

**PERBANDINGAN AKURASI METODE AUTOREGRESSIVE
INTEGRATED MOVING AVERAGE DAN SINGULAR SPECTRUM
ANALYSIS PADA PERAMALAN SUHU RATA-RATA HARIAN
DI STASIUN METEOROLOGI AHMAD YANI**

Zizika Riantika Putri^{*1}, Etik Zukhronah², Isnandar Slamet³

^{1,2,3}Program Studi Statistika, Universitas Sebelas Maret

*Penulis Korespondensi (zizikariantika@student.uns.ac.id)

Abstrak: Suhu udara di kota-kota besar dalam beberapa tahun terakhir mengalami peningkatan yang cukup ekstrim, sehingga diperlukan peramalan yang tepat untuk mengantisipasi terjadinya kerugian akibat suhu yang ekstrim. Penelitian ini bertujuan untuk membandingkan akurasi dari metode *Autoregressive Integrated Moving Average* (ARIMA) dan *Singular Spectrum Analysis* (SSA) dalam meramalkan suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani. Data yang digunakan sebagai data *training* adalah data harian dari Januari 2022 hingga Desember 2022, sedangkan data harian Januari 2023 digunakan sebagai data *testing*. Pemodelan ARIMA diawali dengan menstasionerkan data, selanjutnya menentukan orde ARIMA berdasarkan plot *autocorrelation function* (ACF) dan *partial autocorrelation function* (PACF). Langkah-langkah pemodelan SSA yaitu *embedding*, *singular value decomposition* (SVD), *grouping*, dan *diagonal averaging*. Pemilihan model terbaik ditentukan berdasarkan nilai *mean absolute percentage error* (MAPE) terkecil. Hasil penelitian menunjukkan nilai MAPE data *training* dengan metode ARIMA(0,1,2) sebesar 2,0076%, sedangkan metode SSA sebesar 2,1777%. Nilai MAPE data *testing* dengan metode ARIMA(0,1,2) sebesar 3,2719%, sedangkan metode SSA sebesar 3,9464%. Berdasarkan hasil tersebut dapat disimpulkan bahwa metode ARIMA(0,1,2) lebih tepat digunakan untuk meramalkan data suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani.

Kata kunci: suhu udara, peramalan, ARIMA, SSA

Abstract: Air temperature in big cities in the last few years has increased quite extreme, so accurate forecasting is needed to anticipate losses due to extreme temperatures. This study aims to compare the accuracy of *Autoregressive Integrated Moving Average* (ARIMA) and *Singular Spectrum Analysis* (SSA) methods in predicting the daily average temperature at Ahmad Yani Meteorological Station. The data used as training data is daily data from January 2022 to December 2022, while the daily data for January 2023 is used as testing data. ARIMA modeling begins with stationary data, then determines the ARIMA order based on the *autocorrelation function* (ACF) and *partial autocorrelation function* (PACF) plots. SSA modeling steps are *embedding*, *singular value decomposition* (SVD), *grouping*, and *diagonal averaging*. The selection of the best model is determined based on the most minor *mean absolute percentage error* (MAPE). The results showed that value of MAPE training data using ARIMA(0,1,2) method was 2.0076%, while SSA method was 2.1777%. The MAPE value of testing data using

ARIMA(0,1,2) method is 3.2719%, while SSA method is 3.9464%. Based on these results it can be concluded that the ARIMA(0,1,2) method is more appropriate for forecast daily average temperature data at Ahmad Yani Meteorological Station.

Keywords: *air temperature, forecasting, ARIMA, SSA*

PENDAHULUAN

Iklim merupakan keadaan rata-rata cuaca dalam jangka waktu yang panjang. Perubahan iklim dapat dipengaruhi oleh aktivitas manusia secara langsung maupun tidak langsung (Prasetyo dkk., 2021). Salah satu parameter untuk melihat adanya perubahan iklim adalah suhu udara permukaan bumi. Kenaikan suhu permukaan bumi terjadi karena adanya peningkatan jumlah gas rumah kaca di lapisan atmosfer yang disebabkan karena meningkatnya karbondioksida (CO_2) yang berasal dari pembakaran bahan bakar fosil seperti minyak bumi, gas bumi, dan batubara (Samidjo & Suharso, 2017). Pemanasan global yang terjadi akan menyebabkan perubahan iklim seperti meningkatnya curah hujan sehingga menimbulkan banjir dan erosi, maupun meningkatnya suhu udara sehingga mengakibatkan terjadinya kekeringan yang berkepanjangan.

Kota Semarang tidak lepas dari pengaruh perubahan iklim. Letak geografis Kota Semarang yang memiliki wilayah pesisir dibagian utara dengan panjang garis pantai 13,6 km dan perairan terbuka yang berbatasan langsung dengan laut Jawa menyebabkan Kota Semarang cenderung lebih panas dibandingkan kota-kota lainnya. Pada September 2021, Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) menyatakan bahwa Kota Semarang menduduki posisi kelima wilayah dengan suhu tertinggi di Indonesia dengan suhu mencapai 35°C (Solopos.com, 2022). Menurut BMKG suhu udara rata-rata Kota Semarang pada Januari 2023 berkisar antara 28,5°C dengan suhu terendah berkisar antara 22,8°C dan suhu tertingginya berkisar antara 32,2°C.

Suhu udara berperan penting dalam kehidupan, yaitu dalam bidang pertanian berpengaruh terhadap proses perkembangan tumbuhan, dalam bidang penerbangan berpengaruh terhadap jadwal keperluan *start engine* saat *take off*, dan dalam bidang kesehatan suhu udara yang berfluktuatif dapat menyebabkan timbulnya penyakit (Sakinah, 2018). Dampak lain yang mungkin terjadi akibat suhu udara yang ekstrim adalah meningkatnya intensitas turunnya hujan akibat tingginya penguapan air laut. Kota Semarang merupakan salah satu wilayah di Indonesia yang kerap dilanda banjir. Salah satu penyebabnya adalah karena turunnya hujan berintensitas tinggi akibat cuaca ekstrim. Oleh karena itu, peramalan data suhu penting dilakukan untuk mengantisipasi kejadian suhu ekstrim yang akan berpengaruh terhadap kehidupan sehari-hari manusia.

Peramalan merupakan salah satu ilmu untuk meramalkan suatu peristiwa di masa depan dengan menggunakan data peristiwa masa lalu yang dijadikan sebagai panduan untuk melakukan perencanaan di masa mendatang. Peramalan data suhu rata-rata harian dapat dilakukan dengan menggunakan metode *Autoregressive Integrated Moving Average* (ARIMA) dan *Singular Spectrum Analysis* (SSA).

Anwar (2017) melakukan peramalan data harian suhu udara minimum dan maksimum di Kota Banda Aceh menggunakan metode ARIMA. Hasil penelitian menunjukkan bahwa metode ARIMA cocok untuk diterapkan pada data suhu harian karena memiliki nilai *root mean square error* (RMSE) yang kecil yaitu sebesar 0,969 untuk data suhu udara minimum dan untuk suhu udara maksimum sebesar 1,289. Metode ARIMA merupakan salah satu metode yang sangat baik akurasi. Namun, ARIMA mempunyai kelemahan karena dalam proses analisisnya harus memenuhi berbagai asumsi

seperti stasioneritas dan normalitas residu. Oleh karena itu, diperlukan metode pembandingan lain yang tidak membutuhkan banyak asumsi dalam peramalan yaitu metode SSA. SSA merupakan metode yang penggunaannya terhindar dari berbagai asumsi seperti stasioneritas dan normalitas residu karena merupakan salah satu metode nonparametrik. Sakinah (2018) melakukan peramalan data suhu bulanan di Jakarta menggunakan metode SSA. Hasil penelitian menunjukkan bahwa metode SSA dapat dikatakan baik dalam meramalkan suhu karena menghasilkan nilai *mean absolute percentage error* (MAPE) kurang dari 10% yaitu sebesar 5,0029%.

