

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Meramalkan sesuatu berdasarkan ilmu pengetahuan merupakan sesuatu yang dianjurkan dalam Islam, sebagaimana yang diceritakan dalam Al-qur'an dalam surat Yusuf ayat 47- 48, yaitu:

قَالَ تَزْرَعُونَ سَبْعَ سِنِينَ دَأْبًا فَمَا حَصَدْتُمْ فَذَرُوهُ فِي سُنْبُلِهِ إِلَّا قَلِيلًا مِّمَّا تَأْكُلُونَ ﴿٤٧﴾ ثُمَّ يَأْتِي مِنْ بَعْدِ ذَلِكَ سَبْعٌ شِدَادٌ يَأْكُلْنَ مَا قَدَّمْتُمْ لَهُنَّ إِلَّا قَلِيلًا مِّمَّا تُحْصِنُونَ

Artinya :

"Yusuf berkata "supaya kamu bertanam tujuh tahun (lamanya) sebagaimana biasa, maka apa yang kamu tuai hendaknya kamu biarkan dibulirnya kecuali sedikit untuk kamu makan. Kemudian sesudah itu akan datang tujuh tahun yang amat sulit, yang akan menghabiskan apa yang kamu simpan untuk menghadapinya (tahun sulit), kecuali dari bibit gandum yang kamu simpan".

Ayat di atas tersirat makna bahwa Nabi Yusuf diperintah oleh Allah untuk merencanakan ekonomi pertanian untuk masa lima belas tahun, hal ini dilakukan untuk menghadapi terjadinya krisis pangan menyeluruh atau musim paceklik. Menghadapi masalah ini Nabi Yusuf memberikan usul diadakannya perencanaan pembangunan pertanian yang akhirnya praktik pelaksanaannya diserahkan kepada Nabi Yusuf, berkat perencanaan yang matang itulah Mesir dan daerah-daerah sekelilingnya turut mendapat berkahnya.[4]

Penggalan berita lain yang disampaikan Al Qur'an tentang peristiwa masa depan ditemukan dalam ayat pertama Surat Ar Ruum, yang merujuk pada kekaisaran Bizantium, wilayah timur Kekaisaran Romawi. Dalam ayat-ayat ini, disebutkan bahwa Kekaisaran Bizantium telah mengalami kekalahan besar, tetapi akan segera memperoleh kemenangan.[4]

الْمَرْ ۝ غَلَبَتِ الرُّومُ ۝ فِي أَدْنَى الْأَرْضِ وَهُمْ مِنْ بَعْدِ غَلَبِهِمْ سَيَغْلِبُونَ ۝ فِي بَضْعِ

سِنِينَ ۝ لِلَّهِ الْأَمْرُ مِنْ قَبْلُ وَمِنْ بَعْدِ ۝ وَيَوْمَئِذٍ يَفْرَحُ الْمُؤْمِنُونَ ۝

Artinya:

"Alif, Lam, Mim. Telah dikalahkan bangsa Romawi, di negeri yang terdekat dan mereka sesudah dikalahkan itu akan menang, dalam beberapa tahun (lagi). Bagi Allah-lah urusan sebelum dan sesudah (mereka menang). Dan di hari (kemenangan bangsa Romawi) itu bergembiralah orang-orang yang beriman". (Al Qur'an, 30:1-4)

Ayat-ayat ini diturunkan kira-kira pada tahun 620 Masehi, hampir tujuh tahun setelah kekalahan hebat Bizantium Kristen di tangan bangsa Persia, ketika Bizantium kehilangan Yerusalem. Kemudian diriwayatkan dalam ayat ini bahwa Bizantium dalam waktu dekat menang. Padahal, Bizantium waktu itu telah menderita kekalahan sedemikian hebat hingga nampaknya mustahil baginya untuk mempertahankan keberadaannya sekalipun, apalagi merebut kemenangan kembali.[4]

Peramalan dalam matematika adalah memperkirakan apa yang terjadi di masa yang akan datang, sedangkan ramalan adalah hasil dari perkiraan peramalan untuk menaksir kejadian yang akan datang diperlukan suatu data yaitu data masa lampau ($t - 1$), data masa sekarang (t), dan data di masa yang akan datang ($t + 1$).[4]

Peramalan adalah penggunaan data masa lalu dari sebuah variabel atau kumpulan variabel untuk mengestimasi nilainya di masa yang akan datang. Sedangkan rencana merupakan penentuan apa yang akan dilakukan pada waktu yang akan datang. Peramalan diperlukan karena adanya perbedaan kesenjangan waktu (*time lag*) antara kesadaran akan dibutuhkannya suatu kebijakan baru dengan waktu pelaksanaan kebijakan tersebut. Apabila perbedaan waktu tersebut panjang, maka peran peramalan begitu penting dan sangat dibutuhkan, terutama dalam penentuan kapan terjadinya sesuatu sehingga dapat dipersiapkan tindakan yang perlu dilakukan.[4]

Dalam kehidupan sosial segala sesuatu itu serba tidak pasti, sukar diperkirakan secara tepat. Dalam hal ini perlu diadakan peramalan. Peramalan yang dibuat selalu diupayakan agar dapat meminimumkan pengaruh ketidakpastian ini terhadap sebuah permasalahan. Dengan kata lain peramalan bertujuan mendapatkan peramalan yang bisa meminimumkan kesalahan meramal (*forecast error*) yang biasanya diukur dengan *mean square error*, *mean absolute error*, dan sebagainya.[6]

Kemajuan ilmu pengetahuan telah meningkatkan pengertian mengenai berbagai aspek lingkungan dan akibatnya banyak peristiwa yang dapat diramalkan. Kemampuan menduga berbagai peristiwa kini tampaknya akan sama lazimnya dengan kecermatan peramalan keadaan cuaca ataupun curah hujan dalam berbagai periode. Kecenderungan untuk dapat meramalkan peristiwa secara lebih tepat, khususnya dalam bidang ekonomi, akan terus menerus memberikan dasar yang lebih baik bagi perencanaan.[6]

Kebutuhan peramalan cuaca merupakan kebutuhan utama dalam berbagai sektor, sehingga upaya pengembangan metode peramalan menuju ketepatan dan keakuratan informasi cuaca yang tinggi sangat diperlukan. Berbagai metode peramalan cuaca telah dikembangkan hingga saat ini. Salah satu unsur dari peramalan cuaca adalah curah hujan. Peramalan curah hujan secara terperinci sampai ke wilayah-wilayah masih sangat sedikit. Menurut Wilson faktor yang mempengaruhi banyak curah hujan adalah kelembaban nisbi, tekanan udara, suhu udara, dan kecepatan angin yang dapat dicari korelasinya untuk meramalkan curah hujan. Penelitian tentang peramalan curah hujan untuk suatu wilayah sangat dibutuhkan untuk menunjang pembangunan di berbagai bidang, termasuk bidang pertanian.

Model cuaca sebagian besar menggunakan pendekatan deterministik, khususnya kondisi pada saat t atau $t - 1$ dan $t + 1$ dengan periode harian. Namun untuk meramalkan 2 hari ke depan atau lebih model deterministik tingkat keakuratannya kurang baik sehingga model stokastik dapat dilakukan untuk memprediksi masalah tersebut. Model stokastik ini seringkali digunakan untuk pemodelan dengan periode bulanan dan tahunan. Model stokastik yang biasa digunakan untuk meramalkan cuaca yaitu metode *time series* yaitu *autoregresif*

integrated moving average (ARIMA), fourier regression, fractal analysis, trend surface analysis, dan neural network (NN). Model stokastik lain yang merupakan pengembangan dari metode ARIMA adalah metode ARIMAX.[11]

ARIMAX merupakan suatu perluasan dari ARIMA yang sangat memperhatikan variabel *independent*. Model ARIMAX cukup populer untuk peramalan jangka pendek karena memungkinkan untuk menggunakan dua variabel yaitu variabel terikat dan indikator utama untuk menentukan nilainya di masa depan [13]. ARIMAX telah berhasil diterapkan dalam konteks yang berbeda seperti penyebaran jumlah penyakit malaria (Kinley Wangdi, dkk. 2010) dan penjualan pakaian (Muhammad Hisyam Lee, Suhartono 2010). Hasil penelitian tersebut menunjukkan bahwa model ARIMAX lebih akurat dibanding dengan model ARIMA karena model ARIMAX ini sangat memperhatikan variabel *independent*. Variabel *independent* yang digunakan pada penelitian penyebaran penyakit malaria yaitu faktor iklim di antaranya curah hujan, kelembaban udara dan suhu udara. Pada peramalan jumlah pakaian menggunakan efek variasi kalender sebagai variabel *independent*-nya. [13]

Peramalan curah hujan kebanyakan menggunakan model ARIMA atau Seasonal ARIMA. model ini hanya melibatkan satu variabel saja yaitu satu data runtun waktu dalam sistem. Namun, karena model ini hanya melibatkan satu variabel yaitu variabel *dependent* sehingga model ini tidak bisa mengungkapkan hubungan antar variabel *dependent* dengan variabel *independent*-nya. Oleh karena itu dalam penelitian ini model model ARIMA dan SARIMA diperluas dengan menambahkan variabel *independent* sebagai variabel *eXogen*. Variabel *eXogen* ini merupakan suatu variabel yang mempengaruhi variabel utama.[5]

Pola cuaca di Indonesia cenderung membentuk pola musiman, termasuk curah hujannya juga membentuk pola musiman oleh karena itu perlu diramalkan melalui model musiman [11]. Dengan menambahkan unsur musiman pada model ARIMAX maka model tersebut disebut Seasonal ARIMAX atau SARIMAX. Model ARIMAX dan SARIMAX ini merupakan suatu model yang sangat memperhatikan variabel *independent* sehingga model ini dirasa cukup untuk mengatasi masalah musiman pada curah hujan di Indonesia. Variabel *independent* yang digunakan pada penelitian

ini adalah faktor-faktor yang mempengaruhi curah hujan di antaranya kelembaban udara atau kelembaban nisbi, tekanan udara, suhu udara, dan kecepatan angin. Variabel *independent* ini dalam model ARIMAX dan SARIMAX digunakan sebagai variabel *eXogen*. Dengan alasan itu penulis tertarik untuk memodelkan curah hujan harian dengan menggunakan metode stokastik untuk model ARIMAX dan SARIMAX.

Berdasarkan pemaparan di atas maka pada tugas skripsi ini penulis menarik topik “**Model ARIMAX dan SARIMAX untuk Meramalkan Data Curah Hujan**”.

1.2. Rumusan Masalah

Adapun rumusan masalah yang terkait dalam penelitian ini adalah sebagai berikut.

- a. Bagaimana bentuk model ARIMAX dan SARIMAX?
- b. Bagaimana tahapan dalam pemodelan ARIMAX dan SARIMAX?
- c. Model apa yang cocok digunakan untuk meramalkan data curah hujan?

1.3. Batasan Masalah

Batasan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut.

- a. Model yang digunakan untuk membentuk model curah hujan adalah model ARIMAX dan SARIMAX
- b. Data yang digunakan adalah data curah hujan harian dari 1 Desember 2011 sampai 30 November 2012 untuk wilayah Cemara Kelurahan Pasteur Kecamatan Pasirkaliki Bandung yang merupakan data sekunder dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) Stasiun Geofisika Kelas I Bandung yang beralamat di Jl. Cemara No. 66 Bandung
- c. Kriteria Pemilihan model dilakukan dengan kriteria *Akaike's Information Criteria (AIC)*.

- d. Dari model terbaik yang diperoleh akan dilakukan peramalan curah hujan harian untuk 30 hari ke depan
- e. Data diolah dengan menggunakan software R

1.4. Tujuan dan Manfaat

Tujuan dari penelitian ini adalah:

- a. Mengetahui bentuk model ARIMAX dan SARIMAX
- b. Mengetahui tahapan-tahapan dalam pemodelan ARIMAX dan SARIMAX
- c. Mengetahui model yang cocok untuk peramalan curah hujan di Indonesia menggunakan model ARIMAX dan SARIMAX.

Adapun manfaat dari penelitian ini yaitu dapat dijadikan sebagai informasi khususnya untuk wilayah yang dijadikan sebagai tempat penelitian dapat mengetahui seberapa besar curah hujan untuk 30 hari ke depan.

1.5. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah:

1. Studi Pustaka

Yaitu dengan mempelajari literatur-literatur yang berhubungan dengan masalah yang dibahas yaitu tentang model ARIMAX dan SARIMAX.

2. Pengumpulan data

Dilakukan dengan mengambil sampel data curah hujan harian dan faktor-faktor yang mempengaruhi curah hujan yaitu suhu udara, tekanan udara, kelembaban udara, dan kecepatan angin harian untuk wilayah Cemara Kelurahan Pasteur Kecamatan Pasirkaliki Bandung ke Badan Meteorologi Klimatologi dan Geofisika (BMKG) Stasiun Geofisika Kelas I Bandung yang beralamat di Jl. Cemara No. 66 Bandung.

3. Pengolahan data

Dilakukan dengan menganalisis data curah hujan dengan menggunakan model ARIMAX dan SARIMAX. Dalam analisis ini, data dianalisis dengan dua model yaitu model ARIMAX yang tidak melibatkan unsur musiman kemudian data juga dianalisis dengan model SARIMAX yang melibatkan unsur musiman. Kemudian dari hasil

model ARIMAX dan SARIMAX yang terbentuk keduanya dibandingkan dengan melihat nilai AIC terkecil yang akan dijadikan sebagai keperluan dalam peramalan curah hujan.

1.6. Sistematika Penulisan

Adapun sistematika penulisan dalam penelitian ini adalah:

BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini terdiri dari pengertian peramalan, metode peramalan, ketepatan metode peramalan, estimasi maksimum likelihood, pengertian curah hujan, dan faktor-faktor yang mempengaruhi curah hujan.

BAB III MODEL ARIMAX DAN SARIMAX

Bab ini terdiri dari model peramalan nonstasioner dan tahapan model ARIMAX dan SARIMAX yaitu tahap pembagian data, tahap identifikasi, tahap estimasi parameter, tahap uji diagnostik, tahap peramalan, dan tahap pemilihan model.

BAB IV STUDI KASUS DAN ANALISIS DATA CURAH HUJAN MENGGUNAKAN MODEL ARIMAX DAN SARIMAX

Bab ini terdiri dari sumber data, analisis data menggunakan model ARIMAX, dan analisis data menggunakan model SARIMAX.

BAB V PENUTUP

Bab ini terdiri dari kesimpulan dan saran.

