

Peramalan Kandungan *Particulate Matter* (PM10) dalam Udara Ambien Kota Surabaya Menggunakan *Double Seasonal* ARIMA (DSARIMA)

Bernadeta Chrisdayanti dan Agus Suharsono
Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: agus_s@statistika.its.ac.id

Abstrak— Surabaya merupakan salah satu kota besar di Indonesia yang memiliki jumlah penduduk yang padat dan diiringi oleh banyaknya jumlah kendaraan bermotor. Alat transportasi pribadi sangat mempermudah akomodasi, namun dapat menimbulkan polusi udara. Polusi udara dipantau berdasarkan 5 parameter, salah satunya PM10. PM10 merupakan parameter untuk menyatakan banyaknya kandungan debu dalam udara dan kandungannya paling banyak dalam udara ambien Kota Surabaya pada tahun 2014. Salah satu faktor yang mempengaruhi kandungan PM10 adalah asap kendaraan bermotor. Dampak dari tingginya kandungan PM10 akan langsung dirasakan oleh makhluk hidup. Oleh karena itu, kandungan PM10 perlu dikontrol setiap hari, bahkan setiap jam. Pengontrolan per jam dilakukan karena kandungan PM10 akan sangat tinggi pada jam 10.00 hingga 13.00, dimana nilai ISPUnya dapat melebihi batas udara dinyatakan sehat. Proses pengontrolan untuk beberapa jam dan hari berikutnya dapat dilakukan dengan menggunakan metode time series ARIMA. Metode ini dipilih karena kandungan PM10 dipengaruhi oleh waktu-waktu sebelumnya dan skala waktu yang digunakan konstan yaitu per jam. Stasiun pengamatan yang akan digunakan adalah SUF 6 dan SUF 7. Pemilihan model terbaik didasarkan pada kriteria *out sample Root Mean Square Error* (RMSE). Model terbaik yang dapat digunakan untuk mengontrol kandungan PM10 per jam di SUF 6 adalah ARIMA([1,2,4,7,8,9,10,12,16,17,18],1,0)(1,1,1)²⁴(0,1,1)¹⁶⁸ dan ARIMA ([1,2,4,5,7,8,10,11,12,13,15,16,18,19,20],1,0)(1,1,1)²⁴(0,1,1)¹⁶⁸.

Kata Kunci— ARIMA, Peramalan, PM10, RMSE, Udara Ambien.

I. PENDAHULUAN

Indonesia merupakan negara kepulauan yang terdiri atas lima pulau besar dan ribuan pulau kecil. Salah satu pulau besarnya adalah Pulau Jawa, pulau terpadat di Indonesia. Jawa Timur sebagai salah satu provinsi yang terletak di Pulau Jawa mengakibatkan provinsi ini memiliki jumlah penduduk yang banyak. Surabaya sebagai ibukota provinsi dijadikan sebagai kota perantauan oleh sebagian penduduk Jawa Timur.

Hal tersebut disebabkan oleh banyaknya perguruan tinggi dan lowongan pekerjaan yang tersedia di Surabaya. Peningkatan jumlah penduduk yang berada di Surabaya diiringi dengan peningkatan jumlah alat transportasi pribadi, seperti sepeda motor dan mobil. Kendaraan pribadi mempunyai keuntungan tersendiri bagi setiap penduduk karena mobilitas mereka akan lancar setiap harinya [1]. Namun, hal tersebut turut memberikan dampak pada siklus ekologis berupa polusi serta kemacetan lalu lintas. Salah satu polusi pada lingkungan hidup adalah polusi udara. Sebagai salah satu negara berkembang, maka perubahan dalam berbagai sektor, salah satunya transportasi, akan mempengaruhi kualitas udara yang ada [2]. Berdasarkan hal tersebut, maka diperlukan pemantauan dari pemerintah mengenai kualitas udara di Surabaya. Dinas yang bertugas memantau adalah Badan Lingkungan Hidup (BLH) Kota Surabaya [3].

Kualitas udara dipantau berdasarkan 5 indikator parameter, salah satunya yaitu *Particulate Matter* (PM10). Berdasarkan laporan tahunan BLH Surabaya tahun 2014, kandungan PM10 pada bulan merupakan yang paling banyak dibandingkan dengan parameter yang lain [4]. Selain itu, kandungan PM₁₀ dalam udara ambien merupakan salah satu hal yang penting karena dampaknya dapat dirasakan langsung oleh masyarakat [5], seperti gangguan pernafasan pada manusia dan terhambatnya proses fotosintesis pada tumbuhan jika kandungan PM10 melebihi batas sehat secara terus-menerus. Untuk itu diperlukan pemantauan secara berkala terhadap kandungan PM10. Proses pemantauan dilakukan dengan oleh Dinas setempat melalui 6 stasiun pengamatan yang tersebar di wilayah Surabaya. Namun pada tahun 2014, hanya terdapat 2 stasiun pengamatan yang berfungsi baik yaitu SUF 6 dan SUF 7. SUF 6 terletak di Kebon Bibit Wonorejo dan SUF 7 terletak di Kelurahan Kebonsari. Pemantauan kandungan PM10 beberapa jam kedepan dapat dilakukan dengan menggunakan sebuah model statistik kemudian menggunakan model tersebut untuk meramalkan kandungan PM10.

