

## BAB 2

### TINJAUAN PUSTAKA

#### 1.1 Tinjauan Statistik

##### 1.1.1 Analisis Deskriptif

Statistika deskriptif adalah metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna (Walpole, 1995). Analisis deskriptif merupakan teknik analisis dengan cara mendeskripsikan atau menggambarkan data. Deskripsi data dapat dilihat dari karakter data baik secara visual maupun secara numeris. Secara visual, deskripsi suatu data dapat digambarkan melalui diagram batang, garis, area, *pie*, boxplot, grafik dan diagram. Sedangkan secara numeris, dilakukan dengan menelaah ukuran pusat dan penyebaran data (Pramesti, 2015).

##### 1.1.2 Analisis *Time Series*

*Time series* merupakan serangkaian observasi terhadap suatu variabel yang diambil secara berurutan berdasarkan interval waktu yang tetap (Wei, 2006). Data *time series* merupakan jenis data yang terdiri dari satu objek tetapi meliputi beberapa periode waktu misalnya harian, mingguan, bulanan, tahunan, dan lain-lain. Analisis *time series* merupakan salah satu prosedur statistika yang diterapkan untuk meramalkan struktur probabilitas keadaan yang akan datang dalam rangka pengambilan keputusan (Auryawati and Irawan, 2014). Metode *time series* adalah teknik statistik yang memanfaatkan akumulasi data historis suatu jangka waktu. Metode *time series* mengasumsikan bahwa apa yang terjadi di masa lalu akan terus berlanjut terjadi di masa depan (Taylor, 2013).

##### 1.1.3 Teori Nilai Ekstrem

Kejadian yang melebihi nilai ambang disebut kejadian ekstrem. Pemodelan kejadian ekstrem tidak dapat dilakukan dengan pendekatan sebaran normal, tetapi dengan suatu sebaran yang memiliki ekor yang panjang (*heavy tail*). Teori nilai ekstrem (*extreme value theory/EVT*) bertujuan untuk mengkaji perilaku stokastik dari suatu proses pada suatu nilai ambang (Coles, 2001) dalam (Anik Djuraidah, 2019).

Saat ini dikenal ada dua pendekatan yang digunakan untuk penentuan kejadian ekstrem, yaitu metode blok maksima (*block maxima*/BM) dan metode pelampauan nilai ambang (*peak over threshold* /POT).

#### 1.1.4 Metode *Block Maxima* (BM)

Metode *Block Maxima* (BM) adalah metode penentuan nilai ekstrem yang mengidentifikasi nilai ekstrem berdasarkan nilai tertinggi dari data pengamatan yang dikelompokkan berdasarkan periode tertentu yang disebut blok (Amalia 2017). Pada metode *Block Maxima* (BM) dalam penentuan nilai ekstremnya dibagi dalam blok atau periode waktu tertentu seperti bulanan, triwulan atau tahunan (Kellezi, 2006). Penentuan periode dalam metode BM sangat penting untuk diperhatikan, periode yang terlalu pendek atau terlalu Panjang dapat menyebabkan penduga parameter GEV bias dan ragam yang besar.

Langkah-langkah pengambilan sampel nilai ekstrem dengan metode *Block Maxima* adalah sebagai berikut (Ian Surya Prayoga 2020):

1. Data harian polusi udara DKI Jakarta tahun 2012-2021 dibentuk menjadi data bulanan dengan cara mencari rata-rata data harian per satu bulan.
2. Data polusi udara DKI Jakarta tahun 2012-2021 yang sudah berbentuk bulanan, kemudian dibagi ke dalam blok periode waktu tertentu yaitu bulanan, triwulan, dan tahunan.
3. Data polusi udara DKI Jakarta tahun 2012-2021 dibagi ke dalam blok triwulan dan terbentuk sebanyak 40 blok. Digunakan periode triwulan karena periode yang terlalu pendek atau terlalu panjang dapat menyebabkan penduga parameter GEV bias dan ragam yang besar.
4. Setelah blok periode waktu terbentuk, Langkah selanjutnya yaitu mengambil sampel nilai ekstrem dengan cara menentukan nilai maksimum dari data pengamatan di setiap blok yang sudah terbentuk. Sampel nilai ekstrem tersebutlah yang kemudian digunakan untuk dalam analisis *spatial extreme value theory* dengan pendekatan *gaussian copula*.

Pada metode BM, nilai ekstrem mempunyai sebaran nilai ekstrem terampat (*generalized extreme value/GEV*). Misal  $X_1, X_2, \dots, X_n$  merupakan peubah acak yang menyebar menurut sebaran  $F$ , dan  $M_n = \max(X_1, X_2, \dots, X_n)$  merupakan nilai maksimumnya dan konvergen menuju limit *non-degenerate*, maka limit tersebut merupakan anggota keluarga parametrik. Jika terdapat suatu konstanta  $\{a_n > 0\}$  dan  $\{b_n\}$  maka:

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G^n(a_n x + b_n) \rightarrow G(x) \quad (0.1)$$

saat  $n \rightarrow \infty$ , dengan  $G$  merupakan fungsi sebaran *non-degenerate*, maka fungsi sebaran  $G$  mengikuti salah satu dari tiga sebaran dasar nilai ekstrem, yaitu Gumbel, Frechet, dan Weibull.

