BAB I
PENDAHULUAN

1.1 Latar Belakang

Curah hujan merupakan salah satu unsur iklim yang penting dan memiliki dampak yang besar di berbagi sektor kehidupan. Indonesia merupakan negara yang memiliki iklim tropis dengan suhu yang relatif stabil setiap tahunnya. Wilayah tropis cenderung memiliki curah hujan yang tinggi dibandingkan dengan wilayah lainnya. Curah hujan di Indonesia relatif tinggi, rata-rata nasional sekitar 2.800 mm per tahunnya. Tetapi, tidak menutup kemungkinan terdapat perbedaan yang signifikan antara wilayah–wilayah di Indonesia. Wilayah barat Indonesia seperti Sumatera dan Jawa Barat cenderung memiliki curah hujan tinggi, sementara wilayah timur Indonesia, seperti Nusa Tenggara Timur dan Papua Barat cenderung memiliki curah hujan rendah (BMKG, 2018).

Curah hujan di Indonesia dipengaruhi oleh banyak faktor diantaranya letak geografis, topografi, angin dan sebagainya. Letak geografis Indonesia berada di antara dua samudera yaitu Samudera Hindia dan Samudera Pasifik yang menyebabkan Indonesia menjadi tempat pertemuan antara angin muson barat dari Samudera Hindia dan angin muson timur dari Samudera Pasifik. Angin muson barat membawa uap air dari Samudera Hindia dan mengalami kondensasi saat melewati wilayah pegunungan yang memicu turunnya hujan deras. Oleh karena itu, bagian barat Indonesia yang sebagian besar berupa pegunungan memiliki curah hujan yang tinggi. Sedangkan angin muson timur dari Samudera Pasifik cenderung membawa udara kering karena melewati wilayah Australia dan kering yang mengakibatkan kemarau di Indonesia (Pandia et al., 2019).

Sulawesi Selatan merupakan salah satu provinsi di Indonesia yang curah hujannya beragam. Di bagian utara seperti Enrekang, Tana Toraja, sebagian Luwu dan sekitarnya cenderung memiliki curah hujan yang tinggi karena efek dari angin muson barat yang bertiup ke Indonesia dan daerahnya yang didominasi pegunungan sehingga memicu pembentukan awan dan proses kondensasi yang mengakibatkan turunnya hujan di wilayah tersebut. Di bagian selatan seperti Makassar, Gowa, Barru, Takalar, Bone dan sekitarnya curah hujannya relatif rendah karena wilayahnya yang lebih jauh dijangkau oleh angin muson barat. Selain itu, daerahnya didominasi dengan dataran rendah dan sedikit pegunungan. Hal tersebut mengakibatkan daerah selatan mengalami kemarau yang lebih panjang. Beragamnya curah hujan di provinsi Sulawesi Selatan sehingga diperlukan peramalan di wilayah ini.

Salah satu peramalan pada suatu data deret waktu yang umum digunakan untuk data curah hujan adalah *Autoregressive Integrated Moving Average* (ARIMA). Namun, model ARIMA hanya bersifat *univariate* dan tidak memperhatikan faktor lokasi. Oleh karena itu perlu diterapkan peramalan yang sesuai untuk memprediksi curah hujan yang bersifat *multivariate* yang menggabungkan interdependasi waktu dan lokasi karena curah hujan mempunyai keragaman yang besar, baik dalam skala ruang maupun dalam skala waktu. Pada tahun 1980 Christoper A. Sims memperkenalkan model *Vector Autoregressive* (VAR). Model ini merupakan salah satu model peramalan *time series* yang bersifat *multivariate* (Diani et al., 2014).

VAR merupakan perluasan dari model ARMA. Menurut Wutsqa dan Suhartono (2010) model ini menjelaskan keterkaitan antar pengamatan pada variabel tertentu pada suatu waktu dengan pengamatan pada variabel itu sendiri pada waktu-waktu sebelumnya, dan juga keterkaitannya dengan pengamatan pada variabel lain pada waktu-waktu sebelumnya. Adapun model yang menggabungkan keterkaitan waktu dan lokasi (*space time*) dikenal dengan model *Space Time Autoregressive* (STAR). Model ini diperkenalkan oleh Preifer dan Deutch tahun 1980. Namun, model STAR cenderung tidak fleksibel ketika dihadapkan pada lokasi yang memiliki karakteristik yang berbeda.

