

PEMODELAN ARIMA DAN DETEKSI *OUTLIER* DATA CURAH HUJAN SEBAGAI EVALUASI SISTEM RADIO GELOMBANG MILIMETER

Achmad Mauludiyanto¹ Gamantyo Hendrantoro¹ Mauridhi Hery P.¹ Suhartono²

¹Jurusan Teknik Elektro, FTI, Institut Teknologi Sepuluh Nopember

²Jurusan Statistik, FMIPA, Institut Teknologi Sepuluh Nopember

Email: maulud,gamantyo@ee.its.ac.id

ABSTRACT

The purpose of this paper is to provide the results of Arima modeling and outlier detection in the rainfall data in Surabaya. This paper explained about the steps in the formation of rainfall models, especially Box-Jenkins procedure for Arima modeling and outlier detection. Early stages of modeling stasioneritas Arima is the identification of data, both in mean and variance. Stasioneritas evaluation data in the variance can be done with Box-Cox transformation. Meanwhile, in the mean stasioneritas can be done with the plot data and forms of ACF. Identification of ACF and PACF of the stationary data is used to determine the order of allegations Arima model. The next stage is to estimate the parameters and diagnostic checks to see the suitability model. Process diagnostics check conducted to evaluate whether the residual model is eligible berdistribusi white noise and normal. Ljung-Box Test is a test that can be used to validate the white noise condition, while the Kolmogorov-Smirnov Test is an evaluation test for normal distribution. Residual normality test results showed that the residual model of Arima not white noise, and indicates the existence of outlier in the data. Thus, the next step taken is outlier detection to eliminate outlier effects and increase the accuracy of predictions of the model Arima. Arima modeling implementation and outlier detection is done by using MINITAB package and MATLAB. The research shows that the modeling Arima and outlier detection can reduce the prediction error as measured by the criteria Mean Square Error (MSE). Quantitatively, the decline in the value of MSE by incorporating outlier detection is 23.7%, with an average decline 6.5%.

Keywords: Arima model, outlier detection, Ljung-Box test, Kolmogorov-Smirnov test, MSE

ABSTRAK

Tujuan dari penulisan paper ini adalah untuk memberikan hasil pemodelan ARIMA dan deteksi outlier pada data curah hujan di Surabaya. Dalam paper ini dijelaskan tentang langkah-langkah dalam pembentukan model curah hujan, khususnya prosedur Box-Jenkins untuk pemodelan ARIMA dan deteksi outlier. Tahap awal dari pemodelan ARIMA adalah identifikasi stasioneritas data, baik dalam mean dan varians. Evaluasi stasioneritas data dalam varians dapat dilakukan dengan transformasi Box-Cox. Sedangkan stasioneritas dalam mean dapat dilakukan dengan plot data dan bentuk ACF. Identifikasi bentuk ACF dan PACF dari data yang sudah stasioner digunakan untuk menentukan orde model ARIMA dugaan. Tahapan selanjutnya adalah estimasi parameter dan cek diagnosa untuk melihat kesesuaian model. Proses cek diagnosa dilakukan untuk mengevaluasi apakah residual model sudah memenuhi syarat white noise dan berdistribusi normal. Uji Ljung-Box adalah uji yang dapat digunakan untuk memvalidasi syarat white noise, sedangkan Uji Kolmogorov-Smirnov merupakan uji untuk evaluasi distribusi normal. Hasil uji normalitas residual menunjukkan bahwa residual model ARIMA belum white noise, dan mengindikasikan adanya outlier pada data. Sehingga, langkah selanjutnya yang dilakukan adalah deteksi outlier untuk mengeliminasi efek outlier dan memperbesar ketepatan prediksi dari model ARIMA. Implementasi pemodelan ARIMA dan deteksi outlier dilakukan dengan menggunakan paket MINITAB dan MATLAB. Hasil penelitian menunjukkan bahwa pemodelan ARIMA dan deteksi outlier dapat mereduksi kesalahan prediksi yang diukur dengan kriteria Mean Square Error (MSE). Secara kuantitatif, penurunan nilai MSE dengan memasukkan deteksi outlier adalah 23,7% , dengan rata-rata penurunan 6,5%.

