ESTIMASI DATA YANG HILANG DENGAN MENGGUNAKAN PROSES PENYARINGAN DALAM PEMODELAN DATA TIME SERIES

Rais[1]

[1]Jurusan Matematika FMIPA Universitas Tadulako, email: rais76_untad@yahoo.co.id

Abstrak

Makalah ini mengusulkan sebuah metode baru untuk memperkirakan data yang hilang dengan menggunakan proses penyaringan. Kami menggunakan data asli dan data yang hilang secara acak untuk mengevaluasi metode estimasi baru dengan menggunakan teknik pemodelan Box-Jenkins untuk memprediksi rata-rata curah hujan bulanan untuk Kota Palu. Data curah hujan dikumpulkan dari 1 Oktober 1973 sampai 31 Mei 2011 di Stasiun Badan Meteorologi Kota Palu. Data yang digunakan dalam pengembangan model untuk memprediksi curah hujan ditunjukkan oleh model autoregressive integrated moving average (ARIMA). Model untuk kedua kumpulan data adalah ARIMA(1,1,0)(0,1,1)$_{12}$. Hasil peramalan diperiksa dengan uji sesungguhnnya, dengan menggunakan statistik Thiel's dan diperoleh U = 0.895766 untuk data asli dan U = 0.726352 untuk data yang hilang, ini menunjukkan bahwa keduanya adalah model yang terbaik.

Kata kunci: Model ARIMA, rata-rata curah hujan bulanan, proses penyaringan dan metode peramalan.

I. Pendahuluan

Time series adalah serangkaian pengamatan tercatat dalam suatu waktu. Model Autoregressive Integrated Moving Average (ARIMA) sangat cocok untuk peramalan jangka pendek karena model ARIMA disusun dengan logis dan secara statistik akurat, memasukkan banyak informasi dari data historis, serta menghasilkan kenaikan akurasi peramalan dan pada waktu yang sama menjaga jumlah parameter seminimal mungkin (Jarret, 1991: 317). Sebuah data runtun waktu diamati secara teoritis terdiri dari dua bagian yaitu bagian pertama adalah data runtun waktu yang dihasilkan oleh proses umum dan data runtun waktu kedua yang merupakan hasil dari gangguan luar. Penghapusan kebisingan/gangguan adalah merupakan tujuan utama dari analisis time series. Dimana awal perkembangannya tujuannya untuk menghilangkan gangguan yang timbul dengan pendekatan autoregresif dan moving average teritegrasi atau pendekatan (ARIMA).

Prosedur Box-Jenkins terdiri dari pelaksanaan atau penyelesaian dari beberapa langkah, atau tahap: Rumuskan model umum dan uji stasioner data, identifikasi model tentatif, estimasi parameter atas model tentatif, uji diagnostik apakah model sudah tepat jika ya maka gunakan model untuk peramalan dan jika tidak maka kembali ke identifikasi model tentatif. Dalam menguji kestasioneran suatu data dan menentukan model tentatif Box Jenkins melakukan dengan menganalisis fungsi autokorelasi (ACF) dan fungsi autokorelasi parsial (PACF) (2, 5). Estimasi untuk konstan dan parameter dari persamaan tersebut harus diperoleh. Metode ini menggunakan pendekatan iteratif yang mengidentifikasi kemungkinan model yang bermanfaat. Model terpilih, kemudian, dicek kembali dengan data historis apakah telah mendeskripsikan data tersebut dengan tepat. Model “terbaik” akan diperoleh apabila residual antara model peramalan dan data historis memiliki nilai yang kecil, distribusinya random, dan independen (Hanke & Reitsch, 1998: 408). Tujuan utama dari penelitian ini adalah untuk menganalisis data yang dikumpulkan secara otomatis dan untuk mengevaluasi model prediktif dan kemudian menghasilkan seperangkat perkiraan untuk data stasiun metereologi di mana data yang dikumpulkan. Menurut Pankratz, di ruang kerjanya, metode Box-Jenkins menghasilkan ramalan yang terbaik untuk 74% dari data runtun waktu yang ia dievaluasi (4). Biaya yang terkait dengan pendekatan Box-Jenkins memberikan situasi tertentu umumnya lebih besar daripada banyak metode kuantitatif lainnya. Model Box-Jenkins adalah cara yang paling umum untuk mendekati peramalan dan tidak seperti model lainnya, tidak memerlukan asumsi, awalnya pola tetap dan tidak terbatas pada jenis pola tertentu.

Model ini dapat dipasang untuk setiap himpunan data time series dengan memilih yang sesuai dengan nilai parameter p, d, q sesuai deret/series individu. Permasalah yang sering dihadapi dalam pengumpulan data adalah data observasi yang hilang atau pengamatan yang mungkin hampir mustahil untuk diperoleh, baik karena waktu atau kendala biaya. Dalam rangka untuk menggantikan observasi, ada beberapa pilihan yang tersedia bagi peneliti. Pertama, ganti dengan rata-rata seri. Kedua mengganti dengan bagian tengah ramalan. Bisa juga ganti dengan perkiraan trend sederhana. Atau ganti dengan rata-rata dua pengamatan terakhir yang diketahui dengan pengamatan yang hilang.