#### 1.1.5 Metode *peak over threshold* (POT)

Pada metode *peak over threshold* (POT) dalam penentuan sampel nilai ekstrem ditentukan berdasarkan nilai yang melebihi suatu nilai ambang tertentu. Pada metode POT, seluruh nilai-nilai yang melebihi nilai ambang  $\mu$  disebut sebagai nilai ekstrem (Kellezi, 2006) dalam (Anik Djuraidah, 2019). Sesuai dengan persamaan (2.1), jika penentuan nilai ekstrem diperoleh dari nilai yang melampaui nilai ambang  $\mu$  maka  $y = x - \mu$  akan mengikuti sebaran Pareto terampat (*Generalized Pareto Distribution/GPD*). Dengan fungsi sebarannya yaitu:

$$H(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\theta}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\theta}\right), & \xi = 0 \end{cases} \quad (0.2)$$

dimana  $y > 0$ ,  $\left(1 + \xi \frac{y}{\theta}\right) > 0$ , dan  $\hat{\sigma} = \sigma + \xi(u - \mu)$ . Penentuan nilai  $u$  pada metode POT juga merupakan suatu hal yang cukup sulit. Nilai  $u$  yang terlalu rendah dapat mengakibatkan penduga bias, sedangkan nilai  $u$  yang terlalu tinggi mengakibatkan adanya nilai ekstrem yang tidak terdeteksi dan kurangnya data untuk melakukan pendugaan sehingga ragamnya besar (Coles, 2001).

### 1.1.6 Distribusi *Generalized Extreme Value* (GEV)

Pada metode *block maxima* menggunakan teorima Fisher dan Tippet (1928) dalam (Kellezi 2006), dimana dalam teorema tersebut berbunyi jika data sampel nilai ekstrem yang diperoleh dengan metode *block maxima* akan berdistribusi Gumble, Frechet, dan Weibull. Ketiga distribusi tersebut termasuk kedalam distribusi *Generalized Extreme Value* (GEV). Fungsi sebaran GEV dinyatakan sebagai:

$$G(x) = \begin{cases} \exp \left\{ - \left( 1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right\}, & \xi \neq 0 \\ \exp \left\{ - \exp \left( - \frac{x - \mu}{\sigma} \right) \right\}, & \xi = 0 \end{cases} \quad (0.3)$$

dimana:

$\mu$  : parameter lokasi

$\sigma > 0$  : parameter skala

$\xi$  : parameter bentuk yang menentukan karakteristik ujung sebaran

Jenis distribusi GEV (distribusi Gumble, distribusi Frechet, dan distribusi *Reversed Weibull*) yang memiliki CDF seperti berikut:

a. Distribusi Gumble untuk  $\xi = 0$

$$F(x; \mu, \sigma) = \exp \left\{ - \exp \left( - \frac{x - \mu}{\sigma} \right) \right\}, \quad -\infty < x < \infty \quad (0.4)$$

b. Distribusi Frechet untuk  $\xi > 0$

$$F(x; \mu, \sigma, \xi) = \begin{cases} 0 & x \leq \mu \\ \exp \left\{ - \left( \frac{x - \mu}{\sigma} \right)^{\frac{1}{\xi}} \right\} & x > \mu \end{cases} \quad (0.5)$$

c. Distribusi *Reversed Weibull* untuk  $\xi < 0$

$$F(x; \mu, \sigma, \xi) = \begin{cases} \exp \left\{ - \left( \frac{x - \mu}{\sigma} \right)^{\frac{1}{\xi}} \right\} & x < \mu \\ 1 & x \geq \mu \end{cases} \quad (0.6)$$

Apabila  $\xi < 0$  maka data menyebar Weibull,  $\xi = 0$  maka data menyebar Gumble, dan  $\xi > 0$  maka data menyebar Frechet (Coles, 2001). Nilai  $\xi$  merupakan parameter bentuk ekor (*tail*) dari distribusi. Semakin besar nilai  $\xi$ , maka distribusi akan memiliki ekor yang semakin berat (*heavytail*) sehingga berdampak peluang terjadinya nilai ekstrem semakin besar. Menurut (Finkenstadt 2003) dalam (Amalia

2017) untuk parameter bentuk dengan  $\xi = 0$  dikatakan “*medium tail*”,  $\xi > 0$  dikatakan “*long tail*” dan untuk  $\xi < 0$  dikatakan “*short tail*”. Dari ketiga jenis distribusi GEV di atas dapat diketahui jika distribusi yang memiliki ekor paling berat atau *heavytail* adalah distribusi Frechet ( $\xi > 0$ ).

### 1.1.7 Uji Anderson Darling

Uji Anderson Darling adalah suatu uji yang digunakan untuk mengetahui apakah suatu data mengikuti distribusi tertentu yang dihipotesiskan atau tidak. Pengujian kesesuaian distribusi *Generalized Extreme Value* (GEV) terhadap data ekstrem dapat dilakukan menggunakan uji Anderson Darling dengan prosedur (Engmann, S. dan Denis 2011):

1. Uji Hipotesis:

$H_0: F(x) = F^*(x)$  Data mengikuti distribusi teoritis

$H_1: F(x) \neq F^*(x)$  Data tidak mengikuti distribusi teoritis

2. Statistik Uji:

$$AD = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) (\ln(F^*(x_i))) + \ln(1 - (F^*(x_{n+1-i}))) \quad (0.7)$$

Dimana:

$F(x)$  : fungsi distribusi kumulatif data sampel

$F^*(x)$  : fungsi distribusi kumulatif teoritis

$n$  : ukuran sampel

3. Penentuan Kriteria Uji

Kriteria uji tolak  $H_0$  jika nilai  $AD_{hitung}$  lebih besar dari  $AD_{tabel}$  atau  $p\text{-value} < \alpha$  (taraf signifikansi yang sudah ditentukan).