BAB II LANDASAN TEORI

2.1. Pengertian Peramalan

Peramalan adalah penggunaan data masa lalu dari sebuah variabel atau kumpulan variabel untuk mengestimasi nilainya di masa yang akan datang. Sedangkan rencana merupakan penentuan apa yang akan dilakukan pada waktu yang akan datang. Peramalan diperlukan karena adanya perbedaan kesenjangan waktu (*time lag*) antara kesadaran akan dibutuhkannya suatu kebijakan baru dengan waktu pelaksanaan kebijakan tersebut. Apabila perbedaan waktu tersebut panjang, maka peran peramalan begitu penting dan sangat dibutuhkan, terutama dalam penentuan kapan terjadinya sesuatu sehingga dapat dipersiapkan tindakan yang perlu dilakukan.[6]

Dalam kehidupan sosial segala sesuatu itu serba tidak pasti, sukar diperkirakan secara tepat. Dalam hal ini perlu diadakan peramalan. Peramalan yang dibuat selalu diupayakan agar dapat meminimumkan pengaruh ketidakpastian ini terhadap sebuah permasalahan. Dengan kata lain peramalan bertujuan mendapatkan peramalan yang bisa meminimumkan kesalahan meramal (*forecast error*) yang biasanya diukur dengan *mean square error*, *mean absolute error*, dan sebagainya.

Kegunaan peramalan terlihat pada saat pengambilan keputusan. Keputusan yang baik adalah keputusan yang didasarkan atas pertimbangan-pertimbangan yang akan terjadi pada waktu keputusan itu dilaksanakan.

Keberhasilan dari suatu peramalan sangat ditentukan oleh:

1. Pengetahuan teknik tentang pengumpulan informasi (data) masa lalu, data ataupun informasi tersebut bersifat kuantitatif
2. Teknik dan metode yang tetap dan sesuai dengan pola data yang telah dikumpulkan

Apabila data yang digunakan tidak dapat meyakinkan maka hasil peramalan yang disusun juga akan sukar dipercaya ketepatannya. Oleh karena itu, ketepatan dari ramalan tersebut merupakan hal yang sangat penting. Walaupun demikian perlu

disadari bahwa suatu ramalan adalah tetap ramalan, dimana selalu ada unsur kesalahannya. Sehingga yang penting diperhatikan adalah usaha untuk memperkecil kesalahan tersebut.

Data runtun waktu adalah data yang disusun berdasarkan urutan waktu atau data yang dikumpulkan dari waktu ke waktu. Waktu yang digunakan bisa berupa hari, minggu, bulan, tahun, dan sebagainya. Dengan demikian data *time series* berhubungan dengan data statistik yang dicatat dan diselidiki dalam batas-batas (interval) waktu tertentu, seperti penjualan, harga, persediaan produksi, dan tenaga kerja. [8]

Analisis runtun waktu merupakan suatu metode analisis data yang ditujukan untuk melakukan suatu estimasi maupun peramalan pada masa yang akan datang. Dalam analisis runtun waktu akan diketahui bagaimana proses suatu estimasi dan hasil peramalan dapat diperoleh dengan baik [8]. Beberapa model yang cukup populer untuk melakukan analisis terhadap data runtun waktu yaitu model autoregresif (AR), *moving average* (MA), gabungan AR dan MA (ARMA), ARIMA dan ARIMA dengan tambahan variabel *eXogen* (ARIMAX) [9].

2.1.1 Jangka Waktu Peramalan

Menurut waktunya, peramalan dapat dikelompokkan menjadi 3 yaitu [6]:

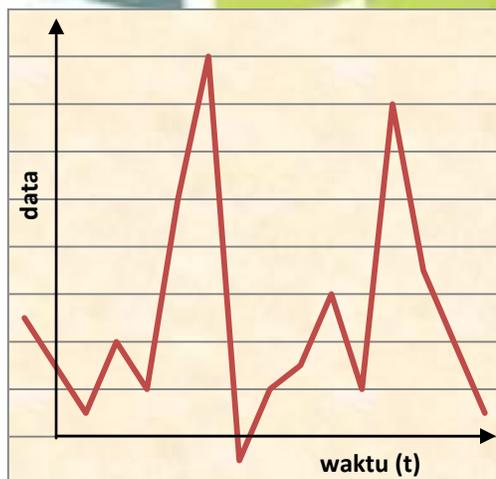
1. Peramalan jangka pendek, yaitu kurang 3 bulan sampai 1 tahun
2. Peramalan jangka menengah, yaitu antara 3 bulan sampai 3 tahun
3. Peramalan jangka panjang, yaitu 3 tahun atau lebih.

2.1.2 Pola Data dalam Peramalan

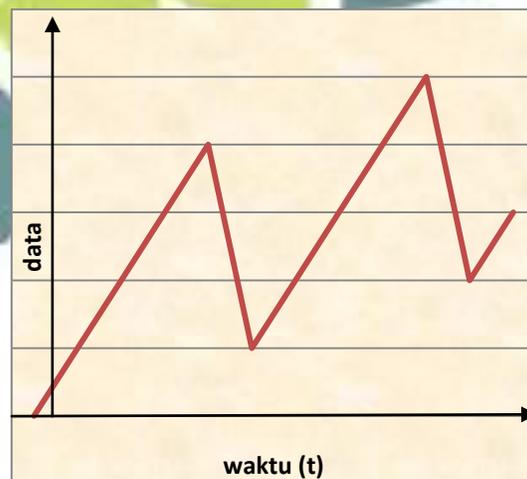
Langkah penting dalam memilih metode peramalan adalah dengan mempertimbangkan jenis pola data. Pola data dalam peramalan dapat dibedakan menjadi empat jenis yaitu pola horizontal, pola musiman, pola siklus, dan pola trend[6].

- a. Pola horizontal, terjadi bilamana nilai data berfluktuasi disekitar nilai rata-rata yang konstan. Deret seperti ini adalah stasioner terhadap rata-ratanya. Suatu proses kontinu yang secara teoritis tidak mengalami perubahan.
- b. Pola musiman, terjadi bilamana suatu deret dipengaruhi oleh faktor musiman (misalnya kuartal tahun tertentu, bulanan, atau hari-hari pada minggu tertentu). Pola musiman menggambarkan pola yang berulang setiap periode. Komponen musiman dapat dijabarkan ke dalam faktor cuaca, liburan, atau kecenderungan perdagangan. Pola musiman berguna dalam meramalkan peramalan jangka pendek. Pola data ini terjadi bila nilai data sangat dipengaruhi oleh musim.
- c. Pola siklis, terjadi bilamana datanya dipengaruhi oleh fluktuasi ekonomi jangka panjang seperti yang berhubungan dengan siklus bisnis. Penjualan produk dapat memiliki siklus yang berulang secara periodik. Banyak produk dipengaruhi pola pergerakan aktivitas ekonomi yang terkadang memiliki kecenderungan periodik. Komponen siklis ini sangat berguna dalam peramalan jangka menengah.
- d. Pola trend, terjadi bilamana terdapat kenaikan atau penurunan sekuler jangka panjang dalam data. Pola data ini terjadi bila data memiliki kecenderungan untuk naik atau turun terus menerus. Dalam meramalkan biaya-biaya yang termasuk di dalam biaya operasi dipergunakan pola trend karena biaya tersebut cenderung naik jika mesin/peralatan semakin tua atau semakin lama jangka waktu pemakaiannya.

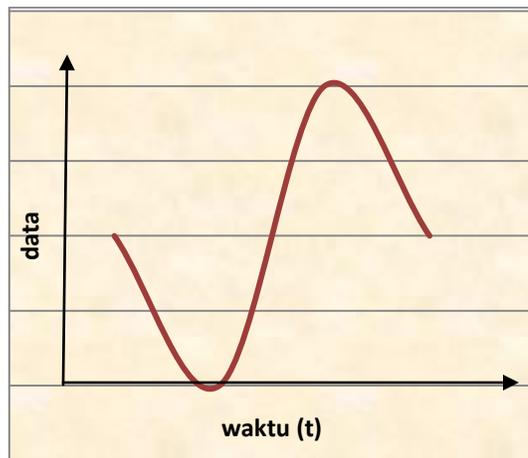
Plot untuk masing-masing pola data disajikan dalam gambar sebagai berikut.



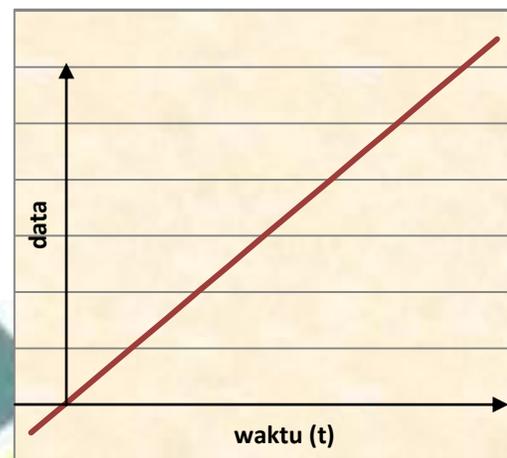
Gambar 2.1 Pola Horizontal



Gambar 2.2 Pola Musiman



Gambar 2.3 Pola Siklus



Gambar 2.4 Pola Trend

2.2. Metode Peramalan

Metode peramalan adalah suatu cara memperkirakan atau mengestimasi secara kuantitatif maupun kualitatif apa yang akan terjadi pada masa depan, berdasarkan data yang relevan pada masa lalu. Kegunaan metode peramalan adalah untuk memperkirakan secara sistematis dan pragmatis atas dasar data yang relevan pada masa lalu. Dengan demikian peramalan diharapkan dapat memberikan objektivitas yang lebih besar. Metode peramalan memberikan urutan dan pemecahan atas pendekatan masalah dalam peramalan, sehingga bila digunakan pendekatan yang sama atas permasalahan, maka akan didapat dasar pemikiran dan pemecahan yang argumentasinya sama [6].

2.2.1 Jenis Metode Peramalan

Berdasarkan sifatnya, peramalan dibedakan atas dua macam yaitu:

1. Peramalan kualitatif

Peramalan kualitatif adalah peramalan yang didasarkan atas data kualitatif pada masa lalu. Hasil peramalan ini sangat bergantung pada orang yang menyusunnya. Hal ini penting karena hasil peramalan tersebut ditentukan berdasarkan pemikiran yang bersifat intuisi, pendapat dan pengetahuan serta pengalaman dari orang-orang yang menyusunnya. Metode ini digunakan dimana tidak ada model matematik, biasanya

dikarenakan data yang ada tidak cukup representatif untuk meramalkan masa yang akan datang (*long term forecasting*).[6]

2. Peramalan kuantitatif

Peramalan kuantitatif adalah peramalan yang didasarkan atas data kuantitatif pada masa lalu. Hasil peramalan ini sangat bergantung pada metode yang dipergunakan dalam peramalan tersebut. Karena dengan metode yang dipergunakan dalam peramalan yang berbeda pula. Baik tidaknya metode yang dipergunakan ditentukan oleh perbedaan atau penyimpangan antara hasil ramalan dengan keyakinan yang terjadi. Semakin kecil penyimpangan antara hasil ramalan dengan hasil ramalan dengan kenyataan yang terjadi berarti metode yang digunakan semakin baik. [6]

Peramalan kuantitatif dapat diterapkan bila terdapat kondisi berikut [6]:

1. Tersedia informasi (data) tentang masa lalu
2. Informasi (data) tersebut dapat dikuantitatifkan dalam bentuk data numerik
3. Dapat diasumsikan bahwa beberapa aspek pola masa lalu akan terus berlanjut pada masa yang akan datang

Peramalan kuantitatif dibedakan atas :

1. Metode peramalan yang didasarkan atas penggunaan analisa pola hubungan antar variabel yang diperkirakan dengan variabel waktu yang merupakan runtun waktu (*time series*). Metode peramalan termasuk dalam jenis ini adalah:
 - a. Metode pemulusan (*smoothing*)
 - b. Metode Box Jenkins (ARIMA)
 - c. Metode proyeksi trend dengan regresi
2. Metode peramalan yang didasarkan atas penggunaan analisa pola hubungan antar variabel yang diperkirakan dengan variabel lain yang mempengaruhinya yang bukan waktunya disebut dengan metode korelasi atau sebab akibat. Metode peramalan yang termasuk jenis ini adalah:
 - a. Metode Regresi dan Korelasi
 - b. Metode Ekonometri
 - c. Metode Input-Output

2.2.2 Model Peramalan Stasioner

Suatu runtun waktu dikatakan stasioner jika tidak terdapat kecenderungan peningkatan atau penurunan pada data tersebut dalam selang yang cukup panjang atau dengan kata lain fluktuasi data berada di sekitar rata-rata dan varians konstan serta tidak tergantung pada waktu.

Pengujian stasioneritas dari suatu data runtun waktu dapat dilakukan dengan beberapa cara berikut [7].

1. Pendeteksian ketidakstasioneran data dalam mean (rata-rata) dapat dilihat dengan menggunakan plot dari data asli, plot *fungsi autokorelasi* (ACF) dan plot fungsi *autokorelasi parsial* (PACF). Jika data mengandung komponen trend, data nonstasioner dalam mean dan plot ACF dan PACF akan meluruh secara perlahan.
2. Pendeteksian ketidakstasioneran dalam varians dapat menggunakan plot ACF dan PACF dari residual kuadrat. Dapat juga dilihat dengan melihat nilai variansinya.

