

# Deteksi *Outlier* pada Model ARIMA Musiman Ganda untuk Peramalan Beban Listrik Jangka Pendek di Jawa Timur

Indah Kurnia Putri dan Suhartono  
Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Institut Teknologi Sepuluh Nopember (ITS)  
Jl. Arief Rahman Hakim, Surabaya 60111, Indonesia  
*Email* : suhartono@statistika.its.ac.id

**Abstrak**—Perencanaan yang baik akan memberikan kontribusi besar terhadap pengembangan sistem distribusi karena merupakan ujung tombak dari pelayanan energi listrik yang langsung berhubungan dengan konsumen. Namun karena banyaknya data dalam peramalan beban listrik jangka pendek, membuat peramalan yang dilakukan tidak memenuhi asumsi distribusi normal. Hal ini dikarenakan *outlier* yang banyak dalam data. Penelitian ini fokus untuk mendeteksi *outlier* yang terjadi supaya asumsi distribusi normal terpenuhi. Apalagi belum ada *software* yang mampu mendeteksi *outlier* pada musiman ganda karena banyaknya *outlier*. Metode yang digunakan prosedur iteratif dengan pembagian data sebagai pengembangannya. Sebagai pengembangan metode maka diperlukan pemvalidasian metode yang dilakukan dari simulasi AR (1) dengan *outlier*. Hasil yang didapatkan yaitu metode pendeteksian *outlier* dari simulasi mampu mengatasi *outlier* lalu diterapkan terhadap data beban listrik jangka pendek sehingga menghasilkan model ARIMA yang memenuhi asumsi distribusi normal. Pada data I diperoleh *outlier* pada periode ke 1698 dan ke 1790 dengan efek 20 data setelahnya, jadi ada 22 *outlier*. Adanya *outlier* pada data I ini bertepatan dengan pemilu Legislatif tanggal 9 April 2014. Kemudian untuk data II diperoleh *outlier* sebanyak 177 dimana pada data II bertepatan dengan pemilu presiden tanggal 9 Juli 2014 dan hari Raya Idul Fitri tanggal 28 Juli 2014.

**Kata Kunci**—ARIMA, Beban Listrik, Deteksi *Outlier*, Musiman Ganda, Simulasi AR (1).

## I. PENDAHULUAN

PERAMALAN beban jangka pendek adalah model perkiraan beban listrik untuk jangka waktu dalam menit/jam/hari/minggu. Kualitas peramalan beban listrik jangka pendek dengan rentang 1 jam sampai beberapa hari ke depan memiliki dampak yang signifikan pada efisiensi operasi sistem tenaga listrik [1]. Perencanaan yang baik akan memberikan kontribusi besar terhadap pengembangan sistem distribusi yang merupakan ujung tombak dari pelayanan energi listrik langsung berhubungan dengan konsumen [2].

Banyak teknik yang dapat digunakan dalam peramalan beban listrik jangka pendek, diantaranya yaitu *Autoregressive Integrated Moving Average* (ARIMA), regresi linear, *Artificial neural network*, *Naive*, *Exponential Smoothing* dan *Adaptive Neuro-Fuzzy Inference System* (ANFIS). Metode-metode tersebut telah banyak digunakan di Indonesia dan luar negeri seperti pada penelitian yang dilakukan oleh [1] dan [3]-[12]. Peramalan beban listrik jangka pendek semakin berkembang dengan adanya model musiman ganda sesuai dengan penelitian yang dilakukan oleh Taylor tentang peramalan permintaan

beban listrik jangka pendek di England dan Wales [10]. Permintaan pemakaian beban listrik yang dilakukan oleh Taylor tidak hanya dipengaruhi oleh musiman harian, tetapi dipengaruhi juga oleh musiman mingguan atau dikenal musiman ganda. Pada penelitian Taylor musiman ganda mempunyai *outlier* yang banyak sehingga asumsi distribusi normal tidak terpenuhi, begitu juga penelitian yang dilakukan oleh [4] dan [8]. Perkembangan analisis deret waktu tentang *outlier* belum secepat dan seluas pada ilmu statistik secara umum. *Outlier* dapat mempengaruhi struktur autokorelasi dari suatu time series sehingga nilai estimasi dari *autocorrelation function* (ACF) dan *partial autocorrelation function* (PACF) yang diperoleh menjadi bias. Keberadaan *outlier* akan menyesatkan ketika dilakukan identifikasi model dan peramalan, oleh karena itu perlu dilakukan penanggulangan dengan cara melakukan deteksi *outlier* terlebih dahulu [13]. Telah banyak penelitian tentang deteksi *outlier* sebagaimana pada penelitian [14]-[17], namun penelitian tentang deteksi *outlier* untuk data *time series* musiman ganda belum diterapkan dalam perangkat lunak statistik manapun. Ini dikarenakan data yang cukup panjang akan mengalami kegagalan dalam pendeteksian karena tahapan iteratif untuk pendeteksian *outlier* memerlukan representasi model *Autoregressive* (AR) yang cukup rumit. Oleh karena itu pada penelitian ini dilakukan deteksi *outlier* pada model ARIMA musiman ganda yaitu pada data beban listrik jangka pendek di Jawa Timur menggunakan prosedur iteratif dengan pembagian data untuk pengembangannya. Namun pada penelitian ini terlebih dahulu menerapkan metode pendeteksian *outlier* pada data simulasi untuk memvalidasi apakah metode yang digunakan mampu mengatasi adanya *outlier*. Pengembangan deteksi *outlier* pada analisis peramalan data beban listrik jangka pendek diharapkan akan memberikan peramalan yang lebih akurat, hal ini akan menjadi sumbangan yang cukup berarti bagi kinerja PT PLN (Persero) dalam memberikan pelayanan penyaluran listrik ke konsumen.