### 1.1.8 Nilai Ekstrem Spasial

Teori nilai ekstrim, juga dikenal sebagai teori nilai ekstrim spasial, membutuhkan pengetahuan tentang suatu wilayah atau lokasi karena pemodelan dilakukan di satu lokasi, seperti ketika menggunakan data berdasarkan fenomena lingkungan seperti hujan lebat, badai, salju, gempa bumi, dan udara. polusi yang

terjadi secara ekstrem di lokasi yang mirip dan sebagian besar berhubungan satu sama lain (Amalia, 2017). Pada data polusi udara diukur berdasarkan lokasi dan waktu sehingga pendekatan teori nilai ekstrem tidaklah cukup, oleh karenanya diperlukan pemodelan teori nilai spasial ekstrem untuk menduga kejadian polusi udara ekstrem.

Pendekatan yang dapat digunakan untuk memodelkan nilai ekstrim spasial atau disebut juga dengan bilangan ekstrim spasial adalah dengan menggunakan bilangan ekstrim multivariat. Data spasial merupakan data multivariat karena dikumpulkan dari berbagai lokasi atau wilayah yang berbeda satu sama lain. Sehingga, ada asumsi-asumsi yang harus dipenuhi, seperti asumsi ketergantungan spasial, untuk memahami apakah ada pengaruh spasial antara satu wilayah dengan wilayah lainnya. Data ekstrim dari beberapa wilayah yang berbeda diasumsikan mewakili variabel multivariat atau distribusi multivariat. Misalkan  $S(i,p)$  adalah data peristiwa ekstrem pada lokasi ke- $i$  dan periode waktu ke  $p$ , pada domain spasial  $D \subset R^2$ . Distribusi dari  $S(i,p)$  adalah:

$$S(i,p) \sim GEV(\mu(i,p), \sigma(i,p), \xi(i,p)) \quad (0.8)$$

dimana:

$\mu(i,p)$ ,  $\sigma(i,p)$  dan  $\xi(i,p)$  : parameter lokasi atau wilayah, skala dan bentuk dari distribusi GEV dimana  $p = 1, 2, \dots, P$  dan  $i = 1, 2, \dots, I$ .

Distribusi parameter GEV mengikuti model *trend surface* yang menggunakan bentuk persamaan titik koordinat lintang (*latitude*) dan bujur (*longitude*) sebagai variabel penjelas (Hakim, 2016) dalam (Ian Surya Prayoga, 2020). Dalam konsep spasial lokasi yang berdekatan cenderung memiliki hubungan yang cukup erat dalam suatu peristiwa atau kejadian dibandingkan dengan lokasi yang lebih jauh (Amalia, 2017). Pada data multivariat seperti polusi udara ekstrem, analisis yang sering digunakan adalah copula dan proses *max-stable*.

### 1.1.9 Max-Stable Process

*Max-Stable Processes* (MSP) adalah pengembangan dari distribusi nilai ekstrem multivariat pada dimensi yang tak terhingga (dimensi tak terbatas). Suatu fungsi distribusi G dikatakan max stable jika dan hanya jika G berdistribusi GEV

(Ramadhani, 2019) dalam (Ian Surya Prayoga, 2020). Dengan metode MSP, nilai maksimum (Maxima) setiap lokasi untuk sampel digunakan proses spasial. (Cooley, D., Nyckah, D., & Naveau, 2007). Dalam konsep spasial ekstrem terdapat dua metode pendekatan yaitu *max-stable* dan *copula* (Davison, Padoan, and Ribatet, 2012). Perbedaan kedua metode tersebut terletak pada proses modeling dan perubahannya. Untuk pemodelan dan estimasi, Copula menggunakan model *Gaussian* dan *student t*, sedangkan Max-Stable menggunakan model *Schlater*, *Smith* dan *Brown-resnick*.

Untuk proses perubahan kedua pendekatan ini menggunakan proses yang sama yaitu *max-stable* karena proses *max-stable* membawa data ke dalam bentuk distribusi frechet, akan tetapi dalam proses perubahan di copula menggunakan perubahan sifat ke-1 dan proses *max-stable* menggunakan perubahan sifat ke-2. Sifat perubahan *max-stable* adalah sebagai berikut:

1. Distribusi marginal satu dimensionalnya mengikuti distribusi GEV  $X \sim GEV(\mu, \lambda, \xi)$  dengan fungsi distribusi sebagai berikut:

$$F(\mu, \lambda, \xi) = \exp \left[ - \left\{ 1 + \frac{\xi(x - \mu)}{\sigma} \right\}^{1/\xi} \right], -\infty < \mu, \xi < \infty, \sigma > 0 \quad (2.9)$$

dimana:

$\mu$  : parameter lokasi

$\lambda$  : parameter skala (*scale*)

$\xi$  : parameter bentuk (*shape*)

2. Distribusi marginal k-dimensional mengikuti distribusi *multivariat extreme value*.  $\{Z(j)\}$  adalah proses *max-stable* yang memiliki margin Frechet unit dengan fungsi distribusi berikut.

$$F(z) = \exp \left( - \frac{1}{z} \right), z > 0 \quad (0.10)$$

Proses ini dapat diperoleh dengan mentransformasi  $\{x(j)\}$  menjadi persamaan (2.7) sebagai berikut.

$$\{Z(j)\} = \left[ \left\{ 1 + \frac{\xi(x - \mu)}{\sigma} \right\}^{1/\xi} \right] \quad (0.11)$$

dimana:

$\mu(x), \xi(x), \sigma(x)$  : suatu fungsi kontinyu

Proses  $Z$  ini juga disebut *max-stable* (Padoan, S., Ribatet, M., & Sisson, 2010).

