

Forecasting the Number of Passengers for the Jakarta-Bandung High-Speed Rail Using SARIMA and SSA Models

Peramalan Jumlah Penumpang Kereta Cepat Jakarta-Bandung dengan Model SARIMA dan SSA

Authors' name* (Font 12)

**Affiliation of authors (italic, font 11)*

Email: email addressess of authors(italic, font 11)

Abstract (Font 11)

(content of abstract, font 10.)

Keywords: at least three keywords (font 11).

Abstrak (font 11)

(for only manuscript in Indonesian language, Font 11)

(content of abstract, font 10.)

Kata kunci: at least three keywords (font 11).

1. PENDAHULUAN

Dalam era modern yang bergerak dengan cepat, transportasi menjadi hal yang sangat penting bagi kehidupan manusia. Transportasi memiliki peran di berbagai aspek bukan hanya sekedar dalam mobilitas individu, tetapi juga berfungsi sebagai pilar utama untuk menunjang ekonomi, sosial, dan budaya di berbagai wilayah. Kemudahan dalam akses transportasi memungkinkan terjalannya konektivitas antarwilayah, mempercepat pertukaran informasi, serta memperlancar distribusi barang dan jasa yang mendukung aktivitas perekonomian.

Di provinsi DKI Jakarta yang memiliki penduduk padat mengakibatkan produktivitas yang tinggi serta mobilitas yang kompleks membutuhkan transportasi cepat untuk dapat menunjang berbagai aspek, selain itu juga dapat untung mempermudah perpindahan masyarakat. Oleh karena itu, pemerintah perlu membangun infrastruktur di koridor Jakarta hingga Bandung guna mendorong dampak positif bagi perkembangan berbagai sektor kehidupan dan perekonomian, baik di ibu kota Indonesia maupun di ibu kota Provinsi Jawa Barat.

Kereta cepat adalah salah satu opsi yang dapat diambil oleh pemerintah untuk memodernisasi transportasi massal di Indonesia. Kebijakan ini bertujuan meningkatkan keterhubungan antarwilayah dan antar kota, yang diharapkan memberikan dampak positif (Ramadhan et al., 2023). Dalam konteks pengelolaan transportasi publik, monitoring dan peramalan jumlah penumpang menjadi penting untuk memahami pola permintaan dan mengoptimalkan alokasi sumber daya, seperti jumlah kereta yang beroperasi dan ketersediaan tempat duduk.

Salah satu metode yang efektif dalam memprediksi pola jumlah penumpang adalah analisis deret waktu. *Seasonal Autoregressive Integrated Moving Average* (SARIMA) merupakan model statistik yang umum digunakan untuk pola musiman yang terstruktur. SARIMA terbukti efektif dalam hal penangkapan pola musiman, misalnya pada berbagai data transportasi. Namun, model ini memiliki keterbatasan dalam penanganan data dengan *noise* tinggi atau pola yang tidak sepenuhnya stasioner (Box et al., 2015). Alternatif lainnya dengan menggunakan *Singular Spectrum Analysis* (SSA) yang dikenal lebih fleksibel dalam menangani data musiman dengan komponen nonlinier dan *noise* yang lebih tinggi. Sehingga, lebih adaptif terhadap dinamika permintaan penumpang yang bervariasi (Golyandina, Nina & Zhigljavsky, 2013).

Penelitian ini bertujuan membandingkan performa model SARIMA dan SSA dalam permalan jumlah penumpang harian kereta cepat “Whoosh” pada suatu rentang waktu tertentu. Diharapkan hasil penelitian ini dapat memberikan rekomendasi model peramalan terbaik yang mendukung pengambilan keputusan operasional guna meningkatkan efisiensi dan responsivitas kereta cepat terhadap perubahan pola penumpang.

2. METODOLOGI

2.1 Bahan dan Data

Data yang digunakan pada penelitian ini adalah data sekunder berupa jumlah penumpang harian yang ada di stasiun kereta cepat Jakarta-Bandung Halim. Data ini diperoleh dari Website Sistem Informasi Angkutan dan Sarana Transportasi Indonesia <https://dashboard-siasati.dephub.go.id/> dengan rentang periode 1 November 2023 hingga 30 September 2024.

