

Autoregressive Integrated Moving Average (ARIMA) for Rainfall Forecasting in Surabaya City

Yonlib W. A. Nanlohy^{1*}, Gabriella Haumahu²

^{1,2}Universitas Pattimura, Jalan Ir. M. Putuhena, Kampus Poka, Kota Ambon, Indonesia.

Email: ywa.nanlohy@gmail.com^{1*}, gbhaumahu@gmail.com²

Manuscript submitted : April 2021;

Accepted for publication : May 2021.

Abstract: Surabaya is the largest city in the province of East Java and is also the center of the provincial government. In the city of Surabaya the dry season is from May to October and the rainy season is from November to April. Heavy rain usually occurs between December and January. One of the negative impacts caused by excessive rainfall in the city of Surabaya is flooding. The method is often used to predict rainfall in the city of Surabaya, it is Autoregressive Integrated Moving Average (ARIMA). ARIMA models is a forecasting model analysis data of single time series or univariate models. The purpose of this study is to forecast the daily rainfall in the city of Surabaya with a ARIMA model.

2010 Mathematical Subject Classification : 62M10, 62-07.

Keywords: ARIMA, rainfall, forecasting.

1. Introduction

Peramalan merupakan teknik untuk memperkirakan keadaan dimasa depan, peramalan juga merupakan bagian integral dari kegiatan pengambilan keputusan [1]. Peramalan dapat membantu manusia dalam merencanakan masa depan dan membuat keputusan yang rasional. Peramalan didasarkan pada data historis/deret waktu untuk memperoleh nilai dugaan dari suatu periode waktu tertentu.

Akhir-akhir ini peranan peramalan telah menjelajah ke berbagai bidang diantaranya bidang meteorologi yang menyangkut peramalan cuaca dan curah hujan. Badan Meteorologi Klimatologi dan Geofisika (BMKG) dalam meramalkan cuaca dan curah hujan telah menggunakan berbagai macam metode statistik diantaranya dengan menggunakan analisis regresi dan analisis deret waktu.

Surabaya merupakan Kota terbesar yang berada di provinsi Jawa Timur dan sekaligus menjadi menjadi pusat pemerintahan provinsi. Secara geografis Surabaya memiliki luas wilayah 326,81 km² dan rata-rata curah hujan berkisar 1,321 mm pertahun. Di Kota Surabaya musim kemarau pada umumnya berada pada bulan Mei sampai Oktober dan musim hujan dari bulan November hingga April. Hujan lebat biasanya terjadi antara bulan Desember dan Januari [2]. Salah satu dampak negatif yang ditimbulkan akibat intensitas curah hujan berlebih di Kota Surabaya adalah banjir. Hujan bukanlah faktor utama yang menyebabkan meluapnya air sungai, tetapi lebih disebabkan kondisi lingkungan yang kurang baik, banyaknya sampah yang menyumbat

aliran air sehingga aliran air hujan tidak bisa mengalir dengan lancar, dan juga disebabkan system drainase yang kurang baik. Namun jika hujan dapat diprediksi hal ini akan dapat membantu pihak pemerintah dan masyarakat untuk mempersiapkan diri sebagai langkah mengantisipasi bila hujan dengan intensitas tinggi tiba di Kota Surabaya.

Metode yang sering kali digunakan untuk meramalkan curah hujan yaitu *Autoregressive Integrated Moving Average* (ARIMA). Model *Autoregresif Integrated Moving Average* (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari deret waktu (*time series*) secara statistik berhubungan satu sama lain (dependen).