### 1.1.10 Madogram

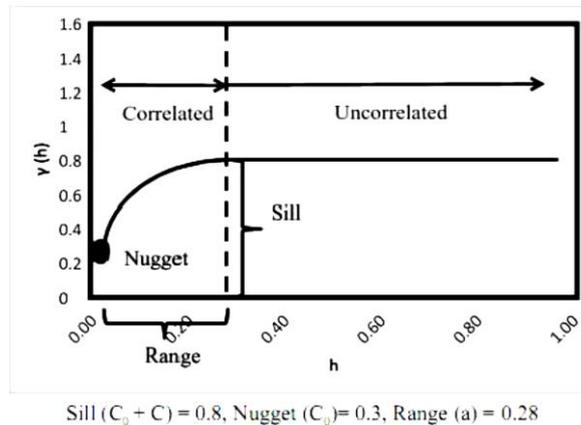
Konsep dasar madogram berasal dari semivariogram yaitu grafik antara semivariansi terhadap fungsi jarak. Semivariogram dapat juga digunakan untuk mengetahui dependensi spasial. Hubungan kebergantungan spasial antara titik lokasi ditentukan oleh jarak antar lokasi, semakin dekat jarak suatu lokasi akan memiliki semivarian yang kecil dan begitupun sebaliknya. Konsep jarak yang digunakan yaitu konsep jarak *Euclid*. Semivariogram dapat didefinisikan pada persamaan berikut:

$$\gamma(h) = \frac{1}{2} E[U(j+h) - U(j)]^2 = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [U(j_i+h) - U(j_i)]^2 \quad (2.12)$$

dimana:

- $\gamma(h)$  : nilai semivariogram dengan jarak  $h$
- $U(j_i)$  : nilai pengamatan di titik  $j_i$
- $U(j_i + h)$  : nilai pengamatan di titik  $(j_i + h)$
- $N(h)$  : banyaknya pasangan titik yang berjarak  $h$

Dalam menentukan model semivariogram, harus dilakukan pendugaan terhadap parameter semivariogram. Parameter tersebut diduga berdasarkan plot semivariogram yang dihasilkan. Plot semivariogram ditunjukkan pada Gambar 2.1.



**Gambar 2.1 Plot Semivariogram**

1. *Nugget Effect* ( $C_0$ )

*Nugget Effect* adalah nilai pendekatan semivariogram pada jarak di sekitar nol.

2. *Sill* ( $C_0 + C$ )

*Sill* merupakan nilai tertentu yang konstan pada semivariogram untuk jarak tertentu sampai dengan jarak yang tak hingga atau nilai semivariogram yang menunjukkan sudah tidak terdapat korelasi antar data.

3. *Range* ( $a$ )

*Range* merupakan jarak maksimum yang masih memiliki korelasi antar data. Semivariogram tidak bisa digunakan dalam kasus data ekstrem. Untuk mengatasi hal itu, (Cooley, D., Naveau, P., dan Poncet, 2006) menggunakan semivariogram orde pertama yaitu madogram yang bisa digunakan untuk data ekstrem. Teori tentang madogram telah dipelajari oleh Matheron pada tahun 1987 dalam (Cooley, D., Naveau, P., dan Poncet, 2006) yang didefinisikan sebagai berikut:

$$v(h) = \frac{1}{2} E[U(j+h) - U(j)] \quad (2.13)$$

Madogram mengharuskan momen pertama terhingga yang tidak selalu terjadi pada kasus ekstrem, untuk mengatasinya (Cooley, D., Naveau, P., dan Poncet, 2006) memperkenalkan madogram yang mentransformasi *variabel random* menggunakan distribusi GEV. Fungsi  $F$  berdistribusi GEV, sehingga  $F$ -madogramnya sebagai berikut:

$$v(h) = \frac{1}{2} E[F(U(j+h)) - F(U(j))] \quad (2.14)$$

Pada semivariogram proses penentuan pola melibatkan banyak titik sehingga, sulit untuk melihat pola tertentu (Amalia 2017). Untuk mengatasinya, madogram dikelompokkan berdasarkan kesamaan jarak. Sehingga, perhitungan  $F$ - madogram dapat dinyatakan sebagai berikut:

$$\hat{v}_F(h) = \frac{1}{2|N(h)|} \sum_{i=1}^{N(h)} [F(U(j_i+h)) - F(U(j_i))] \quad (2.15)$$

dimana:

- $\hat{v}_F(h)$  :  $F$ -madogram pada lag  $h$
- $j_i$  : lokasi titik pengamatan
- $U(j_i)$  : nilai pengamatan pada lokasi ke  $j_i$
- $h$  : jarak antara dua lokasi
- $j_i, j_i + h$  : pasangan data yang berjarak  $h$
- $N(h)$  : banyaknya pasangan lokasi yang berjarak  $h$

Koefisien eksternal dan  $F$ -madogram mempunyai hubungan yang sangat kuat ditunjukkan sebagai berikut:

$$\theta(h) = \frac{1 + 2v_F(h)}{1 - 2v_F(h)} \quad (2.16)$$

### 1.1.11 Variogram

Variogram merupakan plot dua dimensi yang digunakan untuk mengukur keragaman spasial antar amatan pada data yang digunakan. Variogram mencerminkan karakteristik dari korelasi spasial, yaitu korelasi antara dua buah amatan menjadi berkurang atau tidak berkorelasi karena bertambahnya jarak antar amatan tersebut (Cressie, 1993) dalam (Rahmi Lathifah Islami, 2021). Terdapat dua metode estimator dalam *package geoR*.

