

PEMODELAN VARIMA DENGAN EFEK DETEKSI OUTLIER TERHADAP DATA CURAH HUJAN

Achmad Mauludiyanto⁽¹⁾, Gamantyo Hendratoro⁽¹⁾, Mauridhi Hery P.⁽¹⁾, Suhartono⁽²⁾

⁽¹⁾Jurusan Teknik Elektro, FTI; ⁽²⁾Jurusan Statistik, FMIPA

Institut Teknologi Sepuluh Nopember Surabaya

E-mail: {maulud,gamantyo}@ee.its.ac.id

ABSTRAK

Makalah ini menyampaikan hasil pemodelan VARIMA (Vektor Autoregressive Integrated Moving Average) dengan efek deteksi outlier terhadap data curah hujan di Surabaya. Pemodelan ini menggunakan program SAS dan minitab. Tahap awal dari pemodelan ARIMA adalah identifikasi stasioneritas data, baik dalam mean dan varians. Identifikasi bentuk ACF dan PACF dari data yang sudah stasioner digunakan untuk menentukan orde model VARIMA dugaan. Tahapan selanjutnya adalah estimasi parameter dengan MLE (Maximum Likelihood Estimation) dan cek Approx. Pr.>/t /harus lebih kecil dari 0,05. Jika hasil Approx. Pr.>/t /lebih besar dari 0,05 maka orde dari model yang diduga tidak signifikan. Sehingga, langkah selanjutnya yang dilakukan adalah mengeliminasi orde tersebut dari model ARIMA dugaan. Pada bagian akhir dari hasil ditunjukkan rangkuman dari deteksi outlier dan jenisnya. Hasil penelitian menunjukkan Dengan program SAS dapat ditentukan secara otomatis jumlah outlier yang terjadi, sehingga menghindari terjadinya 'spurious outlier'. Untuk jarak antar raingauge > 1 Km maka data curah hujan diantara 2 raingauge tidak saling berpengaruh, sebaliknya jika jaraknya < 1 Km maka hasil perekaman data curah hujan saling mempengaruhi. Hal ini bisa terjadi karena sel hujan yang terjadi mempunyai radius sekitar 1 Km. Artinya diameter sel hujan yang terjadi < 1 Km, sehingga jika jarak pengukuran > 1 Km maka alat ukur (raingauge) yang kedua berada pada sel hujan yang berbeda dengan alat ukur (raingauge) yang pertama.

Kata kunci: model VARIMA, SAS, deteksi outlier, MSE.

1. PENDAHULUAN

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 sekitar 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 AR(Autogressive) [Hendratoro, 2004], model ARMA (Autoregressive Moving Average) [Yadnya, 2008; Yadnya, 2008] dan model ARIMA (Autoregressive Integrated Moving Average) [Mauludiyanto, 2008,2009]. Keluarga model ARMA hanya berlaku khusus pada data stasioner dan tidak berlaku pada data non-stasioner. Sedangkan model ARIMA adalah model yang dapat digunakan untuk data yang tidak stasioner dalam mean.

Pada model AR pada [Hendratoro, 2004] sudah cukup baik dibandingkan dengan data di Barcelona seperti ditunjukkan dengan kurva CCDF (Compliment Cummulative Distribution Function) dari ditribusi lognormal, dan kurva dari fungsi autokorelasi ternormalisasi. Namun model AR ini hanya baik untuk rainrate kurang dari 20 mm/h, seperti ditunjukkan oleh kurva CCDF durasi curah

hujan melampaui threshold kurang dari 40 mm/h. Sedangkan untuk threshold melebihi 40 mm/h model AR sangat jauh dari data ukur di Barcelona.

Model ARMA [Yadnya, 2008; Yadnya, 2008] hanya dapat digunakan untuk data deret waktu yang stasioner, sehingga tidak dapat memodelkan data curah hujan yang tidak stasioner. Karena itu maka diperlukan suatu langkah sehingga data tak stasioner menjadi stasioner yang terdapat pada model ARIMA yaitu proses differencing data.

Model ARIMA sudah dapat memodelkan data tak stasioner, namun model ini belum menunjukkan uji statistik yang diperlukan untuk melihat suatu model sudah signifikan. Oleh karenanya pada makalah ini akan ditunjukkan langkah-langkah yang diperlukan untuk model ARIMA dari data curah hujan 1 even, sekaligus akan ditunjukkan uji statistik yang diperlukan.