Model runtun waktu stasioner diantaranya sebagai berikut.

a. *Autoregresif* (AR)

Bentuk umum dari proses *autoregresif* tingkat p atau $AR(p)$ yaitu [3]:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (2.1)$$

Dimana:

ϕ_p : parameter dari *autoregresif*

Y_t : runtun waktu ke t

e_t : error pada waktu ke t

Pada model ini, peubah Y diregresikan terhadap nilai-nilai sebelumnya dari peubah Y itu sendiri, oleh karenaitulah model ini dikatakan sebagai model *autoregresif*.

b. Moving Average (MA)

Bentuk umum dari proses *Moving Average* tingkat q atau $MA(q)$ yaitu [3]:

$$Y_t = \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + e_t \quad (2.2)$$

Dimana:

θ_q : parameter dari *moving average*

Y_t : runtun waktu ke- t

e_t : error pada waktu ke- t

c. Autoregresif Moving Average (ARMA)

Proses random stasioner kadang tidak dapat dengan baik dijelaskan oleh model *moving average* saja atau *autoregressive* saja, karena proses tersebut mengandung keduanya. Oleh karena itu gabungan dari kedua model tersebut dinamakan model *autoregressive-moving average* dapat lebih efektif dipakai. Pada model ini, *series* stasioner adalah fungsi dari nilai lampaunya serta nilai sekarang dan lampau kesalahannya.[3]

Proses ARMA merupakan suatu proses yang digunakan untuk runtun waktu stasioner. Proses ARMA merupakan suatu perluasan yang diperoleh dari proses AR dan MA. ACF proses ARMA (p, q) menyerupai proses AR(p) yaitu turun secara eksponensial menuju nol. PACF proses ARMA (p, q) menyerupai proses MA(q) yaitu turun menuju nol. Pola acuan ACF dan PACF untuk runtun waktu stasioner dapat dilihat pada tabel berikut.[10]

Tabel 2.1

Pola Acuan ACF dan PACF untuk Proses Stasioner

Model	ACF	PACF
AR(p)	<i>Dies down</i> (Turun secara eksponensial)	<i>Cuts off after lag p</i> (Terpotong setelah lag ke- p)
MA(q)	<i>Cuts off after lag q</i> (Terpotong setelah lag ke- q)	<i>Dies down</i> (Turun secara eksponensial)
ARMA (p, q)	<i>Dies down</i> (Turun cepat secara eksponensial)	<i>Dies down</i> (Turun cepat secara eksponensial)
AR(p) atau MA(q)	<i>Cuts off after lag q</i> (Terpotong setelah lag ke- q)	<i>Cuts off after lag p</i> (Terpotong setelah lag ke- p)

Model umum untuk proses ARMA (p, q) didefinisikan sebagai berikut [4]:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

(2.3)

Dimana:

$\phi_1, \phi_2, \dots, \phi_p$ parameter-parameter *autoregresif* (AR)

$\theta_1, \theta_2, \dots, \theta_q$ parameter-parameter *moving average* (MA)

e_t kesalahan pada waktu ke- t

2.2.3 Model Peramalan Nonstasioner

Dalam dunia industri dan bisnis kebanyakan runtun waktu bersifat tak stasioner dan secara khusus tidak bervariasi disekitar mean yang tetap. Jika sifat dari runtun ini masih tampak homogen dalam arti fluktuasi yang terjadi di sekitar tingkat tertentu mungkin berbeda pada waktu yang berbeda pula, maka jika *difference* pada tingkat tersebut dilakukan maka fluktuasi satu dengan yang lain akan tampak mirip.[3]

a. Autoregresif Integrated Moving Average (ARIMA)

Metode *Autoregressive Integrated Moving Average* (ARIMA) telah dipelajari secara mendalam dan dikembangkan oleh George Box dan Gwilyn Jenkins (1976). Model *Autoregresif* (AR) pertama kali dikembangkan oleh Yule (1926) dan kemudian dikembangkan oleh Walker (1931), sedangkan model *moving average* (MA) pertama kali digunakan oleh Slutsky (1937).[6]

ARIMA sering juga disebut metode runtun waktu Box-Jenkins. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik. Biasanya akan cenderung *flat* (mendatar/konstan) untuk periode yang cukup panjang.

Model ARIMA adalah model yang secara penuh mengabaikan variabel *independent* dalam membuat peramalan dan digunakan untuk suatu proses nonstasioner. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel *dependent* untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari runtun waktu secara statistik berhubungan satu sama lain (*dependent*).

Notasi umum dari model ARIMA nonmusiman adalah [6] :

$$\mathbf{ARIMA}(p, d, q)$$

Sedangkan notasi umum untuk ARIMA musiman adalah:

$$\mathbf{ARIMA}(p, d, q)(P, D, Q)S$$

Dimana:

p merupakan model *autoregresif* nonmusiman order- p

q merupakan model *moving average* nonmusiman orde ke q

P merupakan model *autoregresif* musiman orde- P

Q merupakan model *moving average* musiman orde- Q

D merupakan pembedaan (*differencing*) musiman

d merupakan proses pembedaan (*differencing*) nonmusiman

S banyaknya musiman

Adapun model umum untuk ARIMA nonmusiman adalah [11]:

$$(1 - B)^d \phi_p(B) Y_t = \mu + \theta_q(B) e_t \quad (2.4)$$

Sedangkan model umum untuk Seasonal ARIMA (SARIMA) adalah [11]:

$$(1 - B)^d (1 - B^S)^D \phi_p(B) \Phi_P(B^S) Y_t = \mu + \theta_q(B) \Theta_Q(B^S) e_t \quad (2.5)$$

Dimana:

Y_t variabel *dependent*

B operator back shift atau operator balik dengan $BY_1 = Y_{t-1}$

$\phi_p(B)$ operator *autoregresif*, direpresentasikan sebagai polinomial dalam operator *backshift*: $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$

$\theta_q(B)$ operator *moving average*, direpresentasikan sebagai polinomial operator *backshift*: $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$

e_t error pada waktu ke- t

$(1 - B)^d$ orde pembeda nonmusiman

d pembeda nonmusiman

$(1 - B^S)^D$ orde pembeda musiman

$\Phi_P(B^S)$ operator *autoregresif* musiman dengan:

$$\Phi_P(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{SP}$$

$\Theta_Q(B^S)$ operator *moving average* musiman dengan:

$$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \dots - \Theta_Q B^{SQ}$$

S banyaknya musiman

D pembeda musiman

Sebelum membentuk sebuah model ARIMA pada suatu deret waktu maka harus dilakukan 4 proses yang dijalankan yaitu identifikasi model, estimasi model, pemeriksaan model (*diagnostic checking*), kemudian menggunakan model tersebut untuk prediksi pada masa mendatang.

b. Autoregresif Integrated Moving Average dengan Variabel eXogen (ARIMAX)

Salah satu model runtun waktu yang dapat dipandang sebagai perluasan model runtun waktu ARIMA adalah model ARIMAX. Adapun ARIMAX musiman biasa

disebut dengan Seasonal ARIMAX (SARIMAX). Seasonal ARIMAX (SARIMAX), yaitu model SARIMA dengan variabel *eXogen*. Dalam model ini faktor-faktor yang mempengaruhi variabel *dependent* Y pada waktu ke- t tidak hanya oleh fungsi variabel Y dalam waktu, tetapi juga oleh variabel-variabel *independent* lainnya pada waktu ke- t . Secara umum bentuk model ARIMAX (p, d, q) dapat diberikan dengan persamaan berikut [6]:

$$(1 - B)^d \phi_p(B) Y_t = \theta_q(B) e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \dots + \alpha_k X_{k,t} \quad (2.6)$$

Adapun bentuk model umum untuk SARIMAX (p, d, q)(P, D, Q) S :

$$(1 - B)^d (1 - B^S)^D \phi_p(B) \Phi_P(B^S) Y_t = \theta_q(B) \Theta_Q(B^S) e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \dots + \alpha_k X_{k,t} \quad (2.7)$$

Dimana:

$X_{k,t}$: variabel *independent* atau sebagai variable *eXogen* ke- k pada saat t dengan $k = 1, 2, 3, \dots, k$

Dalam model ini Y_t dan $X_{k,t}$ adalah data runtun waktu dan diasumsikan stasioner. Jika keduanya stasioner, alternatif model lain dapat digunakan adalah model *Autoregresif Distributed Lag* (ADL). Jika hanya Y_t yang tidak stasioner (mengandung trend), maka dapat menggunakan model ARIMAX atau SARIMAX dengan menambahkan komponen model pembeda ke dalam Y_t . [7]

Tahapan untuk memperoleh pemodelan ARIMAX dan SARIMAX akan dijelaskan pada bab berikutnya.

2.3. Ketepatan Metode Peramalan

Dalam peramalan, agar diperoleh model yang terbaik maka dapat dilihat dari nilai kesalahan atau error terkecil. Dalam peramalan, ketepatan peramalan dipandang sebagai kriteria penolakan atau penerimaan untuk memilih suatu metode peramalan. Adapun ketepatan dari peramalan tersebut dapat diukur dengan menggunakan ukuran-ukuran statistik. [7]

2.3.1 Ukuran Statistik Standar

Jika Y_t merupakan data aktual untuk periode t dan \hat{Y}_t merupakan ramalan untuk periode yang sama, maka kesalahan didefinisikan sebagai [6]:

$$e_t = Y_t - \hat{Y}_t \quad (2.8)$$

Jika terdapat nilai pengamatan dan ramalan untuk n periode waktu, maka akan terdapat n buah kesalahan dan ukuran statistik standar berikut dapat didefinisikan sebagai berikut.

Nilai Tengah Kesalahan (*Mean Error*)

$$ME = \sum_{t=1}^n \frac{e_t}{n} \quad (2.9)$$

Nilai Tengah Kesalahan Absolut (*Mean Absolute Error*)

$$MAE = \sum_{t=1}^n \frac{|e_t|}{n} \quad (2.10)$$

Jumlah Kuadrat Kesalahan (*Sum of square error*)

$$SSE = \sum_{t=1}^n e_t^2 \quad (2.11)$$

Nilai Tengah Kesalahan Kuadrat (*Mean square error*)

$$MSE = \sum_{t=1}^n \frac{e_t^2}{n} \quad (2.12)$$

Deviasi Standar Kesalahan (*Standard Deviation of Error*)

$$SDE = \sqrt{\sum_{t=1}^n \frac{e_t^2}{(n-1)}} \quad (2.13)$$

Sebagai suatu ukuran ketepatan peramalan, MSE mempunyai nilai yang terbatas. Ukuran ini tidak memudahkan perbandingan antar deret berkala yang berbeda dan untuk selang waktu yang berlainan karena MSE merupakan ukuran absolut dan menyangkut pengkuadratan sederetan nilai.

2.3.2 Ukuran Statistik Relatif

Karena alasan tersebut hubungannya dengan keterbatasan MSE sebagai suatu ukuran ketepatan peramalan maka diusulkan ukuran-ukuran alternatif yang di antaranya menyangkut kesalahan persentase. Ukuran yang sering digunakan tersebut yaitu:

Kesalahan Persentase (*Percentage error*)

$$PE_t = \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right) \times 100\% \quad (2.14)$$

Nilai Tengah Kesalahan Persentase (*Mean percentage error*)

$$MPE = \sum_{l=1}^n \frac{PE_t}{n} \quad (2.15)$$

Nilai Tengah Kesalahan Persentase Absolut (*Mean absolute percentage error*)

$$MAPE = \sum_{i=1}^n \frac{|PE_t|}{n} \quad (2.16)$$

2.4. Estimasi Maksimum Likelihood

Metode maksimum likelihood atau maksimum likelihood estimation (MLE) adalah metode yang memaksimumkan fungsi likelihood. Penaksir maksimum likelihood merupakan salah satu pendekatan terpenting pada penaksiran dalam semua inferensi statistika. Metode maksimum likelihood menggunakan pengetahuan tentang distribusi diskrit maupun kontinu dalam menentukan penaksir yang sesuai.[12]

Definisi: Bila diketahui $Y_1, Y_2, Y_3, \dots, Y_n$ merupakan pengamatan yang saling bebas, maka penaksir maksimum likelihood $\hat{\theta}$ merupakan perkalian dari pdf atau pmf yang memaksimumkan fungsi likelihood.[12]

$$L(Y_1, Y_2, \dots, Y_n, \theta) = f(Y_1, \theta) f(Y_2, \theta) \dots f(Y_n, \theta) = \prod_{i=1}^n f(Y_i, \theta) \quad (2.17)$$

Dengan θ parameter yang tidak diketahui.

Nilai θ yang memaksimumkan persamaan (2.17) disebut penaksir maksimum likelihood.

Dalam aplikasinya $L(\theta)$ menunjukkan pdf gabungan dari sampel acak. Jika ruang parameter yang merupakan interval terbuka dan $L(\theta)$ merupakan fungsi yang dapat diturunkan serta diasumsikan maksimum maka persamaan maksimum likelihoodnya adalah [12]:

$$\frac{d}{d\theta} L(\theta) = 0 \quad (2.18)$$

Jika penyelesaian dari persamaan (2.18) ada, maka maksimum dari $L(\theta)$ dapat terpenuhi. Apabila pada distribusi normal sulit diselesaikan maka fungsi $L(\theta)$ dapat

dibuat logaritma naturalnya dengan ketentuan memaksimumkan $\ln L(\theta)$, sehingga persamaan logaritma natural likelihoodnya adalah [12]:

$$\frac{d}{d\theta} \ln L(\theta) = 0 \quad (2.19)$$

Penaksir maksimum likelihood dari θ didapat dengan menyelesaikan persamaan $\frac{d}{d\theta} \ln L(\theta) = 0$, misalkan ada k parameter yang tidak diketahui maka penduga parameter likelihood dari θ_i diperoleh dengan menyelesaikan [12]:

$$\frac{d}{d\theta_i} \ln L(\theta_1, \theta_2, \dots, \theta_k) = 0 \quad (2.20)$$

Dengan $i=1, 2, \dots, k$

2.5. Pengertian Curah Hujan

Dalam buku Suyono, curah hujan adalah banyaknya air yang jatuh ke permukaan bumi. Derajat curah hujan dinyatakan dengan jumlah curah hujan dalam suatu satuan waktu. Biasanya satuan yang digunakan adalah mm/jam. Dalam meteorology butiran hujan dengan diameter lebih dari 0.5 mm disebut hujan dan diameter antara 0.1-0.5 mm disebut gerimis. Semakin besar ukuran butiran hujan maka semakin besar pula kecepatan jatuhnya. Ketelitian alat ukur curah hujan adalah 1/10 mm. pembacaan dilakukan satu kali dalam sehari dan dicatat sebagai curah hujan harian terdahulu atau kemarin.[9]

2.6. Faktor-faktor yang Mempengaruhi Curah Hujan

2.6.1 Kelembaban Nisbi

Kelembaban adalah perbandingan antara massa uap dalam suatu satuan volume dengan massa uap yang jenuh dalam satuan volume itu pada suhu yang sama. Secara umum kelembaban menyatakan banyaknya kadar air yang ada di udara. Banyaknya uap yang bergerak di dalam atmosfer berpengaruh terhadap besarnya hujan, lamanya hujan, dan intensitas curah hujan. Kelembaban tertinggi umumnya terjadi pada

musim penghujan dan paling rendah pada musim kemarau. Variasi kelembaban bergantung dari suhu udara, jika pagi hari suhu rendah maka kelembaban akan lebih tinggi jika dibanding pada siang hari saat suhu tinggi.[9]

Umumnya semakin tinggi suatu daerah dari permukaan laut maka kelembaban udaranya semakin tinggi. Makin tinggi kelembaban udara akan dapat menyebabkan bertambah banyak uap air yang dapat diserap awan. Uap air itu akan menghasilkan tekanan yang dinyatakan dengan satuan tinggi air raksa ($1\text{mmHg} = 1,33\text{milibar}$). Tekanan yang diberikan oleh uap air disebut dengan tekanan uap air. [9]

2.6.2. Suhu Udara

Suhu udara adalah keadaan panas atau dinginnya udara. Suhu juga disebut temperatur yang diukur dengan alat termometer. Beberapa faktor yang mempengaruhi suhu udara diantaranya tinggi tempat, daratan atau lautan, radiasi matahari, indeks datang matahari dan angin. Pengukuran biasa dinyatakan dalam skala Celsius (C), Reamur (R), dan Fahrenheit (F). Suhu udara tertinggi di permukaan bumi adalah di daerah tropis (sekitar ekuator) dan makin ke kutub makin tinggi. [9]

2.6.3. Tekanan Udara

Tekanan udara merupakan tenaga yang bekerja untuk menggerakkan massa udara dalam setiap satuan luas tertentu. Diukur dengan menggunakan barometer. Satuan tekanan udara adalah milibar (mb). Garis yang menghubungkan tempat-tempat yang sama tekanan udaranya disebut sebagai isobar. [9]

Tekanan udara dibatasi oleh ruang dan waktu. Artinya pada tempat dan waktu yang berbeda besarnya juga berbeda. Semakin tinggi suatu tempat maka tekanan udaranya semakin menurun, sedangkan tekanan udara pada daerah yang mempunyai rata-rata ketinggian sama maka tekanan udara dipengaruhi oleh suhu udara. Daerah yang suhu udaranya tinggi akan bertekanan rendah dan daerah yang bersuhu udara rendah tekannya tinggi. [9]

2.6.4. Kecepatan Angin

Angin adalah udara yang bergerak akibat adanya perbedaan tekanan udara dengan arah aliran angin dari tempat yang memiliki tekanan tinggi ke tempat yang bertekanan rendah atau dari daerah yang memiliki suhu atau temperatur rendah ke wilayah bersuhu tinggi.