Model yang lebih fleksibel dari model STAR adalah model *Generalized Space Time Autoregressive* (GSTAR). Model GSTAR diperkenalkan oleh Borovka, Lopuhaa dan Ruchjana (2002) yang mana model ini adalah model STAR dengan asumsi bahwa asumsi parameter *autoregressive* dan parameter *time series* tidak harus bernilai sama atau homogen di setiap lokasi sehingga model GSTAR lebih realistis (Adam et al., 2017). Pada kenyataannya juga lebih banyak ditemui model dengan parameter model berbeda untuk lokasi yang berbeda.

Model-model tersebut masih terbatas pada data *time series multivariate* yang stasioner, tetapi belum melibatkan pola musiman atau *seasonal*. Oleh karena itu, untuk menangani hal tersebut diperlukan suatu pemodelan *time series multivariate* dengan pola *seasonal*, model tersebut dikenal *Vector Autoregressive-Generalized Space Time Autoregressive* (VAR-GSTAR). Model VAR-GSTAR merupakan model *Vector Autoregressive* dengan skema respon prediktor yang direpresentasikan dalam skema *Generalized Space Time Autoregressive*.

Penelitian sebelumnya telah dilakukan oleh Wutsqa dan Suhartono (2010) yang mana membandingkan model VARMA dengan model VAR-GSTAR pada data pariwisata. Hasil dari penelitian tersebut menyebutkan bahwa model VAR-GSTAR memberikan hasil yang realitis karena model VARMA tidak dapat mengakomodasi lag-lag *seasonal* pada orde model. Penelitian yang sama juga dilakukan oleh Waqi'ah (2017) dengan membandingkan model VAR-GSTAR dengan model GSTAR pada data pariwisata di Kabupaten Jember. Hasil penelitian menghasilkan bahwa peramalan data pariwisata di Kabupaten Jember mengggunakaan VAR-GSTAR lebih baik daripada model GSTAR. Lalu, Mesyi (2021) yang meramalkan curah hujan di Provinsi Sumatera Barat menggunakan VAR-GSTAR. Hasil dari penelitiannya menyebutkan bahwa tingkat akurasi ramalannya dengan menggunakan model VAR-GSTAR sangat akurat.

Berdasarkan uraian di atas maka penelitian ini berjudul **"Penggunaan Model *Vector Autoregressive-Generalized Space Time Autoregressive* Pada Peramalan Curah Hujan Provinsi Sulawesi Selatan Tahun 2019-2023"**

1.2 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini sebagai berikut:

1. Data yang digunakan merupakan data curah hujan di empat stasiun pengamatan Provinsi Sulawesi Selatan yaitu Stasiun Klimatologi Maros, Stasiun Meteorologi Sultan Hasanuddin, Stasiun Meteorologi Maritim Paotere dan Stasiun Geofisika Gowa tahun 2019-2023
2. Pembobot lokasi yang digunakan adalah bobot lokasi invers jarak dan normalisasi korelasi silang
3. Estimasi parameter model yang digunakan adalah *Ordinary Least Square*

1.3 Tujuan Penelitian

Berdasarkan latar belakang yang telah diuraikan, maka tujuan penelitian dalam penelitian ini adalah:

1. Mendapatkan dugaan model *Vector Autoregressive-Generalized Space Time Autoregressive* pada data curah hujan untuk empat stasiun pengamatan di Provinsi Sulawesi Selatan
2. Meramalkan data curah hujan di Provinsi Sulawesi Selatan tahun 2023 menggunakan model *Vector Autoregressive-Generalized Space Time Autoregressive*

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Menambah wawasan keilmuwan dan pemahaman mengenai pemodelan dengan model *Vector Autoregressive-Generalized Space Time Autoregressive*
2. Dapat dijadikan sebagai salah satu acuan dalam meramalkan curah hujan di Provinsi Sulawesi Selatan untuk periode-periode selanjutnya

1.5 Kajian Teori

1.5.1 Analisis Runtun Waktu

Analisis Runtun waktu diperkenalkan pada tahun 1970 oleh George E. P. Box dan Gwilym M. Jenkins melalui bukunya yang berjudul *Time Series Analysis: Forecasting and Control.* Runtun waktu (*time series*) merupakan serangkaian data pengamatan yang terjadi berdasarkan indeks waktu secara berurutan dengan interval waktu tetap. Analisis runtun waktu adalah salah satu prosedur statistika yang diterapkan untuk meramalkan struktur probabilistik keadaan yang akan terjadi di masa yang akan datang dalam pengambilan keputusan (Aswi & Sukarna, 2006).