1.1 Autoregressive Integrated Moving Avarage (ARIMA)

Model ARIMA merupakan model kombinasi dari *autoregressive* (AR) berordo p dan proses *moving average* (MA) berordo q . Pembeda berordo d dilakukan jika data deret waktu tidak stasioner dalam rata-rata. Pemeriksaan kestasioneran data dapat dilihat dari *Time Series Plot*, *Autocorrelation Function* (ACF) *Plot* dan *Partial Autocorrelation Function* (PACF) *Plot*. Model ARIMA (p, d, q) dapat ditulis sebagai berikut [3] :

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

Model ARIMA yang mengakomodasi efek musiman dalam pengamatan waktu ke- t dinotasikan dengan ARIMA(P, D, Q)^s. Secara umum, model ARIMA musiman dapat ditulis dalam bentuk :

$$\Phi_P(B^S)(1 - B^S)^D Y_t = \Theta_Q(B^S) a_t \quad (2)$$

Jika terdapat efek non-musiman dan efek musiman, maka model yang terbentuk adalah model multiplikatif yaitu ARIMA (p, d, q)(P, D, Q)^s. Bentuk umum model ARIMA (p, d, q)(P, D, Q)^s dapat ditulis sebagai berikut :

$$\phi_p(B)\Phi_P(B^S)(1 - B)^d(1 - B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S) a_t \quad (3)$$

2. Metodologi

2.1 Sumber Data

Data yang digunakan adalah data curah hujan harian di Kota Surabaya. Data yang digunakan adalah data dari Januari tahun 2008 sampai dengan Desember 2015. Data dibagi menjadi data in-sample sebagai data training dan data out-sample sebagai data testing. Data in-sample digunakan data dari Januari 2008 sampai November 2015 dan data out-sample yang digunakan data Desember tahun 2015.

2.2 Variabel Penelitian

Berdasar pada latar belakang dan tujuan penelitian, variabel penelitian yang digunakan yakni curah hujan di Kota Surabaya.

2.3 Tahapan Penelitian

Penelitian ini menggunakan *software* SAS dan Minitab. Tahapan peramalan dengan metode ARIMA sebagai berikut:

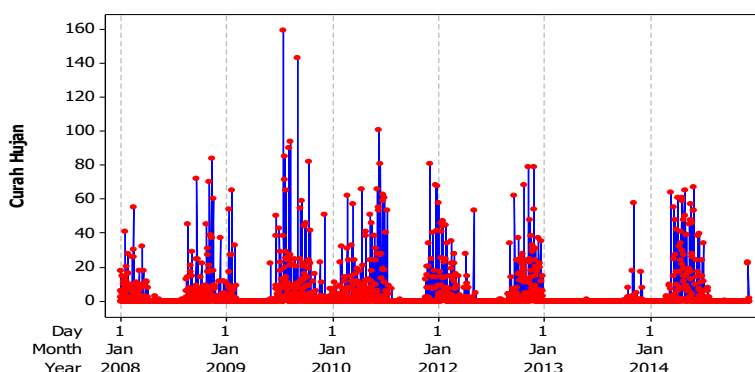
1. Melakukan plot data curah hujan harian di kota Surabaya
2. Jika data belum stasioner pada variansnya maka perlu dilakukan transformasi Box-Cox dan jika data belum stasioner pada rata-rata maka perlu dilakukan *differencing*
3. Jika data telah stasioner, dilihat pola ACF dan PACF nya untuk identifikasi orde model ARIMA.
4. Estimasi parameter dan pengujian signifikansi parameter berdasarkan model terpilih.

5. Melakukan uji dignosis untuk *white noise* dan normalitas residual melalui uji Ljung-Box dan Kolmogorov-Smirnov.
6. Melakukan peramalan curah hujan harian di kota Surabaya dengan model ARIMA
7. Menghitung nilai *Root Mean Square Error* (RMSE).

3. Hasil dan Diskusi

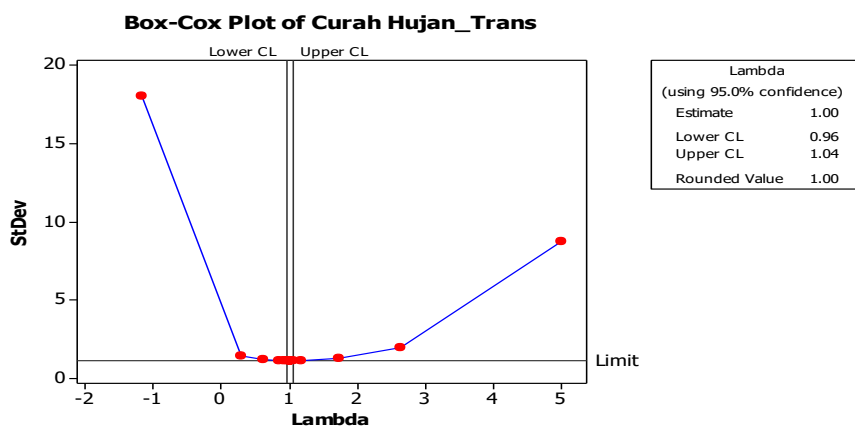
3.1 Peramalan Curah Hujan Harian Di Kota Surabaya