Angin memiliki hubungan yang erat dengan sinar matahari karena daerah yang terkena banyak paparan sinar matahari akan memiliki suhu yang lebih tinggi serta tekanan udara yang lebih rendah dari daerah lain di sekitarnya sehingga menyebabkan terjadinya aliran udara. Angin juga dapat disebabkan oleh pergerakan benda sehingga mendorong udara di sekitarnya untuk bergerak ke tempat lain.[10]



BAB III

MODEL ARIMAX DAN SARIMAX

Curah hujan merupakan salah satu hal penting dalam kehidupan ini. Khususnya dalam bidang pertanian sebab curah hujan dapat mempengaruhi hasil pertanian. Oleh karena itu peramalan mengenai curah hujan sangat dibutuhkan untuk membantu proses tanam-menanam. Menurut Wilson faktor yang mempengaruhi banyak curah hujan adalah kelembaban udara, tekanan udara, suhu udara, dan kecepatan angin yang dapat dicari korelasinya untuk meramalkan curah hujan. [9]

Model peramalan yang biasa digunakan untuk peramalan curah hujan adalah ARIMA dan SARIMA. Namun pada penelitian ini model ARIMA diperluas dengan menambahkan variabel *eXogen* yaitu variabel *independent* ke dalam model sehingga model ARIMA ditambah variabel *eXogen* menjadi ARIMAX. Begitupun model SARIMA ditambahkan dengan variabel *eXogen* menjadi SARIMAX.

3.1 Model Peramalan Nonstasioner

Dalam dunia industri dan bisnis kebanyakan runtun waktu bersifat tak stasioner dan secara khusus tidak bervariasi disekitar mean yang tetap. Runtun waktu yang seperti ini disebut dengan runtun waktu nonstasioner. Jika sifat dari runtun ini masih tampak homogen dalam arti fluktuasi yang terjadi di sekitar tingkat tertentu mungkin berbeda pada waktu yang berbeda pula, maka jika *difference* pada tingkat tersebut dilakukan maka fluktuasi satu dengan yang lain akan tampak mirip.[3]

Berepa model runtun waktu nonstasioner diantaranya adalah ARIMA, SARIMA, ARIMAX dan SARIMAX. Namun yang akan dibahas pada bab ini hanya model ARIMAX dan SARIMAX.

A. Model *Autoregresif Integrated Moving Average* dengan Tambahan Variabel *eXogen* (ARIMAX)

ARIMA merupakan model peramalan yang sangat mengabaikan variabel *independent*. Model ARIMAX merupakan salah satu model runtun waktu yang dapat

dipandang sebagai perluasan model runtun waktu ARIMA. ARIMAX merupakan suatu model peramalan nonstasioner yang sangat memperhatikan variabel *independent*. ARIMAX merupakan perluasan dari ARIMA dengan tambahan variabel *eXogen* yaitu variabel *independent*. Model ARIMAX cukup populer untuk peramalan jangka panjang karena memungkinkan untuk menggunakan dua variabel yaitu variabel terikat dan indikator utama untuk menentukan nilainya di masa yang akan datang [13].

ARIMAX telah berhasil diterapkan dalam konteks yang berbeda seperti penyebaran jumlah penyakit malaria (Kinley Wangdi, dkk. 2010) dan penjualan pakaian (Muhammad Hisyam Lee, Suhartono 2010). Hasil penelitian tersebut menunjukkan bahwa model ARIMAX lebih akurat dibanding dengan model ARIMA karena model ARIMAX ini sangat memperhatikan variabel *independent*. Variabel *independent* yang digunakan pada penelitian penyebaran penyakit malaria yaitu faktor iklim di antaranya curah hujan, kelembaban udara dan suhu udara. Sedangkan pada peramalan jumlah pakaian variabel *independent* yang digunakan adalah efek variasi kalender. [13]

Pada penelitian ini model ARIMAX digunakan untuk meramalkan data curah hujan untuk 30 periode ke depan. Variabel *eXogen* yang digunakan pada penelitian ini adalah faktor-faktor yang mempengaruhi curah hujan yaitu suhu udara, tekanan udara, kelembaban nisbi dan kecepatan angin.

Dalam model ini faktor-faktor yang mempengaruhi variabel *dependent* Y pada waktu ke- t tidak hanya oleh fungsi variabel Y dalam waktu, tetapi juga oleh variabel-variabel *independent* lainnya pada waktu ke- t . Secara umum bentuk model ARIMAX (p, d, q) dapat diberikan dengan persamaan berikut [6]:

$$(1 - B)^d \phi_p(B) Y_t = \theta_q(B) e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \dots + \alpha_k X_{k,t} \quad (2.6)$$

Dimana:

Y_t : variabel *dependent*

$X_{k,t}$: variabel *independent* atau sebagai variable *eXogen* ke- k pada saat t
dengan $k = 1, 2, 3, \dots, k$

Dalam model ini Y_t dan $X_{k,t}$ adalah data runtun waktu dan diasumsikan stasioner. Jika keduanya stasioner, alternatif model lain dapat digunakan adalah model *Autoregresif Distributed Lag* (ADL). Jika hanya Y_t yang tidak stasioner (mengandung trend), maka dapat menggunakan model ARIMAX dengan menambahkan komponen model pembeda ke dalam Y_t . [7]

B. Model *Seasonal Autoregresif Integrated Moving Average* dengan Tambahan Variabel *eXogen* (SARIMAX)

Model ARIMAX merupakan suatu metode peramalan yang sangat memperhatikan variabel *independent* sehingga metode ini akan digunakan untuk memodelkan curah hujan sebagai variabel *dependent* dan faktor-faktor yang mempengaruhi curah hujan digunakan sebagai variabel *independent*. Namun karena pola cuaca di Indonesia cenderung membentuk pola musiman, termasuk curah hujannya juga membentuk pola musiman oleh karena itu perlu diramalkan melalui model musiman. Sehingga model ARIMAX ditambahkan dengan unsur musimannya disebut Seasonal ARIMAX atau SARIMAX.

Adapun ARIMAX musiman biasa disebut dengan Seasonal ARIMAX (SARIMAX). Seasonal ARIMAX (SARIMAX), yaitu model SARIMA dengan variabel *eXogen*.

Adapun bentuk umum untuk model SARIMAX $(p, d, q)(P, D, Q)S$ yaitu:

$$(1 - B)^d(1 - B^S)^D\phi_p(B)\Phi_P(B^S)Y_t = \theta_q(B)\Theta_Q(B^S)e_t + \alpha_1X_{1,t} + \alpha_2X_{2,t} + \dots + \alpha_kX_{k,t} \quad (2.7)$$

Dimana:

Y_t : variabel *dependent*

$X_{k,t}$: variabel *independent* atau sebagai variable *eXogen* ke- k pada saat t dengan $k = 1,2,3, \dots, k$

3.2 Tahapan Analisis Model ARIMAX dan SARIMAX

Tahapan untuk memperoleh model ARIMAX dan SARIMAX sama halnya dengan tahapan ARIMA dan SARIMA. Namun beberapa estimasi parameter untuk variabel *independent*-nya ditambahkan ke dalam model. Berikut akan dijelaskan tahapan untuk memperoleh model ARIMAX dan SARIMAX.

3.2.1. Pembagian data

Sebelum data diolah langkah pertama yaitu proses pembagian data. Data dibagi ke dalam dua bagian yaitu data *in-sampel* dan data *out-sampel*. Data *in-sampel* digunakan untuk proses estimasi model sedangkan data *out-sampel* digunakan untuk proses peramalan.

3.2.2. Tahap Identifikasi

Pada tahap ini ada beberapa hal yang harus dilakukan yaitu sebagai berikut.

a. Pemeriksaan Kestasioneran Data

Suatu data runtun waktu dikatakan stasioner atau menunjukkan kesalahan random jika koefisien autokorelasi untuk semua *lag* secara statistik tidak berbeda signifikan dari nol atau berbeda dari nol untuk beberapa *lag* yang di depan.

Pengujian stasioneritas dari suatu data runtun waktu dapat dilakukan dengan beberapa cara berikut [7].

1. Pendeteksian ketidakstasioneran data dalam mean (rata-rata) dapat dilihat dengan menggunakan plot dari data asli, plot *fungsi autokorelasi* (ACF) dan plot fungsi *autokorelasi parsial* (PACF). Jika data mengandung komponen trend, data nonstasioner dalam mean dan plot ACF dan PACF akan meluruh secara perlahan.
2. Pendeteksian ketidakstasioneran dalam varians dapat menggunakan plot ACF dan PACF dari residual kuadrat. Dapat juga dilihat dengan melihat nilai variansinya.

b. Proses *differencing* (pembedaan) dan transformasi

Proses pembedaan dilakukan apabila data tidak stasioner dalam mean. Notasi yang digunakan yaitu operator *shift* mundur B, yang penggunaannya sebagai berikut [6]:

$$BY_t = Y_{t-1} \quad (3.1)$$

Dengan kata lain, notasi B yang dipasang pada Y_t mempunyai pengaruh menggeser data 1 periode ke belakang. Dua penerapan B untuk *shift* Y_t akan menggeser data tersebut dua periode ke belakang, seperti berikut.

$$B(BY_t) = B^2Y_t = Y_{t-2} \quad (3.2)$$

Jika suatu data runtun waktu tidak stasioner, maka data tersebut dapat dibuat lebih mendekati stasioner dengan melakukan pembedaan pertama dari data, jika setelah melakukan pembedaan pertama data masih belum stasioner maka dilakukan lagi pembedaan kedua. Biasanya pembedaan ini dilakukan sebanyak 2 kali.

Pembedaan pertama

$$Y'_t = Y_t - Y_{t-1} \quad (3.3)$$

Dengan menggunakan operator *shift* mundur, maka persamaan (3.3) menjadi:

$$Y'_t = Y_t - BY_t = (1 - B)Y_t \quad (3.4)$$

Pembedaan orde kedua

$$\begin{aligned} Y''_t &= Y'_t - Y'_{t-1} \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2} \\ &= (1 - 2B + B^2)Y_t \\ &= (1 - B)^2Y_t \end{aligned} \quad (3.5)$$

Perhatikan bahwa pembedaan orde kedua diberi notasi $(1 - B)^2$. Ini berbeda dengan pembedaan kedua yaitu $1 - B^2$.

Proses transformasi dilakukan apabila data tidak stasioner dalam varians. Transformasi yang sering digunakan adalah transformasi Box-Cox. Salah satu transformasi Box-Cox yang sering digunakan dalam analisis runtun waktu adalah transformasi logaritma, yang sering juga digabungkan dengan melakukan pembedaan

terhadap data hasil transformasi logaritma. Berikut beberapa nilai λ dengan transformasi Box-Cox yang bersesuaian disajikan dalam tabel.[7]

Tabel 3.1

Nilai λ dan Transformasinya

Nilai Estimasi λ	Transformasi
2	Y_t^2
0.5	$\sqrt{Y_t}$
0	$\ln Y_t$
-0.5	$\frac{1}{\sqrt{Y_t}}$
-1.0	$\frac{1}{Y_t}$

c. Penentuan Nilai p, d, q Nonmusiman maupun P, D, Q Musiman untuk Memperoleh Model ARIMA atau SARIMA dari Variabel *Dependent*

Langkah selanjutnya setelah data stasioner adalah menetapkan model ARIMA dan SARIMA yang cocok. Untuk menentukan nilai d nonmusiman atau D musiman, jika tanpa perbedaan maka d atau D bernilai 0, jika stasioner setelah *differensi* 1 nonmusiman maka $d = 1$, jika stasioner setelah *differensi* 1 nonmusiman dan *differensi* 1 musiman maka $d = 1$ dan $D = 1$ begitupun seterusnya. Jika hanya mengandung proses AR dan *differencing* maka *series* itu dikatakan mengikuti proses *integrated autoregresif* dan dilambangkan dengan ARIMA ($p, d, 0$), jika hanya mengandung MA dan *differencing* maka *seriesnya* dikatakan mengikuti proses *integrated moving average* dan dilambangkan dengan ARIMA ($0, d, q$).