ARIMA merupakan salah satu metode statistik yang dapat digunakan untuk meramalkan suatu kejadian di masa depan berdasarkan data yang sudah lampau. Salah satu syarat yang harus dipenuhi adalah data bersifat *time series*. Metode ini pernah digunakan oleh Ditago [6] untuk meramalkan konsum-

si energi listrik di Jawa Timur. Kemudian, pada tahun 2014, metode ARIMA digunakan untuk meramalkan penjualan sepeda motor [7][8].

Penelitian terkait kualitas udara telah banyak dilakukan. Salah satu metode yang sering digunakan adalah metode *time series* ARIMA dengan prosedur Box-Jenkins, seperti yang dilakukan oleh Nuraini [9], yang meramalkan kandungan PM10 berdasarkan stasiun pemantau Taman Prestasi. Peneliiian lain pernah dilakukan oleh Desvina [10], yang meramalkan PM10 di Kajang, Malaysia. Selanjutnya, Lee menggunakan model *seasonal* ARIMA dalam melakukan pemantauan [11]. Apabila dilakukan pengamatan menggunakan data pengamatan per jam, maka akan ditemukan 2 pola musiman dalam data, yaitu harian dan mingguan. Sehingga dapat digunakan model *Double Seasonal* ARIMA [12][13].

Penelitian yang akan dilakukan adalah meramalkan kualitas udara ambien berdasarkan parameter PM₁₀ dengan 2 stasiun pemantau yang ada di Surabaya. PM10 dipilih karena dapat memberikan dampak secara signifikan pada makhluk hidup dan merupakan kandungan tertinggi dalam udara ambien Surabaya. Penelitian dilakukan dengan menggunakan analisis *time series* DSARIMA. Analisis tersebut dipilih karena data pengamatan, yaitu kandungan PM10 dalam udara ambien, memiliki skala waktu per jam dan diamati selama satu tahun. Sehingga diduga bahwa data memiliki pola musiman ganda.






II. TINJAUAN PUSTAKA

A. Kualitas Udara Ambien

Kualitas udara berhubungan erat dengan pencemaran udara. Udara bebas di permukaan bumi pada lapisan troposfir disebut udara ambien. Udara ambien dibutuhkan oleh manusia. Selain itu, udara ambien juga mempengaruhi kesehatan manusia, makhluk hidup dan unsur lingkungan hidup lainnya.

Indeks Standar Pencemar Udara (ISPU) merupakan nilai yang digunakan untuk mengukur parameter kualitas udara. Indeks ini dibuat untuk menyeragamkan informasi kualitas udara ambien, sehingga mudah dimengerti oleh masyarakat. Terdapat lima parameter yang digunakan dalam ISPU. Kelima parameter tersebut meliputi PM₁₀, CO, SO₂, NO₂ dan O₃.

Tabel 1. Kategori dan Rentang ISPU

Warna	Kategori	Rentang
	Baik	0 – 50
	Sedang	51 – 100
	Tidak sehat	101 – 199
	Sangat tidak sehat	200 – 299
	Berbahaya	300 – lebih

ISPU tidak memiliki satuan yang menggambarkan kondisi kualitas udara ambien di lokasi dan waktu tertentu, tetapi didasarkan pada dampak yang berpengaruh pada kesehatan manusia, nilai estetika, dan makhluk hidup lainnya [4].

Kualitas udara ambien Surabaya dipantau berdasarkan 6 titik pemantauan yang tersebar di seluruh wilayah Surabaya. Masing-masing titik digunakan untuk mewakili satu wilayah di Surabaya. Hingga tahun 2014, stasiun pemantau yang dinyatakan berfungsi baik oleh Dinas setempat ada dua, yaitu

yang terpasang di Kebon Bibit Wonorejo (SUF 6) dan Kelurahan Kebonsari (SUF 7).

B. Autoregressive Integrated Moving Average (ARIMA)

Model ARIMA merupakan model gabungan atau kombinasi dari model *Autoregressive* (AR) orde *p* dan *Moving Average* (MA) orde *q*, serta diikuti oleh proses *differencing* orde *d*. Model ini biasa digunakan untuk data yang bersifat linier dan univariat. Secara umum, bentuk model ARIMA (*p,d,q*) adalah sebagai berikut [14]

$$\phi_p(B)(1 - B)^d Y_t = \theta_0 + \theta(B)a_t \quad (1)$$

dengan,

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}$$

$$\Theta_q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_q B^{qs}$$

a_t merupakan barisan *white noise* dengan rata-rata nol dan varians konstan sehingga $a_t \sim WN(0, \sigma^2)$, sedangkan *B* merupakan operator *backshift*.

