

Perbandingan Metode ARIMA (*Box Jenkins*), ARFIMA, Regresi Spektral dan SSA dalam Peramalan Jumlah Kasus Demam Berdarah Dengue di Rumah Sakit Hasan Sadikin Bandung

Paridi

Universitas Gunung Rinjani

paridi.faza@gmail.com

Abstrak. Peramalan adalah suatu proses yang bertujuan menduga suatu kejadian yang akan datang dan merupakan alat bantu yang penting dalam perencanaan yang efektif dan efisien. Metode peramalan merupakan metode yang berfungsi untuk memprediksi data runtun waktu beberapa periode yang akan datang berdasarkan data di masa lalu, yang dinyatakan dalam bentuk persamaan matematis. Beberapa metode pemodelan time series telah dikembangkan yaitu metode ARIMA (*Autoregressive Integrated Moving Average*) atau juga dikenal dengan metode Box-Jenkins dan metode ARFIMA (*Autoregressive Fractionally Integreted Moving Average*), Regresi Spektral dan SSA (*Singular Spectrum Analysis*).

Kata Kunci: *Metode Arima, Afrima, Regresi Spektral, SSA*

PENDAHULUAN

Peramalan adalah suatu proses yang bertujuan menduga suatu kejadian yang akan datang dan merupakan alat bantu yang penting dalam perencanaan yang efektif dan efisien (Makridakis *et al.*, 1999). Metode peramalan merupakan metode yang berfungsi untuk memprediksi data runtun waktu beberapa periode yang akan datang berdasarkan data di masa lalu, yang dinyatakan dalam bentuk persamaan matematis.

Metode peramalan melalui analisis suatu variabel yang akan diperkirakan dengan variabel atau fungsi waktu, dikenal sebagai metode hubungan *time series*. *Time series* merupakan serangkaian data pengamatan yang terjadinya berdasarkan urutan waktu (Box, Jenkins dan Reinsel, 1994). Analisis data *time series* dapat digunakan untuk mengetahui perkembangan satu atau beberapa kejadian dan hubungan atau pengaruhnya pada kejadian lainnya, sekaligus menghitung secara kuantitatif besarnya pengaruh suatu kejadian dengan kejadian lainnya apabila berhubungan. Data *time series* juga dapat digunakan untuk membuat ramalan berdasarkan garis regresi atau *trend*.

Beberapa metode pemodelan time series telah dikembangkan. Metode yang paling umum digunakan adalah ARIMA (*Autoregressive Integrated Moving Average*). Menurut Box dan Jenkins (1976), metode ARIMA atau juga dikenal dengan metode

Box-Jenkins merupakan salah satu bentuk analisis deret waktu (*time series analysis*). Model ARIMA dapat menganalisis data secara univariat yang mengandung pola musiman maupun *trend*. Metode ini hanya menganalisis data yang stasioner, sehingga data yang tidak stasioner harus distasionerkan terlebih dahulu dengan transformasi atau pembedaan. Model ARIMA juga bisa digunakan untuk mengatasi masalah sifat keacakan, *trend*, musiman bahkan sifat siklis data *time series* yang dianalisis. ARIMA sangat efektif digunakan untuk memodelkan data yang tidak stasioner, yang ditunjukkan oleh plot ACF yang turun secara eksponensial atau membentuk gelombang sinus. Ada beberapa data yang tidak stasioner dan plot ACF-nya tidak turun secara eksponensial melainkan secara lambat atau hiperbolik. Data seperti inilah yang dikategorikan sebagai *time series* memori jangka panjang (*long memory*). Untuk memodelkan *time series* jangka panjang, Hosking (1981) telah memperkenalkan model ARFIMA (*Autoregressive Fractionally Integreted Moving Average*). ARFIMA dapat mengatasi kelemahan model ARIMA, dimana ARIMA hanya dapat menjelaskan *time series* jangka pendek (*short memory*), sedangkan ARFIMA dapat menjelaskan baik jangka pendek maupun jangka panjang. (Sowell, 1992).

Selain metode ARIMA dan ARFIMA juga dikenal metode Regresi Spektral, yang merupakan metode peramalan trend dan

musiman. Identifikasi musiman yang ada, masih menggunakan plot (Wei, 2006), yaitu dengan menggunakan *Buys Ballot* atau menggunakan plot ACF. Padahal pengamatan musiman melalui gambar ada dua kelemahan, pertama bisa mengandung subjektivitas peneliti dan kedua tidak bisa menyimpulkan suatu fenomena hanya dengan menggunakan plot data deret waktu. Oleh karena itu dengan Regresi Spektral yang merupakan suatu metode yang pasti (*exact*) diharapkan dapat mengidentifikasi *hidden periodicity* dan menentukan besarnya periode dari musiman.

TINJAUAN PUSTAKA

Model ARIMA (*Autoregressive Integrated Moving Average*)

Model ARIMA diaplikasikan untuk deret waktu yang mempunyai mean atau varians yang tidak konstan, atau mempunyai kedua sifat tersebut. Sifat ini disebut sebagai proses yang tidak stasioner. Jika data yang digunakan mengandung pola musiman, maka model yang digunakan adalah SARIMA (p, d, q) (P, D, Q)^s. Sedangkan apabila model ARIMA memiliki pola musiman ganda, maka model yang digunakan adalah DSARIMA (p, d, q) (P_1, D_1, Q_1)^{s1} (P_2, D_2, Q_2)^{s2}, dapat dituliskan sebagai berikut

$$\phi_p(B)(1-B)^d \Phi_{P_1}(B^{s_1})(1-B^{s_1})^{D_1} \Phi_{P_2}(B^{s_2})(1-B^{s_2})^{D_2} Z_t = \theta_0 + \theta_q(B) \Theta_{Q_1}(B^{s_1}) \Theta_{Q_2}(B^{s_2}) a_t \tag{1}$$

dimana:

- $\Phi_{P_1}(B^{s_1})$: koefisien komponen AR periode musiman pertama s_1 dengan orde P_1
- $\Phi_{P_2}(B^{s_2})$: koefisien komponen AR periode musiman kedua s_2 dengan orde P_2
- $\Theta_{Q_1}(B^{s_1})$: koefisien komponen MA periode musiman pertama s_1 dengan orde Q_1
- $\Theta_{Q_2}(B^{s_2})$: koefisien komponen MA periode musiman kedua s_2 dengan orde Q_2
- $(1-B^{s_1})^{D_1}$: differencing musiman pertama s_1 dengan orde D_1
- $(1-B^{s_2})^{D_2}$: differencing musiman pertama s_2 dengan orde D_2

Tahapan analisis data metode ARIMA meliputi: Eksplorasi data, Tahapan Metode ARIMA, dan menghitung *MAPE*, *MAD* dan *MSD*. Tahapan Metode ARIMA sendiri terdiri dari beberapa tahap yaitu: (a) Tahap identifikasi dengan pengidentifikasi kasian model yang dianggap paling sesuai dengan melihat plot ACF dan PACF dari *correlogram*; (b) Tahap estimasi parameter dengan penaksiran

terhadap parameter dalam model tersebut; (c) Tahap uji diagnostik: untuk menguji kesesuaian dari parameter yang didapat pada tahap sebelumnya; (d) Tahap peramalan/penerapan: dilakukan setelah model yang sesuai teridentifikasi.

Autoregressive Fractionally Integrated Moving Average (ARFIMA)

Suatu proses dikatakan mengikuti model *Autoregressive Fractionally Integrated Moving Average* jika operator pembedaan dalam model adalah real. ARFIMA disebut juga ARIMA yang nilai operator pembedaan d tidak hanya berupa nilai integer, melainkan termasuk juga nilai-nilai real yang disebabkan oleh adanya memori jangka panjang. Model ARFIMA (p, d, q) dapat ditulis (Ooms & Doornik, 1999)

$$\phi_p(B)(1-B)^d y_t = \theta_q(B) a_t, t = 1, 2, \dots, T$$

dimana level integrasi d merupakan bilangan real, μ adalah nilai mean yang berupa konstanta dan σ^2 adalah varians. Filter pembedaan $(1-B)^d$ disebut *Long Memory Filter (LMF)* yang menggambarkan adanya ketergantungan jangka panjang dalam deret. Filter ini diekspansikan sebagai deret Binomial : (Hosking, 1981).

$$\begin{aligned} \nabla^d &= (1-B)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j B^j \\ &= \binom{d}{0} (-1)^0 B^0 + \binom{d}{1} (-1)^1 B^1 \\ &\quad + \binom{d}{2} (-1)^2 B^2 \\ &\quad + \binom{d}{3} (-1)^3 B^3 + \dots \\ &= \frac{d!}{0!(d-0)!} - \frac{d!}{1!(d-1)!} \\ &\quad + \frac{d!}{2!(d-2)!} \\ &\quad - \frac{d!}{3!(d-3)!} + \dots \\ &= 1 - dB + \frac{1}{2}(d-1)dB^2 \\ &\quad - \frac{1}{6}(d-2)(d-1)dB^3 + \dots \end{aligned}$$

$$= 1 - dB - \frac{1}{2}d(1-d)B^2 - \frac{1}{6}d(1-d)(2-d)B^3 + \dots$$

Asumsi-asumsi pada deret yang *fractionally integrated* yang harus dipenuhi adalah:

- a. $\phi(B)$ mempunyai order kurang dari atau sama dengan p , $\theta(B)$ mempunyai order kurang dari atau sama dengan q , akar-akar $\phi(B)$ dan $\theta(B)$ diluar *unit circledan* $|z| < 1$, $\epsilon_t \sim IIDN(0, \sigma^2)$
- b. $|d| < 0,5$
- b. akar-akar dari $\phi(B)$ sederhana, atau dengan kata lain akar-akar polynomial *autoregressive* tidak berulang (Sowell, 1992a).

