

**PENERAPAN METODE *ROBUST KRIGING* PADA DATA
CURAH HUJAN WILAYAH SULAWESI SELATAN
UNTUK MENGESTIMASI ADANYA *OUTLIER*
YANG DISEBABKAN OLEH DATA HILANG**

SKRIPSI



MUHAMMAD IBNU SYUKUR

H12116501

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

UNIVERSITAS HASANUDDIN

MAKASSAR

2023

**PENERAPAN METODE *ROBUST KRIGING* PADA DATA
CURAH HUJAN WILAYAH SULAWESI SELATAN UNTUK
MENGESTIMASI ADANYA *OUTLIER* YANG DISEBABKAN
OLEH DATA HILANG**

SKRIPSI

**Diajukan Sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains
pada Program Studi Statistika Departemen Statistika Fakultas Matematika
dan Ilmu Pengetahuan Alam Universitas Hasanuddin**

MUHAMMAD IBNU SYUKUR

H12116501

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

MAKASSAR

AGUSTUS 2023

LEMBARAN PERNYATAAN KEOTENTIKAN


Saya yang bertanda tangan dibawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Penerapan Metode *Robust Kriging* Pada Data Curah Hujan Wilayah Sulawesi Selatan Untuk Mengestimasi Adanya *Outlier* Yang Disebabkan Oleh Data Hilang

Adalah benar hasil karya saya sendiri bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 3 Agustus 2023




Muhammad Ibnu Syukur
Nim. H12116501

PENERAPAN METODE *ROBUST KRIGING* PADA DATA
CURAH HUJAN WILAYAH SULAWESI SELATAN UNTUK
MENGESTIMASI ADANYA *OUTLIER* YANG DISEBABKAN
OLEH DATA HILANG

Disetujui oleh:

Pembimbing Utama



Dr. Dr. Georgina M. Tinungki, M.Si.
NIP. 19620926198702 2 001

Pembimbing Pendamping



Anisa, S.Si, M.Si.
NIP. 19730227199802 2 001

Ketua Program Studi



Dr. Anna Islamiyati, S.si., M.Si.
NIP. 19770808200501 2 002

Pada 3 Agustus 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Muhammad Ibnu Syukur

Nim : H12116501

Program Studi : Statistika

Judul Skripsi : Penerapan Metode Robust Kriging Pada Data Curah Hujan Wilayah Sulawesi Selatan Untuk Mengestimasi Adanya Outlier Yang Disebabkan Oleh Data Hilang

Telah Berhasil Dipertahankan Dihadapan Dewan Penguji Dan Diterima Sebagai Bagian Persyaratan Yang Diperlukan Untuk Memperoleh Gelar Sarjana Sains Pada Program Studi Statistika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

- Tanda tangan
1. Ketua : Dr. Dr. Georgina Maria Tinungki, M.Si. (.....)
 2. Sekretaris : Anisa S.Si, M.Si. (.....)
 3. Anggota : Drs. Raupong, M. Si. (.....)
 4. Anggota : Siswanto, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 3 Agustus 2023

