BAB II
TINJAUAN PUSTAKA

2.1 Data Spasial

Data spasial diperoleh dari hasil pengukuran yang berisi informasi tentang lokasi dan pengukuran. Data ini disajikan dalam bentuk posisi geografis dari objek, lokasi, hubungan dengan objek-objek lainnya, dengan menggunakan titik koordinat dan luasan. Data spasial dapat berupa data diskrit maupun kontinu. Data diskrit adalah data yang diperoleh dengan cara menghitung, tidak merupakan pecahan atau rata-rata. Sedangkan data kontinu adalah data yang bisa mempunyai nilai yang terletak di dalam suatu interval.

Data spasial juga memiliki lokasi spasial yang beraturan (*regular*) maupun lokasi yang tak beraturan (*irregular*). Perbedaan dua lokasi tersebut dilihat dari jarak antara lokasi yang berdekatan. Data spasial dikatakan *regular* jika antara lokasi yang saling berdekatan satu dengan lainnya mempunyai posisi yang beraturan dengan jarak sama besar, sedangkan *irregular* sebaliknya, yaitu jika antara lokasi yang saling berdekatan satu dengan lainnya mempunyai posisi yang tidak beraturan dengan jarak tidak sama besar atau berbeda. Data spasial merupakan salah satu model data dependen (tak bebas), karena data spasial dikumpulkan dari lokasi berbeda yang mengindikasikan ketergantungan antara pengukuran data dan lokasi (Cressie, 1993).

Data spasial mempunyai dua bagian penting yang membuatnya berbeda dari data yang lain, yaitu informasi lokasi (spasial) dan informasi deskriptif (atribut), sebagai berikut :

1. Informasi lokasi (spasial)

   Berkaitan dengan suatu koordinat baik koordinat geografi (lintang dan bujur) dan koordinat XYZ

2. Informasi deskriptif (atribut)

   Suatu lokasi yang memiliki beberapa keterangan yang berkaitan dengannya, contohnya : jenis vegetasi (bermacam jenis tumbuhan dalam suatu wilayah), populasi, luasan, kode pos, dan sebagainya. (Puntodewo, 2003)

2.1.1 Data Geostatistika

Data spasial dibagi menjadi tiga tipe menurut jenis datanya, yaitu : data geostatistika (*geostatistical data*), data area (*lattice data*), dan pola titik (*point pattern*) (Cressie, 1993).

Geostatistika merupakan gabungan dari ilmu statistika dengan geologi, atau mengandung pengertian ilmu statistika yang diterapkan pada ilmu geologi dan beberapa ilmu bumi lainnya. Menurut Cressie (1993), data geostatistika mengarah pada sampel yang berupa titik dari suatu data spasial kontinu, baik yang berbentuk *regular* maupun *irregular*. Karena data geostatistik merupakan tipe mendasar data spasial, pastinya data tersebut memiliki hubungan dengan lokasinya.

2.1.2 Data Area (*Lattice Data*)

Sama halnya dengan data geostatistika, lokasi spasial data area dapat berbentuk beraturan ataupun tidak beraturan. Menurut Cressie (1993) data area merupakan kumpulan data diskrit yang merupakan hasil perhitungan ataupun penjumlahan zona *poligons* pada wilayah tertentu. Data area dapat didefinisikan sebagai sebuah konsep dari garis tepi dan *neighbour* (tetangga sebelah).

Secara umum, data area digunakan dalam studi epidemiologi. Salah satu contoh penggunaannya, untuk mengetahui pertumbuhan suatu penyakit pada suatu wilayah yang terbagi menjadi area-area tertentu. Selain digunakan dalam studi kesehatan, data area juga dapat diaplikasikan pada permasalahan kependudukan, misalnya untuk mengetahui seberapa besar tingkat pertumbuhan penduduk pada area-area dalam sebuah wilayah.