Data dibagi menjadi data *in-sample* dan *out-sample*. Data *in-sample* yang digunakan untuk pemodelan adalah data dari bulan Januari 2008 sampai November 2015. Sedangkan data *out-sample* adalah data bulan Desember 2015. Berikut *Time Series Plot* curah hujan di Kota Surabaya.



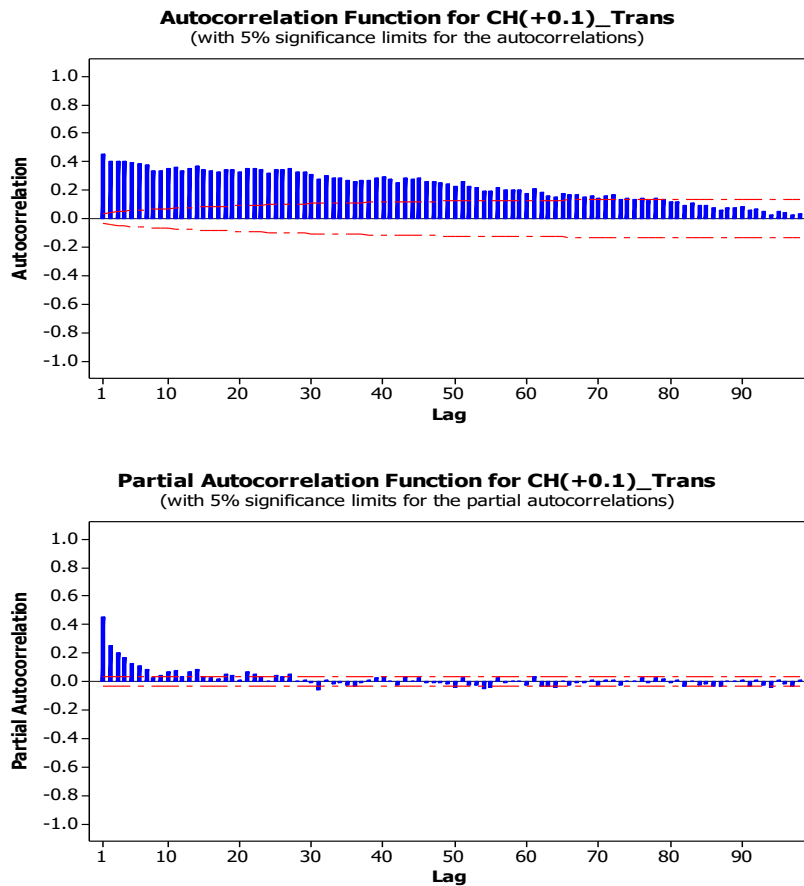
Gambar 1 *Time Series* Curah Hujan Harian di Kota Surabaya

Berdasarkan Gambar 1, curah hujan harian di Kota Surabaya cenderung tidak stasioner dan banyak *outlier*. Asumsi dasar yang harus dipenuhi dalam analisis *time series* adalah stasioneritas data baik dalam varians maupun dalam rata-rata. Pemeriksaan stasioneritas data dalam varians dapat dilakukan dengan menggunakan *Box-Cox plot*. Jika data belum stasioner dalam varians maka perlu dilakukan transformasi pada data. Gambar 2 berikut menampilkan *Box-Cox plot* dari data curah hujan harian di Kota Surabaya yang telah stasioner dalam varians.