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan kasihnya, yang telah menyertai penulis selama proses penyelesaian skripsi dengan judul “**Penerapan Metode *Robust Kriging* Pada Data Curah Hujan Wilayah Sulawesi Selatan Untuk Mengestimasi Adanya *Outlier* Yang Disebabkan Oleh Data Hilang**” sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan. Namun, karena berkat dan kasih Tuhan serta dukungan dari berbagai pihak yang turut membantu sehingga penulisan tugas akhir ini dapat terselesaikan. Penulis menyampaikan ucapan terima kasih yang setinggi-tingginya kepada Ayahanda **Drs. Jabir, MM** dan Ibunda tercinta **Haderia** atas doanya yang tak pernah putus, serta kasih sayang yang melimpah dalam mendidik dan membesarkan penulis dengan begitu banyak pengorbanan yang tak pernah ternilai harganya. Penghargaan yang tulus dan ucapan terima kasih yang sebesar-besarnya juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin, M.Si.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Anna Islamiyati, S.Si, M.Si** selaku Ketua Departemen Statistika dan **Ibu Anisa, S.Si, M.Si**, selaku Sekretaris Departemen Statistika dan segenap dosen pengajar yang telah memberikan ilmu pengetahuan kepada penulis selama menjadi mahasiswa di Departemen Statistika.
4. **Ibu Dr. Dr. Georgina Maria Tinungki, M.Si** selaku pembimbing utama yang dengan tulus meluangkan waktunya dengan penuh kesabaran memberikan bimbingan dan membagi ilmunya kepada penulis.
5. **Ibu Anisa, S.Si, M.Si.** selaku pembimbing pertama yang telah bersedia memberikan bimbingan memotivasi penulis dalam penyusunan tugas akhir ini.
6. **Bapak Dr. Raupong, M.Si.** selaku pembimbing akademik dan anggota penguji yang telah memberikan banyak bantuan selama penulis menjalani pendidikan

dan memberikan kritik dan saran yang membangun untuk penulisan tugas akhir ini.

7. **Bapak Siswanto, S.Si., M.Si.** selaku anggota penguji yang telah memberikan masukan kepada penulis dalam perbaikan skripsi ke arah yang lebih baik.
8. Segenap Staf Pegawai Departemen Statistika yang telah membantu proses administrasi selama penulis menyelesaikan tugas akhir.
9. Sahabat terbaik penulis **Aldi Reza Septiawan S.T, Samsir Aditya Ania, S.Si** dan **Andi Isna Yunita, S.Si., M.Si.** yang telah memberikan semangat dan motivasi selama penyusunan skripsi.
10. Keluarga Besar Statistika UNHAS 2016 terima kasih atas motivasi, ilmu, dan bantuan yang diberikan kepada penulis baik sebelum dan selama penyusunan skripsi.
11. Seluruh warga Statistika UNHAS, sebagai keluarga yang memberi nasehat-nasehat sehingga penulis dapat lebih mengerti arti pentingnya kebersamaan dan semua pihak yang telah banyak membantu penulis dan tak sempat penulis sebutkan satu per satu. Semoga segala bantuan dan partisipasinya dibalas oleh Tuhan Yang Maha Esa.
12. Semua pihak yang telah memberikan bantuan dan dukungan secara langsung maupun tidak langsung yang tidak sempat penulis sebutkan satu-persatu, semoga kita semua selalu dalam lindungan dan berkat Tuhan.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat bagi para pembaca.

Makassar, 3 Agustus 2023



Muhammad Ibnu Syukur

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan dibawah ini:

Nama : Muhammad Ibnu Syukur
NM : H12116501
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty Free Right*)** atas tugas akhir saya yang berjudul:

“Penerapan Metode Robust Kriging Pada Data Curah Hujan Wilayah Sulawesi Selatan Untuk Mengestimasi Adanya Outlier Yang Disebabkan Oleh Data Hilang”

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal diatas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal, 3 Agustus 2023

Yang menyatakan



(Muhammad Ibnu Syukur)

ABSTRAK

Data curah hujan yang kita peroleh seringkali datanya tidak lengkap dan terdapat data yang hilang, agar ukuran sampel tidak berkurang dilakukan *median imputation* pada data yang hilang. Data yang digunakan dalam penelitian ini adalah data curah hujan wilayah Sulawesi Selatan Pada tahun 2021 yang mengalami masalah data hilang. Untuk mengatasi adanya pencilan yang disebabkan oleh data hilang tersebut maka dapat menggunakan metode Robust kriging yaitu metode yang mengakomodir adanya pencilan. Tujuan dari penelitian ini adalah mencari nilai data hilang dengan menggunakan *Median Imputation* dan mendapatkan nilai model *semivariogram* terbaik berdasarkan kriteria *Root Mean Square Error (RMSE)*. Hasil yang diperoleh menunjukkan bahwa terdapat data hilang pada bulan Januari – Desember untuk kelima stasiun Sulawesi Selatan. Model *semivariogram* teoritis terbaik untuk model *spherical* pada bulan Februari (5.68), Juli (4.26), September (6.54), Oktober (3.85) dan desember (1.30), model *exponential* pada bulan Januari (7.87) dan untuk model *gaussian* pada bulan Maret (4.00), Juni (3.31), Agustus (8.05) dan November (3.93).