## 1. Classical Estimator

$$\gamma(h) = \frac{1}{2N_h} \sum_{i=1}^{N_h} (Y(x_i + h) - Y(x_i))^2 \quad (2.17)$$

## 2. Hawkins and Crassie's Modulus Estimator

$$\gamma(h) = \frac{\left( \frac{1}{N_h} \sum_{i=1}^{N_h} (Y(x_i + h) - Y(x_i))^2 \right)^4}{0.914 + \frac{0.988}{N_h}} \quad (2.18)$$

dimana:

- $x_i$  : vektor dari koordinat spasial pada titik ke-  $i$   
 $Y(x_i)$  : variabel pada titik ke-  $i$   
 $h$  : vektor lag antara dua amatan (lokasi)  
 $Y(x_i + h)$  : nilai variabel setelah di lag kan pada titik ke-  $i$   
 $N_h$  : jumlah pasangan pengamatan pada lag  $h$

## 1.1.12 Koefisien Eksternal

Dalam analisis spasial ekstrem yang perlu diperhatikan adalah ukuran dependensi spasial pada lokasi berdasarkan koefisien eksternal. Koefisien eksternal dapat mengukur tingkat dependensi data antara wilayah satu dengan wilayah lainnya. Koefisien eksternal diperkenalkan oleh Smith, dalam (Amalia, 2017) yang didefinisikan pada persamaan (2.15) sebagai berikut:

$$\theta(h_{j,k}) = 2\Phi \left[ \frac{\sqrt{h_{j,k}^T S_{j,k}^{-1} h_{j,k}}}{2} \right] \quad (2.19)$$

dimana:

- $\theta(h_{j,k})$  : nilai koefisien eksternal  
 $\Phi$  : fungsi distribusi kumulatif normal standart  
 $S_{j,k}$  : matriks kovarian dari variabel lokasi ke-  $j$  dan ke-  $k$   
 $h_{j,k}$  : vektor jarak antara lokasi  $j$  dengan  $k$

Jarak euclidean dengan persamaan sebagai berikut  $\sqrt{(lat_1 - lat_2)^2 + (lon_1 - lon_2)^2}$ . Nilai  $\theta(h_{j,k})$  memiliki kisaran nilai  $1 < \theta(h_{j,k}) < 2$ . Nilai  $\theta(h_{j,k})$  semakin mendekati 1 mengindikasikan bahwa antara dua wilayah memiliki hubungan yang dependen. Nilai  $\theta(h_{j,k})$  semakin mendekati 2 mengindikasikan bahwa antara dua wilayah memiliki hubungan yang independent (Davidson, Padoan, and Ribatet, 2012) dalam (Amalia, 2017).

### 1.1.13 Copula

Copula pertama kali diperkenalkan oleh Abe Sklar pada tahun 1959 melalui teorema sklar. Menurut teorema sklar, copula merupakan suatu fungsi yang menghubungkan fungsi distribusi multivariat dengan distribusi marginalnya (Nelsen, R. B., dan Flores, 2005) Copula dapat mengeksplorasi dan mengkarakterisasi struktur dependensi antar variabel random melalui fungsi distribusi marginal (Gudendorf, G., dan Segers, 2010). Copula terbagi menjadi dua macam *families*, yaitu *elliptical* copula dan *archimedian* copula. Untuk kasus *Spatial Extreme Model* copula yang dapat digunakan adalah *elliptical* copula. Copula yang termasuk dalam *elliptical copula* adalah *Gaussian* copula dan *Student's t-copula* (Davison, Padoan, and Ribatet, 2012).

#### 1.1.13.1 Copula Gaussian

Copula Gaussian adalah copula yang sesuai untuk memberikan model dalam *spatial extreme*. Copula Gaussian diperoleh dari transformasi pada variabel random ke dalam distribusi normal standar. Dalam Copula Gaussian untuk kasus *spatial extreme* proses transformasi menggunakan distribusi marginal GEV (Amalia, 2017) dengan persamaan transformasi didefinisikan pada persamaan (2.16) sebagai berikut:

$$u_j = F_{X_j}(x_{ij}) \quad (0.20)$$

dimana:

$F_{X_j}$  : CDF dari distribusi GEV

$x_{ij}$  : data observasi ke- $i$  stasiun ke- $j$

Menurut (Nelsen, R. B., dan Flores, 2005), jika fungsi distribusi marginal dari  $u_j$  kontinu maka  $u_j$  adalah copula unik. CDF copula gaussian mengikuti persamaan (2.17) sebagai berikut:

$$C(u_1, u_2, \dots, u_m) = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_m); \rho) \quad (0.21)$$

dimana:

$\Phi$  : CDF distribusi gaussian

$\rho$  : fungsi korelasi

Jenis korelasi yang digunakan adalah korelasi *whittle-mattern*. Pilihan menggunakan korelasi *whittle-mattern* karena memiliki fungsi yang lebih fleksibel dan sering digunakan, meskipun terkadang sulit untuk memeriksa validitas parameter bentuk ( $\xi$ ). Fungsi korelasi *whittle-mattern* didefinisikan pada persamaan berikut (Davison, Padoan, and Ribatet 2012):

$$\rho(h) = \{2^{(c_0+c)-1} \Gamma(c_0 + c)\}^{-1} (\|h\|/a)^{(c_0+c)} K_{(c_0+c)}(\|h\|/a) \quad (0.22)$$

dimana:

$\Gamma$  : fungsi Gamma

$K_{(c_0+c)}$  : fungsi Bessel dengan derajat  $(c_0 + c)$

$a$  : parameter *range*

$(c_0 + c)$  : parameter *sill*

Dari CDF copula gaussian dibentuk pdf copula gaussian, pdf copula gaussian didefinisikan pada persamaan (2.19) sebagai berikut:

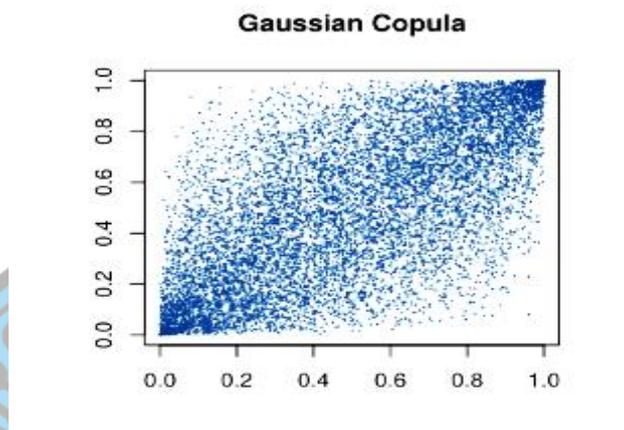
$$c(u_1, u_2, \dots, u_m) = \frac{\partial}{\partial u_1} \cdot \frac{\partial}{\partial u_2} \dots \frac{\partial}{\partial u_m} \cdot C(u_1, u_2, \dots, u_m) \quad (0.23)$$

Menurut teorema sklar peluang bersama copula didefinisikan dengan perkalian antara pdf distribusi marginal dengan fungsi CDF copula, sehingga fungsi peluang bersama didefinisikan pada persamaan (2.20) sebagai berikut:

$$f(x_1, x_2, \dots, x_m) = f_{x_1}(x_1) \cdot \dots \cdot f_{x_m}(x_m) \cdot c(u_1, u_2, \dots, u_m) \quad (0.24)$$

(Schölzel, C. dan Friederichs, 2008)

Berikut merupakan gambar dari Copula Gaussian.



**Gambar 2.2 Ilustrasi Plot Copula Gaussian**

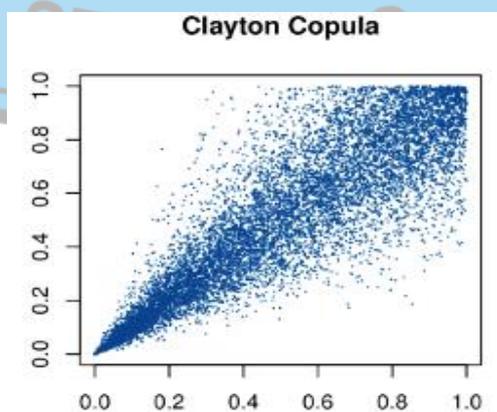
### 1.1.13.2 Copula Clayton

Copula Clayton, merupakan copula yang berlawanan dengan copula Gumbel yaitu memiliki ketergantungan ekor bawah untuk  $\theta > 0$ . Persamaan dari copula Clayton, dengan  $\theta \in (0, \infty)$  adalah sebagai berikut:

(Okhrin, 2015) dalam (Ian Surya Prayoga, 2020)

$$C_{\theta}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \quad (0.25)$$

Berikut merupakan gambar dari Copula Clayton.



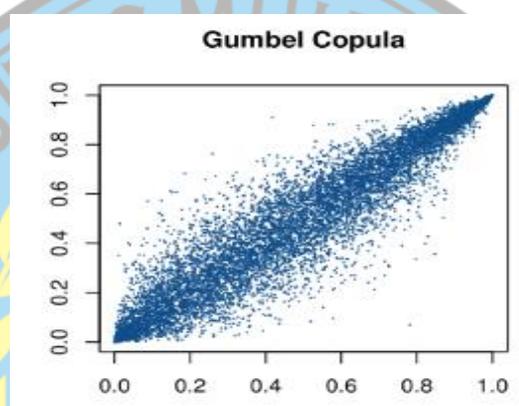
**Gambar 2.3 Ilustrasi Plot Copula Clayton**

### 1.1.13.3 Copula Gumble

Copula Gumble merupakan copula yang memiliki ketergantungan ekor atas. (Embrechts, P., Lindskog, F., & McNeil, 2001) dalam (Ian Surya Prayoga, 2020). Persamaan dari copula Gumble dengan  $\theta \geq 1$  adalah sebagai berikut (Okhrin, 2015):

$$C_{\theta}(u, v) = \exp(-[(-\ln u)^{\theta} + (-\ln v)^{\theta}])^{1/\theta} \quad (0.26)$$

Berikut merupakan gambar dari Copula Gumble.