C. Seasonal ARIMA (SARIMA)

Model ARIMA juga dapat diterapkan pada data yang memiliki pola musiman. Musiman merupakan kejadian berulang selama periode tertentu, artinya data-data yang dipisahkan dalam suatu musim penuh dapat memperlihatkan sifat yang sama pada musim berikutnya. Model SARIMA memiliki bentuk umum [14]

$$\phi_p(B)\Phi_p(B^s)(1 - B)^d(1 - B^s)^D Y_t = \theta_q(B)\Theta_q(B^s)a_t \quad (2)$$

dengan

P = orde dari *seasonal autoregressive* (AR)

Q = orde dari *seasonal moving average* (MA)

D = orde dari *seasonal differencing*

s = periode *seasonal*

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}$$

$$\Theta_p(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_p B^{ps}$$

D. Double Seasonal ARIMA (DSARIMA)

Apabila data mengandung dua pola musiman, maka disebut *Double Seasonal* ARIMA (DSARIMA). Secara matematis, model DSARIMA dapat ditulis sebagai berikut

$$\phi_p(B)\Phi_{p_1}(B^{s_1})\Phi_{p_2}(B^{s_2})(1 - B)^d(1 - B^{s_1})^{D_1}(1 - B^{s_2})^{D_2} Y_t = \theta_q(B)\Theta_{q_1}(B^{s_1})\Theta_{q_2}(B^{s_2})a_t \quad (3)$$

dengan

*P*₁ = orde dari *seasonal 1 autoregressive* (AR)

*P*₂ = orde dari *seasonal 2 autoregressive* (AR)

*Q*₁ = orde dari *seasonal 1 moving average* (MA)

*Q*₂ = orde dari *seasonal 2 moving average* (MA)

*D*₁ = orde dari *seasonal 1 differencing*

*D*₂ = orde dari *seasonal 2 differencing*

*s*₁ = periode *seasonal 1*

*s*₂ = periode *seasonal 2*

$$\Phi_{p_1}(B^{s_1}) = 1 - \Phi_{11} B^{s_1} - \Phi_{12} B^{2s_1} - \dots - \Phi_{1p_1} B^{p_1 s_1}$$

$$\Phi_{p_2}(B^{s_2}) = 1 - \Phi_{21} B^{s_2} - \Phi_{22} B^{2s_2} - \dots - \Phi_{2p_2} B^{p_2 s_2}$$

$$\Theta_{q_1}(B^{s_1}) = 1 - \Theta_{11} B^{s_1} - \Theta_{12} B^{2s_1} - \dots - \Theta_{1p_1} B^{p_1 s_1}$$

$$\Theta_{q_2}(B^{s_2}) = 1 - \Theta_{21} B^{s_2} - \Theta_{22} B^{2s_2} - \dots - \Theta_{2p_2} B^{p_2 s_2}$$

E. Model Subset, Multiplicative dan Additive

Model SARIMA dan DSARIMA merupakan model yang *multiplicative*. Model ini mengasumsikan bahwa parameter signifikan merupakan hasil dari *multiplicative* atau perkalian antara parameter *nonmusiman* dan musiman tanpa harus diuji

secara statistik. Hasil simulasi juga menunjukkan bahwa model musiman ARIMA mengandung model *subset*, *multiplicative* dan *additive* [15].

Misalkan diketahui model ARIMA(0,0,[1,11,12,13] yang diidentifikasi sebagai *subset* SARIMA. Model dapat ditulis sebagai berikut

$$Y_t - \mu = a_t - \theta_1 a_{t-1} - \theta_{12} a_{t-12} - \theta_{13} a_{t-13} \quad (4)$$

dengan θ_1, θ_{12} dan θ_{13} menunjukkan parameter MA. Secara matematis, ACF dari model dapat ditulis sebagai berikut

$$\rho_k = \begin{cases} \frac{-\theta_1 + \theta_{12}\theta_{13}}{(1+\theta_1^2 + \theta_{12}^2 + \theta_{13}^2)}; k = 1 \\ \frac{\theta_1\theta_{12}}{(1+\theta_1^2 + \theta_{12}^2 + \theta_{13}^2)}; k = 11 \\ \frac{-\theta_{12} + \theta_1\theta_{13}}{(1+\theta_1^2 + \theta_{12}^2 + \theta_{13}^2)}; k = 12 \\ \frac{\theta_{13}}{(1+\theta_1^2 + \theta_{12}^2 + \theta_{13}^2)}; k = 13 \\ 0; \text{lainnya.} \end{cases} \quad (5)$$

Misalkan diketahui model ARIMA(0,0,1)(0,0,1)¹² yang diidentifikasi sebagai *multiplicative* SARIMA. Model dapat ditulis sebagai berikut

$$Y_t - \mu = a_t - \theta_1 a_{t-1} - \theta_{12} a_{t-12} + \theta_1 \theta_{12} a_{t-13} \quad (6)$$

dengan θ_1 dan θ_{12} menunjukkan parameter *nonmusiman* dan musiman MA. Model (6) sama dengan model (4) jika $\theta_{13} = -\theta_1 \theta_{12}$. Sehingga dapat disimpulkan bahwa model *multiplicative* merupakan bagian dari model *subset*. Secara matematis, ACF dari model dapat ditulis sebagai berikut

$$\rho_k = \begin{cases} \frac{-\theta_1}{(1+\theta_1^2)}; k = 1 \\ \frac{\theta_1\theta_{12}}{(1+\theta_1^2)(1+\theta_{12}^2)}; k = 11,13 \\ \frac{-\theta_{12}}{(1+\theta_{12}^2)}; k = 12 \\ 0; \text{lainnya.} \end{cases} \quad (7)$$

Persamaan (7) menunjukkan bahwa ACF pada lag 11 sama dengan ACF pada lag 13.