## II. TINJAUAN PUSTAKA

### A. Model ARIMA Musiman Ganda

Mode ARIMA dengan musiman ganda dapat dituliskan sebagai berikut [18]

$$f_{\rho}(B)F_{\rho_1}(B^{S_1})F_{\rho_2}^*(B^{S_2})(1-B)^d(1-B^{S_1})^{D_1}(1-B^{S_2})^{D_2}Z_t \\ = q_{\rho}(B)Q_{\rho_1}(B^{S_1})Q_{\rho_2}^*(B^{S_2})\alpha_t \quad (1)$$

dengan  $S_1$  dan  $S_2$  adalah periode musiman yang berbeda.

## B. Deteksi Outlier

Alasan perlunya mengatasi *outlier* pada time series yaitu supaya karakteristik data time series menjadi lebih baik sehingga menghasilkan peramalan, model dan estimasi yang lebih baik pula, begitu juga ketika dilakukan analisis intervensi akan didapatkan model yang lebih sempurna [9]. Ada empat tipe *outlier* pada data time series, yaitu *Additive Outlier* (AO), *Innovational Outlier* (IO), *Level Shift* (LS), dan *Temporary Change* (TC).

### 1. Additive Outlier (AO)

*Additive Outlier* adalah *outlier* dimana suatu kejadian mempengaruhi data series pada satu waktu tertentu saja. Model *Additive Outlier* (AO) dirumuskan sebagai berikut [18].

$$\begin{aligned} Z_t &= \begin{cases} X_t, & t \neq T \\ X_t + \omega, & t = T \end{cases} \\ &= X_t + \omega_{AO} I_t^{(T)} \\ &= \frac{\theta(B)}{\phi(B)} a_t + \omega_{AO} I_t^{(T)} \end{aligned} \quad (2)$$

dengan

$$I_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases}$$

### 2. Innovational Outlier (IO)

Kemudian *Innovational Outlier* (IO) adalah *outlier* yang mempengaruhi beberapa amatan setelah terjadinya *outlier* ini, sehingga mengakibatkan terganggunya susunan time series. Berikut merupakan persamaan dari *Innovational Outlier* [18]

$$\begin{aligned} Z_t &= X_t + \frac{\theta(B)}{\phi(B)} \omega_{IO} I_t^{(T)} \\ &= \frac{\theta(B)}{\phi(B)} (a_t + \omega_{IO} I_t^{(T)}). \end{aligned} \quad (3)$$

### 3. Level Shift (LS)

*Level Shift* (LS) adalah kejadian yang mempengaruhi deret pada satu waktu tertentu dan efek yang diberikan memberikan suatu perubahan yang tiba-tiba dan permanen. Berikut merupakan persamaan dari *Level Shift* [18]

$$\begin{aligned} Z_t &= X_t + \frac{1}{(1-B)} \omega_{LS} I_t^{(T)} \\ &= \frac{\theta(B)}{\phi(B)} a_t + \frac{1}{(1-B)} \omega_{LS} I_t^{(T)}, \end{aligned} \quad (4)$$

atau dapat dituliskan,

$$Z_t = \frac{\theta(B)}{\phi(B)} a_t + \omega_{LS} S_t^{(T)} \quad (5)$$

dengan

$$S_t^{(T)} = \begin{cases} 1, & t \geq T \\ 0, & t < T \end{cases}$$

### 4. Temporary Change (TC)

Selanjutnya tipe *Temporary Change* (TC) merupakan suatu kejadian dimana *outlier* menghasilkan efek awal  $\omega$  pada waktu  $t$ , dan kemudian efek tersebut berkurang secara lambat laun seiring dengan berkurangnya nilai faktor  $\delta$ . Berikut merupakan persamaan dari *Temporary Change* [18]

$$\begin{aligned} Z_t &= X_t + \frac{1}{(1-\delta B)} \omega_{TC} I_t^{(T)} \\ &= \frac{\theta(B)}{\phi(B)} a_t + \frac{1}{(1-\delta B)} \omega_{TC} I_t^{(T)} \end{aligned} \quad (6)$$

## III. METODOLOGI PENELITIAN

### A. Sumber Data dan Variabel Penelitian

Pada penelitian ini digunakan data sekunder yaitu data beban konsumsi listrik di Jawa Timur pada periode rentang 3 Maret 2014 sampai 14 April 2014 dan 26 Mei 2014 sampai 10 Agustus 2014, yang diperoleh dari PT PLN (Persero) P3B Regional Jawa Timur dan Bali. Variabel Penelitian yang digunakan adalah jumlah beban listrik harian tiap setengah jam dalam satuan Mega Watt (MW) untuk wilayah Jawa Timur.

### B. Langkah Analisis

Langkah analisis data pada penelitian ini secara umum adalah sebagai berikut.