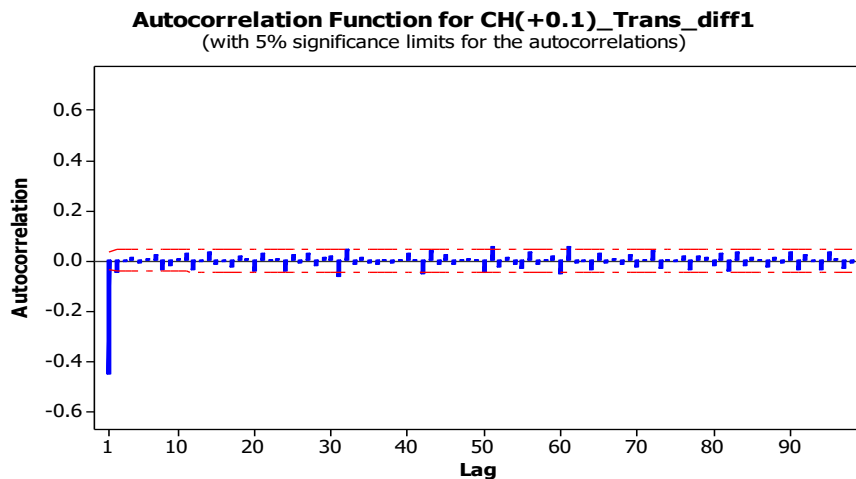
Gambar 2 *Box-Cox plot* Curah Hujan Harian di Kota Surabaya

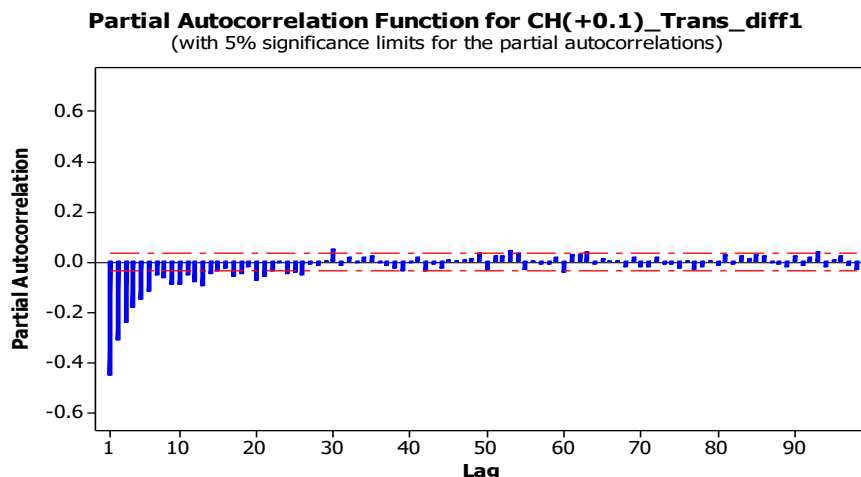
Gambar 2. menunjukkan nilai $\lambda = 1$ maka diasumsikan data stasioner dalam varians. Selanjutnya dilakukan pemeriksaan kestasioneran data dalam rata-rata. Pemeriksaan kestasioneran data dalam rata-rata dapat dilakukan dengan menggunakan *plot* ACF dan PACF.



Gambar 3 ACF dan PACF Plot Curah Hujan Harian di Kota Surabaya

Gambar 3. menunjukkan bahwa data belum stasioner dalam rata-rata karena nilai autokorelasinya cenderung turun secara lambat. Nilai autokorelasi pada suatu lag relatif tidak jauh berbeda dengan lag sebelumnya. Salah satu cara yang dapat digunakan untuk menstasionerkan data dalam rata-rata adalah dengan menggunakan proses differencing. Karena nilai autokorelasi cenderung turun secara lambat, maka dilakukan differencing satu terhadap data hasil transformasi.





Gambar 4. ACF dan PACF Plot Curah Hujan Harian di Kota Surabaya setelah differencing pada lag-1

Berdasarkan Gambar 4, diketahui data hasil transformasi sudah stasioner dalam rata-rata. Langkah selanjutnya adalah menentukan model ARIMA sementara berdasarkan lag-lag yang signifikan pada plot ACF dan PACF. Berdasarkan hasil identifikasi dari lag yang signifikan pada diagram ACF dan PACF, maka model ARIMA sementara curah hujan harian di kota Surabaya, yaitu ARIMA (0,1,1), ARIMA (1,1,[1,2,21]) dan ARIMA ([1,8],1,[1,21,31,43]). Uji signifikansi parameter model ARIMA untuk curah hujan harian di kota Surabaya disajikan dalam Tabel 1.