2.2 Metode Penelitian

Analisis data yang dilakukan menggunakan *software* R. Tahapan-tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Melakukan pembagian data menjadi 3 skenario. Pembagian data menjadi 3 skenario ini bertujuan untuk melakukan validasi pada model yang dibangun pada peramalan deret waktu. Salah satu skenario akan digunakan untuk membangun model terlebih dahulu kemudian akan di ujikan pada skenario lainnya. Skenario pembagian data diringkas pada tabel X.X

Tabel X.X. Skenario pembagian data

Skenario	Rasio pembagian (latih : uji)	Jumlah data latih	Jumlah data uji
1	70:30	236	97
2	80:20	267	66
3	90:10	298	35

2. Melakukan pemodelan SARIMA dengan tahapan sebagai berikut:

- Melakukan eksplorasi data melalui plot data deret waktu, plot *Autocorrelation Function* (ACF), dan plot Box-Cox pada data deret waktu.
- Melakukan pemeriksaan kestasioneran data.
 - Pemeriksaan kestasioneran data dalam ragam dengan uji Box-Cox. Data dikatakan homogen dalam ragam jika selang kepercayaan 95% Box-Cox mencakup nilai 1. Data yang belum stasioner dalam ragam akan dilakukan transformasi.
 - Pemeriksaan kestasioneran data dalam rata-rata dilakukan secara eksplorasi dengan plot ACF dan juga dengan uji *Augmented Dickey Fuller* (ADF). Data yang belum stasioner dalam rata-rata akan dilakukan perbedaan (*differencing*).

- c. Melakukan identifikasi model menggunakan plot ACF, PACF, dan EACF untuk menentukan orde dari komponen ARMA musiman maupun non-musiman. Identifikasi model dilakukan pada beberapa kemungkinan pola *cut-off* dan *tails-off* plot ACF dan PACF serta dengan bantuan EACF.
- d. Melakukan pendugaan parameter dan pemilihan model tentatif terbaik dengan melakukan pemeriksaan signifikansi dari pendugaan parameter pada model yang sudah ditentukan. Kemudian, dipilih model yang terbaik berdasarkan nilai AIC terkecil dan nilai yang signifikan di semua parameter.
- e. Melakukan *overfitting* pada model tentatif terbaik. *Overfitting* merupakan proses menaikkan ordo ARMA secara bergantian untuk memastikan model yang terpilih merupakan model terbaik. Hasil *overfitting* akan dibandingkan dengan model yang terpilih sebelumnya berdasarkan nilai AIC yang lebih kecil dan signifikansi parameter *overfitting*.
- f. Melakukan diagnostik sisaan pada model terbaik yang terpilih. Pengujian yang dilakukan antara lain pengujian kenormalan sisaan, kebebasan sisaan, dan kehomogenan ragam sisaan.

3. Melakukan pemodelan *Singular Spectrum Analysis* (SSA)

1. Dekomposisi
 - a. Tahap *Embedding*. Tahap ini mengonversi data deret waktu satu dimensi menjadi bentuk matriks yang disebut *trajectory matrix* berukuran L dengan $K = N - L + 1$ (jumlah kolomnya). Misalnya, terdapat data deret waktu $X = \{x_1, x_2, \dots, x_n\}$ dengan panjang N . Kemudian, data diubah ke dalam matriks berukuran $L \times K$. Penentuan parameter *Window Length* (L) dapat dilakukan melalui eksplorasi dengan plot *W-Correlation Matrix* dengan kriteria keterpisahan komponen dan juga dapat dilakukan melalui *trial-and-error* yang memenuhi $1 < L < N$. Setelah itu, *trajectory matrix* dapat didefinisikan sebagai:

$$X = [x_1, \dots, x_n] = (x_{ij})_{i,j=1}^{L,K} =$$

- b. Membuat *Singular Value Decomposition* (SVD) untuk mendekomposisi nilai dari matriks lintasan menjadi suatu penjumlahan dari matriks yang terbentuk dari *eigen triple*. Diberikan $S = XX^T$. Matriks S memiliki *eigenvalue* $\lambda_1, \dots, \lambda_L$ dengan urutan yang terbesar ke terkecil $\lambda_1 \geq \dots \geq \lambda_L \geq 0$ dan U_1, \dots, U_L merupakan *eigenvector*. Kemudian, mendefinisikan nilai $d = \max\{i \mid \lambda_i > 0\}$ sebagai rank dari matriks lintasan. Notasi dari komponen utama, yaitu $v_i = \frac{X^T u_i}{\sqrt{\lambda_i}}$, untuk setiap $i = 1, 2, \dots, d$. Maka SVD dapat dinyatakan sebagai persamaan sebagai berikut:

$$\begin{aligned} X &= X_1 + \dots + X_d \\ &= \sqrt{\lambda_1} u_1 v_1^T + \sqrt{\lambda_2} u_2 v_2^T + \dots + \sqrt{\lambda_d} u_d v_d^T \\ &= \sum_{i=1}^d \sqrt{\lambda_i} u_i v_i^T \end{aligned}$$

2. Melakukan rekonstruksi sinyal dengan memilih beberapa komponen yang mampu mewakili data yang ada.
3. Melakukan pemodelan dari model terbaik yang diperoleh dari proses sebelumnya pada data latih ketiga skenario.
4. Melakukan peramalan sebanyak data uji pada ketiga skenario data.
5. Membandingkan performa peramalan pada data uji dari SARIMA dan SSA pada ketiga skenario data dengan menghitung Nilai MAPE.

$MAPE = \frac{\sum_{i=1}^n y_i - \hat{y}_i }{y_i} \times 100\%$	(1)
--	-----

7. Melakukan peramalan dengan hasil model yang lebih terbaik untuk 60 periode ke depan.

3. HASIL DAN PEMBAHASAN

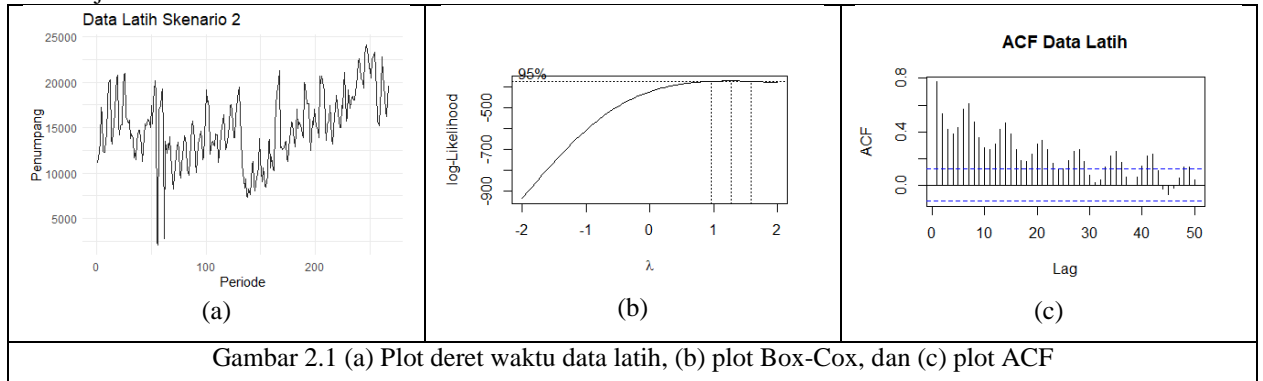
Langkah awal dalam analisis dilakukan dengan melakukan pembagian data menjadi dua bagian yaitu data latih dan data uji. Pembagian ini bertujuan untuk memberikan gambaran dari kemampuan model yang dibangun dalam melakukan peramalan. Pembagian ini dilakukan dengan 3 skenario pembagian data dengan harapan kita mendapatkan kemampuan peramalan. Skenario pembagian data ini disajikan pada tabel 2.1. Skenario 2 akan dijadikan dasar pemodelan awal dari SARIMA dan SSA yang kemudian dicobakan untuk skenario lainnya.