Hosking (1981) menjelaskan bahwa bentuk khusus model ARFIMA adalah proses ARFIMA $(0,d,0)$ yang mempunyai sifat-sifat dasar sebagai berikut :

- a. Jika $d < \frac{1}{2}$, $\{y_t\}$ adalah proses stasioner dan mempunyai orde tak hingga representasi *moving average* orde tak hingga.

$$y_t = \Psi(B)\epsilon_t = \sum_{k=0}^{\infty} \Psi_k \epsilon_{t-k}$$

dimana

$$\Psi_k = \frac{d(1+d)(2+d) \dots (k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}$$

- b. jika $d < \frac{1}{2}$, $\{y_t\}$ invertibel dan mempunyai representasi *autoregressive* orde tak hingga

$$\pi(B)y_t = \epsilon_t = \sum_{k=0}^{\infty} \pi_k y_{t-k}$$

Dimana

$$\pi_k = \frac{-d(1-d)(2-d) \dots (k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!}$$

- c. Fungsi autokovarian dari $\{y_t\}$ adalah

$$\gamma_k = E(y_t, y_{t-k}) = \frac{(-1)^k (-2d)!}{(k-d)! (-k-d)!}$$

sehingga fungsi autokorelasi dari $\{y_t\}$ adalah:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{(-d)!(k+d-1)!}{(d-1)!(k-d)!}, k = 0, \pm 1, \dots$$

Dan $\gamma_0 = \frac{(-2d)!}{\{(-d)\}^2}$ serta $\rho_1 = \frac{d}{(1-d)}$

Pemodelan ARFIMA dapat diperoleh beberapa keuntungan antara lain :

- a. Mampu memodelkan perubahan yang tinggi dalam jangka panjang (*Long Term Persistence*).
- b. Mampu menjelaskan struktur korelasi jangka panjang dan jangka pendek sekaligus.
- c. Mampu memberikan model dengan parameter yang lebih sedikit (*parsimonious*) baik untuk data dengan memori jangka panjang maupun jangka pendek. (Hosking, 1981)

Regresi Spektral

Identifikasi pola musiman pada data tidak hanya dapat dilakukan secara visual, dapat juga dideteksi dengan metode regresi spektral yaitu sebuah metode yang digunakan untuk menelaah periodesitas tersembunyi (*periodesitas yang sulit ditemukan dalam kawasan waktu*).

Persamaan regresi spektral berbentuk sebagai berikut:

$$Z_t = \alpha \cos \omega t + \beta \sin \omega t + e_t \dots \dots \dots (1)$$

dengan,

- Z_t : series data deret waktu pada periode t
- α dan β : parameter (koefisien fourier)
- $\cos \omega t$ dan $\sin \omega t$: fungsi kontinu yang tidak berkorelasi
- e_t : eror pada periode waktu ke-t
- ωt : frekuensi fourier
- t : periode waktu

Tahapan – tahapan untuk melakukan Pengujian musiman dengan metode regresi spektral adalah sebagai berikut :

- 1. Tentukan data yang akan diuji musiman, dalam penelitian ini adalah kasus DBD.
- 2. Karena Persamaan (1), pada dasarnya secara esensi koefisien fourier sama dengan koefisien pada regresi standar maka bila diimplementasikan pada data dapat dituliskan menjadi persamaan fourier sebagai berikut:

$$Z_t = \sum_{k=0}^{[n/2]} (a_k \cos \omega_k t + b_k \sin \omega_k t), \dots \dots \dots (2)$$

Dengan $k = 0, 1, \dots, [n/2]$ dan ω_k merupakan frekuensi fourier yang dapat dihitung dengan rumusan $\omega_k = 2\pi.k / n$.

3. Hitung a_k dan b_k dengan rumusan sebagai berikut :

$$a_k = \begin{cases} \frac{1}{n} \sum_{t=1}^n z_t \cos \omega_k t, & k=0 \text{ dan } k = \frac{n}{2} \text{ jika } n \text{ genap} \\ \frac{2}{n} \sum_{t=1}^n z_t \cos \omega_k t, & k=1, 2, \dots, \frac{(n-1)}{2} \end{cases} \dots \dots \dots (3)$$

$$b_k = \frac{2}{n} \sum_{t=1}^n z_t \sin \omega_k t, \quad k = 1, 2, \dots, \frac{(n-1)}{2} \dots \dots \dots (4)$$

4. Hitung nilai ordinat $I(\omega_k)$ dengan rumusan sebagai berikut :

$$I(\omega_k) = \begin{cases} na_0^2 & k=0 \\ \frac{n}{2}(a_k^2 + b_k^2) & k=1, 2, \dots, \frac{(n-1)}{2}, \\ na_k^2 & \text{dan } k = \frac{n}{2} \text{ ketika } n \text{ genap} \end{cases} \dots \dots \dots (5)$$

5. Uji keberartian terhadap masing – masing frekuensi fourier dengan hipotesis statistik sebagai berikut :

$H_0 : \alpha = \beta = 0$ (data tidak dipengaruhi faktor musiman)

$H_1 : \alpha \neq 0$ atau $\beta \neq 0$ (data dipengaruhi faktor musiman)

Diuji dengan menggunakan statistik uji:

$$T = \frac{I^{(1)}(\omega_{(1)})}{\sum_{k=1}^{[n/2]} I(\omega_k)} \dots \dots \dots (6)$$

dengan,

$I^{(1)}(\omega_{(1)})$: ordinat maksimum dari periodogram pada frekuensi fourier

$I(\omega_k)$: nilai ordinat periodogram pada fekuensi fourier ke – k.

Kriteria uji:

Tolak H_0 jika T-hitung $>g_\alpha$ dengan $\alpha =$ taraf signifikansi. Nilai g_α dapat dilihat dalam tabel 1.

Tabel 1. The critical values for $\alpha = 0,05$ for the ratio of the largest periodogram ordinate to the sum

N*	g_α (by exact formula)	g_α (by first term only)
5	.68377	.68377
10	.44495	.44495
15	.33462	.33643
20	.27040	.27046
25	.22805	.22813
30	.19784	.19794
35	.17513	.17525
40	.15738	.15752
45	.14310	.14324
50	.13135	.13149

*N = (n - 1) / 2 if n is odd and N = (n/2 - 1) if n is even

Diuji dengan statistik uji:

$$F = \frac{(n-3)(a_k^2 + b_k^2)}{2 \sum_{\substack{j=1 \\ j \neq k}}^{[n/2]} (a_j^2 + b_j^2)} \dots \dots \dots (7)$$

Dengan $j = 1, 2, \dots, (n-1)/2$ dan $k=n/2$.

Kriteria uji:

Tolak H_0 jika F-hitung $>F_{\text{tabel}}(2, n-3; \alpha)$ dengan $\alpha =$ taraf signifikansi.