Berdasarkan uraian di atas, terlihat bahwa metode ARIMA dan SSA memiliki keunggulan dalam peramalan karena mampu menghasilkan *error* yang kecil. Oleh karena itu, pada penelitian ini dilakukan peramalan suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani dengan membandingkan metode ARIMA dan SSA untuk mendapatkan metode terbaik berdasarkan nilai MAPE terkecil.

METODE

Penelitian ini menggunakan data sekunder suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani dari Januari tahun 2022 hingga Januari 2023. Data diperoleh melalui portal data *online* BMKG (Pusat Database BMKG, 2015). Data dibagi menjadi dua bagian, yaitu data *training* yang terdiri dari 365 data (1 Januari 2022 hingga 31 Desember 2022) dan data *testing* yang terdiri dari 31 data (1 Januari 2023 hingga 31 Januari 2023). Analisis dilakukan dengan bantuan *software* Minitab, RStudio, dan Microsoft Excel. Penelitian ini menggunakan dua metode yaitu metode ARIMA dan SSA.

Autoregressive Integrated Moving Average (ARIMA)

Model ARIMA dinotasikan sebagai $ARIMA(p, d, q)$ dengan bentuk umum ditulis seperti pada persamaan (1).

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t \quad (1)$$

dengan $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ yang merupakan parameter ke- i model $AR(p)$ dengan $(i = 1, 2, \dots, p)$ dan $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ yang merupakan parameter ke- j model $MA(q)$ dengan $(j = 1, 2, \dots, q)$.

Model ARIMA digunakan untuk meramalkan data runtun waktu yang berpola stasioner. Stasioneritas dibagi menjadi dua, yaitu stasioner dalam variansi dan rata-rata. Stasioner dalam variansi dapat dilihat melalui plot runtun waktu, yaitu apabila plot runtun waktu menunjukkan adanya fluktuasi data yang konstan dari waktu ke waktu. Jika data tidak stasioner dalam variansi yang ditandai dengan nilai $\lambda \neq 1$, maka perlu dilakukan transformasi Box-Cox (Wei, 2006). Stasioner dalam rata-rata dapat diidentifikasi dengan melihat plot *autocorrelation function* (ACF), yaitu ketika plot ACF tidak menurun secara eksponensial. Apabila data tidak stasioner dalam rata-rata yang ditandai dengan pada plot ACF terdapat lebih dari 3 *lag* awal yang keluar dari batas kritis, maka perlu dilakukan *differencing* yang dinyatakan sebagai orde d (Wei, 2006). Kemudian dilakukan identifikasi model untuk menentukan orde p dan q menggunakan plot PACF dan ACF menggunakan data yang telah stasioner (Muis & Setiyadi, 2020).

Estimasi parameter dilakukan terhadap model yang sudah diperoleh dengan menggunakan metode estimasi *maksimum likelihood* dan diperlukan uji signifikansi parameter berdasarkan p_{value} dari setiap model. Model dinyatakan signifikan apabila $p_{value} < \alpha$ dengan $\alpha = 0,05$. Setelah menemukan model yang signifikan, dilakukan uji diagnostik residu yang terdiri dari uji asumsi normal dan uji independensi. Uji asumsi normal dilakukan untuk mengetahui apakah residu berdistribusi normal atau tidak. Uji

asumsi normal dilakukan menggunakan uji Kolmogorov-Smirnov dengan hipotesis sebagai berikut.

Hipotesis:

H_0 : residu berdistribusi normal

H_1 : residu tidak berdistribusi normal

Statistik uji:

$$D_{hit} = \max |S(x) - F_0(x)|$$

dengan $S(x)$ merupakan fungsi distribusi kumulatif dari data asal, $F_0(x)$ merupakan fungsi yang dihipotesiskan berdistribusi normal, dan D merupakan persamaan uji Kolmogorov-Smirnov. Hipotesis H_0 ditolak jika nilai $D_{hit} > D_{(n)}$ atau $p_{value} < \alpha$ dengan $\alpha = 0,05$. Suatu model dikatakan independen apabila bersifat *random* atau tidak memiliki hubungan antar residu. Uji independensi dilakukan menggunakan uji Ljung Box-Q (LBQ) dengan hipotesis sebagai berikut.

Hipotesis:

H_0 : $\rho_1 = \dots = \rho_k = 0$ (tidak terdapat autokorelasi pada residu),

H_1 : minimal terdapat satu $\rho_k \neq 0$ untuk $k = 1, 2, \dots, K$ (terdapat autokorelasi pada residu).

Statistik uji:

$$Q = n(n+2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{n-k}$$

dengan n merupakan jumlah observasi, k merupakan *lag* waktu, K merupakan jumlah *lag* yang diuji, dan $\hat{\rho}_k$ merupakan fungsi autokorelasi residu sampel pada *lag* ke- k . Hipotesis H_0 ditolak jika $Q > \chi_{\alpha; (K-m)}^2$ atau $p_{value} < \alpha$ dengan $\alpha = 0,05$.

Singular Spectrum Analysis (SSA)

SSA bertujuan untuk menguraikan deret aslinya menjadi komponen tren, musiman, dan *noise*. Algoritma SSA terdiri dari dua tahapan, yaitu dekomposisi dan rekonstruksi (Ete, 2017). Dekomposisi terdiri dari dua langkah untuk memperoleh *eigentriple*, yaitu *embedding* dan *singular value decomposition* (SVD). Proses *embedding* dilakukan untuk mengubah data runtun waktu menjadi data berbentuk matriks lintasan berukuran $L \times K$, dengan L merupakan parameter *window length* yang bernilai $2 < L < \frac{N}{2}$ dan $K = N - L + 1$. Matriks lintasan dari deret X dapat ditulis sebagai persamaan (2).

$$X = [X_1, X_2, \dots, X_K] = (x_{ij})_{L,K} = \begin{bmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & x_3 & \dots & x_{K+1} \\ \vdots & \dots & \ddots & \vdots \\ x_L & x_{L+1} & \dots & x_N \end{bmatrix} \quad (2)$$

Matriks lintasan X disebut juga sebagai matriks Hankel karena semua elemen anti diagonalnya bernilai sama (Febrianti dkk., 2021). Prosedur SVD bertujuan untuk memperoleh pemisahan *eigentriple* yang terdiri dari nilai singular ($\sqrt{\lambda_i}$), vektor eigen kiri (U_i), dan vektor eigen kanan (V_i). Setelah memperoleh *eigentriple*, dilakukan rekonstruksi yang terdiri dari *grouping* dan *diagonal averaging*. *Grouping* dilakukan untuk mengelompokkan hasil SVD kedalam kelompok tren, musiman, dan *noise* dengan melihat nilai dari vektor eigen. Vektor eigen yang memiliki pola melambat

dikelompokkan pada kelompok tren, vektor eigen yang mengikuti pola musiman dikelompokkan pada kelompok musiman, sedangkan sisanya dikelompokkan sebagai *noise* (Khaeri dkk., 2018). Pada tahap *diagonal averaging* dilakukan penyusunan ulang hasil yang diperoleh pada tahap *grouping* menjadi data deret baru dengan panjang N menggunakan persamaan (3).

$$\tilde{\mathbf{x}}^{(k)} = \tilde{x}_1^{(k)}, \dots, \tilde{x}_N^{(k)} \quad (3)$$

dengan $\sum_{k=1}^m \tilde{x}_n^{(k)}$ untuk $n = 1, 2, \dots, N$ (Masyitoh et al., 2022). Selanjutnya dilakukan perhitungan nilai koefisien *linear recurrent formula* (LRF) dan peramalan menggunakan persamaan (4).

$$\hat{g}_i = \begin{cases} \tilde{g}_i & \text{untuk } i = 1, \dots, N \\ \sum_{j=1}^{L-1} a_j \hat{g}_{i-j} & \text{untuk } i = N + 1, \dots, N + M \end{cases} \quad (4)$$

dengan $\hat{g}_i$ merupakan hasil peramalan pada periode ke- i , $\tilde{g}_i$ merupakan hasil rekonstruksi *diagonal averaging* pada periode ke- i , dan a_j merupakan koefisien LRF.