Tabel 1. Uji Signifikansi Parameter Model Curah Hujan Harian di Kota Surabaya

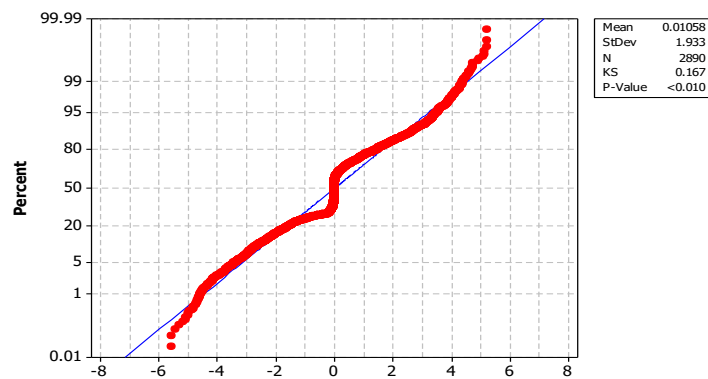
Model	Parameter	Estimasi	S.E	P-value	Keputusan
ARIMA (0,1,1)	θ_1	0,88980	0,00847	<,0001	Signifikan
	ϕ_1	0,34182	0,16810	0,0421	Signifikan
ARIMA (1,1,[1,2,21])	θ_1	1,15857	0,17262	<,0001	Signifikan
	θ_2	-0,21709	0,15678	0,1662	Tidak Signifikan
	θ_{21}	-0,01842	0,00675	0,0064	Signifikan
ARIMA ([1,8],1,[1,21,31])	ϕ_1	0,09026	0,02078	<,0001	Signifikan
	ϕ_8	-0,05587	0,01930	0,0038	Signifikan
	θ_1	0,90635	0,00902	<,0001	Signifikan
	θ_{21}	-0,02509	0,00827	0,0024	Signifikan
	θ_{31}	0,02097	0,00842	0,0128	Signifikan
	θ_{43}	-0,02326	0,00808	0,0040	Signifikan

Tabel 4.18 menunjukkan bahwa model yang memenuhi uji signifikansi parameter (nilai p-value <0,05) adalah model ARIMA (0,1,1) dan ARIMA ([1,8],1,[1,21,31,43]). Untuk menentukan kelayakan model ARIMA dilakukan cek diagnosa residual untuk menguji residual bersifat white noise atau tidak. Hasil dari pengujian white noise residual disajikan dalam Tabel 2.

Tabel 2 Uji Residual Model Curah Hujan Harian di Kota Surabaya

Model	Lag	Chi-Square	DF	P-value	Keputusan
ARIMA (0,1,1)	6	18,23	5	0,0027	Tidak <i>white noise</i>
	12	51,65	11	<,0001	
	18	58,52	17	<,0001	
	24	66,06	23	<,0001	
	30	85,55	29	<,0001	
	36	104,11	35	<,0001	
	42	113,19	41	<,0001	
	48	128,94	47	<,0001	
ARIMA ([1,8],1,[1,21,31,43])	6	.	0	.	<i>White noise</i>
	12	10,58	6	0,1022	
	18	18,54	12	0,1002	
	24	24,48	18	0,1399	
	30	29,80	24	0,1915	
	36	39,10	30	0,1236	
	42	48,93	36	0,0737	
	48	52,91	42	0,1205	

Berdasarkan tabel 2, uji residual untuk model ARIMA (0,1,1) tidak memenuhi asumsi *white noise* karena nilai *p-value* < 0,05 sedangkan uji residual untuk model ARIMA ([1,8],1,[1,21,31,43]) telah memenuhi sifat *white noise* karena nilai *p-value* > 0,05 untuk semua *lag*. Untuk uji normalitas residual dengan uji *Kolmogorov-Smirnov* disajikan pada gambar 5.