6. Buat tabel periodogram yang diperkenalkan oleh Schuster (1898) untuk mencari komponen periodik (musiman) dalam suatu deret, dengan bentuk tabel sebagai berikut:

Tabel 2 Analisis Periodogram

k	Freq(ω_k)	Periode (P)	$I(\omega_k)$	F
1	ω_1	P_1	$I(\omega_1)$	f_1
2	ω_2	P_2	$I(\omega_2)$	f_2
:	:	:	:	:
:	:	:	:	:
n/2	$\omega_{n/2}$	$P_{n/2}$	$I(\omega_{n/2})$	$f_{n/2}$

$$\text{dengan } P = \frac{2\pi}{\omega_k} .$$

7. Berdasarkan tabel periodogram, dapat diketahui hasil pengujian yang signifikan dengan cara membandingkan nilai F hitung dengan F tabel yang dapat dilihat pada tabel distribusi F dengan diketahui $v_1=2$, $v_2=n-3$ dan α = taraf signifikansi. Jika H_0 signifikan berarti mengindikasikan bahwa terdapat pola musiman. Pengujian terhadap masing – masing frekuensi fourier ini hanya memberitahukan ada tidaknya pola musiman pada suatu deret namun tidak memberikan informasi dimana letak pola tersebut.
8. Pengujian untuk melihat dimana letak pola musiman tadi digunakan perumusan statistik sebagai berikut :

$H_0 : \alpha = \beta = 0$ (data tidak dipengaruhi faktor musiman)

$H_1 : \alpha \neq 0$ atau $\beta \neq 0$ (dipengaruhi faktor musiman)

Dan diuji dengan statistik uji yang dapat dilihat pada Persamaan (2) dimana nilai $I^{(1)}(\omega_{(1)})$ diperoleh dari rumusan sebagai berikut:

$$I^{(1)}(\omega_{(1)}) = \max \{ I(\omega_k) \} .$$

.....(8)

Kriteria uji bandingkan T hitung dari Persamaan(2) dengan g_α yang dapat dilihat pada tabel 1.

Nilai g_α pada tabel terdirias dua bagian yaitu :

- a. g_α by exact formula didapat dengan rumusan sebagai berikut:

$$P(T > g) = \sum_{j=1}^m (-1)^{(j-1)} \binom{N}{j} (1 - jg)^{N-1} ,$$

.....(9)

dengan $g > 0$ dan m merupakan integer terbesar kurang dari $1/g$. Jadi untuk setiap α tingkat signifikansi tertentu, dapat digunakan Persamaan (9) untuk menemukan nilai kritis g_α seperti berikut :

$$P(T > g_\alpha) = \alpha .$$

.....(10)

- b. Sedangkan untuk nilai g_α by first term only diperoleh dari pendekatan pada Persamaan (9) sehingga didapat :

$$P(T > g) \cong N(1 - g)^{N-1} .$$

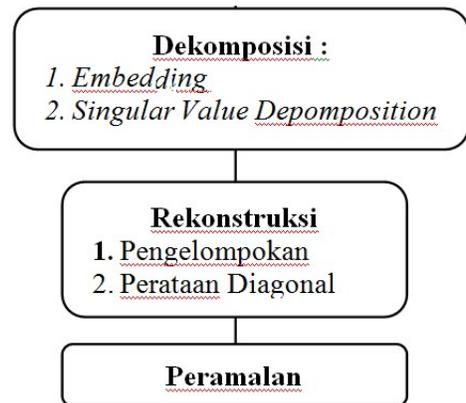
.....(11)

Hasil g_α by exact result dan by first term only nyatanya hampir sama atau tidak berbeda jauh. Bahkan untuk besar sampel yang kecil, hasilnya sama. Sehingga by first term only disarankan untuk sampel yang lebih besar dan juga alasan kepraktisan (Wei, 2006). Setelah membandingkan nilai T hitung dengan g_α , jika hasilnya signifikan (H_0 ditolak) maka dapat disimpulkan bahwa deret data mengandung komponen musiman.

Singular Spectrum Analysis(SSA)

Singular Spectrum Analysis yang disingkat dengan SSA adalah sebuah metode non-parametrik yang modern untuk analisis time series dan gambar digital. Package ini menyediakan serangkaian implementasi yang sepat dan dapat dipercaya atau akurat untuk menghitung varian yang ditunjukkan dari sebuah dekomposisi, rekonstruksi dan peramalan. Penggunaan SSA otomatis bisa dengan software Caterpillar-SSA.

Tahapan SSA :



Proses forecasting dengan metode SSA dilakukan dengan menggunakan bantuan software Caterfilar, dengan langkah sebagai berikut :

- 1) Klik **File**, kemudian **Open**
- 2) Masukan data
- 3) Pilih/tentukan banyak variabel, kemudian **Next** dan **Finish**.

Pada tahapan ini akan muncul plot data Times Series yang membantu dalam identifikasi data.

- 4) Masukkan data *Window Length*, dan klik OK dan kemudian muncul grafik "Moving Average"s and Standards dan grafik "Average Covariations".
- 5) Kemudian *Next*, dan muncul *eigen function* yang merupakan dekomposisi pola data deret waktu.
- 6) Langkah selanjutnya lakukan *grouping parameters*, untuk dilakukan *Reconstruction parameters*. Output dari proses ini diperoleh grafik *initial & Reconstructed series* dan grafik *Residuals*.
- 7) Kemudian dilakukan *Approximation Parameters*.
- 8) Lakukan *Forecasting parameters*, dengan menentukan titik awal peramalan dan jumlah titik peramalan, dengan menggunakan salah satu metode yang tersedia (Vector, Recurrent dan Recurrent (custom)), dan memilih salah satu dari *base series of forecast* (Reconstructed, Initial dan approximating).
- 9) Hasil akhir diperoleh data forecasting berikut grafik data hasil forecast.
- 10) Lakukan kalibrasi dan hitung MAPE.

Kriteria Seleksi Model

Untuk memilih model time series yang terbaik maka digunakan beberapa kriteria pemilihan model. Perbandingan ukuran kebaikan model dilakukan berdasarkan kriteria in sample MSE (Mean Square Error) dan AIC (Akaike Information Criterion) serta out of sample MSE, MAD (Mean Absolute Deviation), dan MAPE untuk kesalahan peramalan.

a. *Mean Square Error* (MSE). MSE untuk model dihitung dari

$$MSE_{insample} = \frac{SSE}{n - n_p}$$

Dimana

$SSE = \sum_{i=1}^n e_i^2$ Adalah dugaan dari residual $e_i = X_i - \hat{X}_i$

n = banyaknya residual

n_p = banyaknya parameter yang diduga

Sedangkan MSE untuk ramalan *out of sampel* dihitung dari

$MSE_{outsample} = \frac{1}{M} \sum_{i=1}^M e_i^2$ dimana M adalah panjang ramalan (Wei, 1990)

b. *Akaike's Information Criterion*

Akaike pada tahun 1973 memperkenalkan suatu pemilihan model terbaik dengan

mempertimbangkan banyaknya parameter yang digunakan dalam model. Metode ini muncul karena menganggap bahwa model yang baik tidak cukup hanya dengan menggunakan nilai MSE terkecil, namun juga harus memenuhi prinsip parsimony. Penghitungan nilai AIC adalah sebagai berikut

$$AIC = n \ln \hat{\sigma}^2 + 2M$$

dimana M adalah banyaknya parameter yang diduga (Wei, 1990)

c. *Mean Absolute Deviation* (MAD) Salah satu ukuran untuk membandingkan akurasi ramalan adalah MAD yang penghitungannya adalah sebagai berikut :

$$MAD = \frac{\sum_{i=1}^n |e_i|}{n}$$

n = panjang ramalan (Wei, 1990)

d. MAPE adalah suatu ukuran yang menggambarkan baik tidaknya data hasil forecasting. Hasil forecasting dikatakan baik apabila mempunyai nilai MAPE maksimum 15 %, semakin kecil nilai MAPE semakin baik. Berikut rumus perhitungan MAPE :

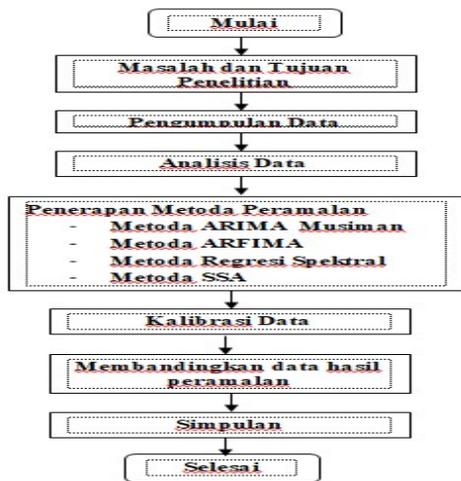
$$MAPE = \frac{\sum_{i=1}^n \left| \frac{(y_i - \hat{y}_i)}{y_i} \right|}{n} \cdot 100 \%$$

METODE PENELITIAN

Penelitian dilakukan dengan mengambil data sekunder kasus DBD di Rumah Sakit Dr. Hasan Sadikin Bandung untuk periode Januari 2006 sampai dengan Desember 2010 yang diamati secara runtut (berkala) setiap bulan. Variabel dependen dalam penelitian ini adalah kasus DBD (Y), sedangkan variabel independennya adalah waktu (X). Analisis data dilakukan secara statistik dengan bantuan program komputer.

Alur Proses Kajian

Proses peramalan dilakukan secara sistematis, dengan alur proses sebagai berikut, Gambar 3.1.