Akurasi Peramalan

Apabila metode yang digunakan telah dianggap benar untuk melakukan peramalan, selanjutnya dilakukan pemilihan metode terbaik dengan melihat nilai *error*. Evaluasi pemilihan metode terbaik pada penelitian ini menggunakan MAPE dengan persamaan (5).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Z_t - \hat{Z}_t|}{|Z_t|} \times 100 \quad (5)$$

dengan Z_t merupakan data aktual pada periode ke- t , $\hat{Z}_t$ merupakan nilai hasil peramalan pada periode ke- t , dan n merupakan jumlah observasi. Suatu model dikatakan memiliki peramalan dengan akurasi yang tinggi jika mempunyai nilai MAPE kecil dari 10% (Lewis, 1982).

Langkah Analisis

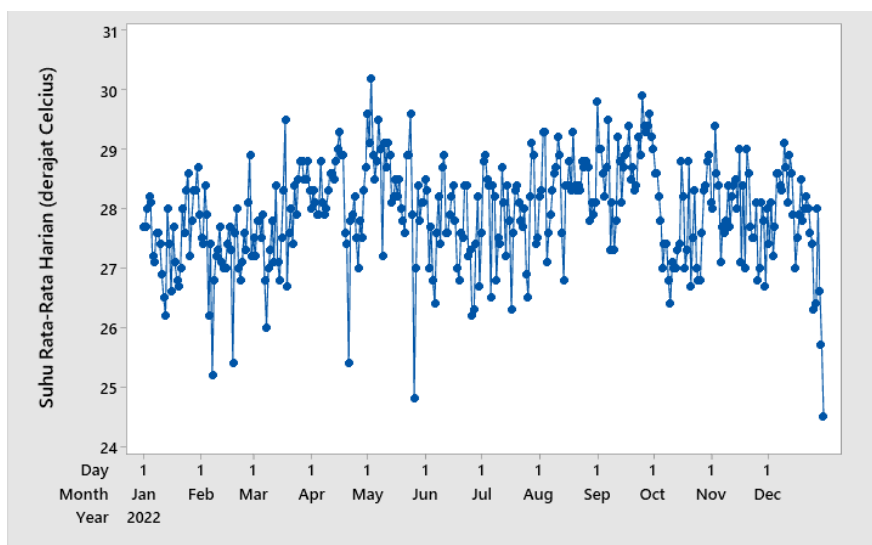
1. Langkah-langkah analisis yang digunakan dalam pembentukan model ARIMA
 - a. Melakukan identifikasi pola data *training* dengan menampilkan plot runtun waktu dan plot ACF.
 - b. Melakukan uji stasioneritas data *training* dalam variansi dan rata-rata dengan membuat plot transformasi Box-Cox dan plot ACF.
 - c. Membuat plot ACF dan PACF dari data yang telah stasioner untuk menentukan orde ARIMA.
 - d. Melakukan uji signifikansi parameter dan uji diagnostik residu yang meliputi uji asumsi normal menggunakan uji Kolmogorov-Smirnov dan uji independensi menggunakan uji LBQ terhadap model yang telah terbentuk.
 - e. Menghitung nilai estimasi parameter dan melakukan peramalan untuk data *testing* menggunakan model ARIMA yang terbentuk.
 - f. Menentukan model ARIMA terbaik dengan melihat nilai MAPE.
2. Langkah-langkah analisis yang digunakan dalam pembentukan model SSA
 - a. *Embedding*, menentukan nilai parameter L dan membentuk matriks lintasan yang berukuran $L \times K$.
 - b. SVD, melakukan dekomposisi dari matriks lintasan untuk membentuk *eigentriple*.
 - c. *Grouping*, melakukan pengelompokan hasil SVD kedalam kelompok tren, musiman, dan *noise*.
 - d. *Diagonal averaging*, menyusun kembali hasil *grouping* menjadi data runtun waktu baru menggunakan persamaan (3).

- e. Menghitung nilai koefisien LRF dan melakukan peramalan untuk data *testing* menggunakan persamaan (4).
3. Membandingkan metode ARIMA dan SSA berdasarkan nilai MAPE.

HASIL DAN PEMBAHASAN

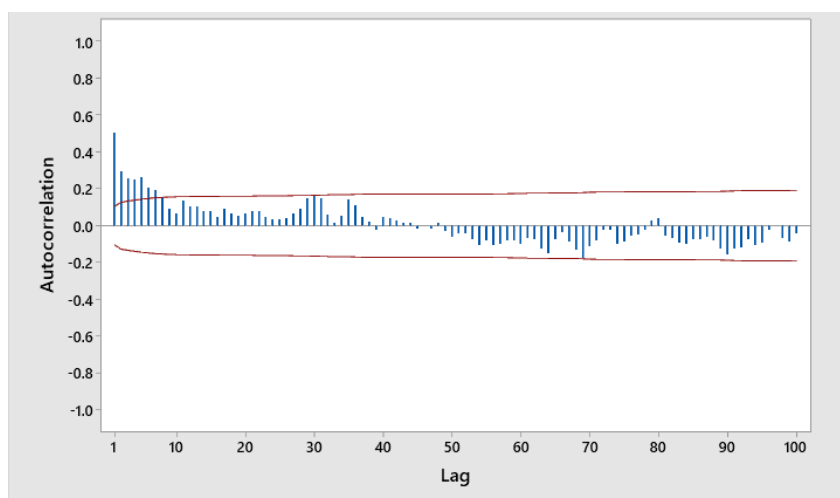
Analisis Metode ARIMA

Identifikasi pola data *training* dilakukan dengan melihat plot runtun waktu dan plot ACF untuk mengetahui karakteristik data yang dianalisis. Gambar 1 menunjukkan plot runtun waktu dan Gambar 2 menunjukkan plot ACF dari data *training* suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani periode Januari 2022 hingga Desember 2022.