II. Deskripsi Pengumpulan Data

Data curah hujan dikumpulkan dari 01 Oktober 1973 sampai dengan 31 Mei 2011 di stasiun Meteorologi Kota Palu. Dalam penelitian ini data curah hujan dikumpulkan dan dicatat setiap hari. Jumlah bulanan dihitung dengan menjumlah semua jumlah curah hujan di bulan tersebut untuk setiap tahun.

III. Metodologi Pengamatan Data yang Hilang

Masalah yang sering dihadapi dalam pengumpulan data adalah serangkaian data yang hilang selama observasi. Dalam rangka untuk menggantikan pengamatan tersebut, ada beberapa pilihan berbeda tersedia untuk para peneliti. Pertama, ganti dengan rata-rata seri. Berarti ini dapat dihitung atas seluruh rentang sampel. Kedua, ganti dengan perkiraan yang sesungguhnya. Model yang sesungguhnya adalah bentuk paling sederhana dari model peramalan univariat, model ini menggunakan nilai periode waktu saat ini untuk peroide waktu berikutnya, yaitu $\hat{Y}_{t-1} = Y_t$. Juga, ganti dengan perkiraan trend sederhana. Ini adalah penyempurnaan dengan mengestimasi persamaan regresi dari bentuk $Y_t = a + b.t$ (dimana t adalah waktu) untuk periode sebelum nilai data yang hilang. Kemudian gunakan persamaan agar sesuai dengan periode waktu yang hilang, atau ganti dengan rata-rata dua pengamatan terakhir yang diketahui dengan pengamatan yang data yang hilang. Makalah ini menyarankan metode baru untuk memperkirakan data yang hilang dengan menggunakan proses penyaringan (1). Proses penyaringan adalah:

$$y_t = \frac{1}{\sum w'_{i+1}} \sum_{i=0}^{M} w'_{i+1} x_{t+i} = \sum_{i=0}^{M} w_{i+1} x_{t+i} \quad (1)$$

Dimana $w_i = \frac{w_{i+1}}{\sum w'_{i+1}}$ adalah bobot/berat dan M adalah jumlah pengamatan pada moving average. Selanjutnya subtitusi $w'_{i+1} = \varphi^i$ dimana φ adalah korelasi keseluruhan data input. selanjutnya model moving average (MA) dapat ditulis sebagai berikut:

$$y_t = \varphi x_t + \varphi^2 x_{t-1} + \cdots + \varphi^M x_{t-m} \quad (2)$$

Dimana x_t adalah data pengamatan. Selanjutnya untuk mengubah data lengkap menggunakan persamaan (2) dan selanjutnya membangun model yang tepat. Selanjutnya diasumsikan bahwa ada jarak lubang secara acak dalam data. Jika y_s yang terlewat (dimana s adalah indeks dari lubang), kita subtitusi rata-rata data lengkap bukan x_s lalu kita menghitung nilai masa depan, berdasarkan persamaan (3) berikut:

$$y_t = \varphi \bar{y} + \varphi^2 x_{s-1} + \cdots + \varphi^M x_{s-m} \quad (3)$$

Kemudian kami membangun model untuk data yang berisi lubang. Menerapkan model yang sama pada data baru, selanjutnya membandingkan hasil model untuk kedua kumpulan data dengan menggunakan model Box-Jenkins ARIMA pada bagian berikutnya.

IV. Model Box-Jenkins Model ARIMA

Metode Box - Jenkins adalah sebuah prosedur untuk menentukan nilai model untuk masa yang lalu pada data time series dan nilai-nilai kesalahan yang lalu. Pendekatan Box-Jenkins terdiri dari prediksi ekstra dari data yang diamati melalui serangkaian iterasi. Model ARIMA yang paling umum mempunyai tiga parameter yaitu p, d, dan q dimana p adalah jumlah parameter autoregressive, d adalah jumlah parameter pembeda dan q adalah jumlah parameter moving average. Secara umum model ARIMA(*p,d,q*) adalah:

$$\tau(B) Z_t = \varphi(B) \Delta^d Z_t = \theta_0 + \varnothing(B) a_t \quad (4)$$

Dimana : $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$ dan $\emptyset(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$

Dalam hal ini:

$\varphi(B)$ adalah operator autoregressive yang diasumsikan stasioner, $\emptyset(B)$ adalah operator moving average diasumsikan dapat diinverskan. Dari persamaan (4) model ARIMA(*p,d,q*) dapat ditulis sebagai berikut:

$$Z_t = C + \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \cdots + \varphi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad (5)$$