Gambar 3.1. Alur Pikir Forecasting

Analisis data deret waktu untuk forecasting data kasus DBD dilakukan dengan menggunakan metode ARIMA musiman, Regresi Spektral, ARFIMA, dan SSA. Berikut ini disajikan secara ringkas langkah-langkah praktis dalam proses Forecasting.

Metode ARIMA musiman

Langkah-langkah perhitungan untuk pembangkitan datakasus DBD menggunakan metode ARIMA adalah sebagai berikut:

1. Melakukan pengecekan kenormalan data dari data kasus DBD tahunan yang diketahui.
2. Mentransformasi data yang tidak terdistribusi normal dengan menggunakan metode Probability Plot of Correlation Coefficient (PPCC).
3. Melakukan analisa statistik dengan menghitung reratanya, standar deviasi, koefisien skewnes.
4. Membuat grafik korelogram autokorelasi dan parsial autokorelasi
5. Menyusun bilangan random dalam distribusi normal menggunakan rumus Box and Muller.
6. Menghitung pemodelan stokastik dengan metode ARIMA dengan rumus sebagai berikut:

$$Y_t = \bar{Q} + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \sigma_\epsilon \cdot \xi_t$$

Metode ARFIMA

Sesuai dengan tujuan penelitian, maka tahapan penelitian dapat dijelaskan secara garis besar sebagai berikut :

a. Tahap Identifikasi

Cara yang digunakan untuk mengetahui pola dalam data adalah :

- Plot *time series*.
- Plot ACF dan PACF
- Transformasi Box-Cox untuk data non stasioner varians. (Wei, 1990)

b. Pendugaan Parameter dan Pemeriksaan Diagnostik

Nilai-nilai dari parameter yang ada diperoleh dengan menentukan beberapa model dugaan sementara (dapat lebih dari satu) untuk ARFIMA dengan metode EML, kemudian menduga parameter model dengan dua cara, antara lain :

- Coba-coba, yaitu menguji beberapa nilai yang berbeda kemudian memilih satu nilai tertentu.
- Perbaiki secara iteratif yaitu memilih penduga awal dan membiarkan program komputer

memperhalus dugaan secara iteratif.

Setelah dilakukan pendugaan parameter, kemudian dilakukan pemeriksaan diagnostic yang meliputi uji signifikan parameter, uji white noise residual dan kenormalan residual untuk membuktikan bahwa model tersebut cukup memadai dan apabila tidak maka dicari nilai dugaan yang lain.

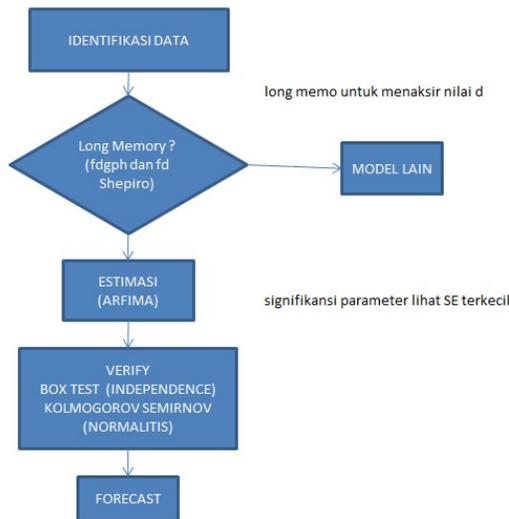
c. Penerapan

Model yang memadai digunakan untuk melakukan paramalan *out of sample* dalam konteks potensi jumlah kasus DBD. Ramalan *out of sample* dilakukan untuk t tahap waktu ke depan

d. Pemilihan Model terbaik

Untuk memilih model hidrologi time series yang terbaik, digunakan beberapa kriteria pemilihan model seperti *in sample* MSE dan AIC terkecil serta *out of sample* MSE dan MAD terkecil untuk kesalahan peramalan, sehingga diperoleh model peramalan jumlah kasus DBD yang cukup handal.

Tahapan proses forecasting dengan metode ARFIMA secara garis besar disajikan dalam alur pikir berikut ini :



Gambar 3.2. Alur Pikir Forecastin Metode Regresi Spektral

Tahapan proses forecasting dengan metode spektral secara garis besar adalah sebagai berikut:

- 1) Tentukan rata-rata data yang telah dilakukan pembedaan musiman $L = 12$, mean $(Y_t - Y_{t-L})$
- 2) Tentukan Nilai Trend bulanan dengan persamaan mean $(Y_t - Y_{t-L})$
- 3) Tentukan nilai Konstanta level dengan persamaan :
- 4) $T_t = \hat{\alpha}_0 + \hat{\beta}_0 \cdot t$
- 5) $T_t = \text{Mean } Y_t$ dengan $t = \frac{(n+1)}{2}$
- 6) Tentukan nilai trend dari $t=1, \dots, t=T$ dengan menggunakan persamaan pada langkah 3
- 7) Tentukan nilai :
 $D_r = Y_t - T_t$
- 8) Tentukan nilai 10 trigonometri pertama dengan persamaan :
 $a_1 \cdot \text{Cos}(\omega t) + b_1 \cdot \text{Sin}(\omega t) + \dots$
 $a_2 \cdot \text{Cos}(2\omega t) + b_2 \cdot \text{Sin}(2\omega t) + \dots$
 $a_5 \cdot \text{Cos}(5\omega t) + b_5 \cdot \text{Sin}(5\omega t) + \dots$
 Tentukan nilai-nilai koefisien $a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4, a_5, b_5$ dengan menggunakan persamaan regresi multiple dengan $Y_t = \text{Defrended}$
- 9) Tentukan nilai frekuensi dan amplitude dari tiap persamaan regresi multiple
 $f_i = i \cdot \frac{n}{2} \quad ; \quad A_i = \sqrt{a_i^2 + b_i^2}$
- 10) Tentukan standar error dari residual (RSE) dari tiap regresi yang dibentuk.
- 11) Jika nilai :

$\frac{A_i}{RSE_i} \geq 0,5$ maka koefisien regresi signifikan

$\frac{A_i}{RSE_i} < 0,5$ maka koefisien regresi tidak signifikan

- 12) Forecast data selanjutnya dengan menggunakan persamaan regresi spectral yang koefisiennya signifikan.

Metode SSA

Proses forecasting dengan metode SSA dilakukan dengan menggunakan bantuan software Caterfilar, dengan langkah sebagai berikut :

1. Klik **File**, kemudian **Open**
2. Masukkan data
3. Pilih/tentukan banyak variabel, kemudian **Next** dan **Finish**. Pada tahapan ini akan muncul plot data Times Series yang membantu dalam dalam identifikasi data.
4. Masukkan data **Window Length**, dan klik **OK** dan kemudian muncul grafik "Moving Average"s and Standards dan grafik "Average Covariations".
5. Kemudian **Next**, dan muncul **eigen function** yang merupakan dekomposisi pola data deret waktu.
6. Langkah selanjutnya lakukan **grouping parameters**, untuk dilakukan **Reconstruction parameters**. Output dari proses ini diperoleh grafik **initial & Reconstructed series** dan grafik **Residuals**.
7. Kemudian dilakukan **Approximation Parameters**.
8. Lakukan **Forecasting parameters**, dengan menentukan titik awal peramalan dan jumlah titik peramalan, dengan menggunakan salah satu metode yang tersedia (Vector, Recurrent dan Recurrent (custom)), dan memilih salah satu dari **base series of forecast** (Reconstructed, Initial dan approximating).
9. Hasil akhir diperoleh data forecasting berikut grafik data hasil forecast.
10. Lakukan kalibrasi dan hitung MAPE

HASIL PENELITIAN

Penelitian dilakukan dengan mengambil data sekunder kasus DBD di Rumah Sakit Dr. Hasan Sadikin Bandung untuk periode Januari 2006 sampai dengan Desember 2010 yang diamati secara runtut (berkala) setiap

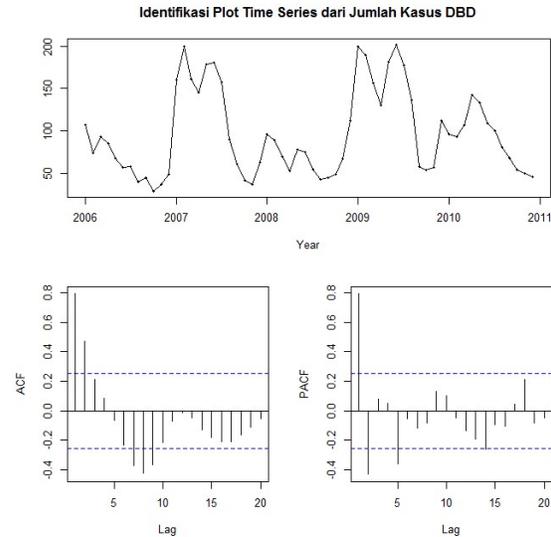
bulan, yang kemudian dilakukan analisis dengan keempat metode peramalan.