2.1.3 Pola Titik (*Point Pattern*)

Cressie (1993) menjelaskan bahwa pola titik akan muncul apabila hal yang akan dianalisis adalah lokasi dari suatu peristiwa. Sebagai contoh, sebuah permasalahan tentang penentuan posisi pohon-pohon dengan ukuran tertentu. Apakah pohon tersebut membentuk cluster, dan bagaimana satu pohon dengan pohon lainnya berinteraksi.

Hal terpenting dari pola titik adalah mengetahui hubungan ketergantungan antar titik. Beberapa contoh kasus yang menggunakan pola titik seperti lokasi tumbuh pohon di hutan, analisis ketersediaan dan pemanfaatan sumber daya, dan sebagainya.

2.2 *Kriging*

Kriging merupakan suatu metode analisis data geostatistika yang digunakan untuk menduga besarnya nilai yang mewakili suatu titik yang tidak tersampel berdasarkan titik tersampel yang berada di sekitarnya dengan menggunakan model struktural semivariogram. Model semivariogram mempresentasikan perbedaan spasial dan nilai diantara semua pasangan sampel data (Fridayani, Kencana, dan Sukarsa, 2012). Istilah *kriging* diambil dari nama seorang ahli, yaitu D. G. Krige, seorang insinyur pertambangan Afrika Selatan. Setelah itu metode ini dikembangkan oleh G. Matheron pada tahun 1960-an dalam bidang geostatistika. *Kriging* biasanya digunakan untuk menganalisis data geostatistika, seperti menduga kandungan mineral berdasarkan data sampel.

Kriging adalah suatu metode geostatistika yang memanfaatkan nilai spasial pada lokasi tersampel dan variogram untuk memprediksi nilai pada lokasi lain yang belum atau tidak tersampel, dimana nilai prediksi tersebut tergantung pada kedekatannya terhadap lokasi tersampel (Matheron, 1963). *Kriging* juga merupakan suatu metode yang digunakan untuk menonjolkan metode khusus yang meminimalkan variansi dari hasil pendugaan (Fridayani, *et al*. 2012).

Jika dilihat secara umum, metode *kriging* adalah suatu metode analisis geostatistik untuk menginterpolasi suatu nilai kandungan sebagai contoh kandungan mineral, berdasarkan data sampel yang diambil di tempat-tempat yang tidak beraturan.

Banyak metode yang dapat digunakan dalam metode *kriging*, namun berdasarkan diketahui atau tidaknya mean, *Kriging* dapat dibedakan menjadi tiga, yaitu *Simple Kriging, Ordinary Kriging, dan Universal Kriging* (Cressie, 1993).

1. *Simple Kriging*

Simple Kriging merupakan metode *kriging* dengan asumsi bahwa rata-rata (mean) dari populasi telah diketahui dan bernilai konstan. Pengolahan dari metode *Simple kriging* adalah dengan cara data spasial yang akan diduga dipartisi menjadi beberapa bagian.

2. *Ordinary Kriging*

   Ordinary kriging merupakan metode yang diasumsikan rata-rata (mean) dari populasi tidak diketahui, dan pada data spasial tersebut tidak mengandung *trend*. Selain tidak mengandung *trend*, data yang digunakan juga tidak mengandung pencilan.

3. *Universal Kriging*

   Universal kriging merupakan metode *kriging* yang dapat diaplikasikan pada data spasial yang mengandung *trend* atau data yang tidak stasioner.

2.3 Pendeteksian Pencilan Spasial

Pencilan spasial dapat didefinisikan sebagai nilai lokasi observasi yang tidak konsisten atau sangat menyimpang (ekstrim) terhadap nilai lokasi observasi yang lainnya. Salah satu metode yang digunakan untuk mendeteksi adanya pencilan adalah *spatial statitics Z test*, yang didefinisikan sebagai berikut (Fridayani, *et al.* 2012):

$$Z_{s(x)} = \left|\frac{s(x)-\mu_s}{\sigma_s}\right| > \theta \tag{1}$$

dengan,

$s(x)$ = selisih antara nilai amatan dari lokasi x dengan rataan nilai amatan lokasi yang dekat dengan x.