Pada makalah lain [Mauludiyanto, 2008,2009] sudah dijelaskan model ARIMA curah hujan dan optimasinya. Secara jelas disampaikan langkah-langkah pemodelan dan optimasi ARIMA sehingga didapatkan MSE yang paling kecil. Model ARIMA ini dapat memodelkan data curah hujan stasioner maupun yang tidak stasioner. Namun model ARIMA ini hanya berlaku untuk 1 lokasi pengukuran curah hujan, tidak berlaku untuk pengukuran lebih dari 1 lokasi pengukuran. Karena itu pada makalah ini akan diuraikan pemodelan dengan menggunakan data curah hujan untuk 3 lokasi yang terpisah. Data intensitas curah hujan ini diukur dengan menggunakan raingauge seperti gambar 1.

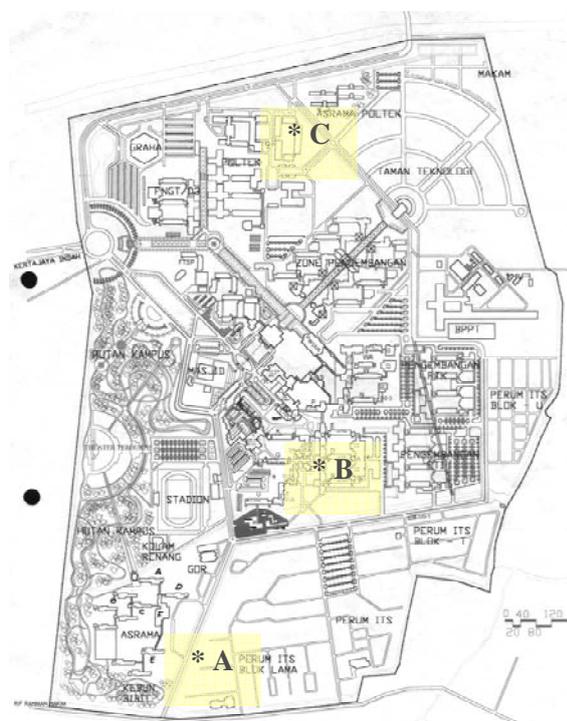
Pemodelan akan dilakukan dengan bantuan software SAS[Brocklebank, 2003]. Tahapan penting dari pemodelan ini adalah penentuan orde AR(p) dan MA(q) dengan tahapan sebagai berikut: tahap identifikasi, tahap estimasi parameter, dan tahap cek diagnosa. Tahap identifikasi ditentukan oleh MACF (Multivariate Aurocorrelation Function), dan MPACF (Multivariate Partial Aurocorrelation Function). Kemudian ditentukan juga AIC Multivariate (Minimum Information Criteria). Tahap estimasi parameter ditentukan oleh besarnya ϕ_{ij} untuk AR dan θ_{ij} untuk MA. Tahap berikutnya cek diagnosa ditentukan oleh multivariate white noise dan multivariate normal.

2. METODOLOGI

Sumber data yang digunakan untuk pemodelan mengacu pada makalah [Mahmudah, 2008]. Pengukuran intensitas curah hujan diperoleh dari pengukuran di Surabaya dengan menggunakan rain gauge di 3 lokasi berbeda seperti pada gambar 1. Program simulasi yang digunakan dalam pemodelan ini adalah program SAS.

Keterangan gambar:

A : Medical Center ITS	A-B = 700 m
B : Dept. of EE	B-C = 950 m
C : EEPIS	A-C = 1550 m



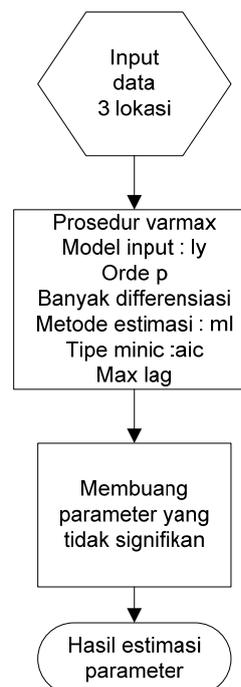
Gambar 1. Peta Pengukuran Curah Hujan dengan Rain gauge

Sebagai langkah awal proses VARIMA adalah dengan pembacaan data data curah hujan untuk 3 lokasi dan mengubah data awal menjadi data yang di lognaturalkan. Misal y_1 , y_2 dan y_3 merupakan parameter curah hujan berturut-turut di lokasi A, B,

dan C. Lokasi A adalah rain gauge yang ditempatkan di Medical Center ITS. Lokasi B adalah rain gauge yang ditempatkan di Jurusan Teknik Elektro ITS. Lokasi C adalah rain gauge yang ditempatkan di PENS ITS.