Kata kunci: Curah Hujan, *Imputation*, *Kriging*, *Missing Data*, *Robust Kriging*, *Semivariogram*.

ABSTRACT

The rainfall data we obtain is often incomplete and there is missing data, so that the sample size does not decrease, the median imputation of the missing data is carried out. The data used in this study is rainfall data for the South Sulawesi region in 2021 which experienced data loss problems. To overcome the outliers caused by the missing data, you can use the Robust kriging method, which is a method that accommodates outliers. The purpose of this study is to find missing data values using Median Imputation and obtain the best semivariogram model values based on Root Mean Square Error (RMSE) criteria. The results showed that there was missing data in January – December for all five South Sulawesi stations. The theoretical semivariogram models are best for spherical models in February (5.68), July (4.26), September (6.54), October (3.85) and December (1.30), exponential models in January (7.87) and for gaussian models in March (4.00), June (3.31), August (8.05) and November (3.93).

Keywords: *Precipitation, Imputation, Kriging, Missing Data, Robust Kriging, Semivariogram.*

DAFTAR ISI

HAIAMAN JUDUL.....	ii
LEMBARAN PERNYATAAN KEOTENTIKAN	iii
HALAMAN PERSUTUJUAN	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR.....	vi
PERNYATAAN PERSETUJUAN KARYA ILMIAH	viii
ABSTRAK	ix
<i>ABSTRACT</i>	x
DAFTAR ISI.....	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar belakang	1
BAB II TINJAU PUSTAKA	4
2.1 Data Spasial.....	4
2.2 Imputasi	4
2.3 Spasial Outlier	6
2.4 <i>Kriging</i>	6
2.5 <i>Robust Kriging</i>	7
2.6 Pemodelan <i>Semivariogram</i>	8
2.6.1. Semivariogram Eksperimental.....	8
2.6.2. Semivariogram <i>Robust</i>	8
2.6.3 Semivariogram Teoritis	9
2.7 <i>Root Mean Square Error</i>	10
2.8 Bobot <i>Kriging</i>	11
BAB III METODELOGI PENELITIAN.....	12

3.1 Sumber Data	12
3.2 Variabel Penelitian	12
3.3 Analisis Data	12
BAB IV HASIL DAN KESIMPULAN.....	14
4.1 Deskripsi Data	14
4.2 <i>Missing Data</i>	15
4.2.1 Lokasi Makassar	15
4.2.2 Lokasi Maros	16
4.2.3 Lokasi Toraja	18
4.2.4 Lokasi Luwu Utara	19
4.3 Imputasi	19
4.3.1 <i>Median Imputation</i>	20
4.4 Pendeteksian <i>Outlier</i>	24
4.5 Menghitung Nilai <i>Semivariogram Robust</i>	26
4.6 Menghitung Model <i>Semivariogram Teoritis</i>	31
4.8 Bobot <i>Kriging</i> Dan <i>Weighted Median</i>	36
BAB IV KESIMPULAN DAN SARAN.....	41
5.1 Kesimpulan.....	41
DAFTAR PUSTAKA	42
LAMPIRAN.....	44