Tabel 2.1. Skenario pembagian data

Skenario	Rasio pembagian (latih : uji)	Jumlah data latih	Jumlah data uji
1	70:30	236	97
2	80:20	267	66
3	90:10	298	35

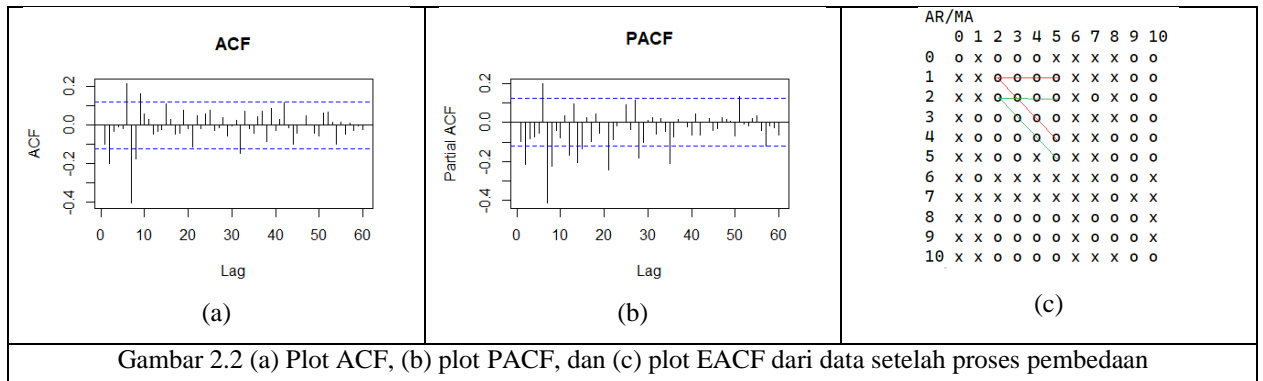
1. Pemodelan SARIMA

Eksplorasi data dilakukan dengan membuat plot deret waktu, plot ACF, dan plot Box-Cox. Plot deret waktu dapat memberikan gambaran terdapat tren yang tidak linear pada data dengan pola musiman. Terjadi penurunan angka yang cukup tinggi di tanggal 25 dan 26 Desember 2023 serta terdapat penurunan setelah terjadi tren kenaikan di pertengahan bulan Maret 2024. Plot ACF dari data menunjukkan nilai autokorelasi yang turun dengan perlahan yang mengindikasikan data cenderung tidak stasioner dalam rata-rata. Plot Box-Cox menunjukkan selang kepercayaan 95% mencakup nilai satu yang mengindikasikan data sudah stasioner dalam ragam. Proses pembedaan dilakukan pada data dengan ordo pembedaan sebesar 1 pada komponen d dan D . Setelah dilakukan pembedaan, nilai uji Dicky-Fuller menunjukkan p -value sebesar <0.01 yang menunjukkan data sudah stasioner dalam rata-rata.



Gambar 2.1 (a) Plot deret waktu data latih, (b) plot Box-Cox, dan (c) plot ACF

Identifikasi model tentatif dilakukan melalui plot ACF, plot PACF, dan plot EACF. Pada plot EACF, Model ARIMA akan cenderung membentuk segitiga-nol (*triangle of zeros*), dengan nilai pojok kiri atas bersesuaian dengan ordo dari ARMA (JURNAL KAK NICKY). Pada plot ACF dan plot PACF, ordo ARMA komponen non-musiman dan komponen musiman dapat ditentukan melalui pola *cut-off* dan *tails-off* yang terlihat. Penentuan ordo musiman dilakukan terlebih dahulu, kemudian ordo dari komponen non-musiman ditentukan melalui kemungkinan-kemungkinan interpretasi dari plot ACF, plot PACF, dan plot EACF.



Gambar 2.2 (a) Plot ACF, (b) plot PACF, dan (c) plot EACF dari data setelah proses pembedaan

Plot ACF dan PACF menunjukan pola mingguan karena nilai autokorelasi signifikan pada *lag* dengan kelipatan 7. Plot ACF mengalami *cut-off* setelah lag ke 7 dan plot PACF cutoff setelah lag ke 35. Jika salah satu plot diasumsikan *tails-off* maka terdapat 2 kemungkinan ordo komponen musiman. Penentuan ordo untuk komponen non-musiman dapat dilakukan dengan proses yang sama seperti mengidentifikasi ordo untuk komponen musiman namun dengan tambahan plot EACF. Model-model tentatif yang diperoleh akan dilanjutkan untuk pendugaan parameter, uji signifikansi parameter, dan perhitungan nilai AIC untuk menentukan model terbaik. Proses ini dilakukan dengan bantuan *software* R. Ringkasan dari proses ini tersaji pada tabel 2.2.