1. Mendapatkan prosedur pendeteksian *outlier*.
2. Memvalidasi prosedur yang didapatkan melalui data simulasi. Simulasi yang digunakan adalah data AR (1) sebanyak 200 dengan  $\mu = 50$  dan  $\phi = 0,6$  dengan residual yang memenuhi IIDN(0,1). Kemudian simulasi tersebut diberikan dua efek *outlier*, baik itu AO, IO, TC atau LS. Sehingga ada 16 kombinasi dari dua *outlier* tersebut dengan *outlier* pada  $T_1 = 75$  dan  $T_2 = 145$  dimana untuk *outlier* pertama disimulasikan dengan efek 15 dan *outlier* kedua disimulasikan dengan efek 20. Secara umum prosedur untuk mendeteksi *outlier* pada data simulasi dengan dua *outlier* adalah sebagai berikut.
  - a. Mengidentifikasi data keseluruhan
  - b. Memodelkan data keseluruhan, berhubung data disimulasikan dari model AR (1) maka dilakukan pemodelan dengan AR (1) meskipun dari identifikasi model tidak sesuai karakteristik AR (1).
  - c. Melakukan cek diagnosa.
  - d. Apabila tidak berdistribusi normal dilakukan deteksi *outlier*.
  - e. Letak *outlier* dapat diketahui melalui plot *standardized residuals*. Titik yang pertama kali keluar dari batas nilai kritis maka dianggap sebagai *outlier* yang pertama.
  - f. Memodelkan data sebelum *outlier* yang pertama, sesuai simulasi maka data awal yang dianggap tidak ada *outlier* yaitu sampai  $t = 74$ . Dari pemodelan data awal tersebut dilakukan perhitungan *fits*, MSE, dan ramalan sampai data sebelum *outlier* yang kedua, sesuai simulasi maka *outlier* yang kedua adalah  $T_2 = 145$ . Sehingga ramalan sampai data ke 144. Kemudian dilakukan perhitungan residual dari selisih data aktual yaitu sebanyak data ke 144 dengan gabungan antara *fits* dan ramalan yang jumlahnya juga 144. Plot *standardized residuals* diidentifikasi ada tidaknya *outlier*.
  - g. Ketika diketahui terdapat *outlier* maka *outlier* tersebut merupakan *outlier* pertama dan diidentifikasi tipe *outlier*.
  - h. Pemodelan untuk data sebanyak data sebelum *outlier* yang kedua, pada data simulasi ini diketahui sebanyak 144, dengan model AR (1) sesuai model pada tahap data

awal yang dianggap tidak ada *outlier* tetapi pada tahap ini ditambah efek *outlier* sesuai tipe *outlier* yang teridentifikasi.

- i. Model dengan efek *outlier* pertama tersebut dilakukan perhitungan *fits*, MSE, dan ramalan sampai akhir periode. Kemudian langkah selanjutnya sama seperti ketika melakukan identifikasi untuk *outlier* yang pertama.
  - j. Selanjutnya dilakukan pemodelan dengan penambahan efek *outlier* yang kedua. Sehingga model yang didapatkan yaitu model AR (1) dengan efek *outlier* yang pertama dan kedua. Kemudian dilakukan cek diagnosa kembali.
3. Menerapkan prosedur deteksi *outlier* yang sudah divalidasi pada data simulasi terhadap data aktual yaitu data beban listrik jangka pendek di Jawa Timur dengan salah satu tahapan yaitu ARIMA Box-Jenkins sebagai berikut.
- a. Membagi data aktual menjadi data *in sample* dan *out sample* dengan data aktual yaitu data I dengan rentang waktu 3 Maret 2014 sampai 14 April 2014 dan data II adalah data dengan rentang waktu 26 Mei 2014 sampai 10 Agustus 2014. Data I memiliki data *in sample* dengan periode 3 Maret 2014 sampai 13 April 2014 dan data *out sample* dengan periode 14 April 2014. Kemudian untuk data II digunakan data *in sample* dengan periode 26 Mei 2014 sampai 9 Agustus 2014 dan data *out sample* dengan periode 10 Agustus 2014.
  - b. Mengidentifikasi model dengan melihat time series plot, ACF dan PACF yang telah dibentuk.
  - c. Estimasi terbaik untuk parameter-parameter dalam model.
  - d. Pemeriksaan diagnostik meliputi pemeriksaan residual apakah telah identik dan independen (*white noise*) serta berdistribusi normal. Apabila distribusi normal tidak terpenuhi maka diindikasikan adanya *outlier*.
  - e. Melakukan deteksi *outlier*.
4. Pemilihan model terbaik melalui nilai MAPE atau RMSE.  
 5. Melakukan peramalan.

#### IV. ANALISIS DAN PEMBAHASAN

##### A. Prosedur Deteksi Outlier

Pada beberapa *software* statistika sebenarnya telah memfasilitasi program untuk mendeteksi *outlier*, namun untuk kasus data dengan pola musiman ganda belum ada *software* yang menyediakan fasilitas tersebut, termasuk SAS. Pada SAS ketika melakukan pendeteksian *outlier* pada data musiman ganda akan mengalami kegagalan karena data yang cukup banyak.

ERROR: The OUTLIER step has been terminated due to insufficient memory.

Gambar 1. Kegagalan SAS Dalam Deteksi *Outlier*

Berikut prosedur analisis deteksi *outlier* yang telah dikembangkan.