Hasil Peramalan dengan Metoda ARIMA Musiman

Tahap Identifikasi Data Model ARIMA

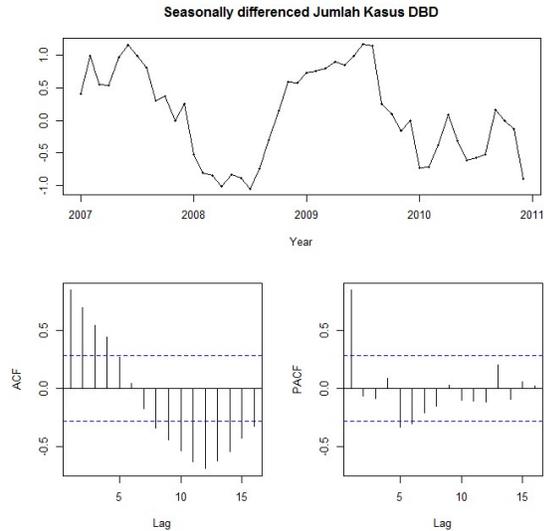
Sebelum data deret waktu dianalisis dengan menggunakan metode ARIMA, terlebih dahulu diidentifikasi apakah data sudah stasioner di dalam mean dan varians. Identifikasi dataderet waktu merupakan langkah awal yang harus dilakukan dalam proses *forecasting*, karena dapat memberi informasi mengenai pola dari data diantaranya pola stasioner, *trend effect*, *seasonal effect* dan *cyclic effect*. Identifikasi dapat dilakukan secara visual menggunakan plot data asli, plot ACF dan plot PACF.

Dari hasil analisis langkah dan strategi pembentukan model dengan metode analisis deret berkala ARIMA nampak secara visual berdasarkan hasil pada Gambar 4.1. bahwaplot data asli jumlah kasus demam berdarah di Rumah Sakit Hasan Sadikin mengikuti pola trend musiman dan belum stasioner dalam mean maupun dalam varian. Pada plot ACF berbentuk sinusoidal dan plot PACF terdapat koefisien yang signifikan (memotong garis selang kepercayaan), yaitu 1, 2 dan 5. Secara uji hipotesis dilakukan *Augmented Dickey-Fuller (ADF) Test* dan *Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test*. Dari hasil kedua uji tersebut dengan taraf nyata 5% diperoleh nilai $p = 0.405$ dan 0.1 lebih besar dari 0.05 , sehingga dapat disimpulkan bahwa data tidak stasioner. Oleh karena itu, harus dilakukan transformasi dan *differencing* untuk menghilangkan ketidakstasionerannya tersebut.



Gambar 4.1 Identifikasi plot data asli sebelum transformasi

Data *time series* yang belum stasioner dalam varians tersebut selanjutnya ditransformasi menggunakan transformasi *log* dengan bantuan software komputer dan kemudian dilakukan proses *differencing*. Untuk melakukan uji hipotesis pada kasus ini sebelum dan sesudah ditransformasi digunakan *Bartlett's test* dan *Levene's test*, kemudian dicari hasil yang terbaik dari kedua uji tersebut. Uji normalitas diperlukan sebelum dilakukan *Bartlett's test*. Dari Gambaran uji normalitas diperoleh bahwa hasil uji normalitas sebelum transformasi mempunyai nilai $p < 0,001$ yang berarti bahwa data tidak terdistribusi normal. Sedangkan pada uji normalitas setelah transformasi diperoleh nilai $p > 0,150$. Dari hasil *Bartlett's test* setelah transformasi diperoleh nilai $p > 0.05$, maka berarti menerima hipotesis nol. Jadi disimpulkan bahwa data setelah ditransformasi homogen. Hasil transformasi dan *differencing* ditunjukkan dalam plot time series (Gambar 4.2).



Gambar 4.2 Identifikasi plot data setelah proses transformasi dan differencing

Setelah data dinyatakan telah stasioner maka dilakukan estimasi parameter sementara dengan berdasarkan plot ACF dan PACF. Dari plot ACF berbentuk sinusoidal. Sedangkan pada plot PACF terdapat koefisien yang signifikan (memotong garis selang kepercayaan), yaitu 1, 5 dan 6, sehingga kemungkinan model yang cocok adalah ARIMA(1,0,0)(1,1,0)¹². Untuk lebih meyakinkan maka dari kedua fungsi tersebut dapat diestimasi beberapa model ARIMA yaitu:

- Model 1 ARIMA(1,0,0)(1,0,0)¹²
- Model 2 ARIMA(1,0,0)(1,1,0)¹²
- Model 3 ARIMA(2,0,0)(1,0,0)¹²
- Model 4 ARIMA(2,0,0)(1,1,0)¹²
- Model 5 ARIMA(2,0,0)(2,0,0)¹²
- Model 6 ARIMA(2,0,0)(2,1,0)¹²

Tahap Estimasi Parameter Model ARIMA

Langkah selanjutnya adalah melakukan estimasi parameter dari beberapa kemungkinan model ARIMA yang telah diidentifikasi berdasarkan plot ACF dan PACF. Selain itu dengan memanfaatkan fasilitas automatic ARIMA dengan perintah “auto.arima()” pada software R yang dapat memberikan model terbaik. Dalam kasus ini didapat dari hasil auto.arima() adalah sebagai berikut:

```
> fit.a <- auto.arima(Yt, lambda=0, d=0, D=1, max.order=9,
+ stepwise=FALSE, approximation=FALSE)
> fit.a
Series: Yt
ARIMA(1,0,0)(1,1,0)[12] with drift
Box Cox transformation: lambda= 0

Coefficients:
          ar1      sar1  drift
0.7658 -0.6295  0.0036
s.e.    0.1283  0.1505  0.0109

sigma^2 estimated as 0.1064: log likelihood=-8.16
AIC=24.32  AICc=25.25  BIC=31.8
```

Dari auto.arima() menghasilkan model ARIMA (1,0,0)(1,1,0)¹² with drift. Selanjutnya masing-masing dilakukan evaluasi terhadap model-model yang telah ada dengan menghitung nilai AIC, BIC dan MSE untuk mendapatkan model yang terbaik.

Tabel. 4.1. Perhitungan nilai AIC, BIC dan MSE

	Model	AI C	BI C	M SE
Mo del 1	ARIMA(1,0,0) (1,0,0) ¹²	31. 4	39. 78	53. 74
Mo del 2	ARIMA(1,0,0) (1,1,0) ¹²	27. 89	33. 5	33. 30
Mo del 3	ARIMA(2,0,0) (1,0,0) ¹²	26. 88	37. 35	56. 54
Mo del 4	ARIMA(2,0,0) (1,1,0) ¹²	29. 26	36. 75	34. 37
Mo del 5	ARIMA(2,0,0) (2,0,0) ¹²	25. 61	38. 18	50. 77
Mo del 6	ARIMA(2,0,0) (2,1,0) ¹²	31. 02	40. 37	-
Mo del 7	ARIMA(1,0,0) (1,1,0) ¹² with drift	24. 32	31. 8	27. 30

Dari hasil perhitungan nilai AIC, BIC dan MSE seperti tercantum pada tabel 4.1, didapat bahwa Model 7 yang memakai mode otomatis dari software R yaitu dengan perintah “auto.arima()” model ARIMA(1,0,0)(1,1,0)¹² with drift menghasilkan nilai AIC, BIC, dan MSE terkecil diantara model lainnya.

Tahap Uji Diagnostik Model ARIMA

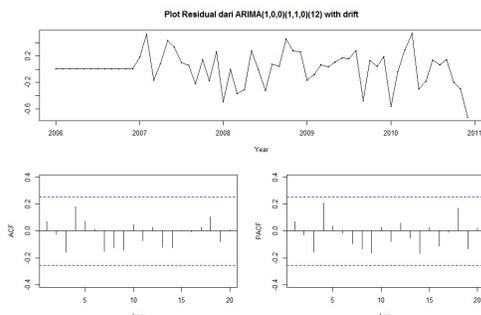
Langkah selanjutnya adalah diagnosis model. Diagnosis model dilakukan untuk mendeteksi adanya korelasi dan kenormalan

antar *residual*. Dalam runtun waktu (*time series*) ada asumsi bahwa *residual* mengikuti proses *white noise* yang berarti *residual* harus independen (tidak berkorelasi) dan berdistribusi normal dengan rata-rata mendekati 0 ($\mu = 0$) dan standar deviasi (σ) tertentu (Iriawan, 2006: 361). Untuk mendeteksi adanya proses *white noise*, maka perlu dilakukan diagnosis model dengan cara uji independensi *residual*. Uji dilakukan untuk mendeteksi independensi *residual* antar lag. Dua lag dikatakan independen (tidak berkorelasi) apabila antar lag tidak ada korelasi cukup berarti.

