA(1,0,1)(0,1,0)<sup>12</sup> dengan nilai MAPE sebesar sebesar 5,978%. Hal ini berarti tingkat akurasi hasil peramalan dari model tersebut dapat dikatakan sangat baik karena memiliki nilai MAPE kurang dari 10%.

## DAFTAR PUSTAKA

- Arumsari, M., Wahyuningsih, S., & Siringoringo, M. (2021). Inflation Forecasting for East Kalimantan Province Using Hybrid Singular Spectrum Analysis - Autoregressive Integrated Moving Average Model. *Jurnal Matematika, Statistika dan Komputasi*, 18(1), 78–92. <https://doi.org/10.20956/j.v18i1.14284>
- Christie, G., Hatidja, D., & Tumilaar, R. (2022). Penerapan Metode SARIMA dalam Model Intervensi Fungsi Step untuk Memprediksi Jumlah Pegunjung Objek Wisata Londa (Application of the SARIMA Method in the Step Function Intervention to Predict the Number of Visitors at Londa Tourism Object). *JURNAL ILMIAH SAINS*, 22(2), 96. <https://doi.org/10.35799/jis.v22i2.40961>
- Darmawan, G., Rosadi, D., & Ruchjana, B. N. (2022). Hybrid Model of Singular Spectrum Analysis and ARIMA for Seasonal Time Series Data. *CAUCHY: Jurnal Matematika Murni Dan Aplikasi*, 7(2), 302–315. <https://doi.org/10.18860/ca.v7i2.14136>
- Fajar, M. (2020). *Comparison of ARIMA, SSA, and ARIMA – SSA Hybrid Model Performance in Indonesian Economic Growth Forecasting*.
- Febrianti, T. K., Sulandari, W., & Pratiwi, H. (2021). *Peramalan curah hujan di kota bandung menggunakan singular spectrum analysis*. 8(2).
- Golyandina, N., Nekrutkin, V., & Zhigljavsky, A.A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall CRC.
- Golyandina, N., & Zhigljavsky, (2020). *Singular Spectrum Analysis for Time Series*. Cardiff: Springer.
- Hassani, H. (2021). Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5(2), 239–257. [https://doi.org/10.6339/JDS.2007.05\(2\).396](https://doi.org/10.6339/JDS.2007.05(2).396)
- Irmawati, D.R., Atok, R.M., and Suhartono. (2018). Singular Spectrum Analysis ARIMA Modelling for Direct and Indirect Forecasting of Farmer's Term of Trade in East Java. *International Conference on Information and Communications Technology*, 889-894. <https://doi.org/10.1109/ICOIACT.2018.8350823>
- Ilahi, E.P., Zukhronah, E., dan Susanti, Y. (2023). Model Hibrida Singular Spectrum Analysis (SSA) dan Autoregressive Integrated Moving Average (ARIMA) untuk Peramalan Indeks Harga Konsumen. *Seminar Nasional Pendidikan Matematika Ahmad Dahlan 2023*, 72- 81.
- Ischak, R., Asrof, A., dan Darmawan, G. (2018). Peramalan Rata-Rata Harga Beras di Tingkat Penggilingan Menggunakan Model Singular Spectrum Analysis (SSA). *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika*, Vol. 4(1), 58- 60.
- Putri, Z. R., Zukhronah, E., & Slamet, I. (2023). Perbandingan Akurasi Metode *Autoregressive Integrated Moving Average* dan *Singular Spectrum Analysis* Pada Peramalan Suhu Rata-Rata Harian Di Stasiun Meteorologi Ahmad Yani. *Prosiding SENPIKA Volume 1 Tahun 2023*. 238-253.
- Setyaningrum, O. A., Zukhronah, E., & Handajani, S. S. (2023). Peramalan Nilai Tukar Petani (Ntp) Di Indonesia Menggunakan Metode Hibrida *Singular Spectrum Analysis* (SSA)-*Seasonal Autoregressive Integrated Moving Average* (SARIMA). *Prosiding SENPIKA Volume 1 Tahun 2023*. 254-266.
- Tokan, L. F., & Hermawan, A. (2023). Implementasi Model SARIMA Untuk Memprediksi Produksi Minyak Kelapa Sawit. *JURNAL FASILKOM*, 13(3), 456–463. <https://doi.org/10.37859/jf.v13i3.6033>
- Zhang, G. P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*. Vol. 50, 159-175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)

Zhang, Q., Wang, B. D., He, B., Peng, Y., dan Ren, M. L. (2011). Singular Spectrum Analysis and ARIMA Hybrid Model for Annual Runoff Forecasting. *Water Resour Manage.* Vol. 25(11), 2683-2703. <https://doi.org/10.1007/s11269-011-9833-y>