Secara garis besar pemodelan data *time series* terbagi atas dua klasifikasi yaitu univariat dan multivariat. Pada model univariat peramalan data suatu variabel hanya didasarkan pada nilai variabel tersebut pada masa lampau, sedangkan model multivariat peramalan data dilakukan dengan menggunakan beberapa variabel secara simultan yang memiliki hubungan atau saling berkorelasi untuk mendapatkan keakuratan peramalan (Eksiandayani, 2016)

Meskipun berhubungan erat dengan urutan waktu, tidak menutup kemungkinan *time series* juga berhubungan erat terhadap dimensi lain seperti ruang. Model *multivariate time series* yang digunakan dalam penelitian ini adalah *Vector Autoregressive*.

1.5.2 Data Spasial

Data spasial merupakan data dependen terhadap lokasi, karena berasal dari lokasi yang berbeda dengan indikasi adanya keterkaitan antara pengukuran data dengan lokasi (Cressie, 1993). Sehingga dapat disimpulkan bahwa data spasial adalah serangkaian data yang diamati di beberapa lokasi tertentu dan memiliki ketergantungan antar lokasi.

Berdasarkan tipe data, pemodelan spasial dapat dibedakan menjadi pemodelan dengan pendekatan titik dan area (Salamah et al., 2012). Pemodelan spasial dengan pendekatan titik antara lain *Geographically Weighted Regression* (GWR), *Geographically Weighted Poisson Regression* (GWPR), *Geographically Weighted Negative Binomial*

Regression (GWNBR), *Space Time Autoregressive* (STAR) dan *Generalized Space Time Autoregressive* (GSTAR). Sedangkan pemodelan spasial dengan pendekatan area yaitu *General Nesting Spatial* (GNS), *Spatial Autoregressive Confused* (SAC), *Spatial Durbin Model* (SDM), *Spatial Durbin Error Model* (SDEM), *Spatial Autoregressive* (SAR), *Spatial Lag of X* (SLX) dan *Spatial Error Model* (SEM) (LeSage, 1998).

1.5.3 *Vector Autoregressive*

Vector Autoregressive (VAR) diperkenalkan oleh Sims (1972) sebagai pengembangan dari pemikiran Granger dan Clive (1969). VAR digunakan untuk pemodelan beberapa variabel endogen yang memiliki hubungan antar variabel endongennya yang dipengaruhi oleh waktu sebelumnya. Model VAR merupakan gabungan dari beberapa model AR (*Autoregressive*), dimana model ini membentuk vektor antara variabel yang saling mempengaruhi (Sihombing & Susilowati, 2019)

Model persamaan Vector Autoregressive (VAR) dengan m variabel dan orde p secara umum dapat dinyatakan seperti pada Persamaan 1 berikut:

$$\boldsymbol{Z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1\boldsymbol{Z}_{(t-1)} + \boldsymbol{\phi}_2\boldsymbol{Z}_{(t-2)} + \cdots + \boldsymbol{\phi}_p\boldsymbol{Z}_{(t-p)} + \boldsymbol{\varepsilon}_t \quad (1)$$

keterangan:

$\boldsymbol{Z}_t$	: vektor berukuran $m \times 1$ pada waktu ke-t
$\boldsymbol{\phi}_0$	: vektor konstanta berukuran $m \times 1$
$\boldsymbol{Z}_{t-i}$	: vektor berukuran $m \times 1$ dengan $i = 1,2,\dots,p$
$\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, $\boldsymbol{\phi}_p$	: matriks koefisien VAR berukuran $m \times m$
$\boldsymbol{\varepsilon}_t$	: vektor sisaan berukuran $m \times 1$ pada waktu ke-t

Persamaan (1) dapat dituliskan sebagai berikut

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{m,t} \end{bmatrix} = \begin{bmatrix} \phi_{1,0} \\ \phi_{2,0} \\ \vdots \\ \phi_{m,0} \end{bmatrix} + \begin{bmatrix} \phi_{1,1(1)} & \phi_{1,2(1)} & \cdots & \phi_{1,m(1)} \\ \phi_{2,1(1)} & \phi_{2,2(1)} & \cdots & \phi_{2,m(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m,1(1)} & \phi_{m,2(1)} & \cdots & \phi_{m,m(1)} \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \\ \vdots \\ Z_{m,t-1} \end{bmatrix} +$$
$$\begin{bmatrix} \phi_{1,1(2)} & \phi_{1,2(2)} & \cdots & \phi_{1,m(2)} \\ \phi_{2,1(2)} & \phi_{2,2(2)} & \cdots & \phi_{2,m(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m,1(2)} & \phi_{m,2(2)} & \cdots & \phi_{m,m(2)} \end{bmatrix} \begin{bmatrix} Z_{1,t-2} \\ Z_{2,t-2} \\ \vdots \\ Z_{m,t-2} \end{bmatrix} + \cdots +$$
$$\begin{bmatrix} \phi_{1,1(p)} & \phi_{1,2(p)} & \cdots & \phi_{1,m(p)} \\ \phi_{2,1(p)} & \phi_{2,2(p)} & \cdots & \phi_{2,m(p)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m,1(p)} & \phi_{m,2(p)} & \cdots & \phi_{m,m(p)} \end{bmatrix} \begin{bmatrix} Z_{1,t-p} \\ Z_{2,t-p} \\ \vdots \\ Z_{m,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{m,t} \end{bmatrix} \quad (2)$$