Keywords: model ARIMA, deteksi outlier, uji Ljung-Box, uji Kolmogorov-Smirnov, MSE

Curah hujan merupakan faktor utama yang mempengaruhi redaman hujan pada gelombang milimeter. Gelombang milimeter adalah gelombang radio yang bekerja di atas frekuensi 10 GHz. Karena redaman hujan sebanding dengan redaman spesifik dikalikan panjang link. Redaman spesifik sebanding dengan dengan curah hujan. Jika diasumsikan panjang link 1 Km maka besar redaman hujan akan bergantung pada curah hujan untuk frekuensi tertentu. Karena itu perlu ditinjau bagaimana model curah hujan yang terjadi setiap saat. Dengan mengetahui model curah hujan, maka dapat ditentukan model redaman hujan.

Pemodelan curah hujan sudah banyak dilakukan dalam beberapa penelitian, antara lain dengan model ARMA (*Autoregressive Moving Average*) [1, 2, 3] dan model ARIMA (*Autoregressive Integrated Moving Average*) [3, 4]. Model ARMA ini hanya berlaku khusus pada data stasioner dan tidak berlaku pada data non-stasioner. Sedangkan model ARIMA yang sudah diteliti belum memberikan hasil-hasil uji statistik untuk evaluasi kesesuaian model. Uji statistik yang dimaksud antara lain adalah deteksi stasioneritas data dalam varians melalui transformasi Box-Cox. Nilai $\lambda=1$ pada hasil transformasi Box-Cox menunjukkan data

stasioner dalam varians.

Uji statistik lain adalah uji signifikansi parameter model ARIMA, yaitu parameter model adalah signifikan jika *p-value* kurang dari 0,05. Beberapa uji pada tahap cek diagnosa juga belum banyak digunakan, yaitu uji Ljung-Box untuk evaluasi apakah residual white noise dan uji Kolmogorov-Smirnov untuk mengetahui normalitas data [5]. Padahal uji statistik ini penting sekali dalam menentukan validitas dari suatu model. Karena itu perlu dibuat model ARIMA yang dilengkapi dengan evaluasi kesesuaian model. Dalam pemodelan ini digunakan software MINITAB, dan MATLAB. Software MINITAB digunakan untuk pengolahan statistik dari pada data curah hujan. Sedangkan MATLAB digunakan untuk mengubah data asli ke data numerik yang dikenal oleh MINITAB.

Dalam paper ini akan diuraikan langkah-langkah pemodelan ARIMA pada data curah hujan. Tahapan penting dalam menentukan model ARIMA adalah identifikasi, estimasi parameter, dan cek diagnosa. Tahap Identifikasi berfungsi untuk menentukan orde model ARIMA melalui bentuk ACF (*Autocorrealation Function*) dan bentuk PACF (*Partial Autocorrelation Function*) dari data yang sudah stasioner. Estimasi parameter berfungsi untuk menduga nilai besaran konstanta dan koefisien dari model AR dan MA. Sedangkan cek diagnosa berfungsi untuk menguji kesesuaian model melalui uji residual apakah sudah memenuhi syarat white noise dan berdistribusi normal.

Tahapan selanjutnya yang diuraikan dalam paper ini adalah cara mendeteksi outlier dan indikasinya pada MSE. Tujuannya adalah agar didapatkan model curah hujan yang tepat dan handal. Ukuran ketepatan ini akan ditunjukkan dengan kriteria atau besaran statistik yang biasa digunakan dalam analisa statistik terhadap persoalan pengukuran di lapangan, seperti hasil ukur curah hujan.

METODOLOGI

Pada bagian ini akan diberikan tentang teori dan metodologi tentang model ARIMA dan deteksi outlier.