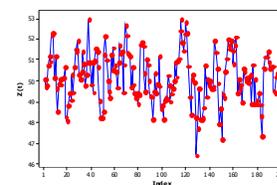
- a. Memodelkan seluruh data dengan anggapan tidak ada *outlier*. Dan dihitung *fits*, residual, dan MSE.
- b. Identifikasi *outlier* dari model di langkah pertama dengan menghitung *standardized residuals* kemudian membuat plot *standardized residuals*. *Standardized residuals* didapatkan

dari pembagian antara residual dan akar MSE. Apabila plot *standardized residuals* ada yang keluar dari batas nilai kritis yaitu  $C=3$  maka dianggap terdapat *outlier*.

- c. Memilih *outlier* pertama kali yang keluar dari  $C=3$ .
- d. Memodelkan kembali data hingga data sebelum terjadi *outlier* pertama dan mendapatkan nilai residual, *fits*, serta ramalan  $k$  tahap kedepan sampai data sebelum *outlier* kedua  $(T_1, T_1 + 1, \dots, T_2 - 1)$ .
- e. Nilai residual yang didapatkan pada langkah keempat, yang diperoleh dengan menggabungkan *fits* dengan hasil ramalan kemudian menghitung selisih antara data aktual dengan gabungan *fits* dan ramalan tersebut. Hasil yang diperoleh digunakan untuk mencari *standardized residuals*. Melalui plot *standardized residuals* tersebut maka dapat diidentifikasi tipe *outlier* pertama yang didapatkan.
- f. Memodelkan data dengan tambahan *outlier* sesuai tipe yang diidentifikasi dan dilakukan cek diagnosa kembali.
- g. Apabila dalam data terdapat  $j$  *outlier* ( $j = 2, 3, \dots, m$ ) maka langkah pendeteksian sama ketika mendeteksi satu *outlier*.
  - Menghitung *fits* dan MSE dari pemodelan sampai data ke  $T_j - 1$ .
  - Menghitung ramalan pada periode  $T_j, T_j + 1, \dots, T_{j+1} - 1$
  - Menggabungkan antara *fits* dan data ramalan, kemudian dilakukan perhitungan residual dari selisih antara data aktual dengan gabungan antara *fits* dan data ramalan tersebut.
  - Menghitung *standardized residuals* dari pembagian antara residual dengan akar MSE, sehingga diperoleh plot *standardized residuals*.
  - Apabila ada data yang keluar dari batas nilai kritis  $C = 3$  maka diketahui sebagai *outlier* dan dilakukan pengidentifikasian tipe *outlier* yang diperoleh.
  - Pemodelan dengan penambahan efek *outlier*.
  - Melakukan cek diagnosa. Langkah pendeteksian dilakukan hingga seluruh data telah dilakukan pendeteksian *outlier*.

##### B. Hasil Simulasi

Simulasi dilakukan dengan membangkitkan data AR (1) yang berjumlah 200 dengan  $\mu = 50$  dan  $\phi = 0,6$  sebagai berikut.



Gambar 2. Simulasi AR (1)

Model AR (1) hasil simulasi tersebut dapat dituliskan yaitu

$$Z_t = 50 + \frac{1}{(1-0,6B)} a_t \text{ atau } Z_t = 20 + 0,6Z_{t-1} + a_t$$

dengan  $a_t$  memenuhi IIDN (0,1). Kemudian AR (1) dimodelkan dengan 2 *outlier* sehingga ada 16 kombinasi.

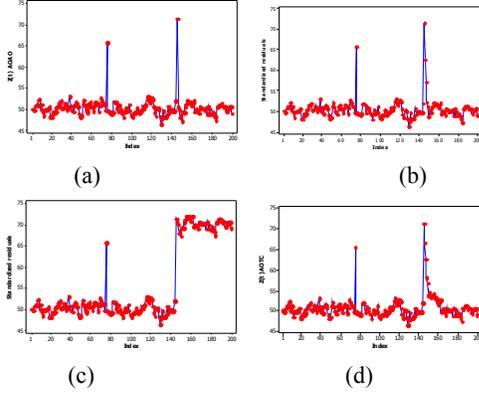
##### C. Outlier Pertama Additive Outlier (AO)

Kombinasi AO dengan empat *outlier* lainnya ditampilkan pada Gambar 2. Plot time series menunjukkan bahwa pada

setiap simulasi memiliki perbedaan dengan simulasi awal. Apabila dilakukan pemodelan dengan *outlier* diperoleh untuk *outlier* tahap pertama yaitu

$$Z_t = 50,2273 + \frac{1}{(1-0,6362B)} a_t + 15,8378I_t^{(75)}$$

dimana *outlier* pertama terletak di  $T_1 = 75$ .



Gambar 3. (a) AO-AO (b) AO-IO (c) AO-LS (d) AO-TC

Sedangkan untuk dua *outlier* dengan *outlier* pertama AO dapat dituliskan dengan persamaan sebagai berikut.