Gambar 5. Plot Kenormalan Residual Model ARIMA ([1,8],1,[1,21,31,43])

Gambar 5. menunjukkan nilai *p-value* < 0,010 sehingga residual model ARIMA ([1,8],1,[1,21,31,43]) tidak berdistribusi normal. Selanjutnya dilakukan deteksi *outlier* untuk mengetahui penyebab residual tidak berdistribusi normal. Hasil deteksi *Outlier* disajikan pada gambar 6 berikut.

The ARIMA Procedure
Outlier Detection Summary

Maximum number searched	50
Number found	50
Significance used	0.05

Outlier Details

Obs	Type	Estimate	Chi-Square	Approx Prob> ChiSq
1813	Shift	-2.26715	48.76	<.0001
1042	Shift	-1.94662	34.73	<.0001
2543	Shift	-1.88691	32.63	<.0001
2917	Shift	-2.72103	27.10	<.0001
158	Shift	1.71197	25.29	<.0001
781	Shift	-1.67273	27.54	<.0001
775	Shift	1.75525	32.97	<.0001
192	Additive	-5.24566	27.02	<.0001
1631	Shift	1.62804	26.60	<.0001
1581	Shift	1.70863	32.73	<.0001
115	Shift	1.62846	29.83	<.0001
1079	Additive	-5.09242	29.61	<.0001
1547	Shift	-1.59864	28.79	<.0001
1881	Additive	4.97710	28.17	<.0001
899	Shift	1.58035	28.35	<.0001
2732	Additive	-4.90909	27.40	<.0001
954	Additive	-4.88228	27.87	<.0001
2731	Additive	-4.86820	28.32	<.0001
1654	Additive	-4.82260	29.45	<.0001
563	Additive	-4.79974	30.42	<.0001
1500	Additive	4.77919	31.08	<.0001
452	Additive	4.75481	31.15	<.0001
427	Shift	-1.55487	32.86	<.0001
2725	Additive	-4.74145	31.05	<.0001
1498	Additive	4.72191	32.57	<.0001
1653	Additive	-4.68522	33.27	<.0001

Gambar 6. Deteksi *Outlier*

Berdasarkan gambar 6 terdapat *outlier* yang menyebabkan residual tidak berdistribusi normal. Oleh karena itu estimasi parameter model ARIMA $([1,8],1,[1,21,31,43])$ dilanjutkan dengan menambahkan deteksi *outlier*. Hasil estimasi parameter model dengan melibatkan deteksi *outlier* dapat dilihat pada Tabel 3.

Tabel 3 Uji Signifikansi Parameter Model ARIMA $([1,8],1,[1,21,31,43])$ dengan Deteksi *Outlier*

Model	Parameter	Estimasi	S.E	P-value	Keputusan
ARIMA $([1,8],1,[1,21,31])$ Dengan Deteksi <i>Outlier</i>	ϕ_1	0.06848	0.01983	0.0006	Signifikan
	ϕ_8	-0.05916	0.01940	0.0023	Signifikan
	θ_1	0.96016	0.00661	<.0001	Signifikan
	θ_{21}	-0.05755	0.00754	<.0001	Signifikan
	θ_{31}	0.05049	0.00804	<.0001	Signifikan
	θ_{43}	-0.02721	0.00706	0.0001	Signifikan
	ω_{LS495}	-3.5699	0.57912	<.0001	Signifikan

	ω_{AO499}	4.73927	1.7908	0.0082	Signifikan

Tabel 3. menunjukkan bahwa semua parameter signifikan karena nilai *p-value* > 0,05. Selanjutnya dilakukan pengecekan kelayakan model dengan menguji asumsi *white noise* dan normalitas dari residual model yang dihasilkan. Hasil uji asumsi seperti pada Tabel 4 berikut.