Model ARIMA

Secara umum model ARIMA diberikan oleh persamaan [6]

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

dengan operator stasioner AR adalah

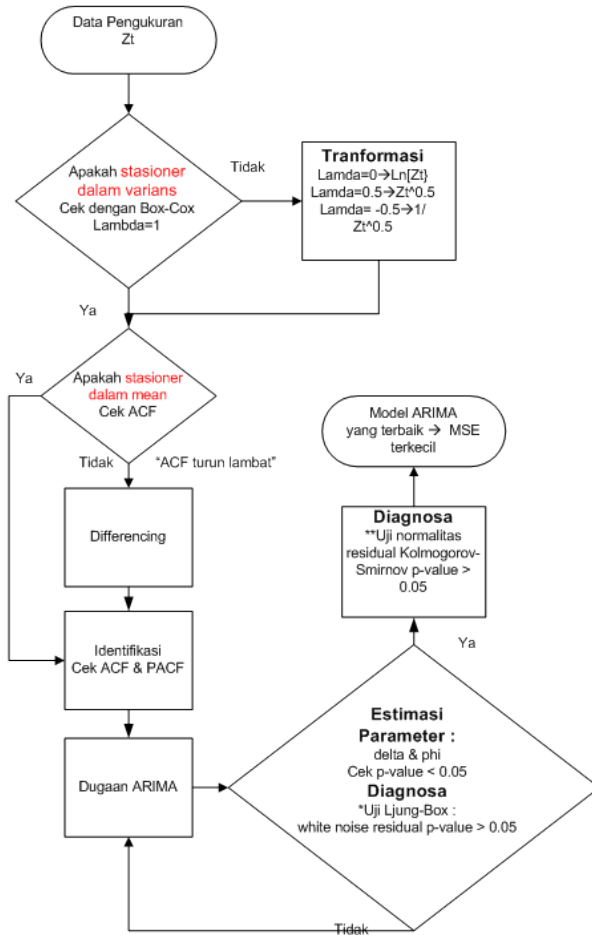
$$\phi_p(B) = (1 - \phi_1 B - \Lambda - \phi_p B^p)$$

dan operator MA adalah

$$\theta_q(B) = (1 - \theta_1 B - \Lambda - \theta_q B^q)$$

Model ARIMA dengan orde (p,d,q) dinyatakan dengan ARIMA(p,d,q), dan d menyatakan orde differencing.

Prosedur Box-Jenkins merupakan prosedur yang populer untuk pemodelan ARIMA. Ringkasan tahap-tahap dalam pemodelan ARIMA dengan prosedur Box-Jenkins dapat dilihat pada Gambar 1 berikut ini. Transformasi Box-



Gambar 1: Flowchart Pemodelan ARIMA

Cox adalah suatu proses yang digunakan untuk mengetahui stasioneritas data dalam varians, yaitu melalui pendugaan nilai lambda (λ) dari data asli $[Z_t]$. Jika $\lambda=1$ berarti data asli sudah stasioner dalam varians, sedangkan $\lambda \neq 0$ berarti data asli belum stasioner dalam varians. Karena itu perlu ditransformasi supaya menjadi stasioner dalam varians. Nilai $\lambda=0$ berarti data ditransformasi dengan $\ln[Z_t]$, $\lambda=0,5$ ditransformasi dengan $[Z_t]^{0,5}$, sedangkan $\lambda=-0,5$ data ditransformasi dengan $1/[Z_t]^{0,5}$. Jika λ tidak sama dengan yang disebutkan maka digunakan nilai pendekatan.

Stasioneritas data dalam mean bisa dilakukan dengan identifikasi plot data dan bentuk ACF data. Jika ACF menunjukkan pola yang turun lambat berarti data belum stasioner dalam mean. Sehingga dibutuhkan differencing agar datanya menjadi stasioner dalam mean. Sebaliknya jika ACF menunjukkan pola yang turun cepat maka data sudah stasioner dalam mean.