μ_s = nilai mean dari $s(x)$

σ_s = standar deviasi dari $s(x)$

θ = nilai Z tabel untuk tingkat signifikansi tertentu

Jika $Z_{s(x)} > \theta$, maka x dideteksi sebagai pencilan.

Untuk melakukan pendeteksian dilakukan uji hipotesis sebagai berikut:

- Hipotesis

  H_0 : x bukan pencilan

  H_1 : x merupakan pencilan

- Taraf siginifikansi

  α = 0,05

- Statistik Uji

$$Z_{s(x)} = \left| \frac{s(x) - \mu_s}{\sigma_s} \right|$$

- Kriteria Uji

  H_0 ditolak jika $Z_{s(x)} > Z_\alpha$

2.4 Variogram dan Semivariogram

Variogram merupakan perangkat statistik yang diperlukan untuk melakukan pendugaan pada data spasial, karena jika ada dua buah nilai spasial yang letaknya berdekatan, maka akan relatif bernilai sama dibandingkan dengan dua buah nilai spasial yang letaknya berjauhan (Cressie, 1993).

Variogram dirumuskan sebagai berikut (Cressie, 1993):

$$2\gamma(h) = E[Z(s) - Z(s+h)]^2 \qquad (2)$$

Untuk melakukan pendugaan pada data spasial, digunakan suatu perangkat untuk menggambarkan, memodelkan, dan menghitung korelasi spasial antara variabel random $Z(s)$ dan $Z(s+h)$, yang disebut dengan semivariogram. Besarnya nilai semivariogram adalah setengah dari nilai variogram (Cressie, 1993). Semivariogram dapat didefinisikan sebagai berikut :

$$\gamma(h) = \frac{1}{2}E[Z(s) - Z(s+h)]^2 \quad (3)$$

2.4.1 Variogram dan Semivariogram Eksperimental

Variogram eksperimental adalah nilai dugaan yang diperoleh dari penarikan sampel di lapangan. Variogram eksperimental dibuat berdasarkan nilai korelasi spasial antara dua buah variabel yang dipisahkan oleh suatu jarak tertentu sebesar h. Variogram eksperimental dirumuskan sebagai berikut (Cressie, 1993):

$$2\gamma(h) = \frac{1}{N(h)}\sum[Z(s_i) - Z(s_i+h)]^2 \quad (4)$$

dimana,

s_i = lokasi titik sampel

$Z(s_i)$ = nilai observasi pada lokasi s_i

h = jarak antara dua titik sampel

s_i , $s_i + h$ = pasangan titik sampel yang berjarak h

$N(h)$ = banyak pasangan data yang memiliki jarak h

Semivariogram eksperimental dirumuskan sebagai berikut (Cressie, 1993):

$$\gamma(h) = \frac{1}{2N(h)}\sum[Z(s_i) - Z(s_i+h)]^2 \quad (5)$$

Untuk mencari nilai semivariogram, banyak pasangan data akan di bagi menjadi beberapa kelas menggunakan pesamaan sturge berikut (Harinaldi, 2005) :

$$k = 1 + 3{,}3 \log n \qquad (6)$$

dimana :

k = banyak interval kelas

n = ukuran sampel

Setelah diperoleh nilai semivariogram eksperimental, maka dapat dihitung parameter-parameter yang akan digunakan untuk perhitungan semivariogram teoritis.

2.4.2 Semivariogram Teoritis

Beberapa parameter yang digunakan untuk mencari nilai dalam semivariogram teoritis adalah *nugget effect*, *sill*, dan *range* (Webster dan Oliver, 2007):

1. *Nugget Effect* (C_0)