Kemudian dalam menetapkan nilai p dan q nonmusiman atau P dan Q musiman dibantu dengan mengamati pola fungsi *autokorelasi* (ACF) dan *autokorelasi parsial* (PACF) dari runtun waktu yang dipelajari. Untuk menentukan nilai p dan q nonmusiman pada model ARIMA maka dilihat dari pola ACF dan PACF yang tanpa *differensi* musiman sedangkan untuk menentukan nilai $p, q, P,$ dan Q musiman dan

nonmusiman pada model SARIMA maka dilihat dari pola ACF dan PACF yang di-*differensi* musiman dan nonmusiman. *Autokorelasi* adalah hubungan antara nilai-nilai yang terurut dari variabel-variabel yang sama. Sedangkan *autokorelasi parsial* digunakan untuk mengukur tingkat keeratan (*association*) antara Y_t dan Y_{t-k} apabila pengaruh dari time lag 1,2,3,..., dan seterusnya sampai $k - 1$ dianggap terpisah dan juga dapat membantu dalam menetapkan model yang tepat untuk peramalan.

d. Pembentukan Model ARIMAX atau SARIMAX sementara

Setelah diperoleh model ARIMA atau SARIMA sementara maka model tersebut digunakan untuk membentuk model ARIMAX atau SARIMAX sementara yaitu dengan menambahkan variabel *independent* ke dalam model ARIMA atau SARIMA.

Secara matematis, model ARIMAX secara umum dituliskan sebagai:

$$(1 - B)^d \phi_p(B) Y_t = \theta_q(B) e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t} \quad (3.6)$$

Dengan mensubstitusikan nilai $\phi_p = 1 - \phi_1 B - \dots - \phi_p B^p$ dan $\theta_q = 1 - \theta_1 B - \dots - \theta_q B^q$, maka persamaan (3.6) menjadi:

$$(1 - B)^d (1 - \phi_1 B - \dots - \phi_p B^p) Y_t = (1 - \theta_1 B - \dots - \theta_q B^q) e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t} \quad (3.7)$$

Dengan menambahkan unsur musiman pada model ARIMAX, maka model ARIMAX menjadi model Seasonal ARIMAX atau SARIMAX. Secara matematis model SARIMAX ditulis sebagai:

$$(1 - B)^d (1 - B^S)^D \phi_p(B) \Phi_P(B^S) Y_t = \theta_q(B) \Theta_Q(B^S) e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \dots + \alpha_k X_{k,t} \quad (3.8)$$

Dengan mensubstitusi $\phi_p = 1 - \phi_1 B - \dots - \phi_p B^p$, $\theta_q = 1 - \theta_1 B - \dots - \theta_q B^q$, $\Phi_P(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{SP}$, dan $\Theta_Q(B^S) = 1 - \Theta_1 B^S - \dots - \Theta_Q B^{SQ}$, maka persamaan (3.8) menjadi:

$$\begin{aligned} (1 - B)^d (1 - B^S)^D (1 - \phi_1 B - \dots - \phi_p B^p) (1 - \Phi_1 B^S - \dots - \Phi_P B^{SP}) Y_t \\ = (1 - \theta_1 B - \dots - \theta_q B^q) (1 - \Theta_1 B^S - \dots - \Theta_Q B^{SQ}) e_t \\ + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t} \end{aligned} \quad (3.9)$$

Dimana:

Y_t : variabel *dependent* yaitu curah hujan

$X_{1,t}$: variabel *eXogen* ke-1 pada saat t yaitu suhu udara

$X_{2,t}$: variabel *eXogen* ke-2 pada saat t yaitu tekanan udara

$X_{3,t}$: variabel *eXogen* ke-3 pada saat t yaitu kelembaban nisbi

$X_{4,t}$: variabel *eXogen* ke-4 pada saat t yaitu kecepatan angin

B : operator *back shift*

d : *differensi* nonmusiman

D : *differensi* musiman

S : banyaknya musiman

3.2.3. Tahap Estimasi parameter

Setelah diperoleh model ARIMAX atau SARIMAX sementara dari hasil identifikasi maka selanjutnya parameter-parameter AR, MA, musiman maupun nonmusiman, dan parameter dari variabel *eXogen*-nya diestimasi dan ditetapkan dengan cara yang terbaik. Estimasi model tersebut dapat diestimasi dengan metode *Maximum Likelihood Estimation* (MLE) dan *Least Square*. [7]

3.2.4. Tahap Uji Diagnostik

Langkah selanjutnya adalah melakukan pemeriksaan diagnostik dari model yang telah diestimasi pada langkah sebelumnya. Pada tahap ini, pemeriksaan diagnostik dilakukan dengan cara sebagai berikut. [7]

a. Uji asumsi *White Noise*

Untuk melihat apakah residual e_t bersifat *white noise* dapat dilakukan dengan melakukan uji korelasi serial, yakni menguji hipotesis $H_0: \rho_1 = \rho_2 = \dots = \rho_k, k < n$. Uji ini dapat dilakukan dengan statistik uji Box-Pierce atau Ljung-Box. Pemeriksaan tersebut meliputi pengujian terhadap residual e_t dengan tahapan sebagai berikut.[7]

i. Hipotesis Uji

H_0 : minimal ada satu $\rho_k \neq 0$ (model belum memenuhi asumsi *white noise*)

H_1 : $\rho_1 = \rho_2 = \dots = \rho_k = 0$ (model memenuhi asumsi *white noise*)

Taraf signifikansi = $\alpha = 5\%$

ii. Statistik Uji menggunakan *Ljung-Box*

$$Q = n(n+2) \sum_{k=1}^m \frac{\rho_k^2}{(n-k)} \quad (3.10)$$

Dimana:

m : banyaknya lag maksimum

n : banyaknya pengamatan asli

ρ_k : koefisien autokorelasi

iii. Kaidah pengambilan keputusan

H_0 ditolak jika $Q > \chi^2_{(1-\alpha, db)}$ atau tolak H_0 jika $p\text{-value} > \alpha$

Dimana $db = m - M$ dengan $M = p + q$

b. Uji Normalitas

Selain residual memenuhi asumsi *white noise*, residual juga harus memenuhi distribusi normal. Pemeriksaan kenormalan residual dapat dilakukan dengan qq-plot. Ciri data yang berdistribusi normal jika diplotkan dengan qq-plot maka titik-titik data tersebut tersebar di sekitar garis lurus. Jika ujung dari titik-titik data tersebut agak menjauh dari garis lurus maka masih dikatakan berdistribusi normal karena yang perlu diperhatikan adalah daerah tengah dari kumpulan titik data tersebut. Apabila ternyata model tidak memenuhi asumsi tersebut, maka harus dirumuskan kembali model yang baru, yang selanjutnya diestimasi dan parameternya diuji kembali.[7]

3.2.5. Tahap Peramalan

Setelah diperoleh model yang telah memenuhi asumsi uji diagnostik dari model ARIMAX atau SARIMAX maka langkah selanjutnya menggunakan model tersebut untuk keperluan peramalan.

Untuk variabel *independent* digunakan 4 variabel *independnet* sebagai variabel *eXogen* yaitu suhu udara, tekanan udara, kelembaban nisbi, dan kecepatan angin. Peramalan untuk model ARIMAX dapat menggunakan rumus (3.7) yaitu:

$$(1 - B)^d(1 - \phi_1 B - \dots - \phi_p B^p)Y_t = (1 - \theta_1 B - \dots - \theta_q B^q) e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t}$$

Sedangkan peramalan untuk model SARIMAX dapat menggunakan rumus (3.9) yaitu:

$$(1 - B)^d(1 - B^S)^D(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^S - \dots - \Phi_P B^{SP})Y_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^S - \dots - \Theta_Q B^{SQ})e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t}$$

3.2.1. Pemilihan Model Terbaik

Setelah model diperoleh yang telah memenuhi asumsi pada uji diagnostik mungkin diperoleh beberapa model yang sesuai. Sehingga dapat dipilih model yang terbaik yang dapat digunakan untuk kebutuhan peramalan. Untuk memilih model yang terbaik dilakukan dengan kriteria berikut.

Uji *Akaike Information Criteria* (AIC)

Persamaan AIC dapat ditulis sebagai [7]:

$$AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{SSE}{n}\right) \quad (3.11)$$

Dimana:

SSE: jumlah residual kuadrat

k : jumlah parameter yang diestimasi

n : jumlah data

BAB IV

STUDI KASUS DAN ANALISIS DATA CURAH HUJAN MENGGUNAKAN MODEL ARIMAX DAN SARIMAX

4.1. Sumber Data

Data yang digunakan untuk keperluan studi kasus ini merupakan data curah hujan harian di wilayah Cemara Kelurahan Pasteur Kecamatan Pasirkaliki Bandung. Data curah hujan ini merupakan data sekunder yang bersumber dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) Stasiun Geofisika Kelas I Bandung yang beralamat di Jl. Cemara No. 66 Bandung.

Curah hujan itu sendiri dijadikan sebagai variabel *dependent* yang akan dilakukan peramalan pada periode berikutnya sedangkan untuk variabel *independent*-nya dipilih beberapa faktor yang mempengaruhi curah hujan seperti kelembaban nisbi, suhu udara, tekanan udara, dan kecepatan angin. Data tersebut merupakan data harian yang masing-masing berjumlah 366 data terhitung dari 1 Desember 2011 sampai 30 November 2012.

4.2. Analisis Data Menggunakan Model ARIMAX

Berikut tahapan untuk memperoleh model curah hujan menggunakan model ARIMAX.

4.2.1. Pembagian Data

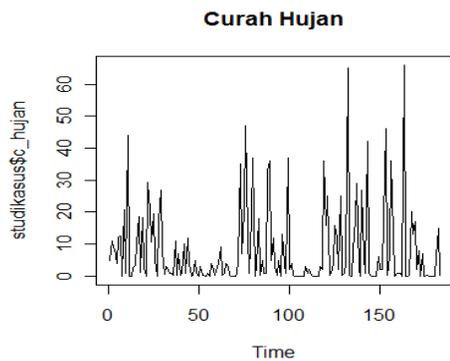
Tahap pertama yang dilakukan adalah pembagian data. Data dibagi ke dalam dua bagian yaitu data *in-sampel* dan data *out-sampel*. Data *in-sampel* digunakan untuk tahap estimasi model yang terhitung mulai 1 Desember 2011 sampai 31 Mei 2012 sebanyak 183 data pengamatan dan data *out-sampel* digunakan untuk tahap peramalan yang terhitung mulai 1 Juni 2012 sampai 30 November 2012 sebanyak 183 data pengamatan.

4.2.2. Tahap Identifikasi

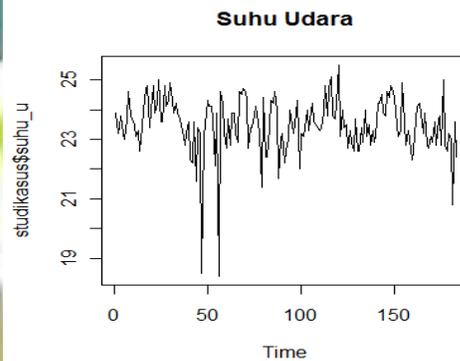
Pada tahap ini ada beberapa hal yang harus dilakukan yaitu sebagai berikut.

a. Pemeriksaan Kestasioneran Data

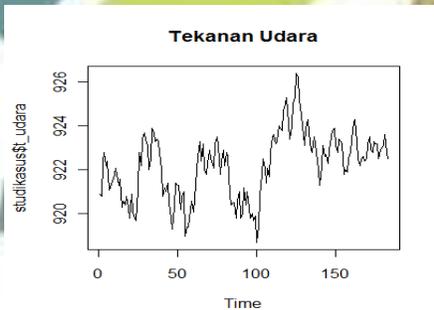
Dapat dilihat dari plot data asli. Hasil plot dari data *dependent* (curah hujan) dan plot data *independent* (kelembaban nisbi, suhu udara, tekanan udara, dan kecepatan angin) disajikan pada gambar berikut.



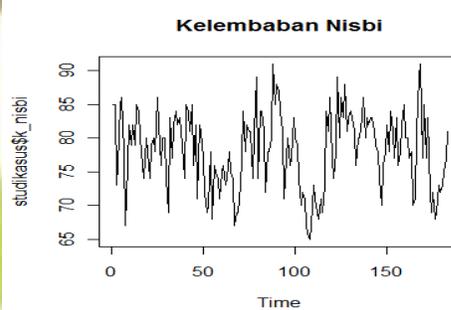
Gambar 4.1 Plot Data Curah Hujan



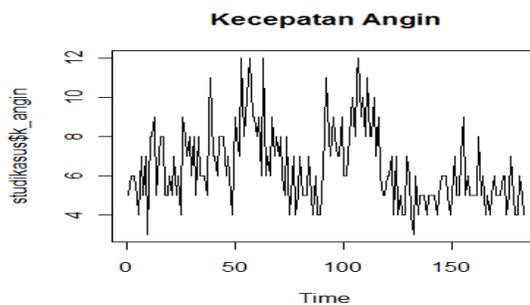
Gambar 4.2 Plot Data Suhu Udara



Gambar 4.3 Plot Data Tekanan Udara



Gambar 4.4 Plot Data Kelembaban Nisbi

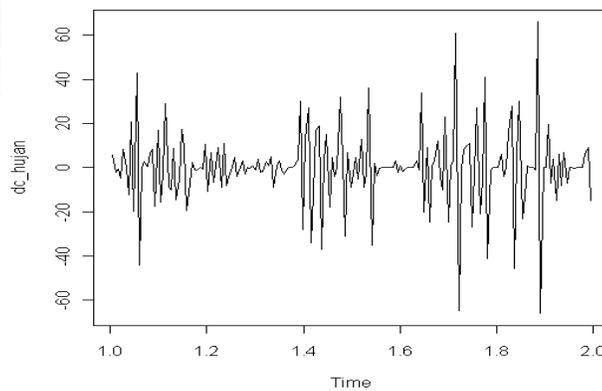


Gambar 4.5 Plot Data Kecepatan Angin

Dari plot data asli untuk curah hujan terlihat bahwa data belum stasioner dalam mean karena data tidak berfluktuasi di sekitar suatu nilai rata-rata yang konstan sehingga perlu dilakukan *differensi* nonmusiman untuk memperoleh data stasioner. Kemudian data curah hujan diasumsikan tidak mengalami adanya unsur musiman sehingga tidak perlu dilakukan *differensi* musiman. Untuk data *independent* terlihat pada plot data asli semuanya sudah stasioner dalam mean dan varians karena nilainya berfluktuasi pada suatu nilai rata-rata yang konstan serta tidak mengalami trend naik maupun turun. Sehingga untuk data *independent* (suhu udara, tekanan udara, kelembaban nisbi, dan kecepatan angin) diasumsikan sudah stasioner dalam mean dan varians sehingga model ARIMAX dapat digunakan untuk analisis curah hujan.

b. Proses *Differencing*

Karena data belum stasioner dalam mean maka dilakukan proses *differensi* untuk memperoleh data yang stasioner agar memenuhi asumsi model ARIMAX. Dan karena data sudah stasioner dalam varians maka tidak perlu dilakukan transformasi. Dengan bantuan software R, didapat hasil plot data yang telah di-*differensi* satu, berikut hasilnya.