#### 1. AO dan AO

$$Z_t = 50,1770 + \frac{1}{(1-0,6209B)} a_t + 15,83886I_t^{(75)} + 20,2415I_t^{(145)}$$

#### 2. AO dan IO

$$Z_t = 50,1939 + \frac{1}{(1-0,6147B)} a_t + 15,8355I_t^{(75)} + \frac{20,1291}{(1-0,5559B)} I_t^{(145)}$$

#### 3. AO dan LS

$$Z_t = 50,2193 + \frac{1}{(1-0,6217B)} a_t + 15,8347I_t^{(75)} + 19,8529S_t^{(145)}$$

#### 4. AO dan TC

$$Z_t = 50,2230 + \frac{1}{(1-0,6252B)} a_t + 15,8353I_t^{(75)} + \frac{20,0161}{(1-0,7754B)} I_t^{(145)}$$

#### D. Outlier Pertama Additive Outlier (IO)

Kombinasi IO dengan empat *outlier* lainnya ditampilkan pada Gambar 3, terlihat bahwa pola plot time series berbeda dengan pola AR (1) tanpa *outlier*. Apabila berada di tahap deteksi *outlier* pertama diperoleh model yaitu

$$Z_t = 50,2387 + \frac{1}{(1-0,63228B)} a_t + \frac{15,7433}{(1-0,5491B)} I_t^{(75)}$$

dimana *outlier* pertama terletak di  $T_1 = 75$ . Sedangkan ketika dilakukan deteksi untuk tahap dua *outlier* diperoleh persamaan sebagai berikut.

#### 1. IO dan AO

$$Z_t = 50,1855 + \frac{1}{(1-0,6188B)} a_t + \frac{15,7558}{(1-0,5505B)} I_t^{(75)} + 20,2421I_t^{(145)}$$

#### 2. IO dan IO

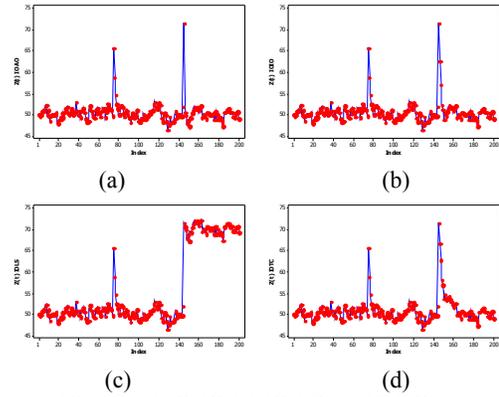
$$Z_t = 50,2110 + \frac{1}{(1-0,6121B)} a_t + \frac{15,7431}{(1-0,5494B)} I_t^{(75)} + \frac{20,1265}{(1-0,5554B)} I_t^{(145)}$$

#### 3. IO dan LS

$$Z_t = 50,2319 + \frac{1}{(1-0,6191B)} a_t + \frac{15,7387}{(1-0,5488B)} I_t^{(75)} + 19,8421S_t^{(145)}$$

#### 4. IO dan TC

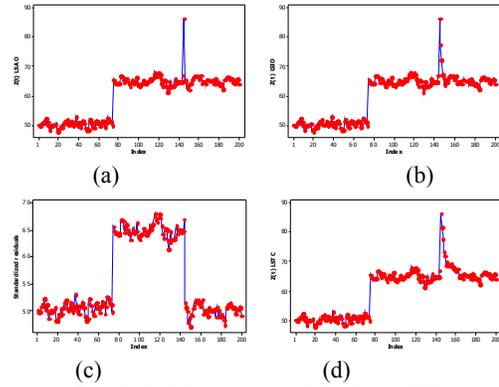
$$Z_t = 50,2412 + \frac{1}{(1-0,6226B)} a_t + \frac{15,7371}{(1-0,5486B)} I_t^{(75)} + \frac{20,0132}{(1-0,7749B)} I_t^{(145)}$$



Gambar 4. (a) IO-AO (b) IO-IO (c) IO-LS (d) IO-TC

#### E. Outlier Pertama Level Shift (LS)

Berikut simulasi TC dengan *outlier* lainnya.



Gambar 5. (a) LS-AO (b) LS-IO (c) LS-LS (d) LS-TC

Apabila dilakukan pendeteksian pada *outlier* pertama diperoleh

$$Z_t = 50,3857 + \frac{1}{(1-0,6157B)} a_t + 14,6778S_t^{(75)}$$

Ketika dilakukan pendeteksian dua *outlier* diperoleh persamaan sebagai berikut.

#### 1. LS dan AO

$$Z_t = 50,3909 + \frac{1}{(1-0,6045B)} a_t + 14,6631S_t^{(75)} + 20,2674I_t^{(145)}$$

#### 2. LS dan IO

$$Z_t = 50,3802 + \frac{1}{(1-0,5997B)} a_t + 14,7072S_t^{(75)} + \frac{20,1931}{(1-0,5582B)} I_t^{(145)}$$

#### 3. LS dan LS

$$Z_t = 50,3903 + \frac{1}{(1-0,6052B)} a_t + 14,6579S_t^{(75)} - 14,9819S_t^{(145)}$$

#### 4. LS dan TC

$$Z_t = 50,3751 + \frac{1}{(1-0,6114B)} a_t + 14,7494S_t^{(75)} + \frac{20,0504}{(1-0,7799B)} I_t^{(145)}$$

#### F. Outlier Pertama Temporary Change (TC)

Simulasi TC dengan *outlier* lainnya ditampilkan pada Gambar 5. Pemodelan pada tahap mendeteksi *outlier* pertama diperoleh persamaan

$$Z_t = 50,3857 + \frac{1}{(1-0,6157B)} a_t + 14,6778S_t^{(75)}$$

Kemudian untuk tahap dua *outlier* diperoleh sebagai berikut.