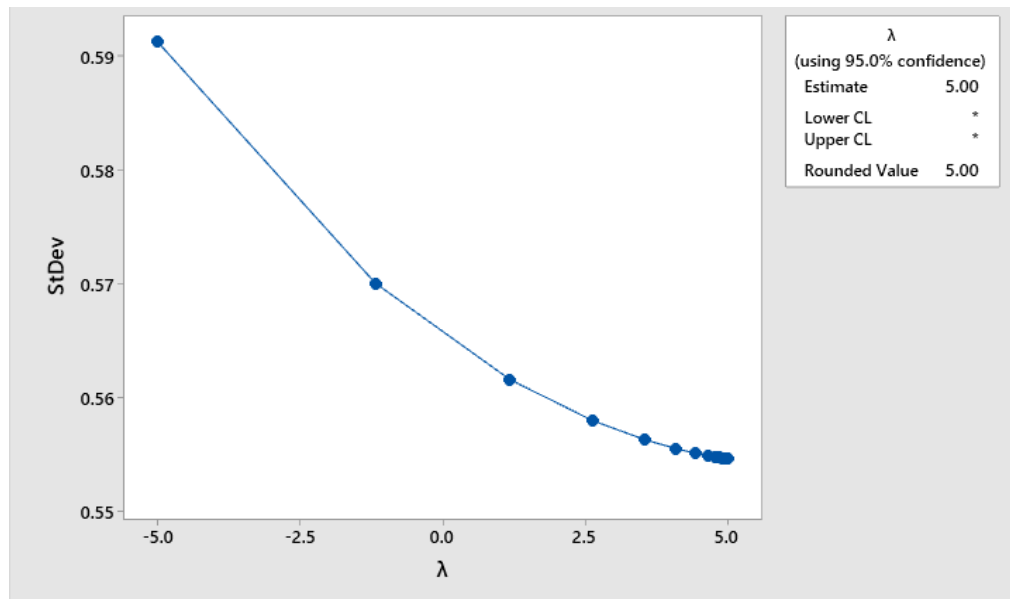
Gambar 1 Plot Runtun Waktu Data Training

Gambar 1 menunjukkan bahwa data *training* suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani berpola tren yang ditandai dengan kenaikan atau penurunan yang fluktuatif dari Januari hingga Desember 2022. Selanjutnya dilihat plot ACF untuk memastikan data tersebut berpola tren.

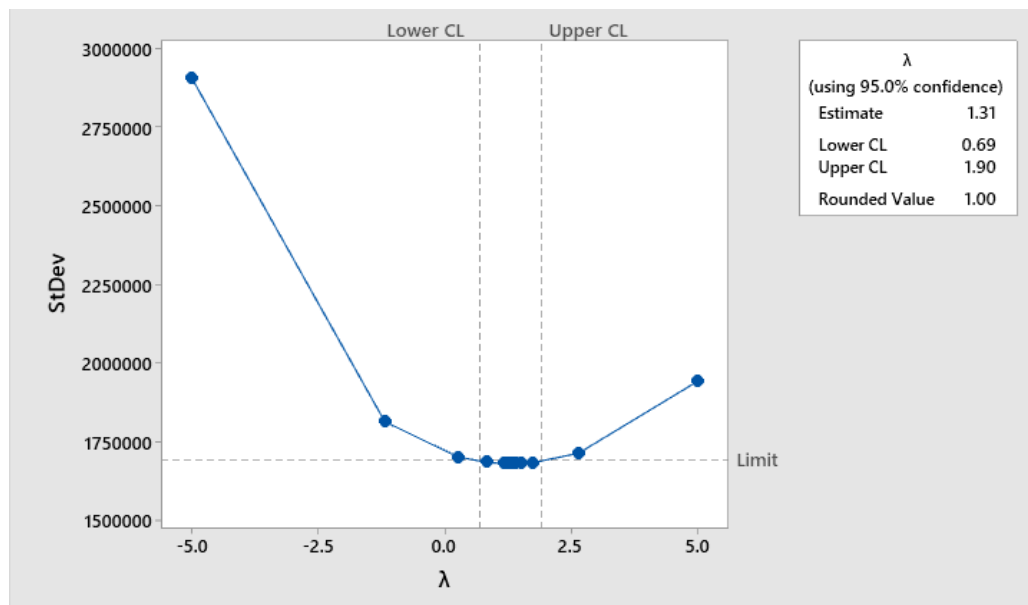


Gambar 2 Plot ACF Data Training

Gambar 2 menunjukkan *lag* mengalami penurunan secara perlahan yang mengindikasikan bahwa data suhu berpola tren. Oleh karena itu, dapat dikatakan bahwa data tidak stasioner, sehingga perlu dilakukan pemeriksaan stasioneritas data dalam variansi dan rata-rata. Kestasioneran data dalam variansi dapat dilihat menggunakan plot transformasi Box-Cox pada Gambar 3. Berdasarkan analisis menggunakan Minitab diperoleh nilai $\lambda = 5$ yang berarti bahwa data belum stasioner dalam variansi, sehingga dilakukan transformasi menggunakan $Z_t^\lambda = Z_t^5$.



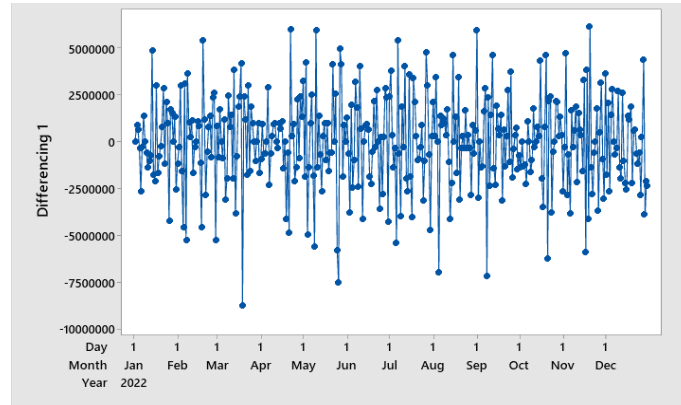
Gambar 3 Sebelum Transformasi Box-Cox



Gambar 4 Setelah Transformasi Box-Cox

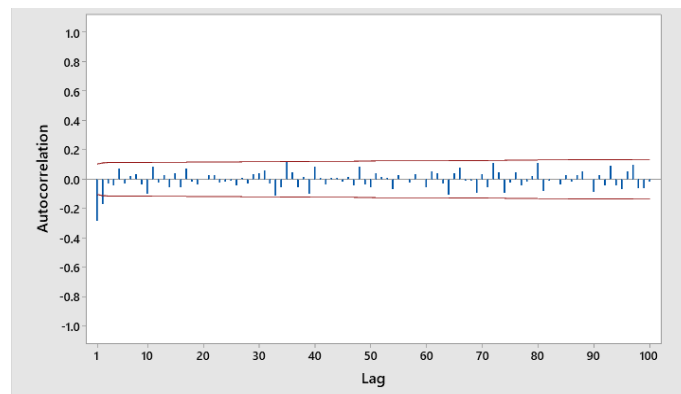
Gambar 4 menunjukkan bahwa nilai $\lambda = 1$ yang berarti bahwa data telah stasioner dalam variansi. Selanjutnya dilakukan pemeriksaan stasioneritas data dalam rata-rata dengan melihat plot ACF. Gambar 2 menunjukkan bahwa pada plot ACF terdapat 7 *lag*

pertama yang melebihi batas kritis. Hal ini menandakan bahwa data suhu tidak stasioner dalam rata-rata, sehingga perlu dilakukan *differencing lag 1*.



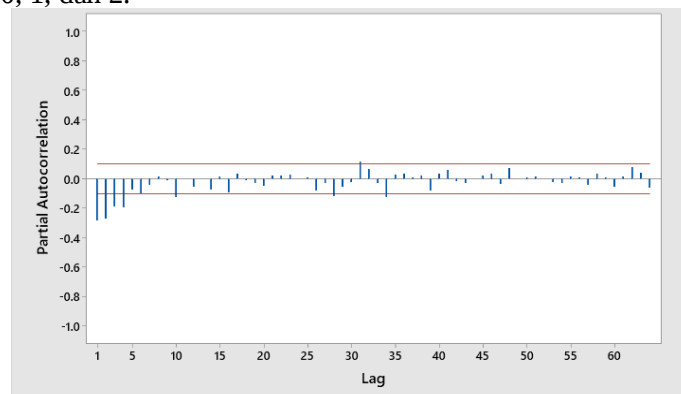
Gambar 5 Plot Runtun Waktu Setelah Differencing

Gambar 5 menunjukkan bahwa data telah stasioner dalam variansi dan rata-rata yang ditandai dengan variansi dan rata-rata dari data suhu telah konstan sepanjang waktu. Selanjutnya dilakukan pengamatan terhadap plot ACF dan PACF setelah dilakukan *differencing*.