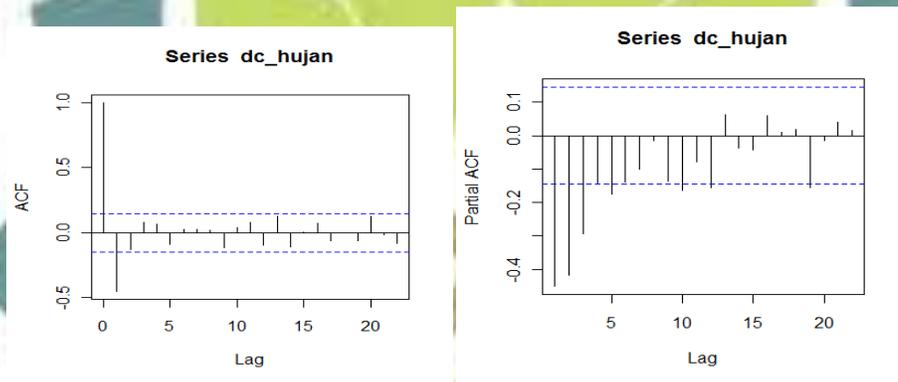


Gambar 4.6 Plot Data yang Telah di-*differensi* 1 Nonmusiman

Setelah data di-*differensi* pertama terlihat bahwa data menjadi stasioner dalam mean hal ini terlihat dari fluktuasi data berada di sekitar suatu rata-rata nilai yang konstan.

c. Penentuan Nilai p, d, q untuk Memperoleh Model ARIMA Sementara

Langkah selanjutnya setelah data stasioner adalah menetapkan model ARIMA sementara. Karena data stasioner setelah *differensi* pertama maka dapat ditetapkan nilai $d = 1$. Untuk menetapkan nilai p dan q dapat dilihat dari plot ACF dan PACF data yang sudah stasioner. Dengan bantuan software R, hasil plot ACF dan PACF dapat dilihat pada gambar di bawah ini.



Gambar 4.7 Plot ACF dan PACF yang Sudah Stasioner

Melalui plot ACF dan PACF di atas maka model yang mungkin untuk data curah hujan adalah:

1. $ARIMA(1,1,1)$
Plot ACF terpotong pada *lag* ke-1
Plot PACF terpotong pada *lag* ke-1
2. $ARIMA(2,1,1)$
Plot ACF terpotong pada *lag* ke-1
Plot PACF terpotong pada *lag* ke-2
3. $ARIMA(3,1,1)$
Plot ACF terpotong pada *lag* ke-1
Plot PACF terpotong pada *lag* ke-3

d. Membentuk Model ARIMAX Sementara

Setelah diperoleh model ARIMA sementara maka selanjutnya membentuk model ARIMAX sementara yaitu dengan menambahkan variabel *independent* pada model. Model sementara yang terbentuk yaitu:

1. ARIMAX (1,1,1)

Model sementara yang terbentuk untuk ARIMAX (1,1,1) yaitu:

$$(1 - B)^1(1 - \phi_1 B)Y_t = (1 - \theta_1 B)e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t} \quad (4.1)$$

2. ARIMAX (2,1,1)

Model sementara yang terbentuk untuk ARIMAX (2,1,1) yaitu:

$$(1 - B)^1(1 - \phi_1 B - \phi_2 B^2)Y_t = (1 - \theta_1 B)e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t} \quad (4.2)$$

3. ARIMAX (3,1,1)

Model sementara yang terbentuk untuk ARIMAX (3,1,1) yaitu:

$$(1 - B)^1(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)Y_t = (1 - \theta_1 B)e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t} \quad (4.3)$$

4.2.3. Estimasi Parameter

Hasil estimasi parameter untuk masing-masing model diperoleh sebagai berikut.

1. ARIMAX (1,1,1)

Dengan bantuan software R, hasil estimasi parameternya bisa dilihat pada halaman lampiran C. Sehingga persamaan (4.1) menjadi:

$$\begin{aligned} (1 - B)^1(1 + 0.0657 B)Y_t &= (1 + 1.0000 B)e_t + 0.7102X_{1,t} + 0.0264 X_{2,t} + \\ &0.5347X_{3,t} - 0.9585X_{4,t} \\ (1 - B)(Y_t + 0.0657 BY_t) &= e_t + 1.0000 Be_t + \\ &0.7102X_{1,t} + 0.0264 X_{2,t} + 0.5347X_{3,t} - 0.9585X_{4,t} \\ Y_t - BY_t + 0.0657BY_t - 0.0657 B^2Y_t &= e_t + 1.0000 Be_t + \\ &0.7102X_{1,t} + 0.0264 X_{2,t} + 0.5347X_{3,t} - 0.9585X_{4,t} \\ Y_t &= BY_t - 0.0657BY_t + 0.0657 B^2Y_t + e_t + 1.0000 Be_t + \\ &0.7102X_{1,t} + 0.0264 X_{2,t} + 0.5347X_{3,t} - 0.9585X_{4,t} \end{aligned} \quad (4.4)$$

2. ARIMAX (2,1,1)

Dengan bantuan software R, hasil estimasi parameternya bisa dilihat pada halaman lampiran C. Sehingga persamaan (4.2) menjadi:

$$\begin{aligned}
 (1 - B)^1(1 + 0.0839B + 0.1597B^2)Y_t &= (1 + 1.0000 B)e_t + 1.0568 X_{1,t} - \\
 &0.1011X_{2,t} + 0.5822 X_{3,t} - 0.9711X_{4,t} \\
 (1 - B)(Y_t + 0.0839BY_t + 0.1597B^2Y_t) &= e_t + 1.0000 Be_t + 1.0568 X_{1,t} - \\
 &0.1011X_{2,t} + 0.5822 X_{3,t} - 0.9711X_{4,t} \\
 Y_t - BY_t + 0.0839BY_t - 0.0839 B^2Y_t + 0.1597B^2Y_t - 0.1597B^3Y_t \\
 &= e_t + 1.0000 Be_t + 1.0568 X_{1,t} - 0.1011X_{2,t} + 0.5822 X_{3,t} - 0.9711X_{4,t} \\
 Y_t = BY_t - 0.0839BY_t + 0.0839 B^2Y_t - 0.1597B^2Y_t + 0.1597B^3Y_t + e_t + \\
 &1.0000 Be_t + 1.0568 X_{1,t} - 0.1011X_{2,t} + 0.5822 X_{3,t} - 0.9711X_{4,t} \quad (4.5)
 \end{aligned}$$

3. ARIMAX (3,1,1)

Dengan bantuan software R, hasil estimasi parameternya bisa dilihat pada halaman lampiran C. Sehingga persamaan (4.3) menjadi:

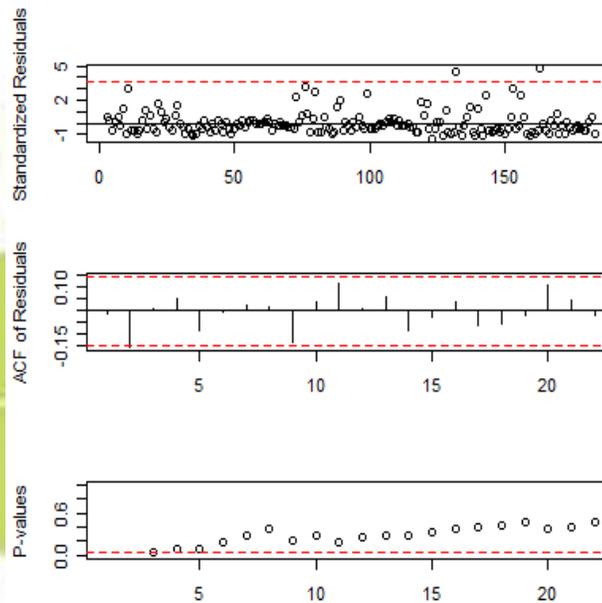
$$\begin{aligned}
 (1 - B)^1(1 - 0.0857 B - 0.1608B^2 + 0.0097B^3)Y_t \\
 &= (1 + 1.0000 B)e_t + 1.0602X_{1,t} - 0.1011X_{2,t} + 0.5822 X_{3,t} - 0.9711X_{4,t} \\
 (1 - B)(Y_t - 0.0857 BY_t - 0.1608B^2Y_t + 0.0097B^3Y_t) \\
 &= e_t + 1.0000 Be_t + 1.0602X_{1,t} - 0.1011X_{2,t} + 0.5822 X_{3,t} - 0.9711X_{4,t} \\
 Y_t - BY_t - 0.0857 BY_t + 0.0857B^2Y_t - 0.1608B^2Y_t + 0.1608B^3Y_t + 0.0097B^3Y_t \\
 &- 0.0097B^4Y_t = e_t + 1.0000 Be_t + 1.0602X_{1,t} - 0.1011X_{2,t} + \\
 &0.5822 X_{3,t} - 0.9711X_{4,t} \\
 Y_t = BY_t + 0.0857 BY_t - 0.0857B^2Y_t + 0.1608B^2Y_t - 0.1608B^3Y_t - 0.0097B^3Y_t + \\
 &0.0097B^4Y_t + e_t + 1.0000 Be_t + 1.0602X_{1,t} - 0.1011X_{2,t} + 0.5822 X_{3,t} - \\
 &0.9711X_{4,t} \quad (4.6)
 \end{aligned}$$

4.2.4. Uji Diagnostik

a. Uji Asumsi *White Noise*

Untuk uji asumsi *white noise* ini dilakukan dengan statistik uji *Ljung-Box* dengan bantuan software R. Dengan kriteria tolak H_0 jika $p - value > \alpha$. Untuk nilai $p - value$ secara keseluruhan dapat dilihat pada lampiran D.

1. *ARIMAX(1,1,1)*

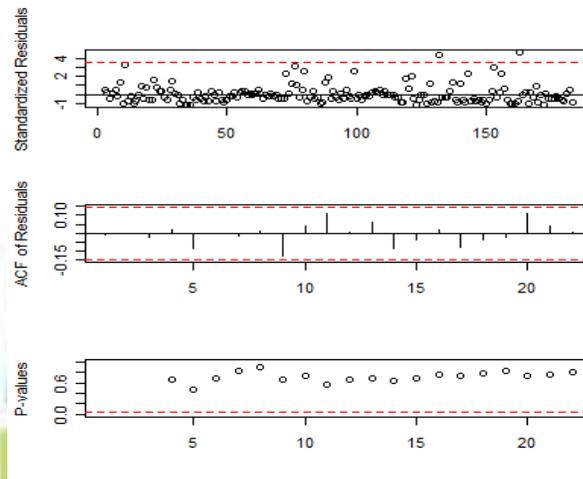


Gambar 4.8 Plot Residual dan $p - value$ Model *ARIMAX(1,1,1)*

Dari hasil plot di atas terlihat bahwa residual ACF sudah memenuhi asumsi *white noise* yang ditandai dengan tidak adanya *lag* (≥ 1) yang keluar dari batas garis interval. Nilai $p - value$ juga memperlihatkan nilai yang cukup besar yang hampir semua di atas garis batas 5%. Hal ini menunjukkan bahwa galat $ke-t$ dan $t - k$ tidak terdapat korelasi.

Dari hasil output diperoleh nilai $p - value > \alpha$ atau $0.8565 > 0.05$. Maka dapat diambil kesimpulan bahwa model *ARIMAX(1,1,1)* memenuhi asumsi *white noise*.

2. $ARIMAX(2,1,1)$

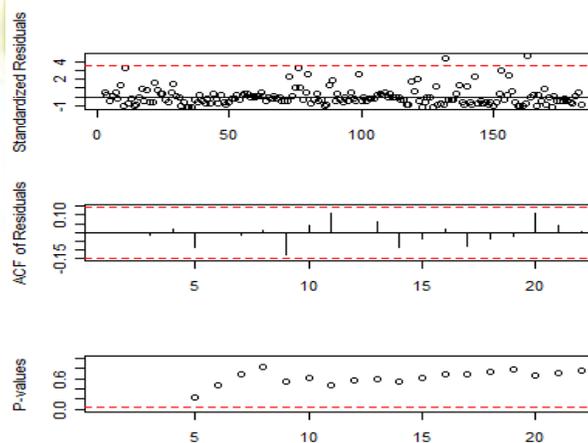


Gambar 4.9 Plot Residual dan p -value Model $ARIMAX(2,1,1)$

Dari hasil plot di atas terlihat bahwa residual ACF sudah memenuhi asumsi *white noise* yang ditandai dengan tidak adanya $lag (\geq 1)$ yang keluar dari batas garis interval. Nilai p -value juga memperlihatkan semua nilai di atas garis batas 5%. Hal ini menunjukkan bahwa galat ke- t dan $t - k$ tidak terdapat korelasi.

Dari hasil output diperoleh nilai $p - value > \alpha$ atau $0.9474 > 0.05$. Maka dapat diambil kesimpulan bahwa model $ARIMAX(2,1,1)$ memenuhi asumsi *white noise*.

3. $ARIMAX(3,1,1)$



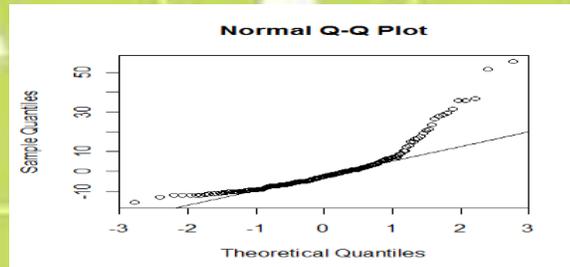
Gambar 4.10 Plot residual dan p -value model $ARIMAX(3,1,1)$

Dari hasil plot di atas terlihat bahwa residual ACF sudah memenuhi asumsi *white noise* yang ditandai dengan tidak adanya *lag* (≥ 1) yang keluar dari batas garis interval. Nilai *p-value* juga memperlihatkan semua nilai di atas garis batas 5%. Hal ini menunjukkan bahwa galat ke-*t* dan *t* – *k* tidak terdapat korelasi.