1. TC dan AO

$$Z_t = 50,1931 + \frac{1}{(1-0,6221B)} a_t + \frac{15,6873}{(1-0,7788B)} I_t^{(75)} + 20,2391 I_t^{(145)}$$

2. TC dan IO

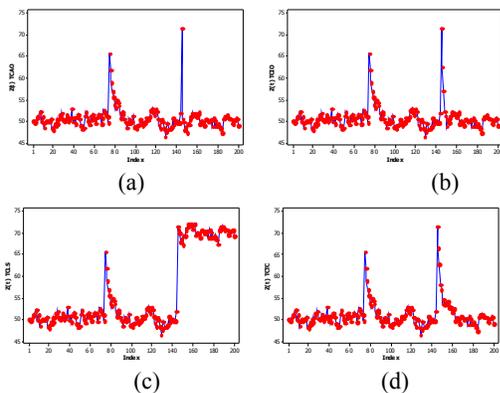
$$Z_t = 50,2198 + \frac{1}{(1-0,6155B)} a_t + \frac{15,6709}{(1-0,7780B)} I_t^{(75)} + \frac{20,1183}{(1-0,5553B)} I_t^{(145)}$$

3. TC dan LS

$$Z_t = 50,2465 + \frac{1}{(1-0,6224B)} a_t + \frac{15,6720}{(1-0,7766B)} I_t^{(75)} + 19,8276 S_t^{(145)}$$

4. TC dan TC

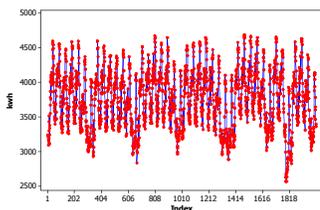
$$Z_t = 50,2536 + \frac{1}{(1-0,6261B)} a_t + \frac{15,6748}{(1-0,7762B)} I_t^{(75)} + \frac{20,0072}{(1-0,77428B)} I_t^{(145)}$$



Gambar 6. (a) TC-AO (b) TC-IO (c) TC-LS (d) TC-TC

G. Pemodelan Data I

Pada prosedur *Box-Jenkins* identifikasi model dilakukan pada data *in sample* yaitu pada tanggal 3 Maret 2014 sampai 13 April 2014. Plot time series yang didapatkan yaitu



Gambar 7. Plot time series Data I

Tabel 1.

Cek Diagnosa Pada Model Dugaan Data I

Model Dugaan	Uji <i>White Noise</i>		Uji Kenormalan
	Sampai Lag	p-value	
ARIMA ([3,18],1,1)(0,1,1) <sup>48</sup> (0,1,1) <sup>336</sup>	6	0,1579	<0,0100
	12	0,4022	
	18	0,1327	
	24	0,1019	
	30	0,0977	
	42	0,0817	

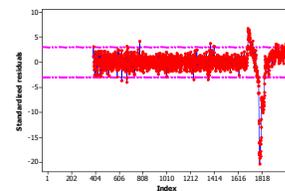
Data dilakukan transformasi *ln* dan dilakukan *differencing* sebanyak 1 lag, 48 lag, dan 336 lag karena memiliki karakteristik musiman ganda. Kemudian dilakukan pendugaan model, sehingga diperoleh model terbaik yaitu ARIMA ([3,18],1,1)(0,1,1)<sup>48</sup>(0,1,1)<sup>336</sup> dengan dilakukan cek diagnosa pada Tabel 1. Berdasarkan Tabel 1 diperoleh model sudah memenuhi asumsi *white noise* tetapi tidak memenuhi asumsi distribusi normal. Hal ini diduga karena adanya *outlier*.

H. Deteksi Outlier Data I

Deteksi *outlier* sudah terdapat pada beberapa *software* statistik. Namun untuk data musiman ganda dengan data yang sangat banyak belum terdapat *software* statistik yang dapat mendeteksi, termasuk SAS. Sehingga dengan langkah seperti data simulasi maka dilakukan pendeteksian *outlier* pada data beban listrik jangka pendek. Pada Data I ini dilakukan dengan 2 tahap untuk membentuk data yang berdistribusi normal. Langkah pertama dari 5 minggu pertama dari data I diperoleh pola yang stabil dan dilakukan pemodelan lalu digunakan mendeteksi *outlier* untuk waktu kedepan. Model 5 minggu pertama memiliki model ARIMA ([4],1,1)(0,1,1)<sup>48</sup>(0,1,1)<sup>336</sup> yang dituliskan dengan

$$Z_t = \frac{(1-0,3765B)(1-0,7229B^{48})(1-0,6236B^{336})}{(1+0,0694B^4)(1-B)(1-B^{48})(1-B^{336})} a_t.$$

Langkah selanjutnya dilakukan pembuatan plot *standardized residuals* untuk 1 minggu kedepan untuk mengetahui letak *outlier* sebagai berikut.

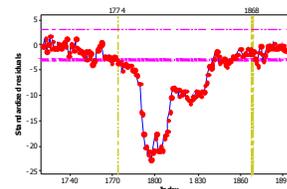


Gambar 8. Plot *Standardized residuals* Data I Sampai Minggu ke 6

Berdasarkan Gambar 7 diperoleh *outlier* pertama yaitu AO pada data ke 1698 dan diperoleh pemodelan yaitu

$$Z_t = \frac{(1-0,3765B)(1-0,7229B^{48})(1-0,6236B^{336})}{(1+0,0694B^4)(1-B)(1-B^{48})(1-B^{336})} a_t + 0,0372 I_t^{(1698)}$$

sedangkan ketika dilakukan pembuatan plot *standardized residuals* berdasarkan tahap 1 diperoleh *outlier* sesuai Gambar 8 sebagai berikut.