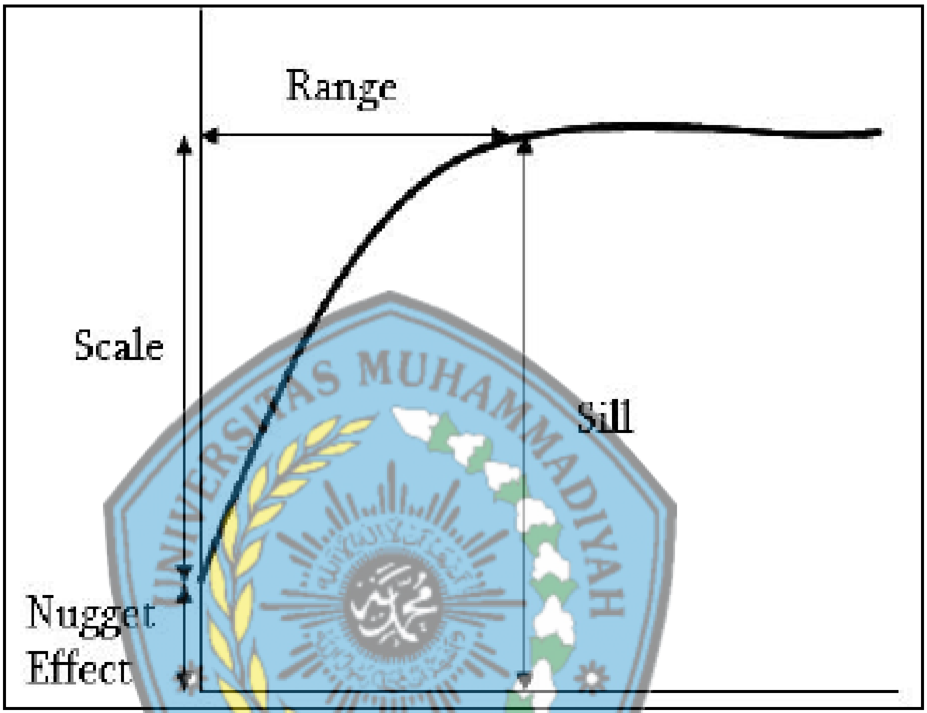
   Nugget effect merupakan pendekatan nilai semivariogram pada jarak sekitar nol.

2. *Sill* (C_0 + C)

   Sill adalah saat dimana nilai semivariogram cenderung mencapai nilai yang stabil. Nilai *sill* sama dengan nilai varian dari data spasial.

3. *Range* (a)

Range merupakan jarak pada saat semivariogram mencapai nilai *sill*. Nilai *sill* diperoleh dengan cara mengambil batas tengah dari nilai semivariogram yang paling mendekati nilai varian dari data.



Gambar 1. Bentuk Umum Variogram

Setelah memperoleh nilai dari ketiga parameter di atas, selanjutnya dilakukan perhitungan nilai semivarogram teoritis. Nilai yang diperoleh dari semivariogram teoritis akan digunakan untuk analisis struktural, yaitu membandingkan nilai *Mean Square Error* (MSE) antara semivariogram eksperimental dengan semivariogram teoritis. Dari perbandingan tersebut akan dipilih model mana yang memiliki nilai *Mean Square Error* (MSE) paling kecil, yang nantinya model tersebut akan digunakan untuk melakukan pendugaan data spasial.

Berikut adalah beberapa model semivariogram teoritis yang digunakan sebagai pembanding (Cressie, 1993):

1. Model *Spherical*

Semivariogram untuk model *Spherical* dirumuskan sebagai berikut :

$$\gamma(h) = \begin{cases} C_0 + C\left[\left(\frac{3h}{2a}\right) - 0{,}5\left(\frac{h}{a}\right)^3\right] & untuk\ h \leq a \\ C_0 + C & untuk\ h > a \end{cases} \tag{7}$$

dengan,

h = jarak lokasi sampel

$C_0 + C$ = *sill*, yaitu nilai semivariogram untuk jarak pada saat besarnya konstan.

a = *range*, yaitu jarak pada saat nilai semivariogram mencapai *sill*.