keterangan:

$Z_{j,t}$	: Data variabel j pada waktu ke-t dengan $j = 1,2,\dots,m$ dan $t = 1,2,\dots,n$
$\phi_{j,0}$	: konstanta persamaan variabel j dengan $j = 1,2,\dots,m$
$Z_{j,t-i}$	: Data variabel j pada waktu ke-$t-i$ dengan $j = 1,2,\dots,m$, $t = 1,2,\dots,n$ dan $i = 1,2,\dots,p$
$\phi_{j,j(i)}$	: koefisien parameter persamaan variabel j untuk variabel ke-j pada lag ke-i dengan $j = 1,2,\dots,m$ dan $i = 1,2,\dots,p$

$\varepsilon_{j,t}$: Sisaan variabel j pada waktu ke-t dengan $j = 1,2,\dots,m$ dan $t = 1,2,\dots,n$

1.5.4 *Generalized Space Time Autoregressive*

Model *Space Time Autoregressive* (STAR) merupakan model yang dikategorikan berdasakan lag yang berpengaruh secara linier pada waktu dan lokasi (Pfeifer & Deutsch, 1980). Model 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 STAR. Model ini dikembangkan oleh Borovka dkk. (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. Di bawah ini merupakan model GSTAR (p_1) dimana p adalah derajat *autoregressive* dan 1 menunjukkan spasial lag dari bentuk AR(p) dengan $\boldsymbol{Z}_t$, $t = 0, \pm 1, \pm 2, \dots.)$ adalah deret waktu *multivariate* dari komponen N yang dinyatakan pada Persamaan 3 sebagai berikut:

$$\boldsymbol{Z}_t = \boldsymbol{\phi}_{10}\boldsymbol{Z}_{(t-1)} + \boldsymbol{\phi}_{11}\boldsymbol{W}^{(l)}\boldsymbol{Z}_{(t-1)} + \boldsymbol{\varepsilon}_t \quad (3)$$

keterangan:

$\boldsymbol{Z}_t$: vektor random berukuran $m \times 1$ pada waktu ke-t

$\boldsymbol{\phi}_{10}$: matriks koefisien parameter waktu berukuran $m \times m$

$\boldsymbol{\phi}_{11}$: matriks koefisien spasial berukuran $m \times m$

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

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

1.5.5 *Vector Autoregressive-Generalized Space Time Autoregressive*

Vector Autoregressive-Generalized Space Time Autoregressive (VAR-GSTAR) dapat digunakan untuk peramalan pada suatu data deret waktu multivariat yang menggabungkan interdependasi waktu dan lokasi. VAR-GSTAR adalah model VAR dengan skema respon prediktor yang direpresentasikan dalam skema pada model GSTAR (Wutsqa & Suhartono, 2010). Untuk membawa model VAR pada Persamaan 1 dalam bentuk model GSTAR diperlukan beberapa notasi seperti berikut. Pertama-tama diasumsikan terdapat n pengamatan bentuk vektor respon $\boldsymbol{Y} = (\boldsymbol{Y}'_1, \boldsymbol{Y}'_2, \dots, \boldsymbol{Y}'_m) = (Y'_1, Y'_2, \dots, Y'_m)'$, matriks $\boldsymbol{X} = \text{diag}(\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_m)$, vektor parameter $\boldsymbol{\beta} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_m)'$, dan vektor sisaan $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \dots, \boldsymbol{\varepsilon}'_m)'$, dengan $\boldsymbol{\phi}_i = (\phi^{(i)}_{1,1}, \dots, \phi^{(i)}_{1,m}, \dots, \phi^{(i)}_{p,1}, \dots, \phi^{(i)}_{p,m})$, $\boldsymbol{Y}_i = \begin{pmatrix} Z_{i,p+1} \\ Z_{i,p+2} \\ \vdots \\ Z_{i,n} \end{pmatrix}$, $\boldsymbol{X}_i = \begin{pmatrix} Z_{i,p} & \dots & Z_{m,p} & \dots & Z_{1,1} & \dots & Z_{m,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{1,n-1} & \dots & Z_{m,n-1} & \dots & Z_{1,n-p} & \dots & Z_{m,n-p} \end{pmatrix}$, dan $\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i,p+1} \\ \varepsilon_{i,p+2} \\ \vdots \\ \varepsilon_{i,n} \end{pmatrix}$