Dalam penelitian ini, uji dilakukan dengan membandingkan *Ljung-Box* χ^2 dan nilai Tabel $\chi^2(\alpha, df)$ pada output proses *Ljung-Box-Pierce*. Hipotesis: $H_0: \rho = 0$ (Ada korelasi antar-lag) dan $H_1: \rho \neq 0$ (Tidak ada korelasi antar lag atau minimal ada 1 lag yang $\neq 0$) Kriteria penolakan H_0 yaitu jika *Ljung-Box* $\chi^2 < \chi^2(\alpha, df)$, di mana distribusi χ^2 yang digunakan mempunyai $df = k-2$. Maka hasil *Ljung-Box-Pierce* menggunakan software R adalah sebagai berikut:

```
> Box.test(residuals(fit.a), lag=36, fitdf=6, type="Ljung")
Box-Ljung test
data: residuals(fit.a)
K-squared = 33.449, df = 30, p-value = 0.3034
```

Selain dengan pengujian hipotesis, independensi antar lag akan ditunjukkan pula oleh grafik fungsi autokorelasi (fak) residual. Suatu *residual* model dikatakan telah independen jika tidak ada satu lag pun pada grafik fungsi autokorelasi (fak) residual yang keluar batas garis seperti terlihat pada gambar 4.3.



Gambar 4.3 Grafik plot residual model fit ARIMA

Dari hasil uji hipotesis menggunakan *Ljung-Box-Pierce*, maupun dengan grafik

terbukti bahwa proses mengikuti *white noise*, sehingga tidak ada pelanggaran asumsi dari model terbaik yang dipilih.

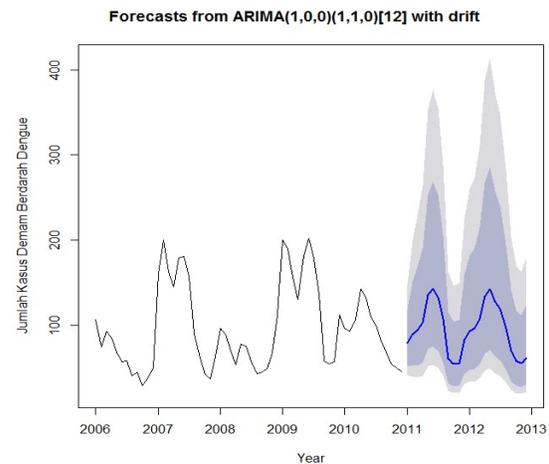
Tahap Peramalan Model ARIMA

Setelah melalui tahap identifikasi, estimasi parameter dan tahap uji diagnostik, maka model terbaik yang didapat adalah model ARIMA(1,0,0)(1,1,0)¹² with drift. Selanjutnya adalah melakukan peramalan dan hasil yang didapat dengan melakukan peramalan 24 bulan ke depan adalah sebagai berikut:

```
> pred.ar <- forecast(fit.a)
> pred.ar
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2011	78.36532	51.59043	119.0361	41.34851	148.5211
Feb 2011	89.07691	52.61301	150.8124	39.81453	139.2915
Mar 2011	94.55395	52.89803	169.0129	38.89660	229.8516
Apr 2011	103.56470	56.25023	190.6774	40.72885	263.4074
May 2011	134.97220	72.09498	252.6874	51.72939	352.1692
Jun 2011	142.05037	75.15176	268.5008	53.64968	376.1124
Jul 2011	132.94887	69.94692	252.6974	49.78739	355.0176
Aug 2011	107.46765	56.35812	204.9269	40.04639	288.3979
Sep 2011	60.48959	31.66203	115.5640	22.47559	162.7984
Oct 2011	54.18696	28.33170	103.6375	20.09976	146.0827
Nov 2011	55.34320	28.91752	105.9175	20.50834	149.3476
Dec 2011	83.08082	43.39431	159.0629	30.76911	224.3296
Jan 2012	92.69159	47.34382	181.4752	33.17461	258.9851
Feb 2012	95.94309	48.38278	190.2552	33.67422	273.3568
Mar 2012	107.71620	53.92047	215.1832	37.38214	310.3830
Apr 2012	133.68450	66.63317	268.2080	46.09083	387.7462
May 2012	141.95828	70.58019	285.5214	48.75637	413.3235
Jun 2012	127.94482	63.51979	257.7130	43.84510	373.3571
Jul 2012	118.46995	58.76549	238.8328	40.54500	346.1618
Aug 2012	96.02311	47.60712	193.6777	32.83760	280.7890
Sep 2012	69.59009	34.49179	140.4039	23.78743	203.8557
Oct 2012	57.83333	28.65970	116.7037	19.76350	169.2359
Nov 2012	55.56602	27.53333	112.1398	18.98575	162.6263
Dec 2012	61.31803	30.38169	123.7555	20.94919	179.4771

Dari hasil peramalan 24 bulan ke depan kemudian dibuat grafik plot dari hasil peramalan berdasarkan model ARIMA terbaik, seperti disajikan pada gambar 4.4.



Gambar 4.4 Grafik plot peramalan

Hasil Peramalan dengan Metode ARFIMA

Setelah melakukan forecasting dengan metode ARIMA musiman, adakalanya, plot ACF dan PACF menunjukkan pola *long memory*, ini terlihat dari nilai-nilai autokorelasi pada plot ACF atau PACF turun secara lambat untuk lag yang semakin

meningkat. Identifikasi ini mengindikasikan bahwa nilai dari d (koefisien pembeda, *differencing*) bernilai pecahan, sehingga model yang paling cocok adalah Model ARFIMA (*Autoregressive Fractionally Integrated Moving Average*). Proses forecasting menggunakan metode arfima seperti halnya menggunakan metode lainnya diawali dengan identifikasi, kemudian estimasi parameter model, uji diagnostik dan peramalan.

Tahap Identifikasi Data model ARFIMA

Pada tahap identifikasi data apakah data termasuk long memory atau tidak digunakan software dengan melihat nilai fdGPH dan fdSperio sebagai berikut:

```
> library(fracdiff)
> fdGPH(Yt,bandw.exp=0.5) $d
[1] 0.6180213
> fdSperio(Yt,bandw.exp=0.5,beta=0.9) $d
[1] 0.2132963
```

Hasil identifikasi menunjukkan bahwa nilai fdGPH dan fdSperio berada sekitar 0,5 dengan demikian bisa dicobakan menggunakan metode ARFIMA dalam proses forecasting.

Tahap Estimasi Parameter Model ARFIMA

Berikut hasil estimasi parameter model menggunakan metode ARFIMA dengan bantuan Software R.

```
> fit2 <- arfima(Yt, order = c(1, 1, 0))
Note: autowed is ON. It is possible, but not likely,
that unique modes may be lost.
Beginning the fits with 4 starting values.
> fit2
Number of modes: 1
Call:
arfima(z = Yt, order = c(1, 1, 0))
Coefficients for fits:
      Coef.1:      SE.1:
phi(1)  0.576147  0.182149
d.f     -0.466429  0.261131
Fitted mean -0.492013  1.97063
logL    -199.532
sigmaC2  886.803
Starred fits are close to invertibility/stationarity boundaries
> AIC(fit2)
[1] 407.063
```

Dari hasil analisis diatas didapat bahwa ARFIMA (1,1,0) memiliki nilai AIC lebih baik dibandingkan ARFIMA (1,0,0).

Tahap Uji Diagnostik Model ARFIMA

```
> Box.test(resid$Model)
Box-Pierce test
data: resid$Model
X-squared = 0.0013, df = 1, p-value = 0.9714
```

Dari hasil uji hipotesis menggunakan *Ljung-Box-Pierce* terbukti bahwa proses mengikuti *white noise*, sehingga tidak ada pelanggaran asumsi dari model terbaik yang dipilih.

Tahap Peramalan Model ARFIMA

Berikut hasil forecast jumlah kasus demam berdarah dengue dengan metode ARFIMA dengan bantuan Software “R”.

