BAB II
TINJAUAN PUSTAKA

2.1 Curah Hujan

Hujan merupakan jatuhnya suatu cairan dari atmosfer yang berwujud cair maupun beku ke permukaan bumi. Menurut Badan Meteorologi, Klimatologi, dan Geofisika (BMKG, 2020) hujan adalah suatu bentuk presipitasi atau endapan dari cairan atau zat padat yang berasal dari kondensasi yang jatuh dari awan menuju permukaan bumi. Hujan turun biasanya tidak lepas dari pengaruh kelembaban udara yang memacu jumlah titik-titik air yang terdapat pada udara.

Curah hujan adalah banyaknya air hujan yang jatuh pada permukaan bumi suatu daerah dalam kurun waktu tertentu. Curah hujan adalah ketinggian air hujan yang terkumpul dalam penakar hujan pada tempat yang datar, tidak menyerap, tidak meresap dan tidak mengalir (BMKG, 2020). Satuan yang digunakan untuk mengukur curah hujan adalah milimeter (mm). Menurut Badan Meteorologi, Klimatologi, dan Geofisika (BMKG, 2020) Satu milimeter hujan berarti air hujan yang turun di wilayah seluas satu meter persegi akan memiliki ketinggian satu milimeter jika air hujan tidak meresap, mengalir, atau menguap.

Sifat hujan adalah perbandingan antara jumlah curah hujan selama rentang waktu yang ditetapkan (satu periode musim hujan atau satu periode musim kemarau) dengan jumlah curah hujan normalnya (rata-rata selama 30 tahun periode 1981 - 2010). Sifat hujan dibagi menjadi 3 kategori, yaitu:

1. Atas Normal (AN), jika nilai curah hujan lebih dari 115% terhadap rata-ratanya.

2. Normal (N), jika nilai curah hujan antara 85%-115% terhadap rata-ratanya
3. Bawah Normal (BN), jika nilai curah hujan kurang dari 85% terhadap rata-ratanya

Normal curah hujan bulanan adalah nilai rata-rata curah hujan masing-masing bulan selama 30 tahun berturut-turut yang periode waktunya dapat ditentukan secara bebas. Standar normal curah hujan bulanan adalah nilai rata-rata curah hujan pada masing-masing bulan selama 30 tahun berturut-turut yang periode waktunya sudah ditetapkan, yaitu:

1. 1 Januari 1901 s/d 31 Desember 1930,
2. 1 Januari 1931 s/d 31 Desember 1960,
3. 1 Januari 1961 s/d 31 Desember 1990,
4. 1 Januari 1971 s/d 31 Desember 2000,
5. 1 Januari 1981 s/d 31 Desember 2010,
6. 1 Januari 1991 s/d 31 Desember 2020, dan seterusnya.

Permulaan musim kemarau ditetapkan berdasarkan jumlah curah hujan dalam satu dasarian (10 hari) kurang dari 50 milimeter dan diikuti oleh beberapa dasarian berikutnya. Permulaan musim kemarau, bisa terjadi lebih awal (maju), sama atau lebih lambat (mundur) dari normalnya (rata-rata 1981-2010). Pemulaan musim hujan ditetapkan berdasarkan jumlah curah hujan dalam satu dasarian (10 hari) sama atau lebih dari 50 milimeter dan diikuti oleh beberapa dasarian berikutnya. Permulaan musim hujan, bisa terjadi lebih awal (maju), sama atau lebih lambat (mundur) dari normalnya (rata-rata dari tahun 1981 - 2010).

2.2 *Vektor Autoregressive* (VAR)

Model *Vector Autoregressive* (VAR) merupakan gabungan dari beberapa model AR (Autoregressive), dimana model ini membentuk vektor antara variabel yang saling mempengaruhi (Sims, 1972). Analisis VAR bisa dipadukan dengan suatu model persamaan simultan karena mempertimbangkan beberapa variabel endogen dan ekogen secara bersama-sama dalam model. Model VAR(1) adalah model VAR berorde 1 artinya variabel bebas dari model tersebut hanyalah satu nilai lag dari variabel tak bebasnya. Model VAR dapat dituliskan sebagai berikut (Tsay, 2014 :349).

$$\boldsymbol{Y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi} \boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

dengan:

$\boldsymbol{\phi}_0$ = Vektor konstanta ke- m

$\boldsymbol{\Phi}$ = matriks $m \times m$

$\boldsymbol{Y}_t$ = $\boldsymbol{Y}_t - \boldsymbol{\mu}$

$\boldsymbol{\mu}$ = $E(\boldsymbol{Y}_t)$

$\boldsymbol{\varepsilon}_t$ = Vektor $m \times 1$ dari residual pada waktu ke-t

$\boldsymbol{Y}_t$ = Vektor $m \times 1$ dari variabel pada waktu ke- t

$\boldsymbol{Y}_{t-1}$ = Vektor $m \times 1$ dari variabel pada waktu ke-$(t-1)$

2.3 *Generalized Space Time Autoregressive* (GSTAR)

Model *Space Time Autoregressive* (STAR) merupakan model yang dikategorikan berdasakan lag yang berpengaruh secara linier pada waktu dan lokasi (Preifer dan Deutsch, 1980). Model *Space Time Autoregressive* (STAR) mengamsusikan lokasi yang digunakan dalam penelitian adalah homogen, sehingga model ini hanya dapat diterapkan pada lokasi yang bersifat seragam. Model *Generalized Space Time Autoregressive* (GSTAR) merupakan generalisasi dari model *Space Time Autoregressive* (STAR). Model ini dikembangkan oleh Borovka, Lopuhaa dan Ruchjana (2002) yang mana model GSTAR adalah model STAR dengan asumsi bahwa asumsi parameter *autoregressive* dan parameter *time series* tidak harus bernilai sama atau homogen di setiap lokasi. Melihat perbedaan antar lokasi pada model GSTAR ditunjukkan dalam bentuk matriks pembobot.

Suatu deret $\{Y(t): t = 0, 1, 2, \dots\}$, merupakan sebuah deret waktu multivariate dari T pengamatan, maka persamaan model GSTAR untuk orde waktu AR$_{(p)}$ dan orde spasial 1 dengan menggunakan 3 lokasi yang berbeda adalah sebagai berikut.

$$\boldsymbol{Y}_{(t)} = \boldsymbol{\phi}_{10}\boldsymbol{Y}(t-1) + \boldsymbol{\phi}_{11}\boldsymbol{W}^{(l)}\boldsymbol{Y}(t-1) + \boldsymbol{\varepsilon}_t$$

dengan:

$\boldsymbol{Y}_{(t)}$ = vektor random berukuran $(T \times 1)$ pada waktu t

$\boldsymbol{\phi}_{10}$ = matriks koefisien parameter waktu

$\boldsymbol{\phi}_{11}$ = matriks koefisien spasial

$\boldsymbol{W}^{(l)}$ = nilai matriks pembobot ukuran $(m \times m)$ pada *lag* spasial ke-l

2.4 VAR-GSTAR

Model *Vector Autoregressive-Generalized Space Time Autoregressive* (VAR-GSTAR) merupakan model peramalan pada suatu data *time series multivariat* yang menggabungkan interdependasi waktu dan lokasi. VAR-GSTAR merupakan model VAR dengan skema respon prediktor yang direpresentasikan dalam skema pada model GSTAR (Wutsqa dan Suhartono, 2010). Model VAR-GSTAR adalah model yang direpresentasikan ke dalam model GSTAR. Model VAR(p) dituliskan sebagai berikut.

$$\boldsymbol{Z}_i(t) = \boldsymbol{\phi}_1\boldsymbol{Z}_{t-1} + \cdots + \boldsymbol{\phi}_p\boldsymbol{Z}_{t-p} + \boldsymbol{\varepsilon}_t$$

dengan:

$\boldsymbol{Z}_i(t)$ = nilai observasi pada daerah i waktu ke-t

$\boldsymbol{\phi}_1$ = matriks koefisien berukuran $m \times m$

$\boldsymbol{\varepsilon}_t$ = vektor sisaan berukuran $m \times 1$ terhadap waktu ke-t

Model tersebut direpresentasikan ke dalam model GSTAR. Model VAR-GSTAR yaitu.