Identifikasi orde model ARMA bisa dilakukan dengan menggunakan bentuk ACF dan PACF data yang sudah stasioner seperti pada Tabel 1. Model-model dugaan yang diperoleh diestimasi nilai-nilai parameternya, dan kemudian diuji apakah *p-value* dari koefisien-koefisien tersebut kurang dari 0,05. Jika *p-value* dari konstanta dan koefisien kurang dari 0,05 maka konstanta atau koefisien tersebut adalah signifikan secara statistik dan valid untuk di-

Tabel 1: Identifikasi ACF dan PACF[6]

Model	ACF	PACF
MA(q) : moving average of order q	Cuts off after lag q	Dies down
AR(p) : autoregressive of order p	Dies down	Cuts off after lag p
ARMA(p,q) : mixed autoregressive-moving average of order (p,q)	Dies down	Dies down
AR(p) or MA(q)	Cuts off after lag q	Cuts off after lag p
No order AR or MA (white noise or random process)	No spike	No spike

gunakan. Jika sebaliknya maka konstanta atau koefisien tersebut dieliminasi dari model. Tahap selanjutnya adalah cek diagnosa. Pada tahap ini, residual model diuji apakah memenuhi syarat kesesuaian model ARIMA. Syarat sesuai tersebut adalah residual yang white noise dan berdistribusi normal. Evaluasi white noise residual dilakukan dengan uji Ljung-Box, yaitu residual white noise jika p-value lebih besar 0,05. Diagnosa berikutnya adalah diagnosa kenormalan residual dengan uji Kolmogorov-Smirnov. Jika p-valuenya lebih besar dari 0,05 maka residual berdistribusi normal. Model-model yang sesuai akan mempunyai nilai MSE yang berbeda. Model terbaik akan mempunyai MSE yang terkecil.

Seringkali ditemukan dari hasil penelitian model-model dugaan tadi belum memenuhi normalitas dari residunya, biasanya karena adanya outlier. Outlier ini dapat dikurangi dengan deteksi outlier, sehingga dapat menaikkan nilai MSE-nya.

Deteksi Outlier

Persamaan outlier-free series diberikan pada persamaan (2) [6]:

$$x_i(t) = \begin{cases} 1, & t = \text{terjadi outlier} \\ 0, & t : \text{yang lain} \end{cases} \quad (2)$$

Persamaan (2) termasuk pada model additive outlier [6]. Dengan adanya deteksi outlier maka persamaan (1) menjadi:

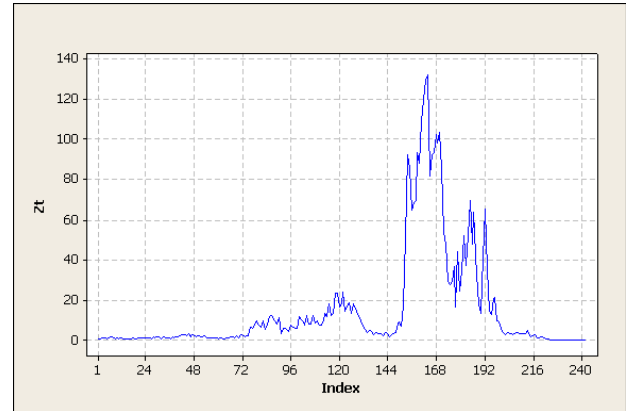
$$\phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t + \beta_i x_i(t) \quad (3)$$

dengan $i = 1, 2, \dots$ sampai mendapatkan residual yang diinginkan, yaitu sampai terdeteksi semua outlier yang ada atau residual telah memenuhi uji normalitas dengan uji Kolmogorov-Smirnov.

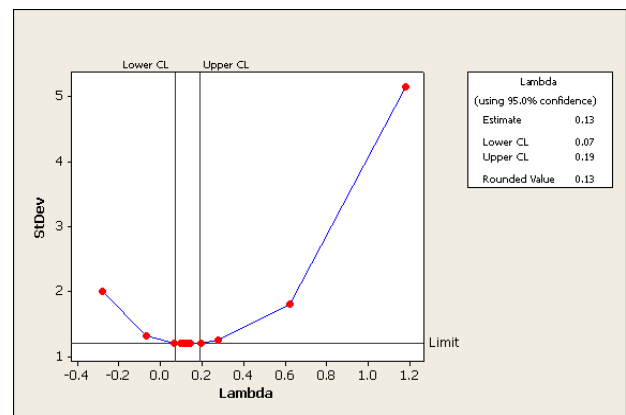
Deteksi outlier dilakukan dengan memplot residual dari model yang ditentukan. Titik-titik data yang mempunyai simpangan yang besar diambil dan bernilai 1 pada fungsi $x_i(t)$, untuk data yang lain bernilai 0 pada fungsi $x_i(t)$.