Dengan demikian model VAR(p) dapat dinyatakan dalam bentuk model pada persamaan 4

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$

yang dalam penelitian ini disebut sebagai model VAR-GSTAR.

1.5.6 Kestasioneran Data

Sebelum melakukan identifikasi model, langkah pertama yang harus dilakukan adalah pengujian stasioneritas untuk melihat pola data yang digunakan dalam penelitian tersebut stasioner atau tidak. Runtun waktu yang stasioner adalah relatif tidak terjadi kenaikan maupun penurunan nilai secara tajam pada data dan fluktuasi data berada pada sekitar nilai rata-rata yang konstan (Wei, 2006).

Kestasioneran pada pemodelan *multivariate time series* secara grafis dapat dilihat dengan plot ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*). Hasil plot yang turun secara lambat menunjukkan bahwa data belum stasioner dalam rata-rata. Apabila plot *time series* menunjukkan data tidak stasioner dalam rata-rata maka hal ini dapat diatasi dengan melakukan *differencing* antar pengamatan (Makridakis et al., 1999).

Selain melihat secara grafis, kestasioneran dalam rata-rata dapat diketahui dengan Uji Augmented Dickey Fuller sebagai berikut:

1. hipotesis:

$H_0 : \gamma = 0$ (data tidak stasioner)

$H_1 : \gamma \neq 0$ (data stasioner)

2. tingkat signifikansi

$\alpha = 5\% = 0.05$

3. statistik uji:

$$t_{hitung} = \left| \frac{\hat{\gamma}}{SE(\hat{\gamma})} \right| \quad (5)$$

4. daerah kritis:

Tolak $H_0 \left| t_{hitung} \right| > t_{(\frac{\alpha}{2};df=n-n_p)}$ atau *p-value* $< \alpha$

keterangan:

$\hat{\gamma}$: nilai taksiran dari parameter γ

$SE(\hat{\gamma})$: standar error dari nilai taksiran $\hat{\gamma}$

n : banyaknya pengamatan

n_p : jumlah parameter

1.5.7 Identifikasi Model

Model VAR-GSTAR mempunyai dua orde yaitu orde waktu yang diperoleh dari model VAR dan orde ruang yang ditentukan dari model GSTAR. Model ini menggunakan orde ruang $(\lambda_s) = 1$, karena orde ruang yang lebih tinggi sulit untuk diinterpretasikan.

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's Information Criterion* (AIC). Lag yang akan dipilih adalah model dengan nilai AIC yang paling kecil. Nilai AIC dapat ditentukan dengan menggunakan perumusan berikut (Tsay, 2010):

$$AIC = \ln\left(\frac{JKS}{n}\right) + \frac{2K^2}{n} \quad (6)$$

keterangan:

JKS : Jumlah Kuadrat Sisaan

n : banyak data

K : jumlah parameter pada model

1.5.8 Pembobot Lokasi

Pada sistem pembobotan lokasi umumnya dilakukan standarisasi, sehingga salah satu syarat dari matriks bobot adalah jumlah semua entri pada setiap baris sama dengan 1 dan diasumsikan bobot suatu lokasi terhadap dirinya bernilai 0 (Suhartono & Atok, 2006).