Gambar 6 Plot ACF Setelah Differencing

Gambar 6 menunjukkan bahwa data telah stasioner dalam rata-rata karena *lag* awal yang keluar dari batas kritis telah kurang dari 3. Gambar 6 menjelaskan bahwa terdapat 2 *lag* awal yang keluar dari batas kritis lalu *cuts off* yang menunjukkan orde q yang mungkin yaitu 0, 1, dan 2.



Gambar 7 Plot PACF Setelah Differencing

Gambar 7 menunjukkan bahwa kemungkinan orde p adalah 0, 1, 2, 3, dan 4 karena pada plot PACF terdapat 4 *lag* awal yang keluar dari batas kritis lalu *cuts off*. Berdasarkan identifikasi awal diperoleh 14 kemungkinan model ARIMA sementara, yaitu ARIMA(0,1,1), ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(3,1,1), ARIMA(4,1,1), ARIMA(0,1,2), ARIMA(1,1,2), ARIMA(2,1,2), ARIMA(3,1,2), ARIMA(4,1,2), ARIMA(1,1,0), ARIMA(2,1,0), ARIMA(3,1,0), dan ARIMA(4,1,0). Setelah diperoleh model awal, dilakukan uji signifikansi parameter untuk melihat apakah semua parameternya signifikan. Hasil uji signifikansi parameter menunjukkan bahwa terdapat 7 model yang semua parameternya signifikan yaitu ARIMA(0,1,1), ARIMA(1,1,1), ARIMA(0,1,2), ARIMA(1,1,0), ARIMA(2,1,0), ARIMA(3,1,0), dan ARIMA(4,1,0).

Langkah selanjutnya yaitu melakukan pengujian terhadap residu dari model yang seluruh parameternya signifikan. Model yang layak harus memenuhi asumsi residu berdistribusi normal dan independen. Uji asumsi normal yang digunakan adalah uji Kolmogorov-Smirnov, residu berdistribusi normal apabila $p_{value} > \alpha = 0,05$. Uji independensi yang digunakan adalah uji Ljung Box-Q, residu independen apabila $p_{value} > \alpha = 0,05$.

Tabel 1 Uji Diagnostik Residu

Model ARIMA	Normalitas	Independensi
ARIMA(0,1,1)	Memenuhi	Tidak memenuhi
ARIMA(1,1,1)	Memenuhi	Memenuhi
ARIMA(0,1,2)	Memenuhi	Memenuhi
ARIMA(1,1,0)	Tidak memenuhi	Tidak memenuhi
ARIMA(2,1,0)	Tidak memenuhi	Tidak memenuhi
ARIMA(3,1,0)	Memenuhi	Tidak memenuhi
ARIMA(4,1,0)	Memenuhi	Memenuhi

Tabel 1 menunjukkan bahwa terdapat 3 model yang memenuhi uji diagnostik residu yaitu model ARIMA(1,1,1), ARIMA(0,1,2), dan ARIMA(4,1,0) yang berarti bahwa ketiga model berdistribusi normal dan tidak terdapat autokorelasi antar residu. Selanjutnya diberikan estimasi nilai dari setiap parameter yang signifikan dan memenuhi uji diagnostik residu pada Tabel 2.

Tabel 2 Estimasi Parameter

Model	Parameter	Koefisien
ARIMA(1,1,1)	AR 1	0,4044
	MA 1	-0,9221
ARIMA(0,1,2)	MA 1	-0,5176
	MA 2	-0,2799
ARIMA(4,1,0)	AR 1	-0,4528
	AR 2	-0,4098
	AR 3	-0,2791
	AR 4	-0,1987

Kemudian dilakukan perhitungan nilai MAPE dari data *training* dan data *testing* untuk mengetahui model ARIMA yang paling tepat dalam meramalkan data suhu.

Tabel 3 Nilai MAPE Metode ARIMA

Model ARIMA	MAPE Data Training	MAPE Data Testing
ARIMA(1,1,1)	2,0095%	3,7009%
ARIMA(0,1,2)	2,0076%	3,2719%
ARIMA(4,1,0)	2,0238%	6,0537%

Tabel 3 menunjukkan bahwa model ARIMA(0,1,2) menghasilkan nilai MAPE terkecil baik pada data *training* maupun data *testing*. Hal ini menunjukkan bahwa model ARIMA(0,1,2) merupakan model ARIMA terbaik dalam meramalkan data suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani. Berdasarkan bentuk umum model ARIMA pada persamaan (1) dan nilai estimasi parameter pada Tabel 2 terbentuk model ARIMA dengan persamaan sebagai berikut.

$$\hat{Z}_t = Z_{t-1} + 0,5176a_{t-1} + 0,2799a_{t-2} + a_t$$

Analisis Metode SSA

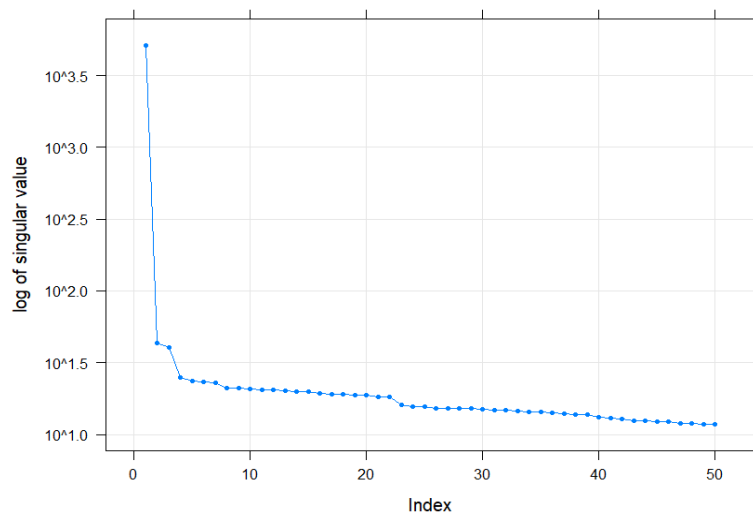
Tahap pertama yang dilakukan dalam analisis menggunakan metode SSA yaitu dekomposisi. Langkah awal dalam tahap dekomposisi adalah *embedding*. Nilai L yang digunakan pada analisis ini adalah sebesar 182. Pemilihan L maksimum bertujuan agar hasil dekomposisi lebih detail dan memiliki keterpisahan yang baik (Golyandina & Korobeynikov, 2014).

Matriks lintasan berukuran $L \times K$ dengan $L = \frac{N}{2} = \frac{365}{2} = 182$ dan $K = 365 - 182 + 1 = 184$ yang terbentuk dapat dituliskan dengan

$$X = (x_{ij})_{182,184} = \begin{bmatrix} x_1 & x_2 & \dots & x_{184} \\ x_2 & x_3 & \dots & x_{185} \\ x_3 & x_4 & \dots & x_{186} \\ \vdots & \vdots & \ddots & \vdots \\ x_{182} & x_{183} & \dots & x_{365} \end{bmatrix} = \begin{bmatrix} 27,7 & 27,7 & \dots & 28,9 \\ 27,7 & 28,0 & \dots & 28,5 \\ 28,0 & 28,2 & \dots & 28,4 \\ \vdots & \vdots & \ddots & \vdots \\ 27,6 & 28,8 & \dots & 24,5 \end{bmatrix}$$

Setelah memperoleh matriks lintasan $L \times K$, langkah selanjutnya yaitu SVD. Pada langkah SVD, matriks lintasan didekomposisi menjadi 182 *eigentriple* yang terdiri dari nilai singular ($\sqrt{\lambda_i}$), vektor eigen kiri (U_i), dan vektor eigen kanan (V_i).