HASIL DAN ANALISIS

Pada bagian ini akan diberikan hasil pemodelan dan analisis data.



Gambar 2: Plot Data curah hujan 1 Maret 2007



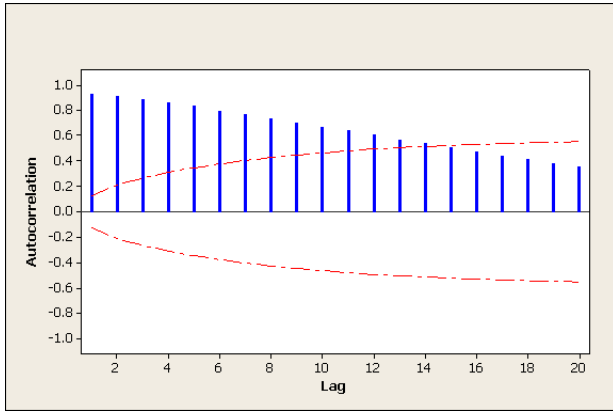
Gambar 3: Hasil transformasi Box-Cox

Langkah-langkah pemodelan

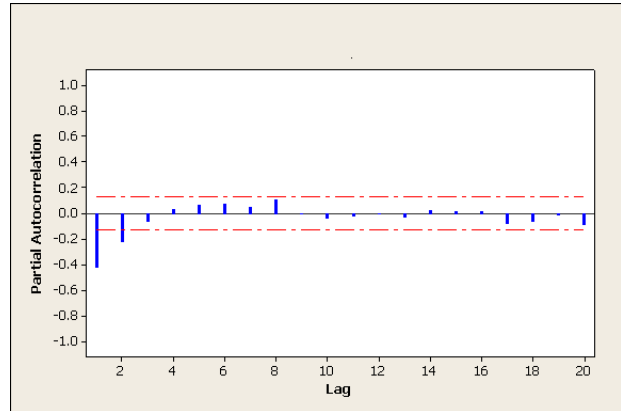
Sebagai studi kasus penelitian diambil satu contoh data, misalnya data curah hujan yang terjadi pada 1 Maret 2007. Berdasarkan Gambar 2 terlihat bahwa maksimum pengukuran sebesar 132 mm/h, dengan sampel data sebanyak 242 data.

Untuk mengetahui apakah data di atas stasioner dalam varians, digunakan uji Box-Cox seperti pada Gambar 3. Dari Gambar 3 terlihat nilai $\lambda = 0,13$ yang jika dibulatkan ke bawah menjadi sama dengan nol. Sehingga data harus ditransformasi dengan me-*logaritmanatural*-kan ($\ln Z_t$) supaya data asli memenuhi kondisi stasioneritas dalam varians.