Prosedur Varima dapat dijelaskan sebagai berikut. Setelah data dibaca dan dikonversi, langkah selanjutnya adalah tahap identifikasi. Pada tahap ini ditentukan data spasial ada berapa, model yang dibentuk dari $\ln y$, orde dicurigai (orde-p) untuk VAR. Metode estimasi MLE (Maximum Likelihood Estimation) sebagai estimasi yang dipilih. Ditentukan juga Minimum Information Criteria (MINIC) misalnya AIC (Akaike information criteria). Estimasi parameter yang akan ditentukan, dengan menentukan besaran p untuk AR dan besaran q untuk MA. Pada bagian ini ditentukan juga maksimum lag yang diharapkan. Prosedur ini dapat digambarkan seperti pada gambar 2.

Dari estimasi parameter didapatkan beberapa parameter yang tidak signifikan (outlier), karena itu diperlukan langkah berikutnya dengan membuat nol nilai parameter outlier yang dimaksud.



Gambar 2. Flowchart prosedur VARIMA

3. HASIL DAN ANALISA

Setelah memasukkan data input curah hujan di 3 lokasi pengukuran yang berbeda, maka hasil Minimum Information Criteria menunjukkan bahwa orde model yang diduga adalah VAR(vektor autoregressive) orde 7.

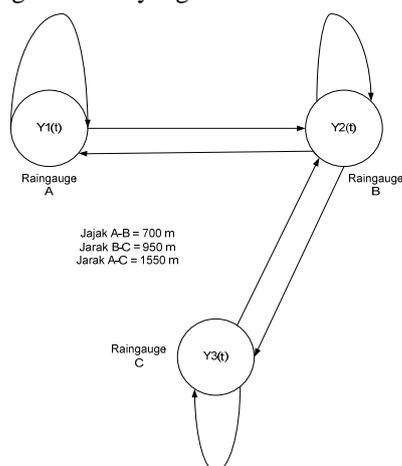
Model yang diperoleh dapat dituliskan dengan mengacu pada hasil estimasi koefisien AR pada persamaan (1), (2), dan (3). Dari persamaan (1), (2), dan (3) maka dapat digambarkan hubungan antara rain gauge A, B, dan C seperti gambar 3.

Tabel 1. Matriks Kovarians

variabel	ly1	ly2	ly3
ly1	0.56455	0.03456	0.02751
ly2	0.03456	0.33919	0.03434
ly3	0.02751	0.03434	0.50968

Dari gambar 3 terlihat bahwa data logaritma natural curah hujan di rain gauge A saling mempengaruhi data logaritma natural curah hujan di rain gauge B. Jarak rain gauge antara A dan B adalah 700 meter. Data logaritma natural curah hujan di rain gauge B saling mempengaruhi data logaritma natural curah hujan di rain gauge C. Jarak rain gauge antara B dan C adalah 950 meter. Data logaritma natural curah hujan di rain gauge A tidak saling mempengaruhi data logaritma natural curah hujan di rain gauge C. Jarak antara rain gauge A dan C adalah 1550 meter. Jadi dapat disimpulkan bahwa untuk jarak rain gauge > 1 Km maka data curah hujan diantara 2 rain gauge tidak saling berpengaruh, sebaliknya jika jaraknya < 1 Km maka hasil perekaman data curah hujan saling mempengaruhi. Hal ini bisa terjadi karena sel hujan yang terjadi mempunyai radius sekitar 1 Km. Artinya diameter sel hujan yang terjadi < 1 Km, sehingga jika jarak pengukuran > 1 Km maka alat ukur (rain gauge) yang kedua berada pada sel hujan yang berbeda dengan alat ukur (rain gauge) yang pertama.

Sebagai validasi benar tidaknya model yang dibentuk dapat dilakukan dengan membangkitkan data log natural curah hujan. Data pembangkitan ini kemudian ditentukan orde dari model VARIMA-nya apakah sama dengan orde hasil dari pengukuran? Jika hasilnya sama maka dapat dikatakan model yang dibentuk sudah sesuai dengan dengan data pengukuran dan model ini valid jika dipakai untuk pembangkitan data yang dimaksud.