Tahap rekonstruksi terdiri dari dua langkah yaitu *grouping* dan *diagonal averaging*. Pada tahap *grouping* dilakukan pengelompokan komponen data menjadi tren, musiman, dan *noise* dengan melihat 182 *eigentriple* yang telah didapatkan dari tahap dekomposisi. Langkah awal pada tahap *grouping* adalah membentuk *scree plot* dari nilai singular yang disajikan pada Gambar 8.

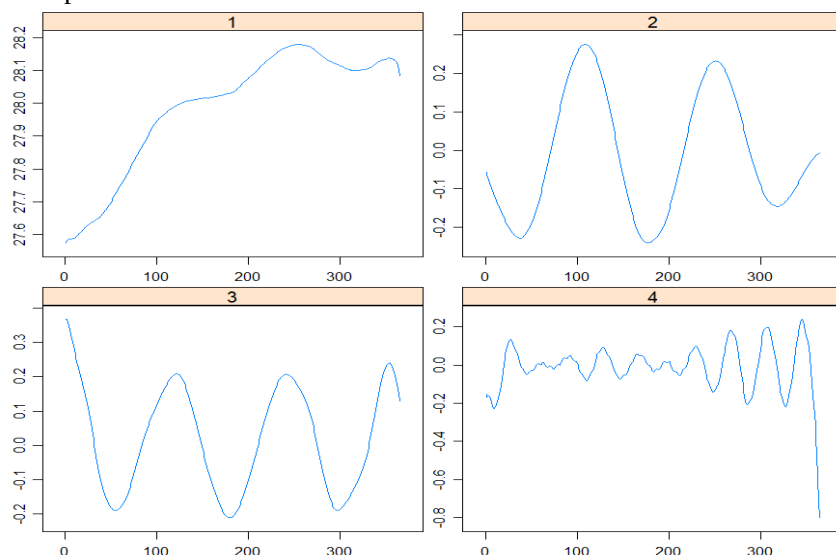


Gambar 8 Scree Plot Nilai Singular

Gambar 8 menjelaskan bahwa nilai singular ke-1 memiliki nilai algoritma yang paling besar. Nilai singular ke-2 dan ke-3 memiliki nilai yang cukup jauh dari nilai

singular ke-1 karena cenderung menurun secara drastis. Sementara itu, nilai singular mulai menurun secara lambat saat nilai ke-4 hingga ke-50 sehingga diidentifikasi sebagai kelompok *noise* (Hassani, 2007).

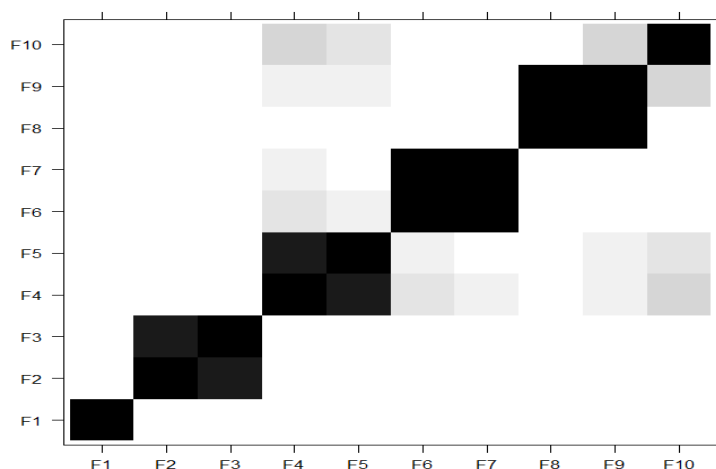
Langkah selanjutnya yaitu membentuk plot vektor eigen dengan hanya melihat 4 vektor eigen pertama karena terlihat bahwa mulai vektor eigen ke-4 telah dikategorikan sebagai kelompok *noise*.



Gambar 9 Plot Rekonstruksi dari 4 *Eigentriple* Pertama

Gambar 9 menunjukkan bahwa *eigentriple* ke-1 bervariasi secara lambat yang diidentifikasi sebagai komponen tren. *Eigentriple* ke-2 dan ke-3 menunjukkan komponen yang bervariasi secara lambat namun diikuti dengan adanya gelombang sinus, sehingga merupakan komponen musiman. Kemudian untuk *eigentriple* ke-4 dikategorikan sebagai komponen *noise* karena tidak memiliki pola.

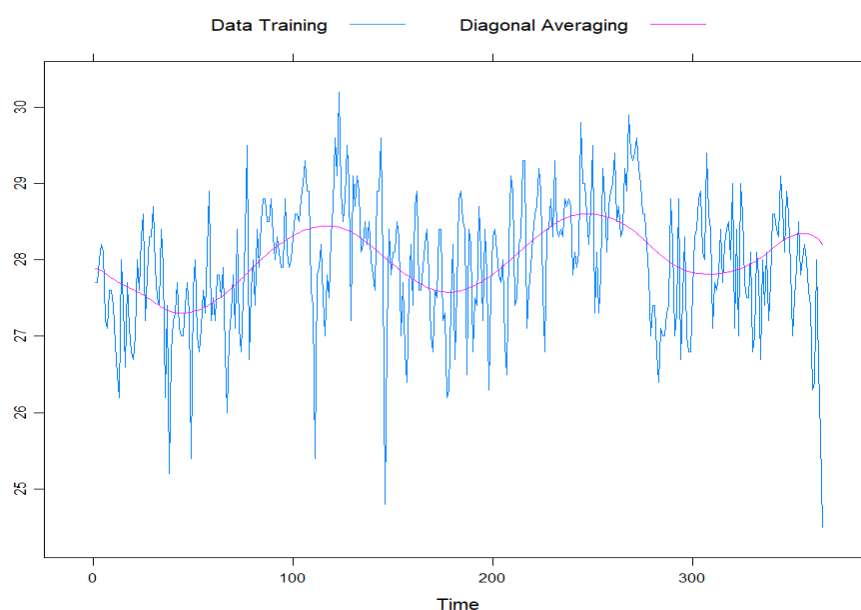
Langkah selanjutnya yaitu melihat besarnya korelasi antar *eigentriple* menggunakan plot matriks korelasi W. Plot matriks korelasi W digunakan untuk mengetahui keterpisahan dari masing-masing komponen. Nilai korelasi ditunjukkan dengan gradasi warna dari hitam (korelasi=1) ke putih (korelasi=0). Semakin gelap gradasi warnanya menunjukkan korelasi antar *eigentriple* yang semakin kuat.



Gambar 10 Plot Matriks Korelasi W

Gambar 10 menunjukkan keterpisahan yang kuat antara komponen sinyal (tren dan musiman) dan *noise* yang terlihat dengan adanya pola yang membentuk batas pada *eigentriple* ke-4. *Eigentriple* ke-1 menunjukkan keterpisahan yang kuat dengan *eigentriple* lainnya yang menandakan komponen tren, sedangkan *eigentriple* ke-2 dan ke-3 memiliki korelasi yang sangat kuat sehingga merupakan komponen musiman. *Eigentriple* ke-4 dan seterusnya merupakan komponen *noise* karena memiliki nilai korelasi beragam antar *eigentriple* yang ditandai dengan warna abu-abu pada plot korelasi W yang bervariasi.

Langkah kedua dalam tahap rekonstruksi adalah *diagonal averaging*. Komponen tren dan musiman yang telah dikelompokkan selanjutnya direkonstruksi ulang menjadi deret waktu yang baru.