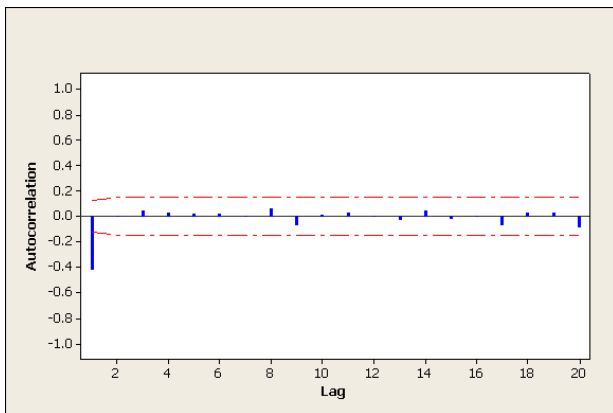
Langkah berikutnya adalah untuk mengetahui apakah data yang sudah ditransformasi sudah stasioner dalam mean, dengan melihat bentuk ACF-nya seperti pada Gambar 4. Dari Gambar 4 terlihat bentuk ACF turun secara pelan, yang berarti data Z_t^* tidak stasioner dalam mean. Sehingga data Z_t^* perlu didifferencing, dan gambar hasil plot ACF dan PACF setelah didifferencing dapat dilihat pada Gambar 5 dan 6. Menurut Tabel 1 dan memperhatikan Gambar 5 dan 6 dapat diduga bahwa model yang sesuai untuk data di atas adalah ARIMA(2,1,0) atau ARIMA(0,1,1). Jika digunakan ARIMA(2,1,0) maka diperoleh MSE = 0,3072, sedangkan ARIMA (0,1,1) diperoleh MSE = 0,3090. Jadi diputuskan untuk menggunakan model



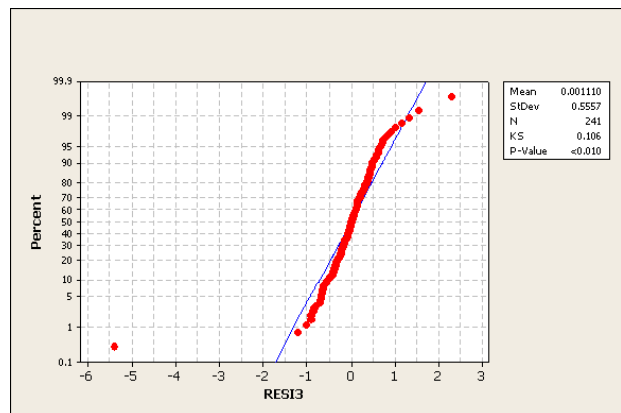
Gambar 4: Plot ACF data Zt^* (transformasi)



Gambar 6: Plot PACF data DZt^*



Gambar 5: Plot ACF data DZt^* (differencing)



Gambar 7: Uji Normalitas Kolmogorov-Smirnov

ARIMA(2,1,0). Hasil uji Kolmogorov-Smirnov pada residual menunjukkan residual belum berdistribusi normal (p-value lebih kecil dari 0,05). Hal ini diduga karena adanya outlier pada data [5]. Output hasil uji normalitas Kolmogorov-Smirnov seperti ditunjukkan pada Gambar 7. Model ARIMA(2,1,0) yang diperoleh dapat ditulis secara matematis dalam persamaan model sebagai berikut:

$$Y(t) = -0,5120Y(t-1) - 0,2165Y(t-2) \quad (4)$$

dengan MSE sebesar 0,3072 (melalui metode Maximum Likelihood Estimation).

Langkah-langkah Deteksi Outlier

Untuk mendeteksi adanya atau terjadinya outlier pada data, digunakan plot residual pada model ARIMA(2,1,0) seperti yang terlihat pada Gambar 8. Dari Gambar 8 terlihat outlier terjadi pada $t = 237$ dan 238 ; sehingga

$$x_1(t) = \begin{cases} 1, & t = 237,238 \\ 0, & t : \text{yang lain} \end{cases} \quad (5)$$

Dengan menggunakan metode least squares dan implementasi perintah regresi di MINITAB didapatkan

$Y(t) = -0,513Y(t-1) - 0,215Y(t-2)$, dengan MSE = 0.309. Dengan ditambahkan deteksi outlier pada model ARIMA akan menjadi

$Y(t) = -0,653Y(t-1) - 0,257Y(t-2) - 2,07X_1t$ dengan MSE = 0.2802. Langkah di atas dilakukan secara terus menerus sehingga didapatkan MSE kecil dan uji kenormalan residual terpenuhi. Berikut ini adalah hasil iteratif deteksi outlier yang ditunjukkan oleh perbaikan nilai MSE.