Dimana:

t : adalah waktu periodik, Z_t: adalah nilai numerik dari pengamatan

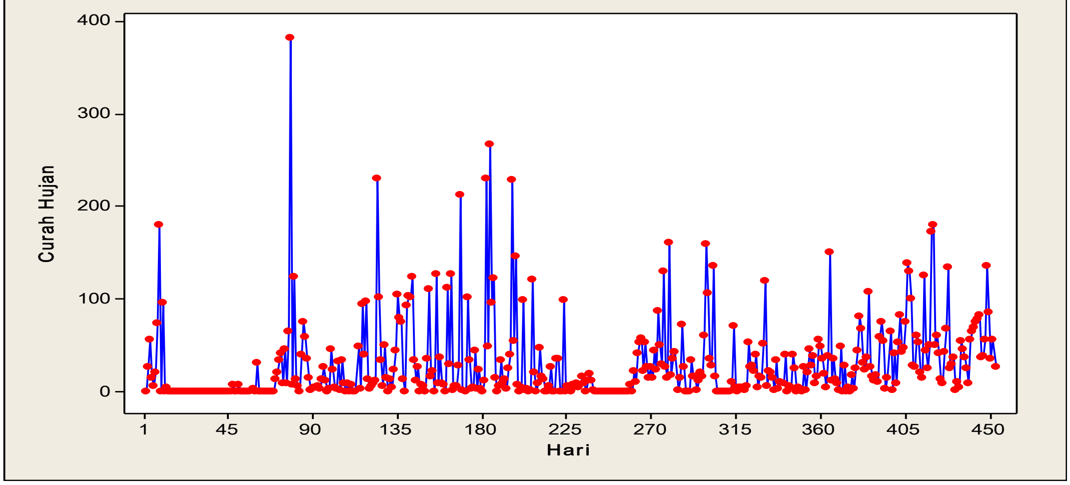
φ_i: untuk i = 1, 2, ... p adalah parameter autoregressive

θ_j: untuk j = 1, 2, ..., q adalah parameter moving average, a_t: adalah nilai eror/galak pada waktu t

Untuk mengestimasi parameter φ_i dan θ_j untuk campuran p tetap dan q kita melakukan persamaan regresi linear berganda dengan metode kuadrat terkecil

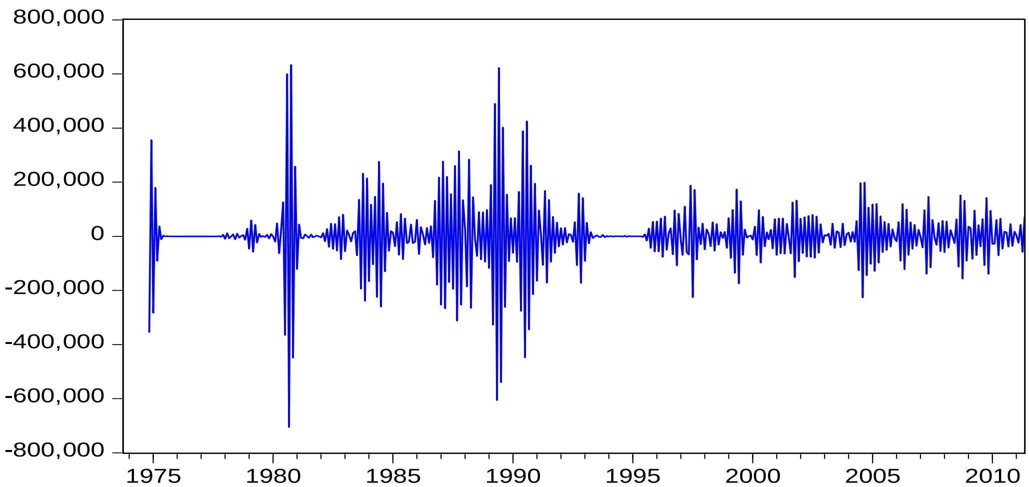
$$\hat{z}_t = \mu + \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_p z_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \qquad (6)$$

Gambar 1: Plot Data Time Series Curah Hujan



Ada dua tahap untuk mengidentifikasi dengan tepat sebuah model Box-Jenkins: mengubah data jika perlu menjadi data time series yang stasioner dan menentukan Model tentatif dengan mengamati perilaku pada fungsi autokorelasi (ACF) dan autokorelasi parsialnya (PACF). Data time series stasioner adalah bahwa data tersebut tidak mengandung nilai tren, yaitu, data tersebut berfluktuasi disekitar nilai rata-rata konstan. Dengan melihat pada suatu kurun waktu plot (lihat gambar 1 plot data tanpa transformasi). Dari gambar 1 terlihat bahwa data curah hujan tidak stasioner sehingga membutuhkan transformasi untuk membuat data menjadi stasioner. Dengan menggunakan transformasi differencing maka data menjadi stasioner seperti seperti terlihat pada gambar (2).

Gambar 2: Plot Data Hasil Transformasi



Perbedaan yang pertama adalah untuk bagian musiman oleh pengurangan nilai dari dua pengamatan berdekatan dalam seri yaitu, $z_t = \Delta Y_t = Y_t - Y_{t-12}$ untuk musiman. Kita dapat menulis perbedaan dengan operator pembeda sebagai berikut $Bz_t = z_{t-1}$. Setelah transformasi, jelas bahwa pengamatan berfluktuasi sekitar rata-rata konstan. Box-Jenkins menunjukkan bahwa jumlah lag yang diperlukan dalam analisis tidak lebih dari n/4 dimana n adalah banyaknya data pengamatan, ukuran koefisien autokorelasi diukur berdasarkan korelasi antara beberapa pengamatan dan sekumpulan pengamatan yang tertinggal dalam sebuah time series. Autokorelasi antara z_t dan z_{t+k} ukuran korelasi antara pasangan $(z_1, z_{1+k}), (z_2, z_{2+k}), \cdots, (z_n, z_{n+k})$. Sampel autokorelasi.