Gambar 11 Plot Hasil Rekonstruksi

Gambar 11 menunjukkan plot perbandingan antara data *training* dan data *diagonal averaging* hasil rekonstruksi. Garis berwarna biru menunjukkan plot untuk data *training*, sedangkan garis berwarna ungu menunjukkan plot deret waktu yang baru hasil rekonstruksi. Gambar 11 menunjukkan bahwa pola data *diagonal averaging* hasil rekonstruksi mengikuti pola data *training*. Langkah selanjutnya yaitu membentuk model SSA menggunakan persamaan (4) dan nilai koefisien LRF. Nilai koefisien LRF yang dihasilkan ditunjukkan pada Tabel 4.

Tabel 4 Nilai Koefisien LRF

j	Koefisien LRF
1	0,00075
2	0,00082
$\vdots$	...
181	0,00963

Kemudian dilakukan pengecekan nilai MAPE dari data *training* dan data *testing* untuk melihat apakah model SSA yang terbentuk dapat meramalkan data suhu dengan baik.

Tabel 5 Nilai MAPE Metode SSA

Model	MAPE Data Training	MAPE Data Testing
SSA ($L = 182$)	2,1777%	3,9464%

Tabel 5 menunjukkan bahwa model SSA memiliki nilai MAPE yang kurang dari 10% baik pada data *training* maupun data *testing*, sehingga dapat dikatakan bahwa model SSA cukup akurat dalam meramalkan data suhu. Model SSA yang terbentuk adalah

$$\hat{g}_i = 0,00075\hat{g}_{i-1} + 0,00082\hat{g}_{i-2} + \dots + 0,00963\hat{g}_{i-181}$$

Pemilihan Metode Terbaik

Pemilihan metode yang lebih baik dapat ditentukan dengan melihat nilai MAPE terkecil dari kedua metode. Perbandingan nilai MAPE untuk metode ARIMA dan SSA ditampilkan pada Tabel 6.

Tabel 6 Perbandingan Nilai MAPE Metode ARIMA dan SSA

Metode	MAPE Data Training	MAPE Data Testing
ARIMA	2,0076%	3,2719%
SSA	2,1777%	3,9464%

Tabel 6 menunjukkan bahwa metode ARIMA lebih baik dibandingkan dengan metode SSA karena memiliki nilai MAPE yang lebih kecil baik pada data *training* maupun data *testing*. Hal ini berarti bahwa peramalan data suhu menggunakan metode ARIMA memiliki akurasi yang lebih tinggi dan menghasilkan peramalan yang lebih akurat dibandingkan dengan metode SSA.

HASIL DAN PEMBAHASAN

Anwar (2017) melakukan peramalan data suhu dengan menggunakan metode ARIMA. Data yang digunakan Anwar (2017) adalah data harian suhu udara minimum dan maksimum di Kota Banda Aceh yang terdiri dari 2.320 data, sedangkan penelitian ini menggunakan data suhu udara rata-rata harian di Stasiun Meteorologi Ahmad Yani yang terdiri dari 396 data. Model ARIMA terbaik yang dihasilkan Anwar (2017) untuk data suhu udara minimum adalah ARIMA(0,2,5) dan untuk data suhu udara maksimum yaitu ARIMA(0,2,3), sedangkan model ARIMA terbaik yang terbentuk pada penelitian ini adalah ARIMA(0,1,2). Nilai ukuran ketepatan peramalan yang digunakan Anwar (2017) adalah RMSE, berbeda dengan penelitian ini yang menggunakan MAPE. Nilai RMSE yang diperoleh Anwar (2017) sebesar 0,969 untuk data suhu udara minimum dan sebesar 1,289 untuk data suhu udara maksimum, sedangkan nilai MAPE yang dihasilkan pada penelitian ini sebesar 3,2719%. Penggunaan ukuran ketepatan peramalan yang berbeda membuat penelitian Anwar (2017) terlihat memiliki *error* yang lebih kecil dibandingkan penelitian ini. Namun, kedua penelitian menghasilkan peramalan dengan akurasi yang tinggi karena menghasilkan *error* yang kecil.

Rumagit dan Azhari (2013) melakukan peramalan pemakaian listrik di Sulluttenggo dengan membandingkan metode ARIMA, Jaringan Saraf Tiruan (JST), dan Hibrid ARIMA-JST. Data yang digunakan merupakan data bulanan pemakaian listrik di wilayah Sulluttenggo dari Januari 2004 hingga Desember 2010 yang terdiri dari kelompok tarif sosial, rumah tangga, bisnis, industri, dan pemerintah. Hasil penelitian Rumagit dan Azhari (2013) menunjukkan bahwa metode ARIMA menghasilkan nilai MAPE terkecil yaitu sebesar 0,00418% untuk tarif sosial, 0,00316% untuk tarif rumah tangga, 0,00608% untuk tarif bisnis, 0,00597% untuk tarif industri, dan 0,01006% untuk tarif pemerintah dengan masing-masing model ARIMA terbaik adalah ARIMA(1,1,2),

ARIMA(1,1,0), ARIMA(2,1,2), ARIMA(2,1,2), dan ARIMA(1,1,0). Hal ini berarti bahwa pada penelitian Rumagit dan Azhari (2013) metode ARIMA lebih baik dibandingkan dengan metode JST dan hibrid ARIMA-JST, sehingga merupakan salah satu alasan digunakan metode ARIMA pada penelitian ini.

Sakinah (2018) melakukan peramalan suhu dengan menggunakan metode SSA. Data yang digunakan Sakinah (2018) adalah data bulanan suhu udara di Jakarta, sedangkan penelitian ini menggunakan data harian suhu udara di Stasiun Meteorologi Ahmad Yani. Nilai L pada penelitian Sakinah (2018) ditentukan berdasarkan *trial and error* dengan nilai L sebesar 204, sedangkan pada penelitian ini digunakan nilai L maksimum yang merupakan setengah dari jumlah data *training* dengan nilai L sebesar 182. Sakinah (2018) mengelompokkan *eigentriple* dalam tiga komponen, yaitu *eigentriple* 1 merupakan komponen tren, *eigentriple* 2, 3, 4, dan 5 merupakan komponen musiman, dan *eigentriple* sisanya merupakan komponen *noise*. Pada penelitian ini, *eigentriple* juga dikelompokkan dalam tiga komponen yaitu *eigentriple* 1 merupakan komponen tren, *eigentriple* 2 dan 3 merupakan komponen musiman, dan *eigentriple* sisanya merupakan komponen *noise*. Nilai MAPE pada penelitian ini sebesar 3,9464%, sedangkan nilai MAPE pada penelitian Sakinah (2018) sebesar 5,0029%. Model SSA yang dihasilkan pada analisis ini lebih baik dibandingkan dengan penelitian yang dilakukan Sakinah (2018) karena menghasilkan nilai MAPE yang lebih kecil.