Gambar 2.4 Ilustrasi Plot Copula Gumble

### 1.1.14 Maximum Pairwise Likelihood Estimation (MPLE)

Menurut (Davison, Padoan, and Ribatet, 2012), Estimasi parameter Copula Gaussian untuk spasial ekstrem dapat menggunakan *Maximum Pairwise Likelihood Estimation* (MPLE). MPLE adalah metode estimasi parameter yang menggunakan fungsi *pairwise*/berpasangan dari dua variabel. Seperti halnya metode MLE, estimasi menggunakan metode ini dilakukan dengan menurunkan satu kali fungsi *ln likelihood* terhadap parameter yang diestimasi dan menyamakannya dengan vektor nol. Metode MPLE menggantikan fungsi  $l(\beta)$  pada MLE dengan fungsi *pairwise likelihood*  $\ell_p(\beta)$  pada persamaan (2.23) sebagai berikut:

$$\ell_p(\hat{\beta}) = \sum_{i=1}^n \sum_{j=1}^{m-1} \sum_{k=j+1}^m \ln(f(u_{ji}, u_{ki}; \hat{\beta})) \quad (0.27)$$

dimana:

$f(u_{ji}, u_{ki}; \hat{\beta})$  : distribusi bersama copula gaussian dengan parameter  $\beta$

$i = 1, 2, \dots, n$  : observasi pada masing-masing variabel

Copula mentransformasikan variabel  $x$  ke unit margin copula  $u$  seperti definisi persamaan (2.16). Estimasi parameter  $\beta_\mu, \beta_\sigma$  dan  $\beta_\xi$ , dapat diperoleh jika

pembentukan fungsi *likelihood* didasarkan pada  $f(u_{ji}, u_{ki}; \hat{\beta})$  dengan  $\beta_\mu = \begin{bmatrix} \beta_{\mu,0} \\ \beta_{\mu,1} \\ \beta_{\mu,2} \end{bmatrix}$ ,  
 $\beta_\sigma = \begin{bmatrix} \beta_{\sigma,0} \\ \beta_{\sigma,1} \\ \beta_{\sigma,2} \end{bmatrix}$  dan  $\beta_\xi$ . (Amalia, 2017).

#### 1.1.15 Pemilihan Model Terbaik

*Akaike Information Criterion* (AIC) digunakan sebagai pemilihan model *trend surface* terbaik, dengan model sebagai berikut:

$$\begin{aligned} \hat{\mu}(j) &= \hat{\beta}_{\mu,0} + \hat{\beta}_{\mu,1} \text{longitude}(j) + \hat{\beta}_{\mu,2} \text{latitude}(j) \\ \hat{\sigma}(j) &= \hat{\beta}_{\sigma,0} + \hat{\beta}_{\sigma,1} \text{longitude}(j) + \hat{\beta}_{\sigma,2} \text{latitude}(j) \\ \hat{\xi}(j) &= \hat{\beta}_{\xi,0} \end{aligned} \quad (0.28)$$

Kriteria dalam memilih model yang terbaik menjadi hal yang sangat penting dalam membentuk model yang baik. Memilih model yang sederhana lebih baik daripada model yang kompleks. Menurut (Ligas, M., dan Banasik, 2012) *Akaike Information Criterion* (AIC) didefinisikan dengan persamaan (2.29) sebagai berikut:

$$AIC = -2\ell_p(\hat{\beta}) + 2q \quad (0.29)$$

dimana  $\ell_p(\hat{\beta})$  adalah fungsi *ln pairwise likelihood* didefinisikan dalam bentuk persamaan (2.26) sebagai berikut:

$$\ell_p(\hat{\beta}) = \text{fungsi } \ln \text{ pairwise likelihood}$$

$$\ell_p = (\hat{\beta}) = \sum_{i=1}^n \sum_{j=1}^{m-1} \sum_{k=j+1}^m \ln(f(u_{ji}, u_{ki}; \hat{\beta})) \quad (0.30)$$

dimana:

$i = 1, 2, \dots, n, j = 1, 2, \dots, m - 1, k = 2, 3, \dots, m$

$q$  = banyaknya parameter yang ditaksir

Nilai AIC yang lebih rendah artinya menunjukkan model yang terbaik.

### 1.1.16 Mean Absolute Percentage Error (MAPE)

Dalam penelitian ini untuk mengukur ketepatan nilai dugaan model yaitu menggunakan MAPE. MAPE dinyatakan dalam bentuk rata-rata persentase absolut kesalahan. MAPE lebih banyak digunakan untuk perbandingan pada data yang mempunyai skala interval waktu berbeda. Perhitungan MAPE yaitu menjumlahkan satu kesalahan absolut pada satu periode dengan mengurangi nilai aktual dengan nilai hasil peramalan dibagi dengan nilai aktual periode tersebut baru kemudian dipresentasikan (Ian Surya Prayoga, 2020).

Skala pengkategorian nilai MAPE yang digunakan pada penelitian dan formula perhitungannya dijelaskan pada Tabel 2.1 dan persamaan (2.31):

$$MAPE = \left(\frac{1}{n}\right) \sum_{i=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (0.31)$$

dimana:

$A_t$  : data aktual pada periode ke-  $t$

$F_t$  : data peramalan pada periode ke-  $t$

$n$  : jumlah periode peramalan yang terlibat

Tabel 2.1 Skala Keباikan MAPE

| MAPE   | Hasil Peramalan |
|--------|-----------------|
| <10%   | Baik Sekali     |
| 10-20% | Baik            |
| 20-50% | Cukup           |
| >50%   | Tidak Baik      |

Pada Tabel 2.1. diatas dapat disimpulkan bahwa suatu model dapat dikatakan baik sekali apabila memiliki nilai MAPE dibawah 10% dan baik apabila nilai MAPE berkisar antara 10% - 20% dikatakan cukup apabila nilai MAPE berkisar antara 20% - 50% dan jika lebih dari itu maka tidak baik (Ian Surya Prayoga, 2020).