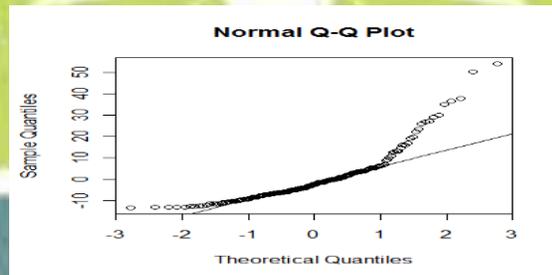
Dari hasil output diperoleh nilai *p* – *value* $> \alpha$ atau $0.9665 > 0.05$. Maka dapat diambil kesimpulan bahwa model *ARIMAX*(3,1,1) memenuhi asumsi *white noise*.

b. Uji Normalitas

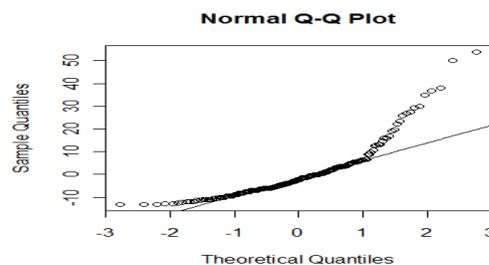
Pemeriksaan uji normalitas ini dapat dilakukan dengan melihat Qq-plot. Dengan bantuan software R, hasilnya dapat dilihat pada gambar di bawah ini.



Gambar 4.11 QQplot model *ARIMAX*(1,1,1)



Gambar 4.12 QQplot model *ARIMAX*(2,1,1)



Gambar 4.13 QQplot model *ARIMAX*(3,1,1)

Dari hasil plot residual terlihat bahwa dengan qqplot yang menyebar mendekati garis linier menunjukkan residual berdistribusi mendekati normal walaupun dalam setiap model terdapat *outlier* (pencilan atau gangguan) namun diasumsikan tetap memenuhi asumsi distribusi normal.

Melalui uji asumsi *white noise* dan uji distribusi normal diperoleh bahwa semua model memenuhi asumsi *white noise* dan uji distribusi normal. Sehingga model bisa digunakan untuk keperluan peramalan.

4.2.5. Peramalan

Setelah diperoleh model yang telah memenuhi asumsi model ARIMAX maka langkah selanjutnya menggunakan model tersebut untuk keperluan peramalan.

Karena pada tahap uji diagnostik diperoleh semua model memenuhi asumsi *white noise* dan distribusi normal maka model tersebut dapat digunakan untuk peramalan. Melalui persamaan (4.4), (4.5), dan (4.6) maka didapat hasil ramalan untuk masing-masing model ARIMAX pada data curah hujan yang hasilnya dapat dilihat pada halaman lampiran F.

4.2.6. Pemilihan Model Terbaik

Dari hasil uji diagnostik diperoleh bahwa semua model memenuhi asumsi *white noise* dan distribusi normal sehingga perlu dipilih model terbaik yang paling mendekati nilai sebenarnya. Pemilihan model terbaik ini dapat dilihat dari nilai AIC terkecil. Dengan bantuan software R diperoleh nilai AIC sebagai berikut:

Tabel 4.1 Nilai AIC Model ARIMAX

Model Peramalan	AIC
<i>ARIMAX</i> (1,1,1)	1424.67
<i>ARIMAX</i> (2,1,1)	1422.17
<i>ARIMAX</i> (3,1,1)	1424.15

4.3. Analisis Data Menggunakan Model SARIMAX

4.3.1. Pembagian Data

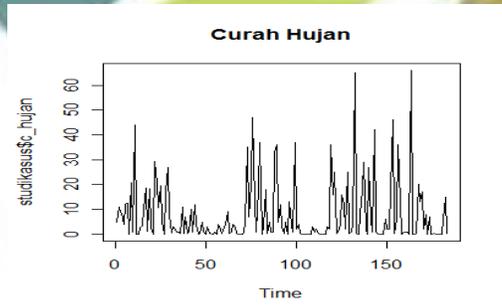
Data dibagi menjadi 2 yaitu data *in-sampel* dan data *out-sampel*. Data *in-sampel* digunakan untuk tahap estimasi model yang dihitung mulai 1 Desember 2011 sampai 31 Mei 2012 sebanyak 183 data pengamatan dan data *out-sampel* digunakan untuk tahap peramalan yang dihitung mulai 1 Juni 2012 sampai 30 November 2012 sebanyak 183 data pengamatan.

4.3.2. Tahap Identifikasi

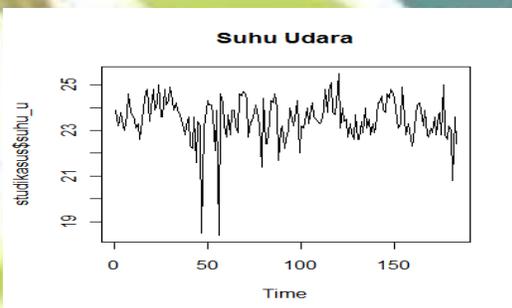
Pada tahap ini ada beberapa hal yang harus dilakukan yaitu sebagai berikut.

a. Pemeriksaan Kestasioneran Data

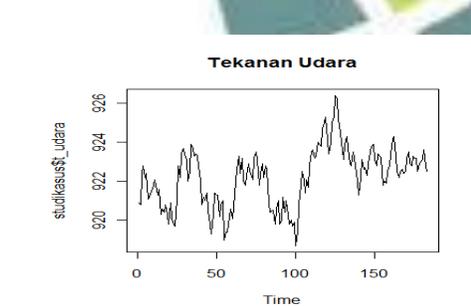
Untuk melihat kestasioneran data dapat dilihat dari plot data asli. Hasil plot dari data *dependent* (curah hujan) dan plot data *independent* (kelembaban nisbi, suhu udara, tekanan udara, dan kecepatan angin) disajikan pada gambar berikut.



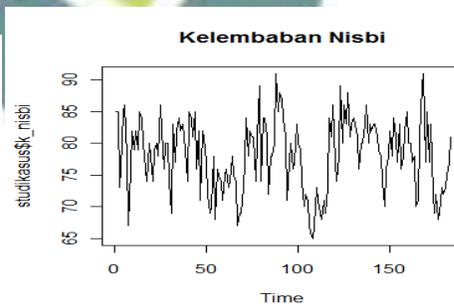
Gambar 4.14 Plot Data Asli Curah Hujan



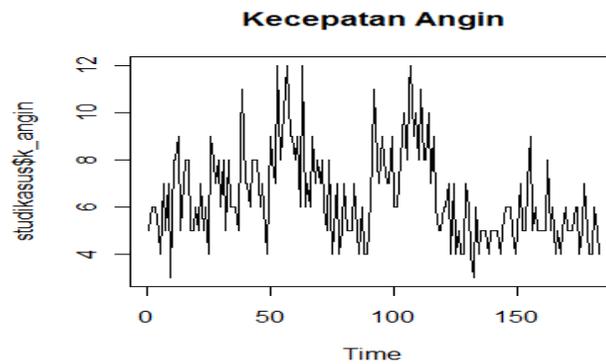
Gambar 4.15 Plot Data Asli Suhu Udara



Gambar 4.16 Plot Data Tekanan Udara



Gambar 4.17 Plot Data Kelembaban Nisbi



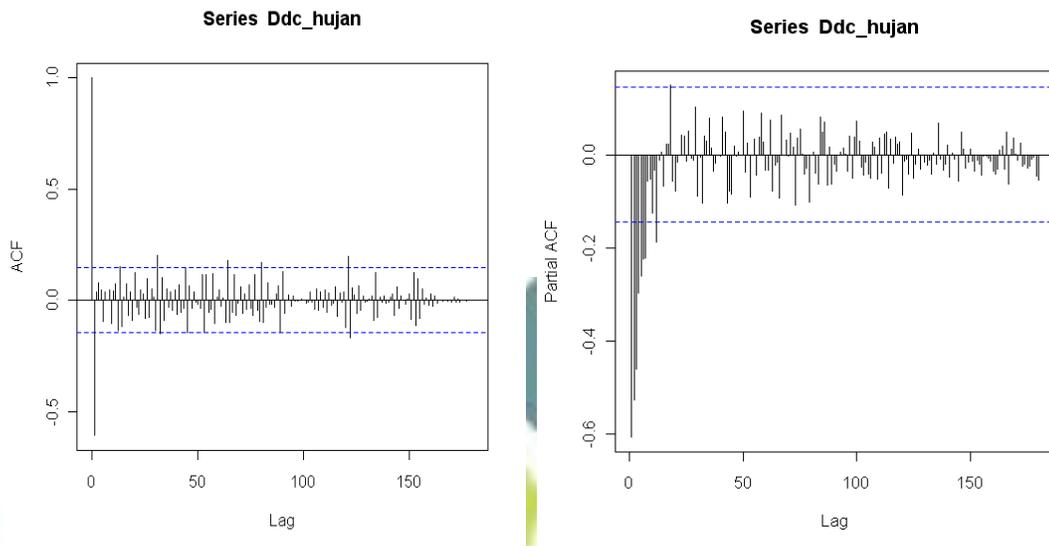
Gambar 4. 18 Plot Data Asli Kecepatan Angin

Dari plot data asli terlihat bahwa data belum stasioner dalam mean dan diasumsikan mengalami unsur musiman. Hal ini terlihat pada setiap bulannya curah hujan mengalami kenaikan dan penurunan yang konstan. Sehingga perlu dilakukan *differensi* nonmusiman dan *differensi* musiman untuk memperoleh data stasioner dalam mean dan musiman.

Untuk plot data *independent* terlihat bahwa data sudah stasioner dalam mean dan varians hal ini terlihat pada nilai data yang berfluktuasi pada suatu nilai rata-rata yang konstan dan data tidak mengalami trend naik dan turun yang begitu berarti. Sehingga dapat diasumsikan bahwa data sudah stasioner dalam mean dan varians sehingga tidak perlu dilakukan proses *differensi* dan transformasi.

b. Proses *Differencing*

Karena data belum stasioner dalam mean dan musiman maka dilakukan proses *differensi* nonmusiman dan musiman untuk memperoleh data yang stasioner agar memenuhi asumsi model SARIMAX. Dan karena data sudah stasioner dalam varians maka tidak perlu dilakukan transformasi. Dengan bantuan software R, didapat hasil plot ACF dan PACF yang telah di-*differensi* 1 musiman dan nonmusiman.

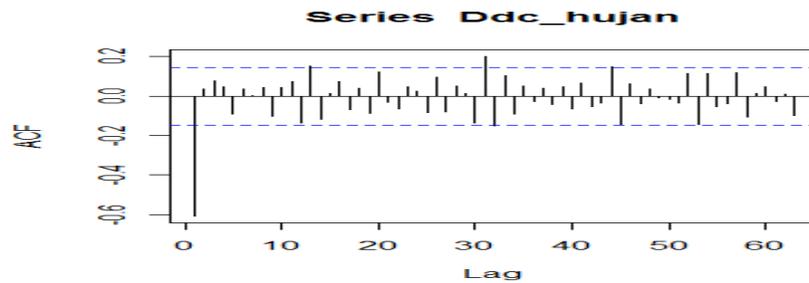


Gambar 4. 19 Plot ACF dan PACF yang Telah di-differensi 1 Musiman dan Nonmusiman

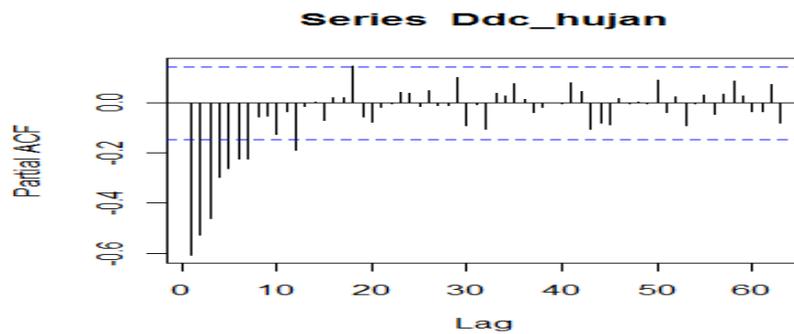
Hasil plot ACF dan PACF setelah di-differensi musiman terlihat bahwa data sudah stasioner. Hal ini terlihat pada lag-lag nonmusiman (lag 1-30) ACF dan PACF cenderung *dies down*. Hal ini juga terlihat pada lag-lag musiman (lag 30, 60) ACF dan PACF juga cenderung *dies down*. Sehingga dapat dikatakan data stasioner setelah $d = 1$ dan $D = 1$.

c. Penentuan Nilai p, d, q Nonmusiman dan P, D, Q Musiman untuk Memperoleh Model SARIMA Sementara

Langkah selanjutnya setelah data stasioner adalah menetapkan model SARIMA sementara. Karena data stasioner setelah di-differensi nonmusiman pertama dan di-differensi musiman pertama maka dapat ditetapkan nilai $d = 1$ dan $D = 1$. Untuk menetapkan nilai p dan q nonmusiman dan P dan Q musiman dapat dilihat dari plot ACF dan PACF data yang sudah stasioner. Dengan bantuan software R, hasil plot ACF dan PACF dapat dilihat pada gambar di bawah ini.



Gambar 4. 20 Plot ACF Data Curah Hujan yang Sudah Stasioner



Gambar 4. 21 Plot PACF Data Curah Hujan yang Sudah Stasioner

Melalui plot ACF dan PACF di atas maka model yang mungkin untuk data curah hujan adalah:

1. SARIMA(1,1,1)(0,1,1)₃₀

Plot ACF nonmusiman terpotong pada *lag* ke-1

Plot PACF nonmusiman terpotong pada *lag* ke-1

Plot ACF musiman terpotong pada *lag* ke-30

2. SARIMA(2,1,1)(0,1,1)₃₀

Plot ACF nonmusiman terpotong pada *lag* ke-1

Plot PACF nonmusiman terpotong pada *lag* ke-2

Plot ACF musiman terpotong pada *lag* ke-30

3. SARIMA(3,1,1)(0,1,1)₃₀

Plot ACF nonmusiman terpotong pada *lag* ke-1

Plot PACF nonmusiman terpotong pada *lag* ke-3

Plot ACF musiman terpotong pada *lag* ke-30

d. Membentuk Model SARIMAX Sementara

Setelah diperoleh model SARIMA sementara maka selanjutnya membentuk model SARIMAX sementara yaitu dengan menambahkan variabel *independent* pada model. Model sementara yang terbentuk yaitu:

1. SARIMAX(1,1,1)(0,1,1)30

Model sementara untuk SARIMAX (1,1,1)(0,1,1)30 yaitu:

$$(1 - B)^1(1 - B^{30})^1(1 - \phi_1 B)Y_t = (1 - \theta_1 B)(1 - \Theta_1 B^{30})e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t} \quad (4.7)$$

2. SARIMAX(2,1,1)(0,1,1)30

Model sementara untuk SARIMAX (2,1,1)(0,1,1)30 yaitu:

$$(1 - B)^1(1 - B^{30})^1(1 - \phi_1 B - \phi_2 B^2)Y_t = (1 - \theta_1 B)(1 - \Theta_1 B^{30})e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t} \quad (4.8)$$

3. SARIMAX(3,1,1)(0,1,1)30

Model sementara untuk SARIMAX (3,1,1)(0,1,1)30 yaitu:

$$(1 - B)^1(1 - B^{30})^1(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)Y_t = (1 - \theta_1 B)(1 - \Theta_1 B^{30})e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t} \quad (4.9)$$

4.3.3. Estimasi Parameter

Dengan bantuan software R, hasil estimasi parameter untuk masing-masing model dapat dilihat pada lampiran H. Dengan mensubstitusi nilai dari masing-masing parameter yang telah diestimasi, maka model SARIMAX sementara dapat dilihat sebagai berikut.