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \varepsilon$$

atau dalam bentuk matriks

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

dimana:

$\boldsymbol{Y}_i = \begin{bmatrix} Z_{i,p+1} \\ Z_{i,p+2} \\ \vdots \\ Z_{i,n} \end{bmatrix}$, Y_i berukuran $(n-p) \times 1$ dan $\boldsymbol{Y}$ berukuran $m(n-p) \times 1$

$$\boldsymbol{X}_i = \begin{bmatrix} Z_{1,p} & \cdots & Z_{m.p} & \cdots & Z_{1,1} & \cdots & Z_{m,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{1,n-1} & \cdots & Z_{m,n-1} & \cdots & Z_{1,n-p} & \cdots & Z_{m,n-p} \end{bmatrix}$$

$$\boldsymbol{\varepsilon}_i = \begin{bmatrix} \boldsymbol{\varepsilon}_{i,p+1} \\ \boldsymbol{\varepsilon}_{i,p+2} \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix}$$

$i = 1,2, \dots, m$, m adalah banyaknya lokasi pengamatan

2.5 Stasioneritas Model

Asumsi stationeritas sangat penting pada analisis *time series*. Sifat-sifat statistik di masa yang akan datang dapat diramalkan berdasarkan data yang telah terjadi di masa lalu. Stationeritas berarti bahwa tidak terdapat terdapat perubahan yang signifikan pada data. Menurut Makridarkis (1995: 351) fuktuasi data berada di sekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan variasi dari fluktuasi tersebut. Pengujian stasioneritas dari model yaitu uji akar unit (*unit root test*) dengan melihat nilai *Augmented Dickey Fuller* (ADF).

Apabila setelah dilakukan uji stasioneritas ternyata diperoleh keseimpulan bahwa data tersebut tidak stasioner, maka selanjutnya dilakukan *Differencing* data. Menurut Box-Jenkins, time series yang bersifat stasioneritas memiliki mean, varians dan kovarians tidak berpengaruh oleh pengamatan. Menurut Aswi & Sukarna (2006:7) kondisi stationer terdiri atas dua hal, yaitu.

1. Stasioneritas dalam varians

Jika data tidak stasioner pada varians, maka harus dilakukan *transformasi*. *Box Cox* (1964) memperkenalkan transformasi pangkat (*power transformations*) dalam menangani data yang tidak stasioner dalam varians (Aswi & Sukarna, 2006:91). Beberapa penggunaan nilai serta kaitannya dengan transformasinya di tampilkan pada Tabel 2.1. sebagai berikut ini.

Tabel 2. 1 Nilai-nilai λ dengan transformasinya

Nilai λ	Transformasi
-1	$\frac{1}{Y_t}$
-0,5	$\frac{1}{\sqrt{Y_t}}$
0	lnY_t
0,5	$\sqrt{Y_t}$
1	Y_t

2. Stasioneritas dalam mean

Data *time series* dikatakan stasioner dalam *mean* jika $E(Y_t) = \mu$ artinya nilai *mean* konstan terhadap waktu. *Time series* yang tidak stasioner dalam *mean* yang berarti $E(Y_t)$ dipengaruhi oleh waktu pengamatan yaitu $E(Y_t) = \mu_t$. Untuk mengatasi *time series* yang tidak stasioner dalam *mean* dapat dilakukan pembedaan atau sering disebut dengan *differencing*. *Differencing* data *time series* dapat dirumuskan seperti berikut:

$$W_t = Y_t - Y_{t-1}$$

dimana:

W_t = *Differencing* orde ke-1

t = Indeks waktu

Y_t = Data Pengamatan ke- t

Y_{t-1} = Data pengamatan ke – $(t-1)$

2.6 Identifikasi Model

Model VAR-GSTAR mempunyai dua orde yaitu orde waktu yang diperoleh dari model VAR dan orde ruang yang ditentukann dari model GSTAR. Model ini menggunakan orde ruang (λ_s), karena orde ruang yang lebih tinggi sulit untuk diinterpretasikan. Orde ruang q menyatakan hubungan keterkaitan antar lokasi. Penentuan orde autoregressive pada model VAR-GSTAR dapat dilakukan dengan menggunakan orde model VAR(p). Penentuan panjang lag optimal yang akan digunakan dalam model VAR dapat ditentukan berdasarkan kriteria *Akaike Information Criterion* (AIC). Lag yang akan dipilih adalah model dengan nilai AIC yang paling kecil. Menurut Tsay (2005: 41) nilai AIC dapat menggunakan rumus berikut.