## **1.2 Tinjauan Non Statistik**

### **1.2.1 Kualitas Udara**

Salah satu komponen penting untuk bertahan hidup adalah udara. Oksigen merupakan komponen utama di dalam udara yang mana sangat penting dalam proses metabolisme makhluk hidup. Udara yang terdapat di atmosfer mengandung gas dengan komposisi permanen, yaitu 21,94% oksigen (O<sub>2</sub>), 0,032% karbondioksida (CO<sub>2</sub>), 78,09% nitrogen (N<sub>2</sub>) dan gas-gas lainnya seperti (CH<sub>4</sub>), (He), (H<sub>2</sub>), (Kr), (Ne), (N<sub>2</sub>O), dan (Xe). Zat-zat tersebut jika masih berada dalam batas tertentu masih dapat dinetralisir, namun jika sudah melampaui ambang batas maka proses netralisir akan terganggu. Udara yang murni dan belum tercampur dengan zat-zat lain, baik dalam bentuk padat, cair, maupun gas yang bersifat merugikan dapat dikatakan udara tersebut termasuk dalam udara bersih. Udara yang bersih dan sehat memiliki ciri tidak berwarna, tidak berbau, tidak berasa, dan terasa segar apabila dihirup (Rahayu, 2021).

### **1.2.2 Polusi Udara**

Polusi udara adalah masuknya zat-zat berbahaya, atau komponen lain ke dalam udara segar yang di sebabkan oleh faktor alami dan buatan, salah satu faktor buatan ialah kegiatan manusia yang melampaui satandar kualitas udara (Waluyo, 2011). Terdapat beberapa sumber pencemaran udara yang dapat merusak kualitas udara segara diantaranya yaitu: sumber perkotaan dan industri yang berasal dari bangunan pabrik-pabrik industri, pembangkit listrik, dan kendaraan pribadi. Selanjutnya sumber pedesaan atau pertanian yang berasal dari penggunaan pestisida yang berlebihan. Kemudian ada sumber alami yang berasal dari abu vulkanik gunung berapi dan bau yang tidak sedap akibat proses pembusukan organik. (Abidin and Hasibuan, 2019).

Baku mutu udara ambien merupakan suatu ukuran pada batas atau kadar zat, energi, dan atau komponen unsur pencemar yang ditenggang keberadaannya dalam udara ambien (Menteri Negara Lingkungan Hidup, 1995). Indeks Standar Pencemar Udara (ISPU) berdasarkan Peraturan Pemerintah Republik Indonesia No.41 Tahun 1999 yaitu suatu angka yang tidak mempunyai satuan yang dimana dapat menggambarkan kondisi mutu udara ambien di suatu lokasi tertentu, yang didasarkan oleh adanya dampak pada kesehatan manusia, nilai estetika dan makhluk hidup lainnya (Kepala Badan Pengendalian Dampak Lingkungan, 1997). Berikut tabel Indeks Standar Pencemaran Udara (ISPU) berdasarkan tingkat pencemaran udara dan dampak kesehatan bagi manusia.

**Tabel 2.2 Indeks Standar Pencemaran Udara (ISPU)**

| <b>ISPU</b> | <b>Tingkat Pencemaran Udara</b> | <b>Dampak Kesehatan</b>   |
|-------------|---------------------------------|---|
| 0-50        | Baik                            | Tingkat kualitas udara yang tidak memberikan efek bagi Kesehatan manusia atau hewan dan tidak berpengaruh pada tumbuhan, bangunan ataupun nilai estetika.               |
| 51-100      | Sedang                          | Kualitas udara yang tidak berpengaruh pada tumbuhan yang sensitive dan nilai estetika.  |
| 101-199     | Tidak Sehat                     | Tingkat kualitas udara yang bersifat merugikan pada manusia ataupun kelompok hewan yang sensitive atau bisa menimbulkan kerusakan pada tumbuhan ataupun nilai estetika. |
| 200-299     | Sangat Tidak Sehat              | Tingkat kualitas udara yang dapat merugikan Kesehatan pada sejumlah segmen populasi yang terpapar.  |
| 300-500     | Berbahaya                       | Tingkat kualitas udara berbahaya yang secara umum dapat merugikan kesehatan yang serius pada populasi.  |

Sumber: [kualitasudara.menlhk.go.id](http://kualitasudara.menlhk.go.id)

### 1.2.3 *Particulate Matter* (PM)

Dalam polusi udara sering muncul istilah PM, campuran dari padatan, larutan, dan larutan partikel yang tersuspensi di udara di sebut dengan *Particulate matter* (PM). (Dwi Lia Oktaviana, 2019). Batas antara partikel kasar dan halus biasanya diantara 1  $\mu\text{m}$  dan 2,5  $\mu\text{m}$ . Aerosol sekunder biasanya terdapat pada partikel yang lebih kecil, misalnya gas ke partikel pembakaran, padatan organic, dan uap logam. Sedangkan untuk material alam yang mengering dan debu dari transportasi dan industri terdapat pada partikel yang lebih besar. Pada partikel halus mengandung komponen asam seperti ion hidrogen (WHO, 2003).

Menurut United States Environmental Protection Agency (2016), partikel terdiri dari PM<sub>10</sub> dan PM<sub>2.5</sub>. PM<sub>10</sub> merupakan partikel yang mudah terhirup dengan ukuran diameter 10 mikrometer, sedangkan PM<sub>2.5</sub> merupakan partikel halus yang mudah terhirup dengan diameter 2,5 mikrometer atau lebih kecil dari PM<sub>10</sub>.