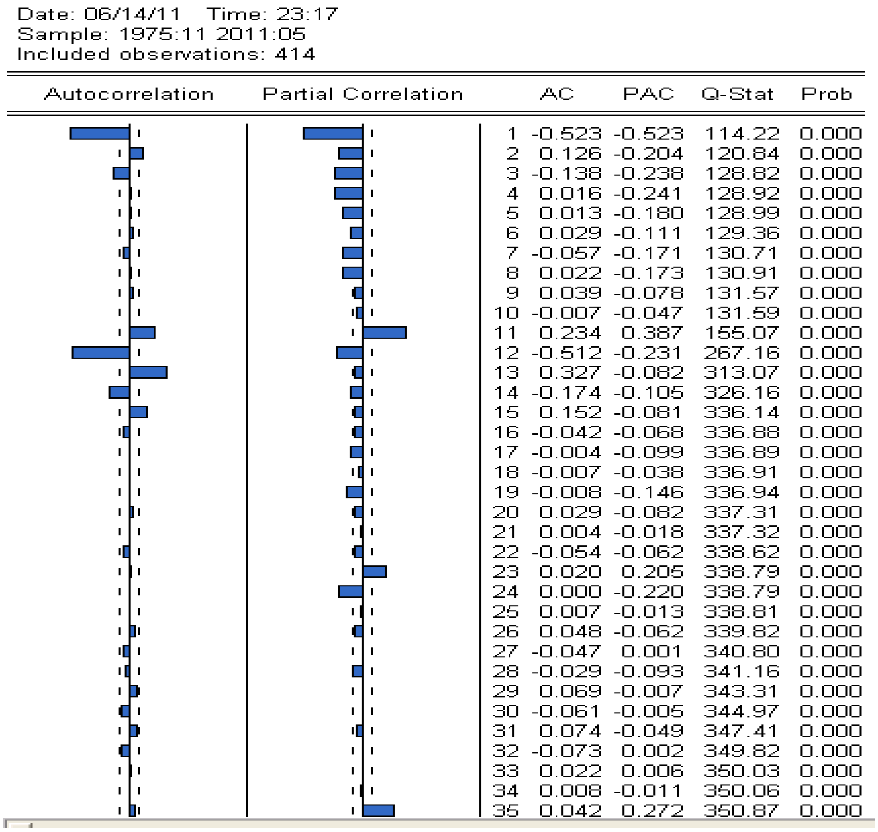
Koefisien r_k adalah sebuah perkiraan ρ_k dimana

$$r_k = \frac{\sum(z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum(z_t - \bar{z})^2} \tag{7}$$

dengan : z_t : data dari time series stasioner., z_{t+k}: data dari periode waktu k didepan t

$\bar{z}$: rata-rata dari time series stasioner.

Tabel 1: Plot ACF dan PACF untuk data asli



Date: 06/14/11 Time: 23:17
Sample: 1975:11 2011:05
Included observations: 414

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.523	-0.523	114.22	0.000
		2	0.126	-0.204	120.84	0.000
		3	-0.138	-0.238	128.82	0.000
		4	0.016	-0.241	128.92	0.000
		5	0.013	-0.180	128.99	0.000
		6	0.029	-0.111	129.36	0.000
		7	-0.057	-0.171	130.71	0.000
		8	0.022	-0.173	130.91	0.000
		9	0.039	-0.078	131.57	0.000
		10	-0.007	-0.047	131.59	0.000
		11	0.234	0.387	155.07	0.000
		12	-0.512	-0.231	267.16	0.000
		13	0.327	-0.082	313.07	0.000
		14	-0.174	-0.105	326.16	0.000
		15	0.152	-0.081	336.14	0.000
		16	-0.042	-0.068	336.88	0.000
		17	-0.004	-0.099	336.89	0.000
		18	-0.007	-0.038	336.91	0.000
		19	-0.008	-0.146	336.94	0.000
		20	0.029	-0.082	337.31	0.000
		21	0.004	-0.018	337.32	0.000
		22	-0.054	-0.062	338.62	0.000
		23	0.020	0.205	338.79	0.000
		24	0.000	-0.220	338.79	0.000
		25	0.007	-0.013	338.81	0.000
		26	0.048	-0.062	339.82	0.000
		27	-0.047	0.001	340.80	0.000
		28	-0.029	-0.093	341.16	0.000
		29	0.069	-0.007	343.31	0.000
		30	-0.061	-0.005	344.97	0.000
		31	0.074	-0.049	347.41	0.000
		32	-0.073	0.002	349.82	0.000
		33	0.022	0.006	350.03	0.000
		34	0.008	-0.011	350.06	0.000
		35	0.042	0.272	350.87	0.000