DAFTAR TABEL

Tabel 3.1 Variabel Curah Hujan Sulawesi Selatan Tahun 2021	12
Tabel 4.1 <i>Stasiun, Lintang dan Bujur</i>	14
Tabel 4.2 Nilai <i>Min, Max, Mean</i> dan <i>Varian</i> Curah Hujan (mm).....	14
Tabel 4.3 Stasiun Meteorologi Maritim Paotere	15
Tabel 4.4 Stasiun Meteorologi Sultan Hasanuddin	16
Tabel 4.5 Stasiun Klimatologi Sulawesi Selatan.....	17
Tabel 4.6 Stasiun Meteorologi Toraja.....	18
Tabel 4.7 Stasiun Meteorologi Andi Jamma	19
Tabel 4.8 Hasil <i>Median</i> Dari 5 Stasiun Tiap Bulannya.....	20
Tabel 4.9 Pendeteksian Data Hilang Dengan Menggunakan <i>Rstudio</i> pada Stasiun Meteorologi Paotere	21
Tabel 4. 10 Pengisian Kembali Data Hilang Dengan Menggunakan Median Imputasi Pada Stasiun Meteorologi Paotere	23
Tabel 4.11 Pendeteksian <i>Outlier</i> dengan <i>Z-test</i>	25
Tabel 4. 12 Nilai Semivariogram Pada Bulan Januari	26
Tabel 4. 13 Nilai Semivariogram Pada Bulan Februari	27
Tabel 4. 14 Nilai Semivariogram Pada Bulan Maret	27
Tabel 4. 15 Nilai Semivariogram Pada Bulan Januari April	27
Tabel 4. 16 Nilai Semivariogram Pada Bulan Januari Mei	28
Tabel 4. 17 Nilai Semivariogram Pada Bulan Juni	28
Tabel 4. 18 Nilai Semivariogram Pada Bulan Juli	28
Tabel 4. 19 Nilai Semivariogram Pada Bulan Agustus.....	29
Tabel 4. 20 Nilai Semivariogram Pada Bulan September.....	29
Tabel 4. 21 Nilai Semivariogram Pada Bulan Oktober.....	29
Tabel 4. 22 Nilai Semivariogram Pada Bulan November	30
Tabel 4. 23 Nilai Semivariogram Pada Bulan Desember.....	30
Tabel 4. 24 Nilai Parameter <i>Nugget, Sill</i> dan <i>Range</i> Untuk Setiap Bulannya serta Nilai Rasionya.....	33
Tabel 4.25 <i>Semivariogram</i> Dengan Kriteria <i>RMSE</i>	35
Tabel 4.26 Nilai Bobot <i>kriging</i>	37
Tabel 4. 27 Nilai <i>Weighted Median Outlier</i>	39

DAFTAR GAMBAR

- Gambar 4. 1** Hasil Plot Dengan Model *Spherical* Pada Data Bulan Januari..... 31
- Gambar 4. 2** Hasil Plot Dengan Model *Exponential* Pada Data Bulan Januari... 32
- Gambar 4. 3** Hasil Plot Dengan Model *Gaussian* Untuk Data Bulan Januari..... 32

DAFTAR LAMPIRAN

Lampiran 1 Data Curah Hujan Tiap Bulan Untuk Stasiun Meteorologi Maritim Paotere Tahun 2021.....	47
Lampiran 2 Data Curah Hujan Tiap Bulan Untuk Stasiun Meteorologi Sultan Hasanuddin Tahun 2021	49
Lampiran 3 Data Curah Hujan Tiap Bulan Untuk Stasiun Klimatologi Sulawesi Selatan Tahun 2021.....	51
Lampiran 4 Data Curah Hujan Tiap Bulan Untuk Stasiun Meteorologi Toraja Tahun 2021	53
Lampiran 5 Data Curah Hujan Tiap Bulan Untuk Stasiun Meteorologi Andi Jemma Tahun 2021	55
Lampiran 6 Hasil Perhitungan Z-tes Dengan <i>Rstudio</i> Untuk Keseluruhan Data	57
Lampiran 7 Hasil Plot Ketiga Model <i>Semivariogram Teoritis</i> Dengan Menggunakan <i>Sofwer Rstudio</i>	58

BAB I

PENDAHULUAN

1.1 Latar belakang

Indonesia merupakan negara yang mempunyai angka curah hujan yang bervariasi dikarenakan daerahnya yang berada di ketinggian yang berbeda-beda. Hal ini menyebabkan daerah di sekitarnya yang tidak memiliki stasiun curah hujan tidak dapat diukur secara pasti. Oleh karena itu dibutuhkan suatu metode *interpolasi spasial* yang sesuai untuk menduga atau memprediksi curah hujan di daerah yang tidak memiliki stasiun curah hujan.