2. Model *Eksponensial*

Semivariogram model *eksponensial* dirumuskan sebagai berikut:

$$\gamma(h) = C_0 + C\left[1 - \exp\left(-\frac{3h}{a}\right)\right] \tag{8}$$

3. Model *Gaussian*

Semivariogram untuk model *gaussian* dirumuskan sebagai berikut :

$$\gamma(h) = C_0 + C\left[1 - \exp\frac{-3h^2}{a^2}\right] \tag{9}$$

2.5 Pendugaan Parameter *Ordinary Kriging*

Ordinary kriging adalah salah satu metode geostatistika yang sederhana. Pada metode ini diasumsikan rata-rata (mean) populasi adalah konstan, tetapi tidak diketahui, sedangkan variogram dari $Z(s)$ diketahui, dan pada data tersebut tidak mengandung *trend*.

Menurut Isaaks dan Srivasta (1989) penduga *kriging* $\hat{Z}(s)$ merupakan kombinasi linier. Kombinasi linier adalah penjumlahan hasil kali anggota himpunan pasangan berurutan. Penduga *kriging* $\hat{Z}(s)$ merupakan kombinasi linier dari variabel sampel $Z(s_i)$ yang diketahui atau ditulis secara matematis sebagai berikut :

$$\hat{Z}(s) = \sum_{i=1}^{n} w_i Z(s_i) \tag{10}$$

dengan,

$\hat{Z}(s)$ = nilai pendugaan pada lokasi tidak tersampel

w_i = koefisien bobot dari $Z(s_i)$, dengan $\sum_{i=1}^{n} w_i = 1$

$Z(s_i)$ = nilai pada lokasi tersampel

n = banyak sampel

Untuk memperoleh suatu pendugaan $\hat{Z}(s)$ di titik P dari 3 titik observasi yang diketahui yaitu, Z_1, Z_2, Z_3 dengan bobot masing-masing untuk persamaan *Ordinary Kriging* yaitu : w_1, w_2, w_3. Untuk memperoleh solusi yang diinginkan, diperlukan 3 persamaan simultan berikut dan ditambahkan dengan 1 persamaan persyaratan yaitu, penjumlahan semua bobot adalah samadengan 1 (Isaaks dan Srivasta, 1989). Sehingga setelah dijabarkan terdapat 4 persamaan sebagai berikut.

$$w_1\gamma(h_{11}) + w_2\gamma(h_{12}) + w_3\gamma(h_{13}) = \gamma(h_{1p}) \tag{11}$$

$$w_1\gamma(h_{21}) + w_2\gamma(h_{22}) + w_3\gamma(h_{23}) = \gamma(h_{2p}) \tag{12}$$

$$w_3\gamma(h_{31}) + w_3\gamma(h_{32}) + w_3\gamma(h_{33}) = \gamma(h_{3p}) \tag{13}$$

$$w_1 + w_2 + w_3 = 1 \tag{14}$$

Untuk menghasilkan solusi yang memiliki galat penduga minimum ditambahkan suatu variabel slag λ pada persamaan (11), (12), dan (13). Dengan demikian, keempat persamaan di atas menjadi:

$$w_1\gamma(h_{11}) + w_2\gamma(h_{12}) + w_3\gamma(h_{13}) + \lambda = \gamma(h_{1p}) \tag{15}$$

$$w_1\gamma(h_{21}) + w_2\gamma(h_{22}) + w_3\gamma(h_{23}) + \lambda = \gamma(h_{2p}) \tag{16}$$

$$w_3\gamma(h_{31}) + w_3\gamma(h_{32}) + w_3\gamma(h_{33}) + \lambda = \gamma(h_{3p}) \tag{17}$$

$$w_1 + w_2 + w_3 + 0 = 1 \tag{18}$$

Sistem persamaan (15), (16), (17), dan (18) dapat ditulis dalam bentuk matriks sebagai berikut :

$$\begin{pmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{13}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \gamma(h_{23}) & 1 \\ \gamma(h_{31}) & \gamma(h_{32}) & \gamma(h_{33}) & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{pmatrix} = \begin{pmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \gamma(h_{3p}) \\ 1 \end{pmatrix} \tag{19}$$