Misalkan diketahui model ARIMA(0,0,[1,12]) yang diidentifikasi sebagai *additive* SARIMA. Model dapat ditulis sebagai berikut

$$Y_t - \mu = a_t - \theta_1 a_{t-1} - \theta_{12} a_{t-12} \quad (8)$$

dengan θ_1 dan θ_{12} menunjukkan parameter *nonmusiman* dan musiman MA. Model (8) sama dengan model (4) jika $\theta_{13} = 0$. Sehingga dapat disimpulkan bahwa model *additive* merupakan bagian dari model *subset*. Secara matematis, ACF dari model dapat ditulis sebagai berikut

$$\rho_k = \begin{cases} \frac{-\theta_1}{(1+\theta_1^2 + \theta_{12}^2)}; k = 1 \\ \frac{\theta_1\theta_{12}}{(1+\theta_1^2 + \theta_{12}^2)}; k = 11 \\ \frac{-\theta_{12}}{(1+\theta_1^2 + \theta_{12}^2)}; k = 12 \\ 0; \text{lainnya.} \end{cases} \quad (9)$$

Persamaan (9) menunjukkan bahwa ACF pada lag 13 adalah nol. Hal inilah yang membedakan persamaan (8) dengan persamaan (4) dan persamaan (6).

F. Pemilihan Model Terbaik

Pemilihan model terbaik pada penelitian ini menggunakan kriteria kesalahan peramalan *Root Mean Square Error* (RMSE) untuk uji data *out sample*. Rumus RMSE pada adalah sebagai berikut [14].

$$RMSE_{out} = \sqrt{\frac{\sum_{l=1}^L (Y_{n+l} - \hat{Y}_n(l))^2}{L}} \quad (10)$$

dengan

Y_{n+l} = data *out sample*

$\hat{Y}_n(l)$ = nilai ramalan

L = jumlah data *out sample*.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Lingkungan Hidup (BLH) Surabaya. Data diperoleh dari alat yang dipasang pada stasiun pemantauan dan disajikan per jam dan diambil pada bulan Januari hingga Desember 2014. Stasiun pemantauan yang digunakan dalam penelitian adalah stasiun yang terletak di Wonorejo (SUF 6) dan Kebonsari (SUF 7).

B. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah sebagai berikut :

1. $Y_{1,t}$ = Kandungan parameter PM₁₀ di stasiun pemantau SUF 6.
2. $Y_{2,t}$ = Kandungan parameter PM₁₀ di stasiun pemantau SUF 7.

Struktur data dalam penelitian ini akan ditampilkan dalam tabel berikut.

Tabel 2. Struktur Data Penelitian

Waktu (per jam)	Variabel	
	PM10 SUF 6	PM10 SUF 7
1	$Y_{1,1}$	$Y_{2,1}$
2	$Y_{1,2}$	$Y_{2,2}$
⋮	⋮	⋮
50	$Y_{1,50}$	$Y_{2,50}$
⋮	⋮	⋮
$n - 47$	$Y_{1,(n-47)}$	$Y_{2,(n-47)}$
⋮	⋮	⋮
n	$Y_{1,n}$	$Y_{2,n}$

C. Langkah Analisis

Berdasarkan metode analisis yang digunakan, maka langkah-langkah analisis yang digunakan adalah sebagai berikut.

1. Menghitung statistika deskriptif, yaitu berupa rata-rata, standar deviasi, nilai maksimum, dan nilai minimum dari masing-masing variabel. Selain itu, analisis juga dilakukan dengan menggunakan plot *time series* dan *boxplot*. Selain dianalisis secara keseluruhan, data kandungan PM10 juga dianalisis berdasarkan jam, hari dan bulan. Hal ini dilakukan untuk mengetahui pola dan karakteristik kandungan PM10 dalam udara ambien Surabaya.
2. Membagi data menjadi *in sample* dan *out sample*. Data *in sample* adalah data kandungan PM10 tanggal 1 Januari hingga 30 November 2014. Data *out sample* adalah data kandungan PM10 tanggal 1 hingga 7 Desember 2014.
3. Membentuk model DSARIMA untuk stasiun pemantau SUF 6 dengan langkah-langkah sebagai berikut.

- a. Mengidentifikasi data yang meliputi uji stasioneritas data, baik dalam *mean* maupun varians. Serta menentukan orde $p, d, q, P_1, P_2, D_1, D_2, Q_1, Q_2, s_1$ dan s_2 berdasarkan plot ACF dan PACF.
 - b. Mengestimasi parameter model ARIMA.
 - c. Menguji signifikansi parameter model ARIMA.
 - d. Melakukan cek diagnosa yang meliputi uji *white noise* dan uji distribusi normal. Untuk menguji *white noise* digunakan plot ACF dari residual model. Sedangkan untuk menguji distribusi normal digunakan uji Kolmogorov-Smirnov.
 - e. Memilih model terbaik dengan membandingkan nilai RMSE *out sample*. Model terbaik adalah model yang memiliki RMSE terkecil.
 - f. Melakukan peramalan menggunakan model terbaik.
4. Mengulangi langkah (3) untuk membentuk model DSARIMA untuk stasiun pemantau SUF 7.
 5. Menarik Kesimpulan.