A. Bobot Invers Jarak

Bobot invers jarak menggunakan jarak sebenarnya antar lokasi, perhitungan bobot diperoleh dari hasil invers jarak sebenarnya yang kemudian dinormalisasi. Bobot invers jarak memberikan nilai koefisien yang kecil untuk jarak yang lebih jauh dan sebaliknya. Bobot invers jarak dapat dihitung dengan rumus sebagai berikut (Siregar, 2015).

$$\boldsymbol{W} = \begin{cases} \dfrac{w_{ij}}{\sum_{j=1}^{N} w_{ik}}, i \neq j \\ 0 \qquad\quad , i = j \end{cases} \tag{7}$$

dengan

$$w_{ij} = \frac{1}{d_{ij}} \tag{8}$$

$$d_{ij} = d_{ji} = \left(\left[x_i(u_i) - x_j(u_j) \right]^2 + \left[x_i(v_i) - x_j(v_j) \right]^2 \right)^{\frac{1}{2}} \tag{9}$$

keterangan:

x_i : lokasi ke-i dengan $i = 1,2, \dots, N$

u_i : garis lintang lokasi ke-i dengan $i = 1,2, \dots, N$

v_i : garis bujur lokasi ke-i dengan $i = 1,2, \dots, N$

d_{ij} : jarak antar lokasi ke-i terhadap lokasi ke-j

B. Bobot Normalisasi Korelasi Silang

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

$$\boldsymbol{W} = \frac{r_{ij}(k)}{\sum_{k=1}^{p} |r_{ik}(k)|} \tag{10}$$

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

$$r_{ij}(k) = \frac{\sum_{t=k+1}^{n} [z_i(t) - \bar{z}_i][z_j(t-k) - \bar{z}_j]}{\sqrt{\left(\sum_{t=1}^{n} [z_i(t) - \bar{z}_i]^2\right)\left(\sum_{t=1}^{n} [z_j(t) - \bar{z}_j]^2\right)}} \tag{11}$$

keterangan:

$r_{ij}(k)$: Nilai korelasi silang antara variabel r_i dan r_j

$r_{ik}(k)$: Nilai korelasi silang antara variabel r_i dan r_k

$z_i(t)$: data waktu ke-t pada daerah i

$z_j(t)$: data waktu ke-t pada daerah j

k : lag waktu ke-k

1.5.9 Estimasi Parameter

Setelah orde model diidentifikasi, langkah selanjutnya yaitu melakukan estimasi pada model VAR-GSTAR. Pada penelitian ini metode yang digunakan untuk mengestimasi parameter adalah metode kuadrat terkecil (*Ordinary Least Square*). Metode kuadrat terkecil merupakan metode parameter yang bertujuan untuk meminimumkan kuadrat sisaan, sehingga nilai estimasinya akan mendekati nilai yang sesungguhnya (Miller, 2006). Sisaan dari model $\boldsymbol{Z} = \boldsymbol{Z}^{*}\boldsymbol{\Phi} + \boldsymbol{\varepsilon}$ dirumuskan sebagai berikut:

$$\boldsymbol{\varepsilon} = \boldsymbol{Z} - \boldsymbol{Z}^{*}\boldsymbol{\Phi} \tag{12}$$

sehingga, jumlah kuadrat sisaan model tersebut adalah

$$\begin{aligned} \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\boldsymbol{Z} - \boldsymbol{Z}^{*}\boldsymbol{\Phi})'(\boldsymbol{Z} - \boldsymbol{Z}^{*}\boldsymbol{\Phi}) \\ &= \boldsymbol{Z}'\boldsymbol{Z} - \boldsymbol{Z}'^{\boldsymbol{Z}^{*}}\boldsymbol{\Phi} - \boldsymbol{\Phi}'^{\boldsymbol{Z}^{*}}\boldsymbol{Z} + \boldsymbol{\Phi}'\boldsymbol{Z}^{*\prime}\boldsymbol{Z}^{*}\boldsymbol{\Phi} \\ &= \boldsymbol{Z}'\boldsymbol{Z} - 2\boldsymbol{\Phi}'^{\boldsymbol{Z}^{*}}\boldsymbol{Z} + \boldsymbol{\Phi}'\boldsymbol{Z}^{*\prime}\boldsymbol{Z}^{*}\boldsymbol{\Phi} \end{aligned} \tag{13}$$

Nilai estimasi $\boldsymbol{\Phi}$ yang meminimukan jumlah kuadrat sisaannya diperoleh dari turunan parsial pertama fungsi $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$ terhadap $\boldsymbol{\Phi}$ dan disama dengankan dengan 0 sehingga diperoleh

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\partial \boldsymbol{\Phi}} &= 0 \\ -2\boldsymbol{\Phi}'^{\boldsymbol{Z}^{*}}\boldsymbol{Z} + \boldsymbol{\Phi}'^{\boldsymbol{Z}^{*}\prime\boldsymbol{Z}^{*}}\boldsymbol{\Phi} &= 0 \\ \hat{\boldsymbol{\Phi}} &= (\boldsymbol{Z}^{*\prime}\boldsymbol{Z})^{-1}(\boldsymbol{Z}^{*\prime}\boldsymbol{Z}) \end{aligned} \tag{14}$$