Tabel 2.2. Ringkasan model tentatif

Model Tentatif	Signifikansi Parameter	Nilai AIC
ARIMA(1,1,2)(0,1,1) ₇	Signifikan	4719,48
ARIMA(2,1,2)(0,1,1) ₇	Tidak Signifikan	4721,15
ARIMA(1,1,1)(0,1,1)₇	Signifikan	4719,96
ARIMA(1,1,2)(5,1,0) ₇	Tidak Signifikan	4724,74
ARIMA(2,1,2)(5,1,0) ₇	Signifikan	4725,54
ARIMA(1,1,1)(5,1,0) ₇	Signifikan	4723,73

Model terbaik yang diperoleh adalah model ARIMA(1,1,1)(0,1,1)₇ dengan nilai AIC 4719,96 dan seluruh parameter signifikan. Ukuran kebaikan model lainnya juga menunjukkan hasil yang cukup baik, dengan nilai MAPE sebesar 12.88% dan RMSE sebesar 2082.95. Model tentatif terbaik tersebut akan dianalisis lebih lanjut melalui proses *overfitting* untuk memastikan bahwa model yang kita peroleh merupakan model terbaik.

Tabel 2.3. Ringkasan proses *overfitting*

Model Overfitting	Signifikansi Parameter	Nilai AIC
ARIMA(2,1,1)(0,1,1) ₇	Tidak Signifikan	4720,56
ARIMA(1,1,2)(0,1,1) ₇	Signifikan	4719,48
ARIMA(1,1,1)(1,1,1)₇	Signifikan	4716,01
ARIMA(1,1,1)(0,1,2) ₇	Signifikan	4716,28

Proses *overfitting* menunjukan model baru yang lebih baik dari model yang sebelumnya kita peroleh. Hal ini ditunjukan dari nilai AIC yang lebih kecil dan juga semua parameter yang juga signifikan. Ukuran kebaikan model baru dari proses *overfitting*, seperti MAPE dan RMSE menunjukan hasil yang lebih baik, dengan nilai MAPE sebesar 12.51% dan RMSE sebesar 2057.83. Hasil diagnostik sisaan menunjukan

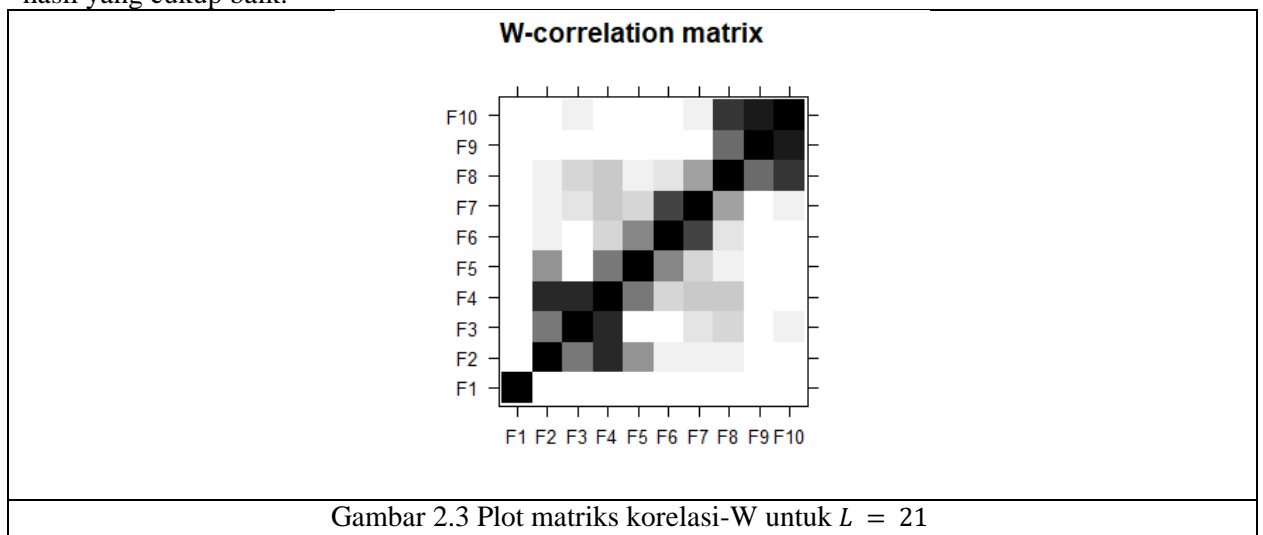
terpenuhinya asumsi sisaan saling bebas dan ragam sisaan homogen namun tidak dengan asumsi kenormalan. Hal ini relatif dapat diabaikan karena tidak memengaruhi hasil peramalan, tetapi memengaruhi pengujian hipotesis dari parameter yang ada (JURNAL MENGABAIKAN ASUMSI KENORMALAN). Oleh karena itu, model yang akan digunakan dalam peramalan adalah model $ARIMA(1,1,1)(1,1,1)_7$.