IV. ANALISIS DAN PEMBAHASAN

A. Statistika Deskriptif

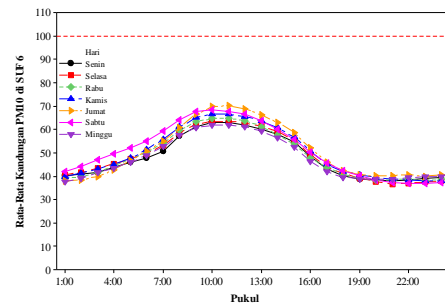
Analisis berdasarkan statistika deskriptif bertujuan untuk mengeksplorasi informasi yang terkandung dalam data dan menggambarkan keadaan yang ada berdasarkan data yang diperoleh. Berikut ini merupakan nilai rata-rata, standar deviasi, nilai minimum dan nilai maksimum dari data kandungan PM10 selama tahun 2014.

Tabel 1. Statistika Deskriptif Kandungan PM10 di SUF 6 dan SUF 7

Stasiun Pemantau	Rata-Rata	Std Deviasi	Minimum	Maksimum
SUF 6	49,285	19,527	1,125	125,943
SUF 7	61,564	23,434	1,23	132

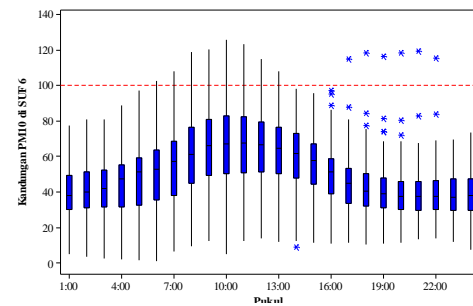
Tabel 1 menunjukkan bahwa rata-rata kandungan PM10 di SUF 7 lebih tinggi dibandingkan dengan SUF 6. Apabila kesimpulan didasarkan pada nilai rata-rata tersebut maka seolah-olah kandungan PM10 masih dalam standar baik. Namun, jika diperhatikan nilai maksimum dari kedua stasiun, kesimpulan yang diperoleh akan berbeda. Kandungan PM10 tertinggi di dua stasiun pemantau menunjukkan nilai yang lebih dari 100 yang berarti bahwa udara ambien pernah mencapai kondisi yang tidak sehat untuk dihirup oleh makhluk hidup. Nilai standar deviasi menunjukkan tingkat keragaman kandungan PM10 dalam udara ambien Kota Surabaya. Keragaman kandungan PM10 yang terpantau melalui SUF 7 lebih tinggi dibandingkan dengan kandungan PM10 yang terpantau melalui SUF 6. Hal ini juga berarti bahwa kandungan PM10 di SUF 7 lebih fluktuatif dibandingkan dengan SUF 6.

Rata-rata kandungan PM10 terendah terjadi pada malam hingga dini hari dan rata-rata kandungan PM10 tertinggi terjadi pada siang hingga sore hari. Kondisi ini terjadi di dua lokasi pengamatan. Rata-rata kandungan PM10 mencapai puncak tertinggi pada pukul 10.00 pagi hingga pukul 13.00 siang berdasarkan dua stasiun pemantau, SUF 6 dan SUF 7. Hal ini terjadi seiring dengan aktivitas lalu lintas yang padat di Surabaya, sehingga menyebabkan kandungan PM10 mencapai nilai tertinggi pada siang hari.



Gambar 1. Plot Rata-rata Kandungan PM10 Per Jam Per Hari di SUF 6

Data kandungan PM10, baik di SUF 6 maupun SUF 7, memiliki banyak *outlier*. *Outlier* di SUF 6 dan SUF 7 bernilai lebih dari interval kepercayaan, sehingga menunjukkan bahwa kandungan PM10 lebih sering mencapai nilai tinggi daripada rendah.



Gambar 2. Box Plot Kandungan PM10 Per Jam di SUF 6

Rata-rata kandungan PM10 tiap hari berbeda. Kandungan PM10 akan rendah pada hari Senin dan Minggu, dimana aktivitas pada hari tersebut tidak sepadat hari lainnya. Sebagian besar masyarakat lebih memilih untuk beristirahat di rumah daripada melakukan aktivitas. Hal inilah yang menyebabkan kandungan PM10 pada hari Senin dan Minggu lebih rendah dibandingkan dengan hari lainnya.

B. Peramalan Menggunakan DSARIMA

Pemodelan dengan DSARIMA menggunakan prosedur Box-Jenkins dimulai dengan identifikasi kestasioneran data dalam *mean* dan varians. Data kandungan PM10 baik di SUF 6 maupun SUF 7 tidak stasioner dalam varians, sehingga dilakukan transformasi. Transformasi yang dilakukan adalah data asli ditambah 20 pada data di SUF 6 dan ditambah 30 pada data di SUF 7, kemudian di-*ln*-kan. Stasioneritas data dalam *mean* dapat diidentifikasi melalui plot ACF. Plot ACF yang memiliki pola turun sangat lambat mengindikasikan bahwa data tidak stasioner dalam mean sehingga perlu dilakukan *differencing*.