Estimasi fungsi autokorelasi parsial PACF digunakan sebagai panduan, selanjutnya dengan fungsi autokorelasi diperkirakan ACF, digunakan untuk memilih satu atau lebih Model ARIMA yang mungkin cocok dengan data yang tersedia. Gagasan analisis autokorelasi parsial adalah bahwa kita ingin mengukur seberapa besar keterkaitan antara z_t dan z_{t+k}. Persamaan yang memberikan perkiraan yang baik dari autokorelasi parsial adalah

$$\hat{\varphi}_{kj} = \hat{\varphi}_{k-1,j} - \hat{\varphi}_{kk}\hat{\varphi}_{k-1,k-j} \qquad (8)$$

untuk *k*, = 3, 4, ...; *j* = 1, 2, ..., *k-1*. Kita bisa menemukan bentuk ACF dan PACF dalam sebuah model musiman seperti pada gambar (3) dan (4). Jadi, perkalian musiman model ARIMA(p, d, q)×(P, D, Q)$_s$ adalah generalisasi dan dianggap sebagai perpanjangan metode untuk barisan dimana pola musiman berulang dari waktu ke waktu, dimana parameter (p,d,q) adalah bukan untuk data musiman dan parameter (P, D, Q)$_s$ adalah untuk data musiman. Setelah data time series stasioner (pemotongan ACF berhenti atau menurun dengan cepat), kita bisa mengidentifikasi model sementara dengan memeriksa perilaku ACF dan PACF. Dalam model campuran baik ACF dan PACF menurun secara eksponensial. Angka-angka dari ACF dan PACF seperti pada tabel 1.

V. Hasil

Statistik-t seperti pada Tabel 2 dan juga dalam Tabel 3. terkait dengan $\varnothing_{12}$ dan φ yang lebih besar dari nilai mutlak 2, hal ini menunjukkan bahwa parameter harus disimpan dalam model untuk kedua kumpulan data. Kami menyimpulkan dari tabel 2 dan 3 bahwa model pertama untuk data lengkap adalah

$z_t = -0{,}418 z_{t-1} - a_t - 0{,}995 a_{t-12}$, dan model untuk data yang hilang adalah $z_t = -0{,}502 z_{t-1} - a_t + 0{,}475 a_{t-12}$.

Tabel 2: Estimasi Parameter Untuk Data Asli

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	-0.417576	0.043695	-9.556586	0.0000
MA(1)	-0.995121	0.003423	-290.6936	0.0000
R-squared	0.496772	Mean dependent var		-0.088033
Adjusted R-squared	0.495588	S.D. dependent var		76.05743
S.E. of regression	54.01749	Akaike info criterion		10.82117
Sum squared resid	1240103.	Schwarz criterion		10.84017
Log likelihood	-2308.319	Durbin-Watson stat		1.585316
Inverted AR Roots	.90 -.24i	.90+.24i	.66+.66i	.66 -.66i
	.24+.90i	.24 -.90i	-.24 -.90i	-.24+.90i
	-.66+.66i	-.66+.66i	-.90+.24i	-.90 -.24i
Inverted MA Roots	1.00			

Pada tahapan estimasi, kita mendapatkan perkiraan yang tepat dari jumlah parameter. Berdasarkan plot ACF terlihat nilai puncak pada lag 1, sementara PACF meluruh secara uniform (tabel 1), sehingga model yang tepat untuk menggambarkan data ini adalah MA(1) . sementara jika diamati lag 12, 24 36 terlihat adanya komponen musiman, menidikasikan perlunya komponen musiman dalam model, baik dalam bentuk perkalian atau penjumlahan. Disimpulkan bahwa model yang tepat adalah ARIMA(1,1,0) (0, 1, 1)$_{12}$. Kami cocok model ini untuk data untuk mendapatkan estimasi parameter yang tepat: φ_1 untuk bagian AR nonmusiman, dan $\varnothing_{12}$ untuk koefisien moving average untuk parameter musiman. Nilai rata-rata μ dari model sejak rata-rata seri bekerja sebesar -

0,00207 dan standar deviasi sebesar 0.897664 untuk data lengkap dan untuk kumpulan data yang hilang rata-rata sebesar -0,00051 dan standar deviasi sebesar 0,889339.