Kemudian ditambahkan outlier kedua, yaitu

$$x_2(t) = \begin{cases} 1, & t = 153,154,237,238 \\ 0, & t : \text{yang lain} \end{cases} \quad (6)$$

Dengan regresi didapat

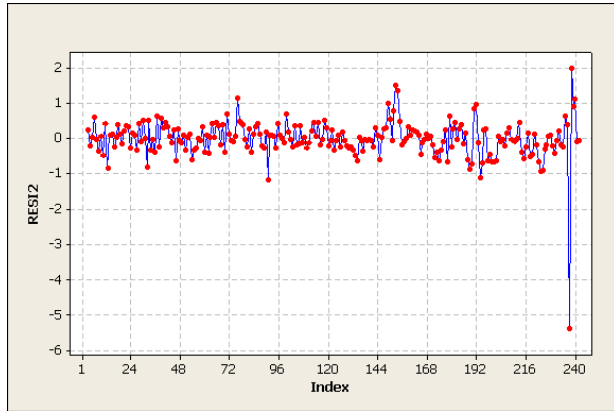
$$Y(t) = -0,712Y(t-1) - 0,292Y(t-2) - 3,88X_1t + 1,66X_2t \text{ dengan MSE}=0.2591.$$

Outlier ketiga ditambahkan, yaitu

$$x_3(t) = \begin{cases} 1, & t = 237,238,239,240 \\ 0, & t : \text{yang lain} \end{cases} \quad (7)$$

Dengan regresi didapat

$$Y(t) = -0,827Y(t-1) - 0,326Y(t-2) - 5,97X_1t + 1,78X_2t + 1,67X_3t \text{ dengan MSE} = 0.2399.$$



Gambar 8: Plot Residual model ARIMA

Outlier keempat ditambahkan, yaitu

$$x_4(t) = \begin{cases} 1, & t = 76, 91, 149, 189, 192, 194, \\ & 195, 223, 224, 237, 238 \\ 0, & t : \text{yang lain} \end{cases} \quad (8)$$

Dengan regresi didapat

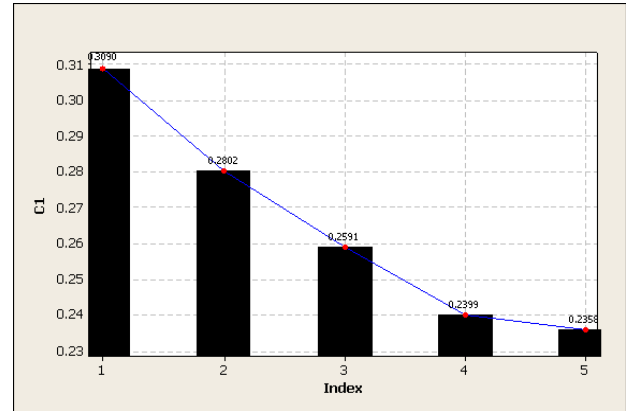
$$Y(t) = -0,842Y(t-1) - 0,335Y(t-2) - 5,69X_1t + 1,80X_2t + 1,70X_3t - 0,369X_4t \text{ dengan MSE} = 0.2358.$$

Kemudian untuk memperlihatkan perbaikan atau penurunan dari nilai MSE dilakukan plotting nilai MSE terhadap indeks ke- i ($i = 0,1,2,\dots,5$), seperti pada Gambar 9. Dari gambar tersebut terlihat adanya perbaikan dalam hal nilai MSE sebagai hasil adanya deteksi outlier yang ditambahkan pada model ARIMA(2,1,0). Perbaikan nilai MSE yang ditunjukkan dari indeks 1 (sebelum ada deteksi outlier) ke indeks 2 (deteksi outlier 1) adalah 9,3%. Perbaikan MSE dari indeks 2 ke indeks 3 (deteksi outlier 2) sebesar 7,5%. Perbaikan dari indeks 3 ke indeks 4 (deteksi outlier 3) sebesar 7,4%. Perbaikan dari indeks 4 ke indeks 5 (deteksi outlier 4) sebesar 1,7%.