Sebelumnya, berdasarkan plot ACF diketahui bahwa data memiliki pola setiap kelipatan 168, sedangkan berdasarkan plot PACF diketahui bahwa data memiliki nilai yang signifikan setiap kelipatan 24. Hal ini menunjukkan bahwa data memiliki dua pola musiman, yaitu musiman harian dan musiman mingguan. Data yang memiliki dua pola musiman dapat diselesaikan dengan menggunakan metode DSARIMA.

Untuk memenuhi stasioner dalam mean, maka perlu dilakukan *differencing*. Data didifferencing pada kedua lag musiman, yaitu lag 24 dan lag 168, dan pada lag *non*musiman, yaitu lag 1. Setelah dilakukan *differencing*, maka dibuat plot ACF dan PACF kembali. Plot menunjukkan bahwa data sudah stasioner dalam *mean*.

Secara teoritis, plot ACF dari data yang telah stasioner memiliki pola *cut off* pada lag musiman 24 dan 168 dan *dies down* pada lag *non*musiman. Sedangkan plot PACF memiliki pola *dies down* pada lag musiman dan *non*musiman. Berdasarkan plot ACF dan PACF dari data yang telah stasioner, diperoleh dua model dugaan untuk data kandungan PM10 di SUF 6 dan SUF 7. Berikut ini model dugaan yang diperoleh.

Tabel 2. Model Dugaan DSARIMA Kandungan PM10

Stasiun Pemantau	Model Dugaan
SUF 6	Model 1 ARIMA([1,2,4,7,8,9,10,12,16,17,18],1,0) (1,1,1) ²⁴ (0,1,1) ¹⁶⁸
	Model 2 ARIMA([1,2,3,4,7,8,9,10,11,12,16,17,18],1,0) (1,1,1) ²⁴ (0,1,1) ¹⁶⁸
SUF 7	Model 1 ARIMA([11,2,7,9,10,11,12,16,17,19,23],1,4,5,8,20) (0,1,1) ²⁴ (0,1,1) ¹⁶⁸
	Model 2 ARIMA([1,2,4,5,7,8,10,11,12,13,15,16,18,19,20],1,0) (1,1,1) ²⁴ (0,1,1) ¹⁶⁸

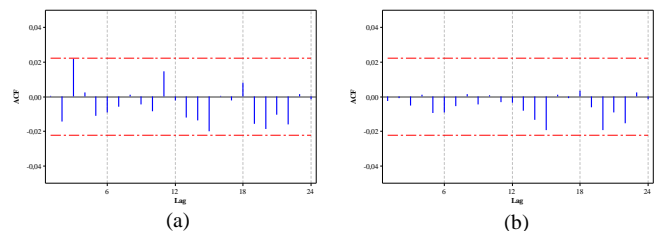
Orde-orde dalam model tersebut diperoleh dengan melihat lag-lag pada plot ACF dan PACF data yang sudah stasioner yang nilainya lebih besar dari batas signifikansi. Kemudian dilakukan pengujian pada lag-lag tersebut, dengan cara memasukkannya dalam model. Lag dikatakan signifikan pengaruhnya jika *p-value* lebih kecil dari taraf signifikansi α yang digunakan. Lag-lag yang signifikan tersebut kemudian disebut *subset*. Lag-lag yang diuji signifikansinya adalah lag *non*musiman antara lag 1 hingga lag 24, lag musiman 24 dan 168.

Setelah model dugaan diperoleh, maka langkah selanjutnya adalah mengestimasi dan menguji signifikansi parameter model. Semua *p-value* yang diperoleh bernilai kurang dari taraf signifikansi α sebesar 0,1. Hal ini menunjukkan bahwa semua parameter dalam model dugaan, baik model dugaan untuk SUF 6 dan SUF 7, sudah signifikan.

Apabila parameter dalam model sudah signifikan, maka langkah selanjutnya adalah pengujian diagnostik residual. Pengujian dilakukan untuk mengetahui apakah residual telah memenuhi asumsi, baik *white noise* maupun berdistribusi normal. Pengujian asumsi *white noise* dilakukan dengan melihat plot ACF residual. Apabila tidak ada lag yang nilainya melebihi batas signifikansi, maka dapat disimpulkan bahwa asumsi *white noise* terpenuhi. Sedangkan pengujian asumsi distribusi normal dilakukan dengan menggunakan uji Kolmogorov-Smirnov.

ACF dari residual model untuk SUF 6 dan SUF 7 tidak ada lag yang nilainya melebihi batas signifikansi sehingga dapat disimpulkan bahwa residual, baik residual model untuk SUF 6 maupun SUF 7, sudah memenuhi asumsi *white noise*.

Selanjutnya adalah pengujian asumsi distribusi normal pada residual model. Tabel 3 menunjukkan hasil dari pengujian residual menggunakan uji Kolmogorov-Smirnov. *P-value* uji Kolmogorov-Smirnov untuk keempat model bernilai kurang dari taraf signifikansi 0,05. Artinya, residual dari keempat model tidak berdistribusi normal. Residual tidak berdistribusi normal dikarenakan data kandungan PM10, baik di SUF 6 maupun SUF 7, memiliki banyak *outlier*.