Tabel 3: Estimasi Parameter Untuk Data Yang Dibangkitkan

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	-0.502176	0.042349	-11.85794	0.0000
MA(1)	0.047559	0.049881	0.953442	0.3409
R-squared	0.260034	Mean dependent var		0.031336
Adjusted R-squared	0.258189	S.D. dependent var		1.745889
S.E. of regression	1.503707	Akaike info criterion		3.658695
Sum squared resid	906.7152	Schwarz criterion		3.678540
Log likelihood	-735.2269	Durbin-Watson stat		1.990371
Inverted AR Roots	.91+.24i	.91 -.24i	.67 -.67i	.67+.67i
	.24 -.91i	.24+.91i	-.24+.91i	-.24 -.91i
	-.67 -.67i	-.67 -.67i	-.91 -.24i	-.91+.24i
Inverted MA Roots	-.05			

V.1 Pengujian Diagnostik

Tahap selanjutnya adalah pemeriksaan diagnostik, kami menggunakan statistik Ljung-Box dinotasikan dengan *Q** seperti pada Persamaan (9) untuk memeriksa kecocokan/keakuratan model dengan memeriksa residual autokorelasi ACF dan residual autokorelasi parsial PACF.

$$Q^* = n'(n'+2)\sum_{i=1}^{K}(n'-l)r_l^2(a') \qquad (9)$$

dimana *n' = (n-d)* dengan *n* adalah jumlah observasi dalam data time series asli, $r_l\ (a')$ adalah sampel autokorelasi dari residual pada lag *l* dan *d* adalah derajat differencing nonmusiman digunakan untuk mengubah nilai-nilai data time series nonstasioner menjadi data time series stasioner. Nilai p-value mempunyai hubungan dengan *Q** menunjukkan bahwa model.

$$z_t = -0{,}418z_{t-1} - a_t - 0{,}995a_{t-12}$$

adalah cukup untuk data lengkap selanjutnya karena nilai p lebih besar dari 0,05 dan kurang dari chi square untuk nilai K sebesar 6,12, 18, 24 dan 36.

Nilai p adalah area di bawah kurva distribusi chi-kuadrat yang memiliki 5 derajat kebebasan di sebelah kanan *Q** = 7,2364 dengan nilai p adalah 0,065, ini menunjukkan bahwa p-value = 0,065 > 0,05 = α, kita tidak dapat menolak kecukupan model dengan mengambil nilai α=0,05 ini menunjukkan bahwa membandingkan p-value dengan α menghasilkan kesimpulan yang sama seperti membandingkan *Q** dengan $\chi^2_{(\alpha)}(K-n_c)$. Disini terlihat ACF dan PACF tidak signifikan, yakni residual dari model bersifat white noise, dan statistik *Q** (amati lag > 3 dengan taraf signifikan $\alpha = 0{,}05$) pada tabel 4 di atas, bersifat tidak signifikan karena nilai probabilitasnya lebih dari 0,05, kecuali pada lag 3 dan 4. Menurut hipotesis H_0 diterima, artinya tidak terdapat korelasi serial dalam residual dari hasil estimasi dengan model yang diamati. Demikian pula kita mendapatkan *Q** untuk model dengan data yang hilang. Dalam rangka untuk memperkirakan logaritma alami dari jumlah curah hujan bulanan dalam 2 tahun ke depan (bulan 451 sampai 470), kami mencatat bahwa sejak $z_t = y'_t - y'_{t-12}$ dimana $y'_t = \ln y_t$ kita dapat mengekspresikan model untuk data lengkap

$$y'_t = y'_{t-12} - 0{,}418(y'_{t-1} - y'_{t-13}) - a_t - 0{,}995a_{t-12}$$

Dan model

$$y'_t = y'_{t-12} - 0{,}502\,(y'_{t-1} - y'_{t-13}) - a_t + 0{,}475a_{t-12}$$

untuk data dengan lubang. Untuk mengestimasi parameter model digunakan metode kuadrat terkecil, ramalan ditunjukkan dalam tabel (5). Beberapa model diperiksa. Hasil estimasi jumlah curah hujan bulanan peramalan curah hujan dengan interval kepercayaan 95% adalah disajikan dalam tabel (5).

Tabel 4: Plot ACF dan PACF of residual untuk data asli

Q-statistic probabilities adjusted for 2 ARIMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.037	0.037	0.5839	
		2	0.109	0.107	5.5228	
		3	-0.063	-0.072	7.1905	0.007
		4	0.010	0.004	7.2361	0.027
		5	-0.001	0.013	7.2364	0.065
		6	0.043	0.037	8.0019	0.092
		7	-0.033	-0.037	8.4643	0.132
		8	0.059	0.055	9.9651	0.126
		9	0.030	0.038	10.337	0.170
		10	0.060	0.040	11.856	0.158
		11	0.058	0.055	13.308	0.149
		12	-0.028	-0.041	13.636	0.190
		13	0.032	0.032	14.068	0.229
		14	-0.042	-0.037	14.838	0.250
		15	0.055	0.049	16.153	0.241
		16	0.003	0.005	16.156	0.304
		17	-0.004	-0.025	16.164	0.371
		18	-0.013	-0.005	16.233	0.437
		19	-0.018	-0.028	16.373	0.498
		20	0.016	0.022	16.483	0.559
		21	-0.019	-0.033	16.643	0.614
		22	-0.011	-0.009	16.701	0.672
		23	0.033	0.042	17.191	0.699
		24	-0.016	-0.025	17.303	0.746
		25	0.009	0.006	17.338	0.792
		26	0.082	0.082	20.305	0.679
		27	0.011	0.014	20.362	0.728
		28	0.048	0.028	21.412	0.720
		29	0.050	0.063	22.517	0.711
		30	-0.017	-0.024	22.646	0.751
		31	0.034	0.023	23.182	0.768
		32	-0.053	-0.045	24.474	0.750
		33	-0.029	-0.037	24.857	0.774
		34	0.030	0.040	25.261	0.795
		35	-0.005	-0.017	25.272	0.830
		36	-0.062	-0.088	27.038	0.796
		37	-0.123	-0.134	33.972	0.518
		38	0.059	0.088	35.560	0.489
		39	-0.017	-0.025	35.697	0.530
		40	0.098	0.076	40.143	0.375
		41	-0.047	-0.039	41.160	0.376
		42	-0.015	-0.030	41.261	0.415
		43	-0.018	0.024	41.413	0.453
		44	-0.036	-0.055	42.006	0.471
		45	-0.068	-0.036	44.169	0.422
		46	0.029	0.041	44.553	0.448
		47	-0.062	-0.022	46.370	0.416
		48	0.064	0.059	48.285	0.381