Sodiqin *et al.* (2021) melakukan peramalan dengan membandingkan metode SSA *Recurrent* dan SSA *Vector*. Data yang digunakan Sodiqin *et al.* (2021) merupakan data bulanan jumlah kunjungan wisatawan asing ke DKI Jakarta dari Januari 2011 hingga Desember 2019. Nilai L yang digunakan pada penelitian Sodiqin *et al.* (2021) adalah 40 yang ditentukan melalui proses *trial and error*, sedangkan penelitian ini menggunakan nilai L maksimum yaitu 182. Penelitian Sodiqin *et al.* (2021) menghasilkan 3 kelompok *eigentriple*, yaitu *eigentriple* 1 sebagai komponen tren, *eigentriple* 2, 3, 4, dan 5 sebagai komponen musiman, dan *eigentriple* 6 hingga 40 sebagai *noise*. Penelitian Sodiqin *et al.* (2021) menghasilkan nilai MAPE metode SSA menggunakan algoritma *recurrent* sebesar 16,6% dan menggunakan algoritma *vector* sebesar 18,1%. Hal ini menunjukkan bahwa pada penelitian Sodiqin *et al.* (2021) metode SSA dengan algoritma *recurrent* lebih baik daripada menggunakan algoritma *vector*, sehingga pada penelitian ini digunakan metode SSA dengan algoritma *recurrent*. Hasil yang diperoleh penelitian ini lebih akurat dibandingkan dengan penelitian Sodiqin *et al.* (2021) karena menghasilkan nilai MAPE yang lebih kecil dan kurang dari 10%.

Cristian (2018) melakukan peramalan suhu rata-rata di Romania dengan membandingkan metode SSA, ARIMA, dan *Neural Network* (NN). Data yang digunakan Cristian (2018) merupakan data suhu rata-rata bulanan di Romania dari tahun 1990 hingga 2015 yang terdiri dari 300 data. Pada metode SSA, Cristian (2018) menggunakan nilai L sebesar 144 yang menghasilkan 3 kelompok *eigentriple* yaitu *eigentriple* 1 sebagai tren, *eigentriple* 2 dan 3 sebagai musiman, dan sisanya sebagai *noise*. Hasil penelitian Cristian (2018) menunjukkan bahwa metode SSA lebih baik dibandingkan dengan metode ARIMA dan NN karena menghasilkan nilai RMSE dan MAPE terkecil yaitu sebesar 2,02 dan 73,29%. Hal ini berbeda dengan hasil penelitian ini yang memperoleh kesimpulan bahwa metode ARIMA lebih baik dibandingkan dengan metode SSA karena menghasilkan nilai MAPE terkecil.

PENUTUP

Model ARIMA terbaik untuk meramalkan data suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani adalah menggunakan model ARIMA(0,1,2) yang merupakan model dengan semua parameternya signifikan dan memenuhi uji asumsi residu. Analisis

metode SSA menggunakan nilai L sebesar 182 menghasilkan 3 kelompok komponen, yaitu komponen tren terdiri dari 1 *eigentriple*, komponen musiman terdiri dari 1 pasang *eigentriple*, dan sisanya sebagai komponen *noise*. Model ARIMA(0,1,2) menghasilkan nilai MAPE untuk data *training* sebesar 2,0076% dan data *testing* sebesar 3,2719%, sedangkan metode SSA menghasilkan nilai MAPE untuk data *training* sebesar 2,1777% dan data *testing* sebesar 3,9464%. Jadi, dapat disimpulkan bahwa metode yang lebih tepat digunakan untuk meramalkan data suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani adalah menggunakan metode ARIMA. Metode ARIMA menghasilkan nilai MAPE kurang dari 10% yang berarti bahwa hasil peramalan memiliki akurasi yang sangat tinggi. Saran untuk penelitian selanjutnya diharapkan dapat menghasilkan peramalan suhu rata-rata harian di Stasiun Meteorologi Ahmad Yani yang lebih baik dan menghasilkan MAPE yang lebih kecil dengan menggunakan parameter yang lebih bervariasi untuk digunakan dalam metode SSA.

DAFTAR RUJUKAN

- Anwar, S. (2017). Peramalan Suhu Udara Jangka Pendek di Kota Banda Aceh dengan Metode Autoregressive Integrated Moving Average (ARIMA). *Malikussaleh Journal of Mechanical Science and Technology*, 5(1), 6–12.
- Cristian, M. (2018). Average Monthly Temperature Forecast in Romania By Using Singular Spectrum Analysis. *Academica Brancusi Publisher*, 3, 48–57.
- Ete, A. A. (2017). *Forecasting The Amount Of Foreign Tourist Travelers to Indonesia by Entrance Guide Using Singular Spectrum Analysis and ARIMA* (Thesis). Institut Teknologi Sepuluh Nopember, Surabaya.
- Febrianti, T. K., Sulandari, W., & Pratiwi, H. (2021). Peramalan Curah Hujan di Kota Bandung Menggunakan Singular Spectrum Analysis. *Jurnal Ilmiah Matematika*, 8(2), 56–65.
- Golyandina, N., & Korobeynikov, A. (2014). Basic Singular Spectrum Analysis and forecasting with R. *Computational Statistics and Data Analysis*, 71, 934–954. doi: <https://doi.org/10.1016/j.csda.2013.04.009>.
- Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5(2), 239–257. doi: [https://doi.org/10.6339/JDS.2007.05\(2\).396](https://doi.org/10.6339/JDS.2007.05(2).396).
- Khaeri, H., Yulian, E., & Darmawan, G. (2018). Penerapan Metode Singular Spectrum Analysis (SSA) Pada Peramalan Jumlah Penumpang Kereta Api di Indonesia Tahun 2017. *Jurnal Euclid*, 5(1), 8–20. doi: <https://doi.org/10.33603/e.v5i1.496>.
- Lewis, C. D. (1982). *International and Business Forecasting Methods*. London: Butterworths.
- Masyitoh, A. D., Zukhronah, E., & Slamet, I. (2022, March 24). *Forecasting for Monthly Rainfall in Gondang Reservoir Lamongan Using Singular Spectrum Analysis (SSA) Method*. doi: <https://doi.org/10.1063/5.0075313>.
- Muis, S., & Setiyadi, D. (2020). Model Statistik Arima Dalam Meramal Pergerakan Harga Saham. *Information System for Educators and Professionals*, 4(2), 154–167.
- Prasetyo, S., Hidayat, U., Haryanto, Y. D., & Riama, N. F. (2021). Variasi dan Trend Suhu Udara Permukaan di Pulau Jawa Tahun 1990-2019. *Jurnal Geografi : Media Informasi Pengembangan Dan Profesi Kegeografian*, 18(1), 60–68. doi: <https://doi.org/10.15294/jg.v18i1.27622>.
- Pusat Database BMKG. (2015). *Data Harian Temperatur Rata-Rata di Stasiun Meteorologi Ahmad Yani*. Diakses pada 30 Januari 2023 dari http://dataonline.bmkg.go.id/data_iklim.

- Rumagit, S. E., & Azhari. (2013). Prediksi Pemakaian Listrik Menggunakan Jaringan Syaraf Tiruan dan ARIMA Di Wilayah Sulluttenggo. *Berkala Ilmiah MIPA*, 23(2), 139–148.
- Sakinah, A. M. (2018). Akurasi Peramalan Long Horizon dengan Singular Spectrum Analysis. *Kubik: Jurnal Publikasi Ilmiah Matematika*, 3(2), 93–99. doi: <https://doi.org/10.15575/kubik.v3i2.4111>.
- Samidjo, J., & Suharso, Y. (2017). Memahami Pemanasan Global dan Perubahan Iklim. *Journal IKIP Veteran Semarang*, 24(2), 36–46.
- Sodiqin, M. A., Sulandari, W., & Respatiwan. (2021). The Application of Singular Spectrum Analysis Method in Forecasting The Number of Foreign Tourists Visit to Special Capital Region of Jakarta. *Jurnal Riset Dan Aplikasi Matematika*, 5(2), 92–102.
- Solopos.com. (2022). *Tak Hanya Jateng, Semarang Juga Termasuk Kota Terpanas di Indonesia*. Dipetik pada 30 Januari 2023 dari <https://www.solopos.com/tak-hanya-jateng-semarang-juga-termasuk-kota-terpanas-di-indonesia-1321732>.
- Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.). New York: Pearson Addison Wesley.