Gambar 5. Plot ACF Residual (a) Model 1 dan (b) Model 2 DSARIMA untuk SUF 7

Tabel 3. Uji Asumsi Distribusi Normal Residual

Stasiun Pengamatan	D	P-value
SUF 6	Model 1	0,069 < 0,01
	Model 2	0,071 < 0,01
SUF 7	Model 1	0,095 < 0,01
	Model 2	0,096 < 0,01

Keempat model untuk SUF 6 dan SUF 7 yang memiliki parameter signifikan dan residual *white noise* selanjutnya digunakan untuk meramal hingga 168 tahap kedepan untuk periode 1 hingga 14 Desember 2015. Data *out sample* dikurangi dengan hasil ramalan, kemudian dihitung RMSEnya.

Tabel 4. RMSE *Out Sample* Model DSARIMA di SUF 6 dan SUF 7

Stasiun Pengamatan	RMSE <i>Out Sample</i>
SUF 6	Model 1 16,372
	Model 2 16,531
SUF 7	Model 1 14,284
	Model 2 14,159

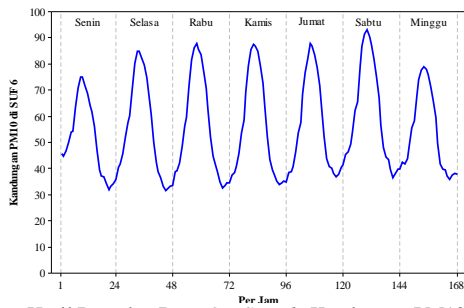
RMSE *out sample* digunakan sebagai kriteria untuk memilih model terbaik. Nilai RMSE model 1 di SUF 6 lebih kecil dibandingkan dengan model 2. Sedangkan di SUF 7, model 2 memiliki RMSE yang lebih kecil dibandingkan dengan model 1. Sehingga disimpulkan bahwa model 1 merupakan model terbaik untuk meramalkan dan mengontrol kandungan PM10 di SUF 6 dan model 2 merupakan model terbaik untuk meramalkan dan mengontrol kandungan PM10 di SUF 7.

Hasil ramalan data *out sample* yang diramalkan 168 tahap kedepan menunjukkan bahwa kandungan PM10 pada hari Senin dan Minggu lebih rendah dibandingkan dengan hari lainnya. Hal ini sesuai dengan analisis statistika deskriptif yang telah dilakukan sebelumnya.

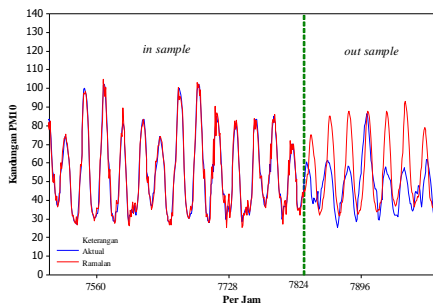
Selain itu, jika dibuat plot rata-rata per jam menunjukkan hasil yang sesuai dengan analisis statistika deskriptif bahwa kandungan PM10 pada siang hari lebih tinggi dibandingkan pada malam hari.

Selanjutnya adalah membandingkan hasil peramalan dengan data *in sample* dan *out sample*. Gambar 7 menunjukkan bahwa secara grafik hasil ramalan data *in sample* lebih baik dibandingkan dengan data *out sample*. Hal ini disebabkan oleh peramalan data *in sample* merupakan peramalan satu tahap kedepan sehingga diperoleh RMSE yang kecil, sedangkan peramalan pada data *out sample* merupakan peramalan *k*-tahap kedepan secara langsung. Hasil ramalan

yang kurang baik pada data *out sample* disebabkan oleh semakin banyak ramalan yang dilakukan maka semakin besar pula RMSE yang dihasilkan.



Gambar 6 Hasil Ramalan Data *Out Sample* Kandungan PM10 di SUF 6



Gambar 7. Plot *Time Series* antara Data Aktual dengan Hasil Ramalan Kandungan PM10 di SUF 6

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan analisis data kandungan PM10 tahun 2014, maka didapatkan beberapa kesimpulan sebagai berikut.

1. Data kandungan PM10 yang dipantau oleh stasiun pemantau SUF 7 lebih tinggi dibandingkan dengan yang dipantau oleh SUF 6. Perbedaan ini disebabkan oleh karakteristik kedua lokasi yang berbeda, dimana di sekitar stasiun pemantau SUF 6 lebih banyak terdapat pohon rindang dan hutan bakau yang mampu mereduksi kadar polutan PM10, sedangkan di SUF 7 tidak. Rata-rata kandungan PM10 tinggi pada saat siang hari, dimana aktivitas masyarakat sangat padat. Tingginya kandungan PM10 dimulai pukul 10.00 hingga 13.00. Apabila dianalisis berdasarkan rata-rata per hari, maka kandungan PM10 pada hari Senin dan Minggu lebih rendah dibandingkan dengan hari lainnya. Kandungan PM10 yang dipantau oleh stasiun pemantau SUF 7 memiliki *outlier* yang lebih banyak dibandingkan dengan SUF 6. *Outlier* merupakan nilai yang berada diluar interval, sehingga menyebabkan data kandungan PM10 tidak berdistribusi normal.
2. Model ARIMA terbaik berdasarkan kriteria *out sample* RMSE untuk meramalkan kandungan PM10 per jam di SUF 6 adalah ARIMA $([1,2,4,7,8,9,10,12,16,17,18],1,0)(1,1,1)^{24}(0,1,1)^{168}$ dan model terbaik untuk di SUF 7 adalah ARIMA $([1,2,4,5,7,8,10,11,12,13,15,16,18,19,20],1,0)(1,1,1)^{24}(0,1,1)^{168}$.