V.2 Statistik Theil's untuk Akurasi Ramalan

Ketepatan ramalan itu diperiksa dengan menggunakan uji Theil's U yang membandingkan akurasi model ARIMA dengan model sesungguhnya. Untuk menguji akurasi model digunakan nilai aktual untuk jangka waktu terkahir Y_t sebagai perkiraan untuk $\hat{Y}_{t+1}$, untuk menghitung nilai Theils U dapat menggunakan persamaan (10) sebagai berikut

$$UU = \frac{RMSE(ARIMA)}{RMSE(naive)} \qquad (10)$$

dimana (RMSE) adalah Akar dari Rata-rata Squared Error, secara matematis didefinisikan seperti persamaan. (11). Berikut ini

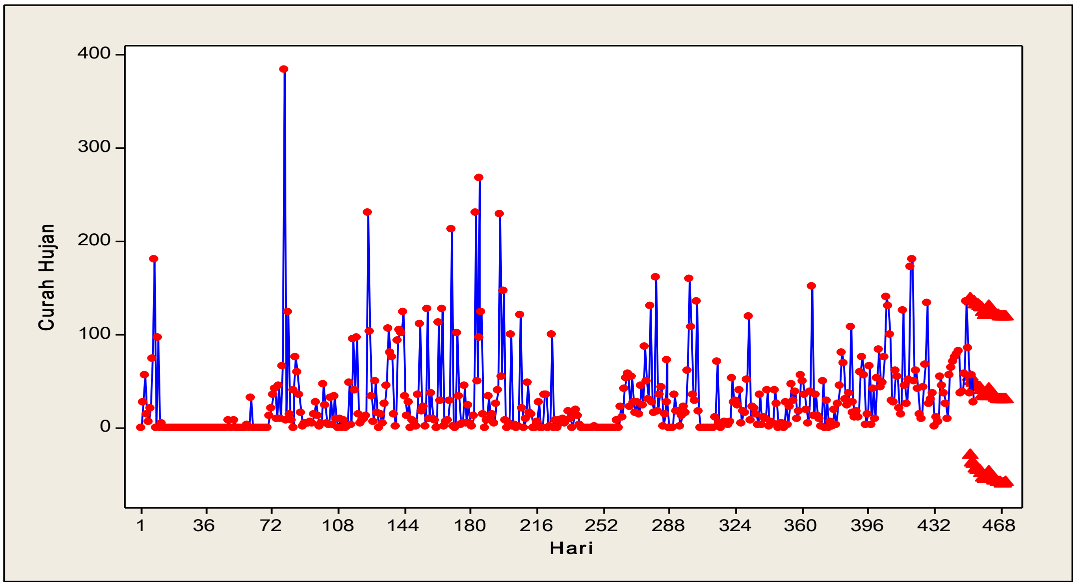
$$RMSERMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(e_t)^2} \qquad (11)$$

dimana n adalah jumlah observasi dan e adalah kesalahan error,

Tabel 5 : Hasil ramalan data asli dan data yang hilang

Priode	Data Asli (95% Limits)				Data yang Hilang (95% Limits)			
	Ramalan	Lower	Upper	Aktual	Ramalan	Lower	Upper	Aktual
451	54.676	-29.788	139.141	36.58	39.653	-43.345	122.652	36.58
452	49.508	-36.866	135.882	56.64	35.734	-47.827	119.294	56.64
453	48.131	-39.261	135.524	27.43	40.654	-42.906	124.215	27.43
454	44.79	-43.151	132.731		40.937	-42.623	124.497	
455	43.455	-44.782	131.692		43.341	-40.22	126.901	
456	42.818	-45.579	131.215		45.438	-38.122	128.998	
457	40.514	-47.971	128.998		43.868	-39.692	127.429	
458	35.286	-53.246	123.817		39.365	-44.195	122.925	
459	33.032	-55.525	121.589		37.152	-46.408	120.712	
460	34.145	-54.426	122.716		38.843	-44.717	122.403	
461	41.44	-47.138	130.019		47.497	-36.063	131.057	
462	37.075	-51.508	125.657		43.61	-39.95	127.17	
463	34.27	-54.725	123.265		39.128	-44.856	123.112	
464	33.065	-55.956	122.086		38.474	-45.516	122.464	
465	31.651	-57.384	120.687		36.932	-47.058	120.922	
466	31.113	-57.93	120.156		36.843	-47.147	120.833	
467	30.473	-58.575	119.52		36.09	-47.9	120.08	
468	29.926	-59.124	118.975		35.432	-48.558	119.422	
469	29.92	-59.131	118.971		35.924	-48.066	119.914	
470	30.68	-58.372	119.731		37.336	-46.654	121.326	