Tabel 2.3. Ringkasan uji diagnostik sisaan

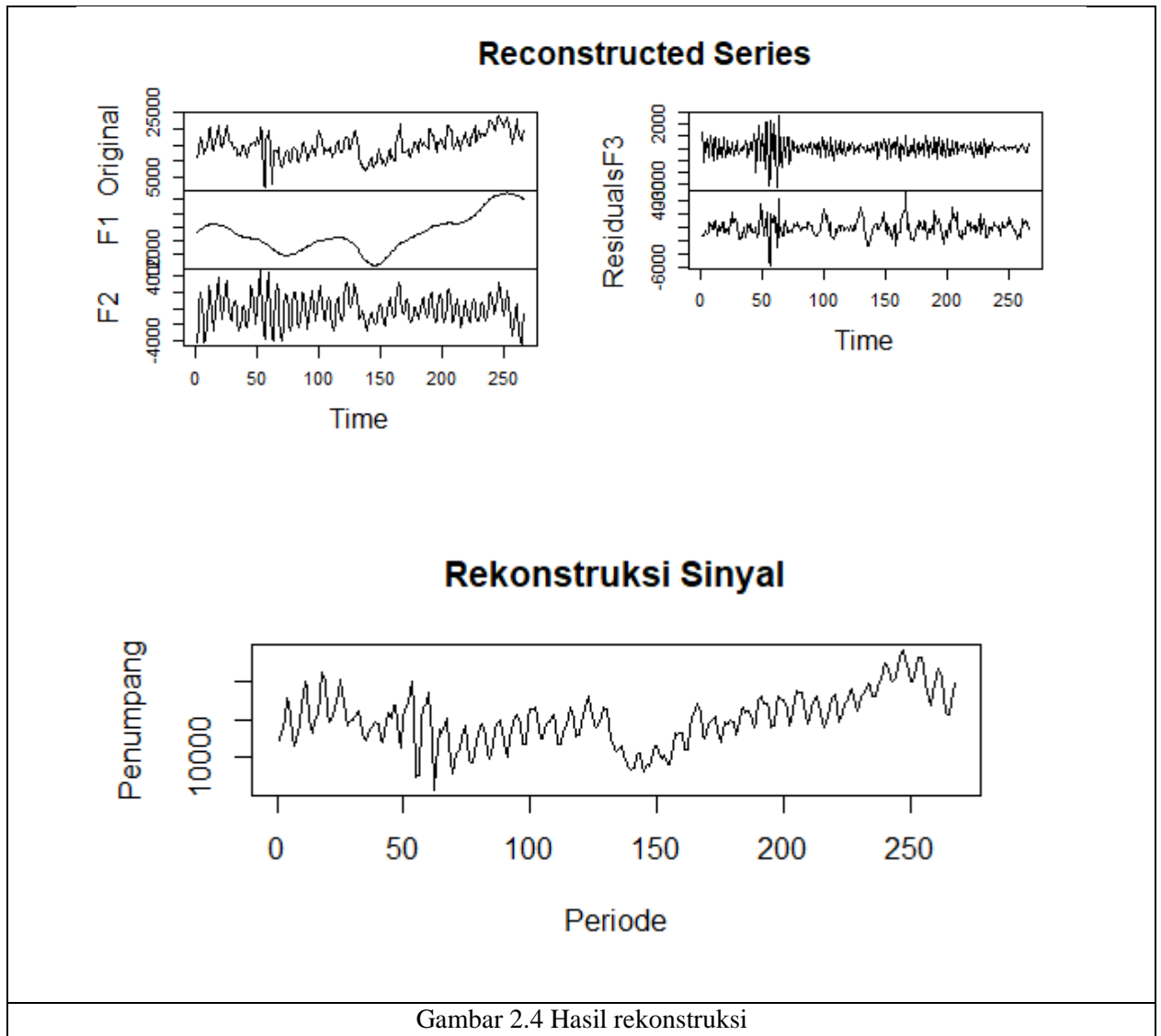
Uji diagnostik sisaan	<i>p-value</i>	Keterangan
Jarque-Bera <i>Test</i>	0,0000	Sisaan tidak menyebar normal
Box-Ljung <i>Test on Residuals</i>	0,4387	Sisaan saling bebas
Box-Ljung <i>Test on Squared Residuals</i>	0,0736	Ragam sisaan homogen