B. Saran

Setelah melakukan penelitian ini, maka beberapa hal yang perlu diperbaiki pada penelitian berikutnya adalah pemodelan kandungan PM10 menggunakan *Double Seasonal ARIMA* (DSARIMA) dengan menambahkan pendekatan *percentile error bootstrap* untuk mengatasi interval hasil ramalan yang semakin lebar jika dilakukan beberapa tahap kedepan. Selain itu, untuk mengatasi residual yang tidak berdistribusi normal dapat digunakan deteksi *outlier* pada residual model ARIMA yang dihasilkan.

Saran yang dapat diberikan untuk Pemerintah Kota Surabaya adalah menambah jumlah pohon rindang di kawasan Kebonsari untuk mereduksi kandungan PM10 dalam udara. Selain itu, Pemerintah juga dapat meramalkan kandungan PM10 satu jam kedepan menggunakan model ARIMA yang diperoleh.

DAFTAR PUSTAKA

- [1] Tamin, O. Z. (1992). *Metodologi Peramalan Lalu Lintas Perkotaan untuk Negara Berkembang*. Jakarta: PT Perentjana Djaja.
- [2] Bowman, K. W. (2013). Toward the Next Generation of Air Quality Monitoring: Ozone. *Elsevier: Atmospheric Environment* 80, 571-583.
- [3] Syafei. (2012, Juni 5). *Data Kualitas Udara Kota Surabaya*. Dipetik Desember 26, 2014, dari Catatan Teknik Lingkungan: <http://envirodiary.com/id/data-kualitas-udara-kota-surabaya.php>
- [4] Tamin, O. Z. (1992). *Metodologi Peramalan Lalu Lintas Perkotaan untuk Negara Berkembang*. Jakarta: PT Perentjana Djaja.
- [5] BLH. (2014). *Laporan Kegiatan Operasional dan Pengadaan Stasiun Monitoring Udara Ambien (Kode Kegiatan 1.08.16.0023)*. Surabaya: Badan Lingkungan Hidup.
- [6] Yu, H.-L., Lin, Y.-C., & Kuo, Y.-M. (2015). A Time Series Analysis of Multiple Ambient Pollutants to Investigate the Underlying Air Pollution Dynamics and Interactions. *Elsevier: Chemosphere*
- [7] Ditago, A. P., Suharsono, A. & Suhartono. (2013). Perbandingan Model ARIMAX dan Fungsi Transfer untuk Peramalan Konsumsi Energi Listrik di Jawa Timur. *Jurnal Sains dan Seni POMITS Vol. 2 No. 2, 2337-3520*.
- [8] Anggraeni, A. D. & Suharsono, A. (2014). Peramalan Penjualan Sepeda Motor Tiap Jenis di Wilayah Surabaya dan Blitar dengan Model ARIMA Box-Jenkins dan *Vector Autoregressive* (VAR). *Jurnal Sains dan Seni POMITS Vol. 3 No. 2, 2337-3520*.
- [9] Harahap, M. R. P. & Suharsono, A. (2014). Analisis Peramalan Penjualan Sepeda Motor Di Kabupaten Ngawi dengan Arima dan Arimax. *Jurnal Sains dan Seni POMITS Vol. 3 No. 2, 2337-3520*.
- [10] Nuraini. (2004). *Analisis Statistik Terhadap Polutan Udara Partikulat (PM10) Berdasarkan Pengukuran Stasiun Pemantau Taman Prestasi dengan Metode Bootstrap Time Series*. Surabaya: Jurusan Statistika ITS.
- [11] Desvina, A. P. (2011). *Peramalan Particulate Matter (PM₁₀) di Kajang Malaysia dengan Menggunakan Model Autoregressive*. Riau: Jurusan Matematika Fakultas Sains dan Teknologi UIN Suska.
- [12] Lee, M. H., dkk. (2012). Seasonal ARIMA for Forecasting Air Pollution Index: A Case Study. *American Journal of Applied Sciences* 9,570-578.
- [13] Khusna, H. (2015). *Pendekatan Percentile Error Bootstrap pada Model Double Seasonal Holt-Winters, Double Seasonal ARIMA, dan Naive untuk Peramalan Beban Listrik Jangka Pendek Area Jawa Timur-Bali*. Surabaya: Jurusan Statistika ITS.
- [14] Putri, I. K. (2015). *Deteksi Outlier pada Model ARIMA Musiman Ganda untuk Peramalan Beban Listrik Jangka Pendek di Jawa Timur*. Surabaya: Jurusan Statistika ITS.
- [15] Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods, Second Edition*. United States: Pearson Education, Inc.
- [16] Suhartono, & Lee, M. H. (2011). Forecasting of Tourist Using Subset, Multiplicative or Additive Seasonal ARIMA Model. *MATEMATIKA Vol. 27 Department of Mathematical Sciences, UTM*, 169-182.