1. SARIMAX(1,1,1)(0,1,1)30

Dengan bantuan software R, maka didapat hasil estimasi untuk model ini yang dapat dilihat pada halaman lampiran H. Sehingga persamaan (4.7) menjadi:

$$(1 - B)^1(1 - B^{30})^1(1 + 0.0466 B)Y_t = (1 + 1.0000B)(1 + 0.9975B^{30})e_t + 0.9015X_{1,t} + 0.0610.061X_{2,t} + 0.4700X_{3,t} - 1.2150X_{4,t} \quad (4.10)$$

2. SARIMAX(2,1,1)(0,1,1)30

Dengan bantuan software R, maka didapat hasil estimasi untuk model ini yang dapat dilihat pada halaman lampiran H. Sehingga persamaan (4.8) menjadi:

$$(1 - B)^1(1 - B^{30})^1(1 + 0.0636B + 0.1339B^2)Y_t = (1 + 1.0000B)(1 + 0.8714 B^{30})e_t + 1.3110 X_{1,t} - 0.0811X_{2,t} + 0.5297 X_{3,t} - 1.1789X_{4,t} \quad (4.11)$$

3. SARIMAX(3,1,1)(0,1,1)30

Dengan bantuan software R, maka didapat hasil estimasi untuk model ini yang dapat dilihat pada halaman lampiran H. Sehingga persamaan (4.9) menjadi:

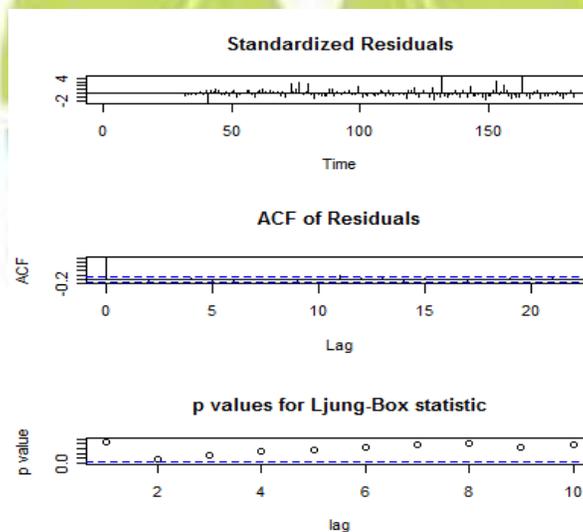
$$(1 - B)^1(1 - B^{30})^1(1 + 0.0659 B + 0.1352B^2 + 0.0141 B^3)Y_t = (1 + 1.0000B)(1 + 0.8734B^{30})e_t + 1.3183 X_{1,t} - 0.0891X_{2,t} + 0.5341X_{3,t} - 1.1808X_{4,t} \quad (4.12)$$

4.3.4. Uji Diagnostik

a. Uji Asumsi White Noise

Untuk uji asumsi *white noise* ini dilakukan dengan statistik uji *Ljung-Box* dengan bantuan software R. Untuk nilai *p-value* secara keseluruhan dapat dilihat pada halaman lampiran I. Dengan kriteria tolak H_0 jika $p - value > \alpha$.

1. SARIMAX(1,1,1)(0,1,1)30

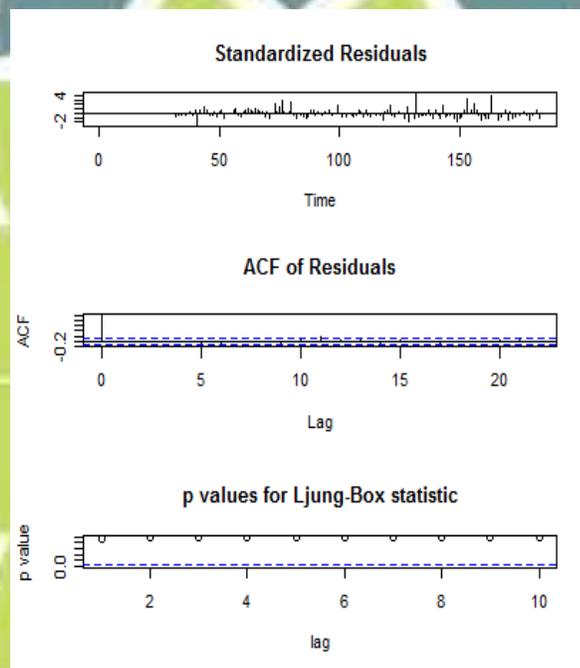


Gambar 4.22 Plot Residual dan *p-value* Model SARIMAX(1,1,1)(0,1,1)30

Dari hasil plot di atas terlihat bahwa residual ACF sudah memenuhi asumsi *white noise* yang ditandai dengan tidak adanya *lag* (≥ 1) yang keluar dari batas garis interval. Nilai *p-value* juga memperlihatkan semua nilai di atas garis batas 5%. Hal ini menunjukkan bahwa galat ke-*t* dan *t* – *k* tidak terdapat korelasi.

Dari hasil output diperoleh nilai *p* – *value* $> \alpha$ atau $0.8673 > 0.05$. Maka dapat diambil kesimpulan bahwa model *SARIMAX*(1,1,1)(0,1,1)₃₀ memenuhi asumsi *white noise*.

2. *SARIMAX*(2,1,1)(0,1,1)₃₀

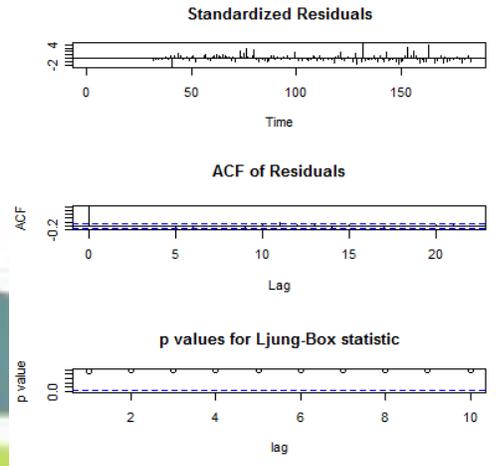


Gambar 4.23 Plot Residual dan *p-value* Model *SARIMAX*(2,1,1)(0,1,1)₃₀

Dari hasil plot di atas terlihat bahwa residual ACF sudah memenuhi asumsi *white noise* yang ditandai dengan tidak adanya *lag* (≥ 1) yang keluar dari batas garis interval. Nilai *p-value* juga memperlihatkan semua nilai di atas garis batas 5%. Hal ini menunjukkan bahwa galat ke-*t* dan *t* – *k* tidak terdapat korelasi.

Dari hasil output diperoleh nilai *p* – *value* $> \alpha$ atau $0.9144 > 0.05$. Maka dapat diambil kesimpulan bahwa model *SARIMAX*(2,1,1)(0,1,1)₃₀ memenuhi asumsi *white noise*.

3. SARIMAX(3,1,1)(0,1,1)30



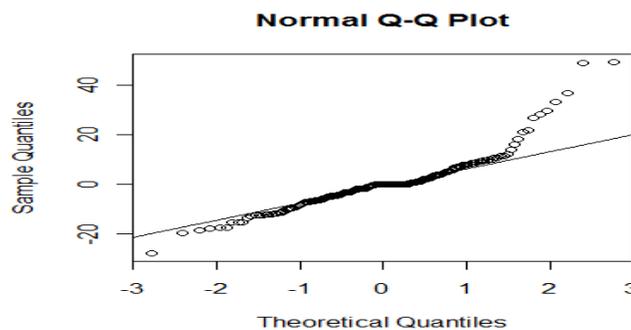
Gambar 4.24 Plot Residual dan p -value Model SARIMAX(3,1,1)(0,1,1)30

Dari hasil plot di atas terlihat bahwa residual ACF sudah memenuhi asumsi *white noise* yang ditandai dengan tidak adanya $lag (\geq 1)$ yang keluar dari batas garis interval. Nilai p -value juga memperlihatkan semua nilai di atas garis batas 5%. Hal ini menunjukkan bahwa galat ke- t dan $t - k$ tidak terdapat korelasi.

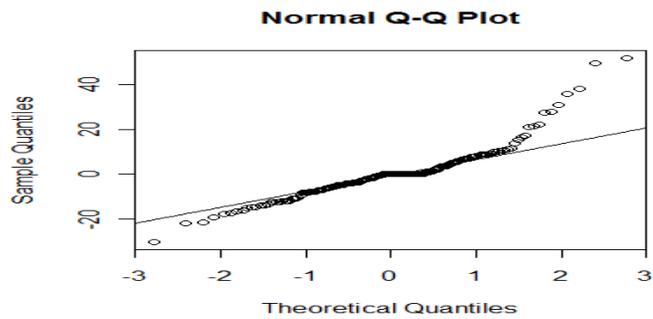
Dari hasil output diperoleh nilai $p - value > \alpha$ atau $0.938 > 0.05$. Maka dapat diambil kesimpulan bahwa model SARIMAX(3,1,1)(0,1,1)30 memenuhi asumsi *white noise*.

b. Uji Normalitas

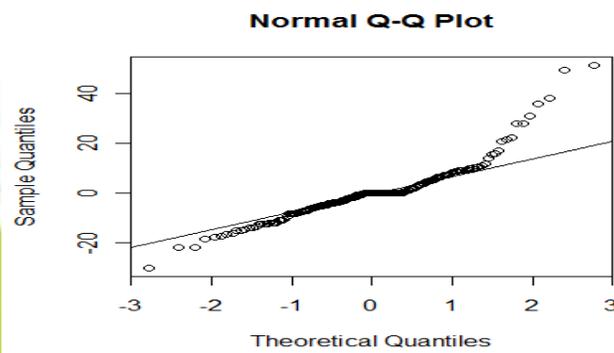
Pemeriksaan kenormalan residual dapat dilakukan dengan qq-plot. Dengan bantuan software R, hasilnya dapat dilihat sebagai berikut.



Gambar 4.25 Plot Qqplot Model SARIMAX(1,1,1)(0,1,1)30



Gambar 4.26 Plot Qqplot Model SARIMAX(2,1,1)(0,1,1)30



Gambar 4.27 Plot Qqplot Model SARIMAX(3,1,1)(0,1,1)30

Dari hasil plot residual terlihat bahwa dengan qqplot yang menyebar mendekati garis linier menunjukkan residual berdistribusi mendekati normal walaupun dalam setiap model terdapat *outlier* (pencilan atau gangguan) namun diasumsikan tetap memenuhi asumsi distribusi normal.

Melalui uji asumsi *white noise* dan uji distribusi normal diperoleh bahwa semua model memenuhi asumsi *white noise* dan uji distribusi normal. Sehingga model bisa digunakan untuk keperluan peramalan.

4.3.5. Peramalan

Setelah diperoleh model yang memenuhi asumsi model SARIMAX maka langkah selanjutnya menggunakan model tersebut untuk keperluan peramalan.

Karena pada tahap uji diagnostik diperoleh semua model memenuhi asumsi *white noise* dan distribusi normal maka model tersebut dapat digunakan untuk peramalan. Melalui persamaan (4.10), (4.11), dan (4.12)) maka didapat hasil ramalan untuk

masing-masing model SARIMAX pada data curah hujan yang hasilnya dapat dilihat pada halaman lampiran K.

4.3.6. Pemilihan Model Terbaik

Dari hasil uji diagnostik memungkinkan diperoleh model lebih dari satu sehingga harus dipilih dengan cara yang terbaik. Pemilihan model terbaik ini dapat dilihat dari AIC terkecil. Model yang memiliki nilai AIC terkecil adalah model terbaik yang dipilih dan dapat digunakan untuk keperluan peramalan. Perbandingan nilai AIC dari ketiga model dapat dilihat pada tabel berikut.

Table 4.2 Nilai AIC Model SARIMAX

Model Peramalan	AIC
<i>SARIMAX(1,1,1)(0,1,1)30</i>	1251.37
<i>SARIMAX(2,1,1)(0,1,1)30</i>	1250.86
<i>SARIMAX(3,1,6)(0,1,1)30</i>	1252.83

BAB V PENUTUP

5.1. Kesimpulan

Dari hasil pembahasan yang telah dipaparkan dapat diambil kesimpulan bahwa:

- a. Model ARIMAX yang diperoleh untuk meramalkan data curah hujan harian yaitu:

$$(1 - B)^d(1 - \phi_1 B - \dots - \phi_p B^p)Y_t = (1 - \theta_1 B - \dots - \theta_q B^q)e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t}$$

Sedangkan model SARIMAX untuk meramalkan data curah hujan harian diperoleh:

$$(1 - B)^d(1 - B^S)^D(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^S - \dots - \Phi_P B^{SP})Y_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^S - \dots - \Theta_Q B^{SQ})e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \alpha_4 X_{4,t}$$

- b. Untuk memperoleh model ARIMAX atau SARIMAX ada 6 tahap yang harus dilakukan yaitu tahap pembagian data, tahap identifikasi, tahap estimasi parameter, tahap uji diagnostik, tahap peramalan, dan tahap pemilihan model
- c. Dari hasil analisis studi kasus diperoleh bahwa model SARIMAX lebih cocok digunakan untuk meramalkan data curah hujan harian dibanding dengan model ARIMAX dengan perolehan nilai AIC terkecil model SARIMAX(2,1,1)(0,1,1)30 sebesar 1250.86 dan untuk model ARIMAX(2,1,1) nilai AIC terkecil sebesar 1422.17. Hal ini sesuai pernyataan yang menyatakan bahwa pola cuaca di Indonesia cenderung membentuk pola musiman termasuk curah hujannya juga membentuk pola musiman.

5.2. Saran

Penulis menyarankan untuk mencari model lain (seperti VARIMAX) untuk meramalkan data curah hujan kemudian bandingkan dengan model ARIMAX dan SARIMAX.