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

1.5.10 Uji Asumsi *White Noise*

Uji *White Noise* bertujuan untuk melihat apakah sisaan dari model sudah saling independen (saling bebas) antara satu dengan lain. 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. tingkat signifikansi

$\alpha = 5\% = 0.05$

3. statistik uji:

$$Q_{LB} = n(n+2)\sum_{k=1}^{n}\left(\frac{1}{n-k}\right)\hat{P}_k^2 \tag{15}$$

4. daerah kritis:

Tolak H_0 jika nilai $Q_{lB} > X^2_{1-a;k}$ atau *p-value*> α

keterangan

n : banyaknya pengamatan
k : banyaknya lag

$\hat{P}_k^2$: autokorelasi duga pada lag ke-k

1.5.11 Pemilihan Model Terbaik

Pemilihan model terbaik *time series* dilakukan pada data *testing*. Kriteria pemilihan model terbaik yang digunakan pada penelitian ini adalah *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). Model yang dipilih adalah model yang memiliki nilai RMSE terkecil.

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

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Z_t - \hat{Z}_t)^2} \tag{16}$$

keterangan:

Z_t : data aktual

$\hat{Z}_t$: data prediksi dengan suatu sistem pembobotan lokasi yang dipilih

n : banyaknya data

1.5.12 *Forecasting*

Setelah data sudah stasioner dan memenuhi asumsi-asumsi sisaan berarti model sudah cocok dan dapat merepresentasikan data dengan baik. Selanjutnya dapat langsung melakukan peramalan. Peramalan adalah proses untuk memperkirakan kebutuhan dimasa yang akan datang meliputi kebutuhan dalam ukuran kuantitas, kualitas, waktu, dan lokasi yang dibutuhkan dalam rangka memenuhi permintaan barang atau jasa (Nasution & Prasetyawan, 2008)

Prakiraan pada dasarnya merupakan suatu dugaan atau prediksi mengenai terjadinya suatu kejadian atau peristiwa dimasa yang akan datang. Prakiraan dapat disebut juga dengan peramalan yang ilmiah (Assauri, 1984). Dengan kata lain, peramalan dapat diartikan sebagai proses untuk memperkirakan kebutuhan dimasa yang akan datang dengan menggunakan data di masa lalu yang dapat memberikan hasil peramalan yang dipercaya ketetapannya menggunakan metode-metode peramalan yang bertujuan untuk meminimalkan resiko kesalahan.

1.5.13 Curah Hujan

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. Sedangkang 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 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 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 diikui oleh beberapa dasarian berikutnya. Permulaan musim hujan bisa terjadi lebih awal (maju), sama atau lebih lambat (mundur) dari normalnya (rata-rata 1981-2010 (BMKG, 2020).

BAB II
METODOLOGI PENELITIAN

2.1 Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh melalui lembaga pemerintahan resmi yakni dari Badan Meteorologi, Klimatologi dan Geofisika (BMKG) Wilayah IV. Data yang digunakan adalah data curah hujan pada empat stasiun pengamatan Provinsi Sulawesi Selatan dari periode Januari 2019 hingga Desember 2023 dengan jumlah data sebanyak 60 data, data terlampir (**Lampiran 1**).

2.2 Variabel Penelitian dan Struktur Data

Variabel penelitian yang digunakan adalah data curah hujan pada empat stasiun pengamatan Provinsi Sulawesi Selatan yaitu Stasiun Klimatologi Maros (Stasiun 1), Stasiun Meteorologi Sultan Hasanuddin (Stasiun 2), Stasiun Meteorologi Maritim Paotere (Stasiun 3) dan Stasiun Geofisika Gowa (Stasiun 4).

Data curah hujan di bagi menjadi data *training* dan data *testing* dengan proporsi 80% untuk data *training* dan 20% untuk data *testing*. Data *training* digunakan untuk membentuk model dan data *testing* digunakan untuk memeriksa daya ramal atau validasi model yang dibentuk dari data *training*. Data *training* meliputi dari bulan Januari 2019 hingga Desember 2022, sedangkan data *testing* meliputi dari bulan Januari 2023 hingga Desember 2023. Data *training* terdiri dari 48 data dan data *testing* terdiri dari 12 data. Adapun struktur data disajikan pada Tabel 2 sebagai berikut.