Hasil peramalan dapat dilahat pada gambar 3 untuk kedua model ARIMA(1,1,0) $(0, 1, 1)_{12}$ dan untuk model sesungguhnya MSE dan RMSE.statistik Theil adalah U = 0.895766 untuk data asli dan U = 0.726352 untuk data yang hilang. Ini adalah kurang dari 1, berarti model yang dipilih adalah model yang baik. Statistik Theil's U jika nilainya besar dari 1 menunjukkan bahwa model peramalan lebih buruk daripada model sesungguhnya, dan jika nilai kurang dari 1 mengindikasikan bahwa lebih baik. U dekat dengan 0 model yang terbaik yang kita miliki. Kami mengamati bahwa nilai-nilai sekitar dekat satu sama lain yang berarti bahwa metode yang digunakan untuk memperkirakan data yang hilang lebih baik setidaknya pada data yang digunakan dalam jurnal ini.



Gambar 3. Grafik hasil peramalan untuk data asli

VI. Kesimpulan

Makalah ini menyelidiki penerapan teknik Box dan Jenkins untuk meramalkan bulanan rata-rata curah hujan di stasiun Pinang dengan menggunakan saran baru metode untuk memperkirakan nilai hilang. Model parameter tersebut diestimasi dengan menggunakan model Autoregresif Integrated Moving Average (ARIMA) dalam suatu periode dari 1 Oktober 1973 sampai 31 Mei 2011. Model diuji dalam peramalan dengan mengamati jumlah curah hujan bulanan pada periode yang sama. hasil estimasi parameter model ARIMA untuk peramalan jumlah curah hujan bulanan adalah model ARIMA(1,1,0) (0, 1, 1)$_{12}$. Selanjutnya dibandingkan hasil dari model tersebut kedalam dua kumpulan data yaitu data asli dan data yang hilang. Persamaan untuk model data asli adalah $z_t = -0{,}418z_{t-1} - a_t - 0{,}995a_{t-12}$, dan model dengan data yang hilang adalah $z_t = -0{,}502z_{t-1} - a_t + 0{,}475a_{t-12}$

Hasilnya diperiksa sehubungan dengan uji sesungguhnya, yaitu nilai Theil U = 0.895766 untuk data asli dan untuk data yang hilang nilai Theil U = 0,726352 itu berarti hasilnya tertutup satu sama lain, artinya model ARIMA(1,1,0) (0, 1, 1)$_{12}$ adalah model yang baik. Hasil penelitian menunjukkan bahwa teknik deret waktu dapat digunakan untuk mengembangkan akurat perkiraan jangka pendek dari jumlah curah hujan bulanan tergantung pada pengamatan terakhir untuk stasiun Meteorologi Kota Palu.

VII. Daftar Pustaka

1 Gencay R., Selcuk F. and Whitcher B. (2002). *An Introduction to Wavelets and other filtering methods in finance and economics*, Permissions Department, Harcourt, Inc.

2 James W. T. and Kurtz, T. G. (2007). *A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center*. Said Business School, University of Oxford. *Mahir* & *Al-Khazaleh /Estimation of missing data by using the filtering process in a time serie*1*s*2

3 John C. B. and David A. D. (2003). *SAS for Forecasting Time Series*, Second Edition. Cary, NC: Institute Inc.

4 Pankratz A. (1983). *Forecasting with Univariate Box-Jenkins Models*, Wiley New York.

5 Patricia E G. (1994). *Introduction to Time Series Modeling and Forecasting in Business and Economics*, Cgraw-Hill M. Inc.

6 Richard T. B.,Sfetsos A. and Sang-Kuck Ch.(2002). *Modeling and forecasting from trend-stationary long memory models with applications to climatology 1*. International Journal of Forecasting, Vol. 18, issue 2 pp. 215-226.

7 Rais, 2005, *Kriteria Kesesuaian Model Untuk Penentuan Arsitektur Optimal Pada Neural Network Untuk Pemodelan Time Series*, Tesis Yogyakarta.

9 .... 2008, *Kriteria Kesesuaian Model Untuk Penentuan Arsitektur Optimal Model Neural Network Backpropagation Untuk Indeks Harga Saham Gabungan*, Jurnal Ilmiah Matematika dan Terapan Vol. 5 No. 1 hal 1 – 11.

8 Sabry M.,Abd - El - 1 1 Latif H. ,Yousif S. and Badra N.(2007). *Use of Univariate Box and Jenkins Time Series Technique in Rainfall Forecasting 1*. Australian Journal of Basic and Applied Sciences, (4) pp. 386-394.