```
> predict(fit2, n.ahead = 24)
> predict(fit2)
Mode 1
Mode 1's Forecasts and SDs
Forecasts  1  2  3  4  5  6
Exact SD   29.8210 44.6376 54.6274 61.6808 66.8705 70.8412
Limiting SD 29.7792 44.4846 54.3486 61.2655 66.3153 70.1466
Forecasts  7  8  9 10 11 12
Exact SD   73.9900 76.5669 78.7338 80.5975 82.2311 83.6893
Limiting SD 73.1884 75.6014 77.6377 79.3741 80.8825 82.2162
Forecasts 13 14 15 16 17 18
Exact SD   84.9963 86.1909 87.2889 88.3055 89.2528 90.1400
Limiting SD 83.4086 84.4869 85.4710 86.3759 87.2124 87.9929
Forecasts 19 20 21 22 23 24
Exact SD   90.9748 91.7634 92.5110 93.2218 93.8996 94.5475
Limiting SD 88.7218 89.4063 90.0514 90.6615 91.2400 91.7901
Mode 1's Bootstrap Replicates
[1] 1000
Mode 1's Bootstrap Predictions and Intervals
Upper 95% 129.57848 151.8107 163.6208 187.5390 186.7292
Prediction (Mean) 45.13704 43.7043 42.5987 41.1092 41.1494
Lower 95% 7.06817 -32.3265 -55.5124 -68.0766 -75.4855
Upper 95% 204.0101 158.8653 212.5371 211.9421
Prediction (Mean) 40.3340 40.5550 39.3711 37.9797
Lower 95% -91.3868 -97.4279 -101.1753 -101.2666
Upper 95% 214.3253 214.2771 218.8097 224.3797
Prediction (Mean) 37.2351 35.8321 35.3402 36.2685
Lower 95% -112.1285 -123.2928 -136.5701 -132.5087
Upper 95% 216.7511 234.250 244.6623 255.9301
Prediction (Mean) 36.7721 37.546 36.3563 36.3035
Lower 95% -132.0802 -138.664 -137.3342 -143.9671
Upper 95% 250.653 250.6287 263.4141 246.4799
Prediction (Mean) 35.679 35.0713 35.5271 35.7653
Lower 95% -157.605 -164.9844 -159.2623 -163.4006
Upper 95% 252.4852 254.3305 252.9369
Prediction (Mean) 36.0737 37.4872 37.1597
Lower 95% -157.3288 -159.7604 -158.9099
```

Hasil Peramalan dengan Metode Regresi Spektral

Pada peramalan berdasarkan model regresi, yang ditelaah adalah karakteristik dari komponen trend.Sedangkan telaahan komponen siklis, sulit untuk dilakukan.Padahal dengan diketahui karakteristik komponen siklis, dapat dianalisis kondisi ekstrim dri data. Telaahan komponen siklis dapat dilakukan dengan analisis spektral, yang merupakan pengembangan dari analisis deret Fourier. Dari tahapan proses forecasting menggunakan metode spectral untuk jumlah kasus DBD di Rumah Sakit

Hasan Sadikin memberikan hasil identifikasi awal bahwa data merupakan data musiman. Dari hasil statistik pengujian keberartian periodesitas dengan bantuan software R, diperoleh periodesitas tersembunyi yang signifikan adalah 12 bulanan.

Persamaan regresi yang didapat dari hasil perhitungan pada data deret waktu tersebut mempunyai orde 2.

$$Z_t = \alpha_0 - \alpha_1 \cos \omega t + \beta_1 \sin \omega t + \alpha_2 \cos \omega t + \beta_2 \sin \omega t + e_t$$

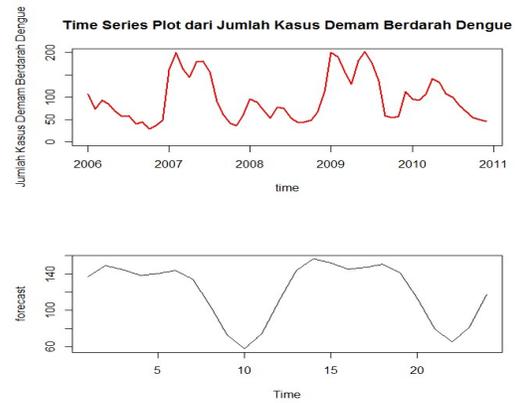
$$Z_t = 78.6 - 17.6 \cos \omega t + 37.96 \sin \omega t + 9.01 \cos \omega t + 17.38 \sin \omega t + e_t$$

dengan

- Z_t : series data deret waktu pada periode t
- α dan β : parameter (koefisien fourier)
- $\cos \omega t$ dan $\sin \omega t$: fungsi kontinu yang tidak berkorelasi
- e_t : eror pada periode waktu ke-t
- ωt : frekuensi fourier = 0.523
- t : periode waktu = 12

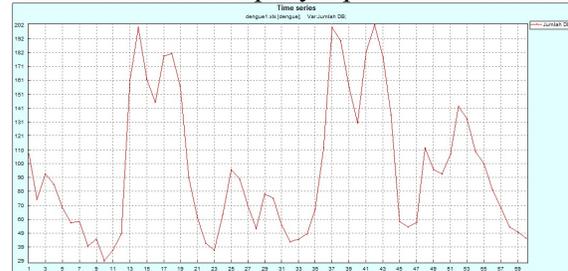
Hasil forecasting menggunakan metode Regresi Spektral memberi hasil sebagai berikut :