Gambar 3 Bentuk Keterkaitan antar rain gauge A, B, dan C

Beberapa langkah pembangkitan yang perlu disampaikan dalam makalah ini adalah sebagai berikut (untuk 3 lokasi A, B, dan C):

1. Dengan program simulasi ditentukan mean untuk ketiga lokasi A, B, dan C
2. Kemudian ditentukan matriks covariance-nya (diambil dari hasil pemodelan VARIMA data ukur), misalkan gunakan *MLE* untuk metode estimasinya
3. Bangkitkan error di lokasi A, B, dan C secara bersama-sama (multivariate) dengan program simulasi
4. Hasil pembangkitan error kemudian digunakan untuk pembangkitan data awal log natural curah hujan di lokasi A, B, dan C
5. Cek validasi data dengan melihat hasil minimum information criteria, jika ordenya sama dengan hasil orde model data ukur maka data yang dibangkitkan adalah valid

4. KESIMPULAN

Berdasar hasil analisis dan pembahasan pada bagian sebelumnya, ada beberapa kesimpulan yang dapat diperoleh dari penelitian ini, antara lain :

Diperoleh model dari log natural curah hujan untuk 3 lokasi yang terpisah, dimana lokasi A saling mempengaruhi dengan lokasi B, lokasi B saling mempengaruhi dengan lokasi C, lokasi A tidak saling mempengaruhi dengan lokasi C.

Diperoleh data hasil pembangkitan dari model, setelah divalidasi dengan melihat orde VARIMA sama dengan orde model data ukur. Pemodelan VARIMA ini dapat digunakan baik data stasioner maupun data tidak stasioner.

PUSTAKA

- Brocklebank, J.C., Dickey, D.A., (2003), "SAS for Forecasting Time Series", SAS Institute and Wiley,.
- Hendrantoro, G., Mauludiyanto, A., Handayani, P., (2003), "An Autoregressive Model For Simulation Of Time-Varying Rain Rate", ANTEM 2004/URSI, July 20-23, Ottawa, Canada.
- Mahmudah, H., Wijayanti, A., Mauludiyanto, A., Hendrantoro, G., Matsushima, A., (2008), "Analysis of Tropical Attenuation Statistics using Synthetic Storm for Millimeter-Wave Wireless Network Design", WOCN 2008, Surabaya, Mei.
- Mauludiyanto, A., Hendrantoro, G., Purnomo, M.H., Suhartono, (2008), "Pemodelan ARIMA dan Deteksi Outlier Data Curah Hujan Sebagai Evaluasi Sistem Radio Gelombang Milimeter", submit ke JUTI (Jurnal Teknologi Informasi), FTI ITS Surabaya, Desember.
- Mauludiyanto, A., Hendrantoro, G., Purnomo, M.H., Suhartono, (2009), "Optimasi Pemodelan ARIMA dengan Efek Deteksi Outlier terhadap Data Curah Hujan untuk Evaluasi Sistem Radio Gelombang Milimeter", Prosiding Seminar Radar Nasional III, Bandung, 30 April.
- Yadnya, M.S., Mauludiyanto, A., Hendrantoro, G., (2008) "Pemodelan ARMA untuk Curah Hujan di

Surabaya”, Prosiding SITIA, Mei, Jurusan Teknik
Elektro, ITS.
Yadnya, M.S., Mauludiyanto, A., Hendratoro, G.,
(2008), “ARMA Modelling from rain rate

Measurement to Simulation Communication
Channel Modelling for Millimeter Wave in
Surabaya”, The 6th Kumamoto University Forum,
November 5-6, , Surabaya, Indonesia.

$$ly_1^*(t) = -0,35852ly_1^*(t-1) + 0,17610ly_2^*(t-1) - 0,19331ly_1^*(t-2) \\ + 0,20637ly_2^*(t-3) + a_1(t) \quad (1)$$

$$ly_2^*(t) = -0,49574ly_2^*(t-1) - 0,16237ly_3^*(t-1) - 0,20883ly_2^*(t-2) - 0,17046ly_3^*(t-2) \\ - 0,10077ly_1^*(t-4) + 0,43686ly_3^*(t-4) - 0,14320ly_2^*(t-5) + 0,46472ly_3^*(t-5) + \\ 0,39134ly_3^*(t-6) + a_2(t) \quad (2)$$

$$ly_3^*(t) = ly_3^*(t) = 0,21434ly_2^*(t-1) - 0,39287ly_3^*(t-1) - 0,16470ly_2^*(t-7) + a_3(t) \quad (3)$$