$$AIC = \ln\left(\frac{JKS}{n}\right) + \frac{2K^2}{n}$$

dimana:

JKS = jumlah kuadrat sisaan

n = banyak data

K = jumlah parameter pada model

2.7 Bobot Lokasi pada Model VAR-GSTAR

Sistem pembobotan lokasi umumnya dilakukan standardisasi, sehingga salah satu syarat dari matriks bobot adalah jumlah semua entri pada setiap baris sama dengan satu dan diasumsikan bobot suatu lokasi terhadap dirinya bernilai nol. Cara penentuan nilai bobot lokasi, yaitu bobot seragam, bobot normalisasi korelasi silang dan bobot inverse jarak.

1. Bobot Normalisasi Korelasi Silang

Bobot normalisasi korelasi silang menggunakan hasil normalisasi korelasi silang antar lokasi pada lag waktu yang bersesuaian (Suhartono & Atok, 2006). Pembobot normalisasi korelasi yang dikenalkan oleh Suhartono dan Subanar (2006) dituliskan sebagai berikut.

$$W_{ij}(k) = \frac{r_{ij}(k)}{\sum_{k=1}^{p} |r_{ik}(k)|}$$

dimana $i \neq j, k = 1,2, \dots, p$ dan taksiran dari korelasi silang pada data sampel dirumuskan sebagai berikut.

$$r_{ij}(k) = \frac{\sum_{t=k+1}^{n}[z_i(t) - \bar{z}_i][z_j(t-k) - \bar{z}_j]}{\sqrt{\left(\sum_{t=1}^{n}[z_i(t) - \bar{z}_i]^2\right)\left(\sum_{t=1}^{n}[z_j(t) - \bar{z}_j]^2\right)}}$$

dimana $z_i(t)$ merupakan data waktu ke-t pada daerah i, $z_j(t)$ merupakan data waktu ke-t pada daerah j dan k adalah lag waktu ke-k. Untuk memenuhi ketentuan bahwa jumlah elemen dalam matriks korelasi harus bernilai satu, maka perlu dilakukan normalisasi.

2. Bobot Inverse Jarak

Nilai dari bobot inverse jarak diperoleh dengan menggunakan perhitungan berdasarkan jarak sebenarnya antara lokasi dengan cara menginverskan jaraknya. Lokasi yan berdekatan mendapatkan nilai bobot yang lebih besar. Jarak yang digunakan pada bobot inverse jarak adalah satuan derajat lintang dan derajat bujur. Bobot inverse jarak dirumuskan sebagai berikut.

$$w_{ij} = \frac{w_{ij}^*}{\sum_{k=1}^{p} W_{ik}^*}$$

Dimana

$$w_{ij}^* = \begin{cases} \frac{1}{d_{ij}}, i \neq j \\ 0 \quad , i = j \end{cases}$$

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}$$

dengan w_{ij} adalah nilai bobot dari lokasi i dan j, d_{ij} adalah jarak dari lokasi i ke j, (u_i, u_j) adalah koordinat dari garis lintang dan (v_i, v_j) adalah koordinat dari garis bujur.

2.8 Estimasi Parameter

Menurut Wutsqa & Suhartono (2010) proses penentuan estimasi parameter untuk memperoleh nilai sisaan yang bersifat *white noise* secara multivariate dapat digunakan metode kuadrat terkecil (*least square*), 2SLS (*Two Stage Least Squares*) atau SUR (*Seemingly Unrelated*). Pada penelitian ini metode yang digunakan untuk mengestimasi parameter adalah metode kuadrat terkecil (*least square*). Metode kuadrat terkecil (*least square*) merupakan metode parameter yang bertujuan untuk meminimumkan kuadrat sisaan, sehingga nilai estimasinya akan mendekati nilai yang sesungguhnya (Miller, 2006).