Atau dapat ditulis sebagai berikut :

$$\boldsymbol{A}\,\boldsymbol{w} \quad = \quad \boldsymbol{B} \tag{20}$$

Jadi untuk memperoleh $\boldsymbol{w}$ digunakan persamaan berikut :

$$\boldsymbol{w} = \boldsymbol{A}^{-1}\boldsymbol{B} \tag{21}$$

Terdapat n data hasil pengukuran, yaitu $Z(s_1), Z(s_2), Z(s_3), \dots, Z(s_n)$. Akan diduga $Z(s)$ pada lokasi yang tidak tersampel berdasarkan data tersampel yang akan dinyatakan dengan $Z(s_0)$. Berdasarkan persamaan (10) diperoleh penduga *error* sebagai berikut :

$$\hat{e}(s_0) = \hat{Z}(s_0) - Z(s_0) = \sum_{i=1}^{n} w_i Z(s_i) - Z(s_0) \tag{22}$$

Ordinary Kriging akan menghasilkan penduga yang tak bias (*Unbiased Estimator)*. Berdasarkan persamaan (10) diperoleh pendugaan sebagai berikut :

$$E\big(\hat{e}(s_0)\big) = E\left(\sum_{i=1}^{n} w_i Z(s_i) - Z(s_0)\right)$$

$$= \sum_{i=1}^{n} w_i E(Z(s_i)) - E(Z(s_0)) \tag{23}$$

Dengan asumsi bahwa metode *Ordinary Kriging* bersifat stasioner, maka setiap nilai ekspektasi boleh dituliskan sebagai $E(Z)$, sehingga diperoleh persamaan sebagai berikut :

$$E\big(\hat{e}(s_0)\big) = \sum_{i=1}^{n} w_i E\big((Z)\big) - E\big((Z)\big)$$

Karena $E\big(\hat{e}(s_0)\big) = 0$, maka

$$E\big(\hat{e}(s_0)\big) = 0 = \sum_{i=1}^{n} w_i E\big((Z)\big) - E\big((Z)\big)$$

$$0 = \sum_{i=1}^{n} w_i E\big((Z)\big) - E\big((Z)\big)$$

$$\sum_{i=1}^{n} w_i E\big((Z)\big) = E\big((Z)\big)$$

$$\sum_{i=1}^{n} w_i = 1$$

Sehingga

$$E\big(\hat{Z}(s)\big) = E\left(\sum_{i=1}^{n} w_i\, Z(s)\right)$$

$$= \sum_{i=1}^{n} w_i\; E\big(Z(s)\big)$$

$$E(\hat{Z}(s)) = E(Z(s))\sum_{i=1}^{n} w_i$$

$$E\left(\hat{Z}(s)\right) = Z(s)$$

Berdasarkan penjabaran di atas dapat disimpulkan bahwa *Ordinary Kriging* menghasilkan penduga tak bias.

2.6 *Inverse Distance Weighted* (IDW)

Menurut Pasaribu dan Haryani (2012), metode *Inverse Distance Weighted* (IDW) memiliki asumsi bahwa setiap titik *input* mempunyai yang bersifat lokal yang berkurang terhadap jarak. Metode interpolasi ini dapat menyesuaikan pengaruh relative dari titik-titik sampel. Nilai power atau parameter kuasa adalah nilai positif yang dapat diubah-ubah. Nilai *power* pada interpolasi IDW menentukan pengaruh titik-titik yang lebih dekat sehingga menghasilkan permukaan yang lebih detai.

Menurut Azpurua dan Ramos (2010) dalam Pasaribu dan Haryani (2012), pembobot *Inverse Distance Weighted* (IDW) dapat rumuskan sebagai berikut :

$$w_i = \frac{h_i^{-p}}{\sum_{i=1}^{n} h_i^{-p}} \tag{24}$$

Untuk nilai dugaan menggunakan *Inverse Distance Weighted* dirumuskan sebagai berikut :

$$Z^* = \sum_{i=1}^{n} w_i Z_i \tag{25}$$

dengan,

h_i = jarak antara titik dugaan dengan sampel ke-i

p = parameter kuasa

Z^* = nilai pada titik dugaan

Z_i = nilai sampel pada titik ke-i