```
> forecast
[1] 136.84176 149.44557 144.68482
138.26157 140.16485 143.89125
134.33063
[8] 106.06700 72.95267 58.27584
74.27244 110.70768 143.71953
156.45856
[15] 151.76008 145.30204 147.15778
150.91906 141.47400 113.30079
80.14499
[22] 65.29818 81.11660 117.49576
```

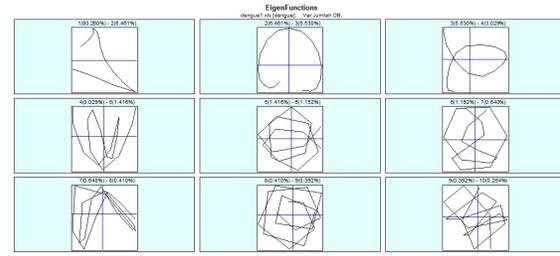


Hasil Peramalan dengan Metode SSA

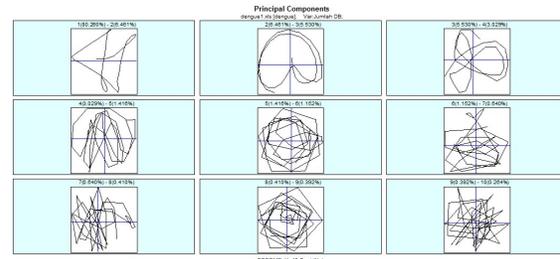
Berikut hasil perhitungan forecasting data jumlah kasus DBD di Rumah Sakit Hasan Sadikin menggunakan metode SSA, dengan menggunakan software CaterpillarSSA dengan langkah-langkah seperti telah dijelaskan pada Bab 3. Gambar 4.5. menunjukkan plot data asli, nampak jelas bahwa data ini mempunyai pola musiman.



Gambar 4.5. Plot data kasus DBD

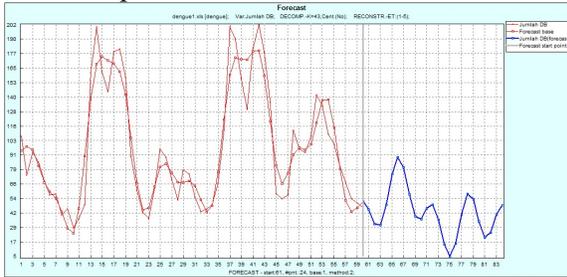


Gambar 4.6. Decomposisi dari pola data kasus DBD



Gambar 4.7. komponen dari pola data kasus DBD

Gambar 4.6 dan Gambar 4.7, menunjukkan komponen dan dekomposisi pola data kasus DBD. Berdasarkan gambar tersebut nampak bahwa komponen pola jenis 1, 2 dan 3, memiliki pola tertentu.



Gambar 4.8. Plot data hasil peramalan jumlah kasus DBD

Gambar 4.8 merupakan plot data asli dan hasil forecast dengan Metoda SSA pada *Window Length* sebesar 18 untuk 24 bulan kedepan. Terlihat bahwa jumlah kasus DBD ada kecenderungan menurun pada 24 bulan kedepan.

Perbandingan hasil peramalan antara keempat metode

Untuk mencari model terbaik dalam meramalkan jumlah kasus DBD 24 bulan kedepan, yaitu dengan cara membandingkan nilai MAPE, MAD, MSD. Berikut hasil perbandingan nilai MAPE, MAD, MSD dari keempat model berdasarkan keempat metode analisis peramalan yang telah dilakukan.

Tabel 4.1 Perbandingan Nilai MAPE, MAD dan MSD

Model	Insample			Outsample		
	MAPE	MAD	MSD	MAPE	MAD	MSD
ARIMA (1,1,0) ¹²	1.12	46.2	2470	9.02	84.4	7874
ARFIMA (1,1,0)	0.70	24.7	824	8.78	71.2	5206
Spektral	1.73	70.1	5609	11.4	109.8	12853
SSA	0.40	17.6	491	3.90	32.4	1407

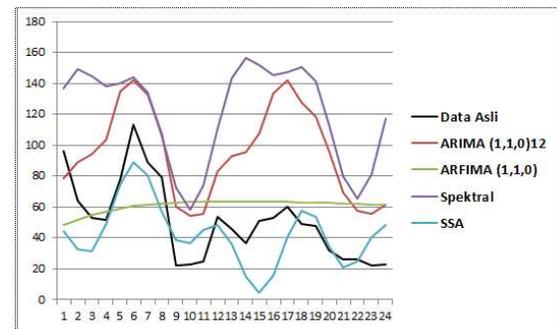
Berdasarkan Tabel. 4.1 menunjukkan bahwa peramalan menggunakan metode SSA menghasilkan nilai MAPE, MAD dan MSD terkecil baik pada insample maupun outsample. Hal membuktikan bahwa metode SSA merupakan metode peramalan yang terbaik dalam membuat peramalan jumlah kasus DBD di Rumah Sakit Hasan Sadikin Bandung.

Tabel 4.2 Perbandingan data insample dan hasil peramalan dari keempat metode

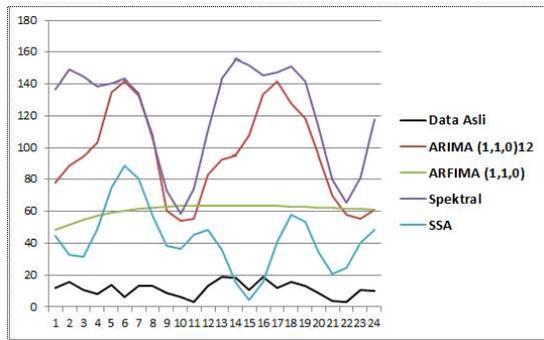
Data Asli	ARIMA (1,1,0) ¹²	ARFIMA (1,1,0)	Spektral	SSA
96	78.36532	48.5008	136.8418	44.4
64	89.07691	51.8191	149.4456	32.5
53	94.55395	54.8322	144.6848	31.3
52	103.5647	57.2886	138.2616	48.9
78	134.9722	59.1966	140.1649	74.9
113	142.05037	60.6386	143.8913	89.1
89	132.94887	61.7065	134.3306	80.6
79	107.46765	62.4811	106.067	57.3
22	60.48959	63.0278	72.95267	38.6
23	54.18696	63.3972	58.27584	36.5
25	55.3432	63.6279	74.27244	45.4
54	83.08082	63.7491	110.7077	48.6
46	92.69159	63.7829	143.7195	35.9
37	95.94309	63.7461	156.4586	14.9
51	107.7162	63.6517	151.7601	4.6
53	133.6845	63.5098	145.302	15.9
60	141.95828	63.328	147.1578	40.8
49	127.94482	63.1128	150.9191	57.8
48	118.46995	62.8691	141.474	53.6
32	96.02311	62.6009	113.3008	34.7
26	69.59009	62.3116	80.14499	20.7
26	57.83333	62.0039	65.29818	24.8
22	55.56602	61.6801	81.1166	40.6
23	61.31803	61.3421	117.4958	48.5

Tabel 4.3 Perbandingan data Outsample dan hasil peramalan dari keempat metode

Data Asli	ARIMA (1,1,0) ¹²	ARFIMA (1,d,0)	Spektral	SSA
12	78.36532	48.5008	136.8418	44.4
16	89.07691	51.8191	149.4456	32.5
11	94.55395	54.8322	144.6848	31.3
8	103.5647	57.2886	138.2616	48.9
14	134.9722	59.1966	140.1649	74.9
6	142.05037	60.6386	143.8913	89.1
13	132.94887	61.7065	134.3306	80.6
13	107.46765	62.4811	106.067	57.3
9	60.48959	63.0278	72.95267	38.6
6	54.18696	63.3972	58.27584	36.5
3	55.3432	63.6279	74.27244	45.4
13	83.08082	63.7491	110.7077	48.6
19	92.69159	63.7829	143.7195	35.9
18	95.94309	63.7461	156.4586	14.9
11	107.7162	63.6517	151.7601	4.6
19	133.6845	63.5098	145.302	15.9
12	141.95828	63.328	147.1578	40.8
16	127.94482	63.1128	150.9191	57.8
13	118.46995	62.8691	141.474	53.6
9	96.02311	62.6009	113.3008	34.7
4	69.59009	62.3116	80.14499	20.7
3	57.83333	62.0039	65.29818	24.8
11	55.56602	61.6801	81.1166	40.6
10	61.31803	61.3421	117.4958	48.5



Gambar 4.9. Plot data insample dan hasil peramalan jumlah kasus DBD dari keempat metode peramalan



Gambar 4.10. Plot data Outsample dan hasil peramalan jumlah kasus DBD dari keempat metode peramalan

Dari hasil plot data insample maupun outsample dengan hasil peramalan dari keempat metode yang telah dilakukan, menunjukkan bahwa metode peramalan dengan SSA memberikan hasil yang paling mendekati, hal ini terlihat pada Gambar 4.9 dan 4.10.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang dilakukan pada bab 4, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Identifikasi data menunjukkan bahwa data jumlah kasus DBD mengikuti pola musiman dengan periode 12.
2. Hasil peramalan menggunakan ARIMA menghasilkan model ARIMA $(1,0,0)$ $(1,1,0)^{12}$ dengan nilai MAPE sebesar 1.12%, kemudian ARFIMA $(1,1,0)$ dengan nilai MAPE sebesar 0.7, metode spectral sebesar 1.73 dan yang terkecil metode SSA sebesar 0.4
3. Hasil forecasting data jumlah kasus DBD menggunakan keempat metodetersebut memberikan hasil yang cukup baik dengan nilai MAPE lebih kecil dari nilai maksimal MAPE yang masih diperbolehkan yaitu sebesar 15 %.
4. Setelah melakukan peramalan dengan metode ARIMA, ARFIMA, Spektral, dan SSA maka dapat disimpulkan bahwa Metode SSA adalah metode terbaik untuk menggambarkan dan meramalkan pola atau fluktuasi jumlah kasus DBD dibandingkan dengan metode yang lainnya.

DAFTAR PUSTAKA

- Beran, J. (1994), "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models", *Journal of the Royal Statistical Society*, Vol. 57, hal.659-672.
- Box, G.E.P., & Jenkins, G.M. 1976. *Time Series Analysis, Forecasting and Control*. HoldenDay:San Francisco.
- Cryer, J.D. (1986), "Time Series Analysis", PWS-KENT Publishing Company, Boston, USA.
- Darmawan, G. *Pengujian Pola Musiman Pada Data Deret Waktu Dengan Menggunakan Regresi Spektral*. UNPAD. Bandung 2012.
- Depkes RI. 2006. *Pemberantasan Sarang Nyamuk Demam Berdarah Dengue (PSN DBD) oleh Juru Pemantau Jentik (Jumantik)*. DitjenPP & PL: Jakarta.
- Depkes RI. 2005. *Pencegahan dan Pemberantasan Demam Berdarah Dengue di Indonesia*. DitjenPP & PL: Jakarta.
- Depkes RI. 2004. *Kajian Masalah Kesehatan Demam Berdarah Dengue*, Jakarta. www.litbang.depkes.go.id, (sitasi 11 Desember 2009).
- Granger, C. W. J. dan Joyeux, R. (1980), "An Introduction to Long-Memory Time Series Models and Fractional Differencing", *Journal of Time Series Analysis*, Vol. 1, hal. 15-29.
- Hosking, J.R.M. (1981), "Fractional Differencing", *Biometrika*, Vol. 68, hal. 165-176.
- Ispriyanti, D. 2004. *Pemodelan Statistika dengan Transformasi Box Cox*, *Jurnal Matematik dan Komputer*, Vol. 7. No. 3, 8-17.
- Makridakis, S., Wheelwright, S. & Victor, E.M. edition, Erlangga: Jakarta. 1999. *Metode dan Aplikasi Peramalan*, 2
- Makridakis, S., Syeven C.W & Victor, E.M. 1995. *Metode dan Aplikasi Peramalan*, Terjemahan Hari Suminto, Binarupa Aksara: Jakarta.
- Sowell, F. (1992), "Maximum Likelihood Estimation of Stationary Univariate

Fractionally Integrated Time Series Models”, *Journal of econometrics*, Vol.53, hal.165 – 188.

Terasvirta, T., Lin, C.F & Granger, C.W.J. 1993. Power of The Neural Network Linearity Test, *Journal of Time Series Analysis*, Vol. 14, p.159–171.

White, H. 1989. An additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks. In *Proceedings of The International Joint Conference on Neural Networks*, pp. 451-455, San Diego, CA: SOS Printing, Washington, DC.