Gambar 9. *Outlier* Tahap 2 Untuk Data I

Berdasarkan Gambar 8 didapatkan *outlier* kedua adalah AO pada data ke 1790 dimana efeknya hingga 20 data setelahnya dan diperoleh model yang *white noise* dan berdistribusi normal. *Outlier* tersebut diprediksi karena pada waktu tersebut terjadi pemilu legislatif tanggal 9 April 2014. Selanjutnya pemodelan yang dapat dituliskan adalah

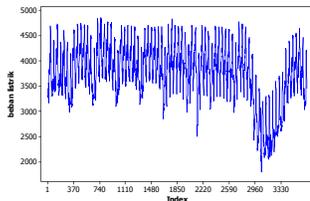
$$Z_t = \frac{(1-0,3802B)(1-0,7239B^{48})(1-0,6125B^{336})}{(1+0,0611B^4)(1-B)(1-B^{48})(1-B^{336})} a_t + 0,0344 I_t^{(1698)} - 0,1342 I_t^{(1790)} - 0,1407 I_t^{(1791)} - 0,2329 I_t^{(1792)} - 0,2902 I_t^{(1793)} - 0,2886 I_t^{(1794)} - 0,3169 I_t^{(1795)} - 0,2929 I_t^{(1796)} - 0,3187 I_t^{(1797)} - 0,2819 I_t^{(1798)} - 0,2734 I_t^{(1799)} - 0,2146 I_t^{(1800)} - 0,1963 I_t^{(1801)} - 0,2518 I_t^{(1802)} - 0,2619 I_t^{(1803)} - 0,2302 I_t^{(1804)} - 0,2490 I_t^{(1805)} - 0,1927 I_t^{(1806)} - 0,1889 I_t^{(1807)} - 0,1633 I_t^{(1808)} - 0,1171 I_t^{(1809)} - 0,0853 I_t^{(1810)}.$$

Berdasarkan pemodelan tersebut kemudian dilakukan validasi terhadap data *out sample* tanggal 14 April 2014 sebanyak 48 data. Perhitungan yang dilakukan menggunakan MAPE dan RMSE sebagai berikut. Berdasarkan MAPE dan RMSE model terbaik yaitu ARIMA ([4],1,1)(0,1,1)<sup>48</sup>(0,1,1)<sup>336</sup>

dengan *outlier* karena didapatkan MAPE dan RMSE terkecil sebesar 2,499% dan 106,778.

### I. Pemodelan Data II

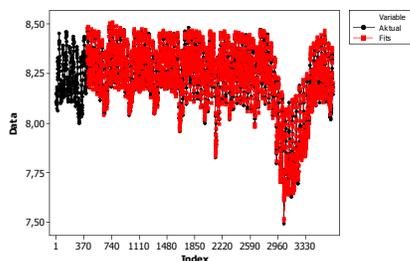
Pada data II yaitu tanggal 26 Mei 2014 hingga 9 Agustus 2014 memiliki plot time series sebagai berikut.



Gambar 10. Plot time series Data II

Gambar 9 menunjukkan plot yang melonjak kebawah diakhir data, hal tersebut dikarenakan efek hari Raya Idul Fitri dan pemilu presiden tanggal 9 Juli 2014. Dengan melihat plot time series data II sudah diduga adanya *outlier*. Namun dalam penelitian ini akan dilakukan pendeteksian *outlier* dengan asumsi penyebab tidak diketahui. Langkah awal adalah memodelkan data keseluruhan untuk mengetahui apakah data memang terdapat *outlier* dan pemodelan yang sesuai dilakukan transformasi 0,5 karena nilai lambda sebesar 0,5 dan *differencing* sebanyak 1 lag, 48 lag, dan 336 lag. Hasil pemodelan dengan estimasi parameter yang signifikan yaitu ARIMA  $([2,3,7,16,17],1,1)(0,1,1)^{48}(0,1,1)^{336}$  dengan pengujian *white noise* didapatkan asumsi *white noise* hanya sampai lag 12 dan 18 sedangkan *p-value* untuk uji *Kolmogorov-Smirnov* sebesar  $<0,0100$  yang berarti data tidak normal dan perlu dilakukan deteksi *outlier*.

Deteksi *outlier* pada data II dengan jumlah data *in sample* sebanyak 3648 dilakukan dengan langkah yang sama pada data I dan didapatkan model terbaik dari pendeteksian tersebut adalah ARIMA  $([3,5,8,16],1,[1,2])(0,1,1)^{48}(0,1,1)^{336}$  dengan *outlier* sebanyak 177. Ketika dilakukan pendeteksian tersebut diperoleh model yang telah *white noise* dan berdistribusi normal. Kemudian ketika dilakukan pembuatan plot time series data *in sample* dengan ramalan *out sample* dan data aktual diperoleh plot dengan pola yang sama seperti Gambar 10. Berdasarkan *out sample* diperoleh MAPE sebesar 4,503% dan RMSE sebesar 178,834.