2. Pemodelan SSA

Pemodelan SSA dilakukan dengan menentukan nilai L yang akan digunakan untuk melakukan dekonstruksi sinyal. Nilai L diperoleh melalui proses *trial-and-error* dengan mencari L yang dapat memisahkan gugus data deret waktu menjadi beberapa komponen yang saling berkaitan dan juga kemampuan dalam memprediksi. Pada penelitian kali ini ditemukan bahwa $L = 21$ menunjukkan hasil yang cukup baik.



Plot matriks korelasi-W menunjukan hubungan antara komponen-komponen dari deret waktu yang telah di dekonstruksi sebelumnya, semakin gelap petak, maka semakin tinggi korelasi antara komponen-komponen. Petak yang memiliki korelasi tinggi akan digabung menjadi satu kesatuan komponen pada proses rekonstruksi sinyal. Berdasarkan plot pada gambar 2.3, proses rekonstruksi sinyal akan menggunakan komponen 1,2,3,4,6,7,8,9, dan 10 dengan penggabungan pada komponen 2,3, dan 4 serta 8,9, dan 10.



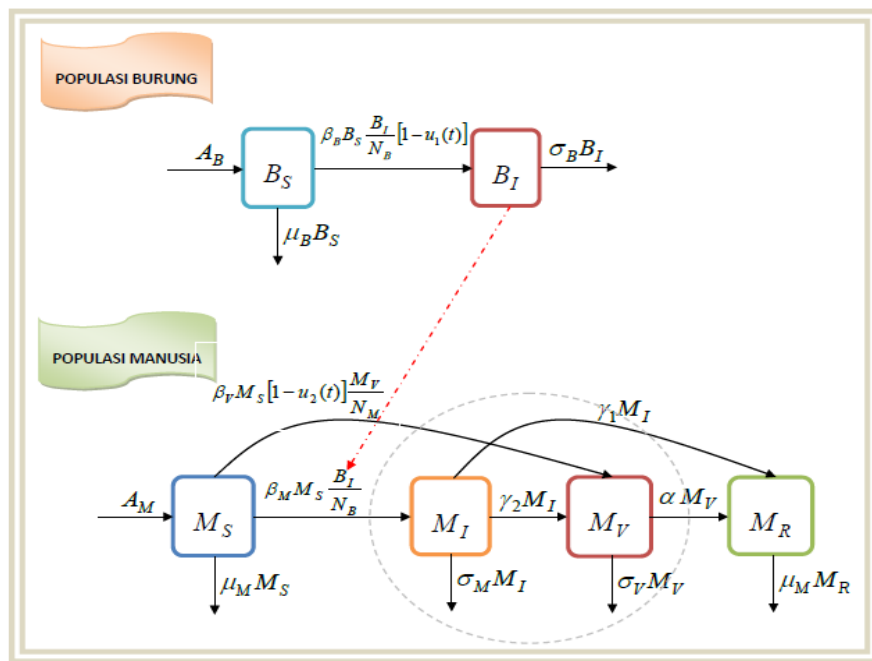
Komponen F1 menjelaskan tren non linear, F2 menjelaskan tren seasonal yang cenderung berubah, dan F3 mengakomodir fluktuasi nilai.