Tabel 1. Struktur Data

Tahun	Bulan	Stasiun 1	Stasiun 2	Stasiun 3	Stasiun 4
2019	Januari	$Z_{1,1}$	$Z_{2,1}$	$Z_{3,1}$	$Z_{4,1}$
	Februari	$Z_{1,2}$	$Z_{2,2}$	$Z_{3,2}$	$Z_{4,2}$
	Maret	$Z_{1,3}$	$Z_{2,3}$	$Z_{3,3}$	$Z_{4,3}$
	April	$Z_{1,4}$	$Z_{2,4}$	$Z_{3,4}$	$Z_{4,4}$
	Mei	$Z_{1,5}$	$Z_{2,5}$	$Z_{3,5}$	$Z_{4,5}$
	Juni	$Z_{1,6}$	$Z_{2,6}$	$Z_{3,6}$	$Z_{4,6}$
	Juli	$Z_{1,7}$	$Z_{2,7}$	$Z_{3,7}$	$Z_{4,7}$
	Agustus	$Z_{1,8}$	$Z_{2,8}$	$Z_{3,8}$	$Z_{4,8}$
	September	$Z_{1,9}$	$Z_{2,9}$	$Z_{3,9}$	$Z_{4,9}$
	Oktober	$Z_{1,10}$	$Z_{2,10}$	$Z_{3,10}$	$Z_{4,10}$
	November	$Z_{1,11}$	$Z_{2,11}$	$Z_{3,11}$	$Z_{4,11}$
	Desember	$Z_{1,12}$	$Z_{2,12}$	$Z_{3,12}$	$Z_{4,12}$
⋮	⋮	⋮	⋮	⋮	⋮
2023	September	$Z_{1,57}$	$Z_{2,57}$	$Z_{3,57}$	$Z_{4,57}$
	Oktober	$Z_{1,58}$	$Z_{2,58}$	$Z_{3,58}$	$Z_{4,58}$
	November	$Z_{1,59}$	$Z_{2,59}$	$Z_{3,59}$	$Z_{4,59}$
	Desember	$Z_{1,60}$	$Z_{2,60}$	$Z_{3,60}$	$Z_{4,60}$

2.3 Tahapan Analisis

Tahapan analisis yang digunakan dalam penelitian ini adalah:

1. Mengidentifkasi data dengan menghitung nilai *mean*, standar deviasi, variansi, minimum dan maksimum serta membuat plot *time series* untuk masing-masing stasiun pengamatan
2. Membagi data menjadi data *training* dan data *testing*. Data *training* berisi data yang akan digunakan untuk membuat model. Data *testing* berisi data yang akan digunakan untuk memvalidasi model yang dibentuk dari data *training*
3. Memeriksa kestasioneran data menggunakan uji Augmented Dickey Fuller sesuai dengan persamaan (4). Melakukan *differencing* apabila data tidak stasioner
$$t_{hitung} = \left|\frac{\hat{\gamma}}{SE(\hat{\gamma})}\right|$$
4. Mengidentifikasi model VAR-GSTAR dengan melihat plot ACF dan PACF. Identifikasi model juga dapat dilihat berdasarkan nilai *Akaike's Information Criteria* yang paling kecil sesuai dengan persamaan (7)
$$AIC = \ln\left(\frac{JKS}{n}\right) + \frac{2K^2}{n}$$
5. Menghitung pembobot lokasi yaitu invers jarak dan normalisasi korelasi silang sesuai persamaan (8) dan (11)
$$W = \begin{cases} \dfrac{w_{ij}}{\sum_{j=1}^{N} w_{ik}}, & i \neq j \\ 0 & , i = j \end{cases}$$
$$W = \frac{r_{ij}(k)}{\sum_{k=1}^{p} |r_{ik}(k)|}$$
6. Mengestimasi parameter model VAR-GSTAR menggunakan metode kuadrat terkecil (*Ordinary Least Square*) sesuai persamaan (15)
$$\hat{\Phi} = (Z^{*\prime}Z)^{-1}(Z^{*\prime}Z)$$
7. Melakukan uji asumsi *White Noise* menggunakan uji Ljung Box untuk memeriksa apakah model memiliki sisaan yang bersifat *white noise* atau tidak sesuai persamaan (16)
$$Q_{LB} = n(n+2)\sum_{k=1}^{n}\left(\frac{1}{n-k}\right)\hat{P}_k^2$$
8. Pemilihan model terbaik dapat diperoleh dengan menghitung nilai RMSE sesuai persamaan (17). Nilai RMSE dapat mengindikasi adanya ketidakcocokan pada model
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Z_t - \hat{Z}_t)^2}$$
9. Melakukan peramalan menggunakan model terbaik