Secara keseluruhan dari perbaikan dari indeks 1 ke indeks 5 sebesar 23,7%. Akhirnya diperoleh nilai MSE terkecil dan uji residual memenuhi syarat distribusi normal. Jika residual sudah memenuhi syarat distribusi normal, maka model ARIMA(2,1,0) dengan koefisien yang diperoleh dari regresi merupakan model yang valid.

Pemodelan ARIMA yang benar adalah pemodelan dengan tahapan yang melibatkan adanya indikator uji statistik seperti yang diuraikan pada paper ini. Karena keterbatasan kemampuan dari MINITAB, maka pemodelan ARIMA dan deteksi outlier secara simultan tidak dapat dilakukan untuk semua model, khususnya pada model yang mengandung orde MA. Setelah mengetahui titik-titik outlier maka fungsi dari pada outlier ditambahkan dengan model ARIMA yang diperoleh dan kemudian dievaluasi lagi residual model. Demikian seterusnya sehingga mencapai residual yang berdistribusi normal.

Penggunaan deteksi outlier dengan MINITAB hanya dapat digunakan untuk model-model AR. Sehingga untuk model MA harus menggunakan paket program yang lain,



Gambar 9: Nilai Penurunan MSE

seperti SAS. Dengan demikian, proses deteksi outlier dapat digunakan untuk memperbaiki model ARIMA yang diperoleh agar mendapatkan model yang valid untuk prediksi. Model yang valid akan memberikan nilai prediksi yang lebih efisien dalam merekonstruksi perilaku curah hujan atau perilaku redaman hujan. Sehingga model *anti fading* yang akan dibangun dapat bekerja dengan optimal.

SIMPULAN

Berdasar hasil penelitian ini dapat diambil beberapa kesimpulan antara lain pemodelan ARIMA dapat dilakukan dengan tahapan seperti pada *flowchart* di Gambar 1 menggunakan software MINITAB dan MATLAB. Prosedur ini seringkali belum mampu memberikan residual yang memenuhi uji normalitas Kolmogorov-Smirnov, karena adanya outlier. MINITAB dapat digunakan untuk deteksi outlier khusus pada model ARIMA(p,d,q) dengan $q = 0$, dan tidak berlaku untuk ARIMA (p,d,q) dengan $p = 0$. Deteksi outlier dapat memperbaiki MSE dari model yang ditentukan. Perbaikan MSE pada model memberikan residual yang cenderung memenuhi syarat distribusi normal.

DAFTAR PUSTAKA

- [1] Yadnya, M., Mauludiyanto, A., Hendratoro, G.: *ARMA Modelling from rain rate Measurement to Simulation Communication Channel Modelling for Millimeter Wave in Surabaya*. In: The 6th Kumamoto University Forum, Surabaya, Indonesia. (November 2008)
- [2] Yadnya, M., Mauludiyanto, A., Hendratoro, G.: *Pemodelan ARMA untuk Curah Hujan di Surabaya*. In: SITIA, Jurusan Teknik Elektro, ITS. (Mei 2008)
- [3] Mauludiyanto, A., Hendratoro, G., Yadnya, M., D.Kalfarosi: *Pemodelan Spektral, ARMA Dan ARIMA Untuk Curah Dan Redaman Hujan Tropis Menggunakan Data Surabaya Sebagai Evaluasi Sistem Radio Gelombang Milimeter*. submit ke Makara Seri Sains, Direktorat Riset dan Pengabdian kepada Masyarakat Universitas Indonesia (Oktober 2008)

- [4] Mauludiyanto, A., Hendranto, G., D.Kalfarosi: *Pe-modelan Curah dan Redaman Hujan dengan Model ARIMA Sebagai Evaluasi Sistem Radio Gelombang Milimeter*. submit ke Jurnal Penelitian dan Pengembangan Telekomunikasi, PPLMI, Institut Teknologi Telkom (Oktober 2008)
- [5] Iriawan, N., Astuti, S.: *Mengolah Data Statistik dengan Mudah Menggunakan Minitab 14*. Andi Yogyakarta (2006)
- [6] Wei, W.: *Time Series Analysis*. Addison-Wesley Publishing Company, USA (2006)