Gambar 11. Data Aktual dan *fits*

## V. KESIMPULAN DAN SARAN

Berdasarkan pembahasan yang dilakukan maka disimpulkan bahwa pendeteksian *outlier* pada simulasi AR (1) sudah mampu mengatasi *outlier* pada data. Kemudian metode deteksi *outlier* yang dilakukan pada data simulasi diterapkan pada data beban listrik jangka pendek dengan dua data. Deteksi *outlier* pada data I dan data II menghasilkan asumsi residual

yang telah terpenuhi baik *white noise* dan distribusi normal, dengan model terbaik untuk data I yaitu model ARIMA  $([4],1,1)(0,1,1)^{48}(0,1,1)^{336}$  dengan *outlier* AO pada data ke 1698 dan AO pada data ke 1790 dimana efeknya hingga 20 data, jadi diperoleh 22 *outlier* pada data I. Selanjutnya untuk data II didapatkan model terbaik adalah ARIMA  $([3,5,8,16],1,[1,2])(0,1,1)^{48}(0,1,1)^{336}$  dengan *outlier* sebanyak 177 *outlier*.

Kemudian saran untuk penelitian selanjutnya dapat melakukan pengembangan pendeteksian *outlier* melalui data simulasi yang memang tidak berdistribusi normal. Sehingga adanya *outlier* pada data simulasi merupakan efek *outlier* sebenarnya dari data bangkitan dan dapat diketahui apakah metode pada penelitian ini dapat divalidasi dari data simulasi tersebut.

## DAFTAR PUSTAKA

- [1] Hinojosa, V. H., & Hoese, A. (2010). Short-Term Load Forecasting Using Fuzzy Inductive Reasoning and Evolutionary Algorithms. *IEEE Transactions on Power Systems*, 25 (1), 565-574.
- [2] Suswanto, D. (2009). *Sistem Distribusi Tenaga Listrik, Edisi Pertama*. Padang: Jurusan Teknik Elektro Universitas Padang.
- [3] Abdullah, A. G. (2008). Short Term Load Forecasting (STLF) Melalui Pendekatan Logika Fuzzy. *Electrans*, 7, 1-6.
- [4] Endharta, A. J., & Suhartono. (2009). Peramalan Konsumsi Listrik Jangka Pendek Dengan ARIMA Musiman Ganda dan ELMAN-Recurrent Neural Network. *Jurnal Ilmiah Teknologi Informasi (JUTI)*, 7 (4), 185-192.
- [5] Kotillova, A. (2011). Very Short-Term Load Forecasting Using Exponential Smoothing And ARIMA Models. *Journal of Information, Control and Management Systems*, 9 (2), 85-92
- [6] Mohamed, N., Ahmad, M. H., Suhartono, & Ismail, Z. (2011). Improving Short Term Load Forecasting Using Double Seasonal Arima Model. *World Applied Sciences Journal*, 15 (2), 223-231.
- [7] Rothe, J. P., Wadhwani, A. K., & Wadhwani, S. (2009). Short Term Load Forecasting Using Multi Parameter Regression. *(IJCSIS) International Journal of Computer Science and Information Security*, 6 (2), 303-306.
- [8] Sa'diyah, H. (2008). *Model ARIMA Musiman Ganda Untuk Peramalan Beban Listrik Jangka Pendek di PT. PLN (Persero) Gresik*. Tugas Akhir, Institut Teknologi Sepuluh Nopember
- [9] Taylor, J. W., & McSharry, P. E. (2008). Short-Term Load Forecasting Methods: An Evaluation Based on European Data. *IEEE Transactions on Power Systems*, 22, 2213-2219.
- [10] Taylor, J. W., Menezes, L. M., & McSharry, P. E. (2006). A Comparison of Univariate Methods for Forecasting Electricity Demand Up to a Day Ahead. *International Journal of Forecasting*, 22, 1-16.
- [11] Trapsilasiwi, R. K. (2011). *Peramalan Beban Listrik Jangka Pendek Menggunakan Hybrid Improved Particle Swarm Optimization-Support Vector Machine di PT. PLN Region Jawa Timur Bali*. Tugas Akhir, Institut Teknologi Sepuluh Nopember.
- [12] Tsay, R. S. (1988). Outliers, Level Shifts, and Variance Changes in Time Series. *Journal of Forecasting*, 7, 1-20.
- [13] Gaspersz, P. (2009). *Deteksi Outlier Pada Pemodelan Indeks Harga Konsumen Kota Ambon*. Thesis, Institut Teknologi Sepuluh Nopember.
- [14] Chen, C., & Liu, L. (1993). Joint Estimation of Model Parameters and Outliers Effects in Time Series. *Journal of The American Statistical Association*, 88 (421), 284-297.
- [15] Fox, A. J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society*, 3, 350-363.
- [16] Lin, F., Le, W., & Bo, J. (2010). Research on Maximal Frequent Pattern Outlier Factor for Online High-Dimensional Time-Series Outlier Detection. *Journal of Convergence Information Technology*, 5 (10), 66-71.
- [17] Liu, L.-M., & Hudak, G. B. (1992-1994). *Forecasting And Time Series Analysis using The SCA Statistical System*. U.S.A: Scientific Computing Associates Corp.
- [18] Wei, W. W. S., (2006). *Time Series Analysis*. New York: Addison Wesley.