3. Perbandingan performa peramalan dengan beberapa skenario data

Skenario	SARIMA		SSA	
	MAPE Train	MAPE Test	Data uji	Data latih
Skenario 1	13.56 %	14.19%	9.33%	12.93%
Skenario 2	12.51 %	22.19 %	8.67%	19.92%
Skenario 3	11.67%	10.83%	8.08%	10.72%

Table 2.1. Meaning of the symbols

$N(t)$	Number of Prey at time t
$P(t)$	Number of Predator at time t
r	Growth rate of Prey
a	decrements of prey
K	Carrying capacity of Prey



(Font 11) Figure 2.1. Compartment diagram

Theorem 2.1. *(Font 11italic))The stability of equilibrium points E_1 , E_2 of the system (1) is given by*

.....

Proof. From the Jacobian matrix of System (1.1)...

Theorem 2.2. (*Font 11italic*) *The stability of equilibrium points E_3 , E_4 of the system (2.1) is given by*
.....

Proof. From the Jacobian matrix of System (2.1)...

Example 2.3. (*Font 11italic*) *This is an example ..*

3. CONCLUSION (if any) (Font 14)

(Font 11) The conclusion should be concise and clear

ACKNOWLEDGEMENT (optional)

(Font 11) Give informations of financial support institutions (if any) and greeting of thank.

CONFLICT OF INTEREST

(Font 11) The authors declare that there is no conflict of interest

REFERENCES (Font 14)

(Alphabetically ordered references), Times new roman, font 11)

- [1] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. & Ljung, G.M., 2015. *Time Series Analysis: Forecasting and Control*, Fifth Edition. John Wiley & Sons Inc., New Jersey.
- [1] Golyandina, Nina & Zhigljavsky. 2013. *Singular Spectrum Analysis for Time Series*. Springer. doi: 10.1007/978-3-642-34913-3.
- [2] Hattaf, K., 2009. Optimal Control of Treatment in a Basic Virus Infection Model. *Applied Mathematical Sciences*, Vol. 3, No. 20, 949 – 958.
- [3] Jung, E., Lenhart S. & Feng Z., 2002. Optimal Control of Treatments in a Two-Strain Tuberculosis Model. *Disc. & Cont. Dynamical Systems–Series B*, Vol. 2, No. 4, 473-482.
- [4] Kasbawati & Nurwahyu, B., 2010. Model Matematika Penyebaran Virus Flu Burung H5N1 pada Populasi Burung dan Manusia. *Prosiding Sem-Nas Matematika UNPAR*, Vol.5 thn 2010, 103-111. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Katolik Parahyangan, Bandung.
- [5] Lewis, F. & Syrmos, V., 1995. *Optimal Control*, Second Edition. John Wiley & Sons Inc., New York.
- [6] Mulyadi, B. & Prihatini, 2005. Diagnosis Laboratorik Flu Burung (H5N1) (Laboratoric Diagnosis of Avian Influenzae (H5N1)). *Indonesian Journal of Clinical Pathology and Medical Laboratory*, Vol. 12, No. 2, 71-81.

- [7] Ramadhan, B., Savitri, D. P., Rahman, A. S., & Arifin, M. A. (2023). Pengembangan Sarana Transportasi Perkotaan Berkelanjutan Pada Kereta Cepat (Whoosh). In *Prosiding Seminar Rekayasa Teknologi (SemResTek)* (pp. 88-93).
- [8] Renee, F., K., Lenhart, S. & Scott, J., 1998. Optimizing Chemotherapy in an HIV Model. *Electronic Journal of Differential Equations*, Vol. 1998(1998), No. 32, 1-12.
- [9] Steven, R.T., 1989. *Fractal Programming in C*. M&T Publishing, California. <http://www.geocities.com/CapeCanaveral/4257/Paper.html>. [17 Maret 2018]

[]