Sisaan dari model $Z = Z^*\Phi + \varepsilon$, dirumuskan sebagai berikut.

$$\varepsilon = Z - Z^*\Phi$$

sehingga jumlah kuadrat sisaan model tersebut adalah

$$\varepsilon'\varepsilon = (Z - Z^*\Phi)'(Z - Z^*\Phi)$$

$$= Z'Z - Z'^{Z^*}\Phi - \Phi'Z^*Z + \Phi'Z^{*\prime}Z^*\Phi$$

$$= Z'Z - 2\Phi'^{Z^*}Z + \Phi'Z^{*\prime}Z^*\Phi$$

Nilai estimasi Φ yang meminimumkan jumlah kuadrat sisaannya diperoleh dari turunan parsial pertama fungsi $\varepsilon'\varepsilon$ terhadap Φ dan disama dengankan dengan 0 sehingga diperoleh.

$$\frac{\partial\varepsilon'\varepsilon}{\partial\Phi} = 0$$

$$-2\Phi'Z^*Z + \Phi'{Z^*}'Z^*\Phi = 0$$

$$\hat{\Phi} = \left({Z^*}'Z\right)^{-1}({Z^*}'Z)$$

Berdasarkan hasil estimasi parameter model VAR-GSTAR dengan metode kuadrat terkecil diperoleh bahwa estimator untuk $\hat{\Phi}$ adalah $\left({Z^*}'Z\right)^{-1}\left({Z^*}'Z\right)$.

2.9 Uji White Noise

Nilai sisaan pada model VAR-GSTAR yang diperoleh harus bersifat *white noise*. Nilai sisa *white noise* adalah sisaan yang mengikuti distribusi identik independen yang diditeksi dengan uji *white noise*. Uji *white noise* digunakan untuk mengetahui ada tidaknya korelasi nilai sisa antar lag. Pemeriksaan sisaan yang bersifat *white noise* dapat dilakukan dengan menggunakan uji *Ljung and Box*. Langkah-langkah pengujian *white noise* dengan uji *Ljung-Box* adalah sebagai berikut.

1. Hipotesis

   H_0 : sisaan tidak bersifat *white noise*

   H_1 : sisaan bersifat *white noise*

2. Statistik Uji

$$Q_{LB} = n(n+2)\sum_{k=1}^{n}\left(\frac{1}{n-k}\right)\hat{p}_k^2$$

dimana:

n = banyaknya pengamatan

k = banyaknya lag

$\hat{p}_k^2$ = autokorelasi duga pada lag ke- k

3. Kriteria Pengujian

   Dengan mengambil taraf signifikan $(\alpha = 0{,}05)$ dan ukuran sampel n, tolak H_0 jika $Q_{LB} > \chi^2_{1-\alpha;k}$

2.10 Verifikasi Model VAR-GSTAR

Verifikasi model digunakan memastikan ukuran ketepatan model, setelah dinyatakan mempunyai sisaan yang bersifat *white noise*. Salah satu cara untuk menentukan model terbaik dapat dilakukan dengan menggunakan *Root Mean Square Error* (RMSE). RMSE digunakan untuk melihat seberapa besar nilai sisa dalam model yang digunakan. RMSE akan meningkatkan bersama dengan total square error. Selain itu RMSE dapat untuk mengindikasi adanya ketidakcocokan dalam pemodelan (Willmott & Matsuura, 2005).

Root Mean Square Error (RMSE) dapat ditentukan dengan perumusan sebagai berikut.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}\left(Z_t - \hat{Z}_t\right)^2}{n}}$$

dimana Z_t adalah data actual, $\hat{Z}_t$ adalah data prediksi dengan suatu sistem pembobotan lokasi yang dipilih, dan n adalah banyaknya data. Model yang dipilih adalah model yang memiliki nilai RMSE terkecil.

2.11 Validasi Model

Validasi Model dengan menggunakan nilai MAPE atau Mean Absolute Percentage Error merupakan ukuran kesalahan nilai dugaan model yang dinyatakan dalam bentuk rata-rata persentase absolut residual. Persamaan MAPE di bawah ini sebagai berikut.

$$MAPE = \frac{100\%}{M}\sum_{t=1}^{M}\left|\frac{Z(t) - \hat{Z}(t)}{Z(t)}\right|$$

dimana:

M = banyak data ramalan yang dilakukan

$Z(t)$ = data yang sebenarnya pada waktu ke$-t$

$\hat{Z}(t)$ = data hasil ramalan pada waktu ke$-t$

Menurut Zainun dan Majid (2003) kemampuan peramalan sangat baik jika memiliki nilai MAPE kurang dari 10% dan mempunyai kemampuan peramalan yang baik jika nilai MAPE kurang dari 20%.