Tabel 4 Uji *White Noise* Residual Model ARIMA $([1,8],1,[1,21,31,43])$ dengan Deteksi *Outlier*

Model	Lag	Chi-Square	DF	P-value	Keputusan
ARIMA $([1,8],1,[1,21,31,43])$ Dengan Deteksi <i>Outlier</i>	6	.	0	.	<i>White noise</i>
	12	8.85	6	0.1820	
	18	16.73	12	0.1602	
	24	22.68	18	0.2031	
	30	25.91	24	0.3579	
	36	38.04	30	0.1487	
	42	49.84	36	0.0623	
	48	53.24	42	0.1146	

Berdasarkan Tabel 4, model ARIMA $[[1,8],1,[1,21,31,43]]$ dengan deteksi *outlier* sudah memenuhi asumsi *white noise* dengan nilai *p-value* $> 0,05$ pada semua lag. Model ARIMA $[[1,8],1,[1,21,31,43]]$ dengan deteksi *outlier* yang terbentuk adalah :

$$\begin{aligned}
 Y_t = & 1.06848Y_{t-1} - 0.06848Y_{t-2} - 0.05916Y_{t-8} + 0.05916Y_{t-9} + a_t + 0.96016a_{t-1} - 0.05755a_{t-21} + \\
 & 0.05049a_{t-31} - 0.02721a_{t-43} - \frac{3.56990}{1-B}I_t^{T=495} + \frac{3.43978}{1-B}I_t^{T=1500} + \frac{3.96046}{1-B}I_t^{T=521} + \frac{3.00759}{1-B}I_t^{T=1232} + \\
 & \frac{2.77369}{1-B}I_t^{T=438} + \frac{3.06386}{1-B}I_t^{T=1945} - \frac{3.09949}{1-B}I_t^{T=712} - \frac{2.38085}{1-B}I_t^{T=1403} - \frac{2.71608}{1-B}I_t^{T=300} - \frac{2.75711}{1-B}I_t^{T=1060} + \\
 & \frac{2.13577}{1-B}I_t^{T=1045} - \frac{3.20880}{1-B}I_t^{T=1881} + \frac{1.90248}{1-B}I_t^{T=771} - \frac{1.95784}{1-B}I_t^{T=2659} + \frac{2.69655}{1-B}I_t^{T=1906} + \frac{2.56116}{1-B}I_t^{T=1861} - \\
 & 5.83365I_t^{T=1247} - 5.53531I_t^{T=2394} - 5.64247I_t^{T=689} + \frac{2.29481}{1-B}I_t^{T=2639} + \frac{2.64769}{1-B}I_t^{T=2683} + \\
 & 5.90714I_t^{T=1887} - 5.34462I_t^{T=1633} - \frac{1.71085}{1-B}I_t^{T=1841} - 4.92942I_t^{T=2708} + 4.91497I_t^{T=1185} - \\
 & 4.88782I_t^{T=282} + 5.44553I_t^{T=502} - 4.96412I_t^{T=526} - 4.68312I_t^{T=1746} + 5.71383I_t^{T=1477} - \\
 & \frac{1.98261}{1-B}I_t^{T=1460} - \frac{1.40058}{1-B}I_t^{T=2543} - \frac{1.55258}{1-B}I_t^{T=2573} + \frac{1.23072}{1-B}I_t^{T=899} + 4.73927I_t^{T=499}
 \end{aligned}$$

Nilai RMSE yang diperoleh dari hasil peramalan data *out-sample* untuk jangka waktu 31 hari ke depan adalah 11,6056 yang berarti bahwa peramalan curah hujan di Kota Surabaya dengan model ARIMA hampir mendekati data aktualnya.

4. Kesimpulan

Berdasarkan hasil analisis diperoleh model ARIMA untuk peramalan curah hujan di Kota Surabaya dan nilai RMSE hasil peramalan data *out-sample* untuk jangka waktu 31 hari ke depan adalah 11,6052 sehingga dapat disimpulkan model ARIMA untuk peramalan curah hujan di kota Surabaya harus menjadi model peramalan curah hujan di Kota Surabaya.

References

- [1] Aswi dan Sukarna, (2006). *Analisis Deret Waktu*. Makassar, Andira Publisher.
- [2] Badan Pusat Statistik Kota Surabaya, (2014). Kota Surabaya dalam Angka. Surabaya.
- [3] Wei, W.W.S, (2006). *Time Series Analysis Univariate and Multivariate Methods*. New York: Addison-Wesley Company Inc.