Menurut Anderson (2001) *interpolasi spasial* sendiri merupakan suatu metode atau fungsi matematika yang menduga nilai atribut pada lokasi-lokasi yang datanya tidak diketahui. *Interpolasi spasial* mengasumsikan bahwa atribut data bersifat kontinyu dalam ruang (*space*) dan atribut ini saling berhubungan (*dependence*) secara *spasial*. Kedua asumsi tersebut yang mendasari bahwa pendugaan atribut pada suatu lokasi dapat dilakukan berdasarkan data dari lokasi-lokasi disekitarnya dan nilai pada titik-titik yang berdekatan akan lebih mirip dari pada titik-titik yang berjauhan (Prasasti, Wijayanto dan Christanto, 2005).

Metode penaksiran yang digunakan untuk menduga nilai pada suatu titik disebut dengan metode kriging. Metode ini ditemukan oleh Daniel Gerhardus Krige untuk memperkirakan nilai dari bahan tambang. Kemudian metode ini dikembangkan oleh Georges Matheron dalam bidang *geostatistika*. Menurut Oliver dan Webster (2015) *kriging* adalah teknik *geostatistik* yang digunakan untuk mengestimasi besar nilai yang mewakili suatu titik yang tidak tersample berdasarkan titik-titik tersample yang berada di sekitar dengan mempertimbangkan *korelasi spasial* yang ada dalam data tersebut. *Kriging* sendiri bekerja berdasarkan pembobot yang tergantung pada *variogram* dan susunan dari lokasi di lingkungan sekitar target. *Kriging* bersifat *best linear unbiased predictor*, baik dalam meminimumkan ragam galat prediksi.

Kriging umumnya terbagi menjadi *simple kriging*, *ordinary kriging* dan *universal kriging*. Ketiga metode ini membutuhkan asumsi bahwa data yang dimiliki harus berdistribusi normal. Namun pada kenyataannya data yang sering diperoleh tidak berdistribusi normal dan memiliki *outlier*. *Outlier* bisa terjadi

karena berbagai alasan, seperti kesalahan dalam pengukuran, data yang kurang lengkap, ataupun karena faktor lain yang tidak terduga. Oleh karena itu Untuk mengatasi adanya pencilan tersebut maka dapat menggunakan metode *Robust kriging* yaitu metode yang mengakomodir adanya *outlier*.

Beberapa cara yang dapat dilakukan untuk mengatasi data yang hilang seperti *Listwise deletion*, *Pairwis deletion* dan *Imputation*. penanganan data hilang juga dapat dilakukan dengan cara menghapus data yang tidak lengkap dengan asumsi bahwa jika data yang hilang jumlahnya relatif kecil dibandingkan dengan keseluruhan data maka dapat menghapus data tersebut. Tetapi umumnya hal ini akan memberikan kesimpulan yang valid, hanya terjadi ketika *missing* data secara acak, dalam arti bahwa probabilitas respon tidak tergantung pada nilai-nilai data yang diamati atau hilang. Sebagai solusi yang tepat agar ukuran sampel tidak berkurang, maka jika terdapat *missing* data dalam survey maka dapat melakukan *imputation* pada data yang hilang.

Penelitian sebelumnya tentang metode *Robust kriging* terhadap pencilan data telah dilakukan oleh Firdayanti dkk (2012) yang meneliti tentang “Perbandingan Interpolasi Spasial Dengan Metode *Ordinary* dan *Robust Kriging* pada Data Spasial Berpencilan” dengan studi kasus yaitu curah hujan dikabupaten Karangasem, Aini (2015) yang meneliti tentang “Metode *Robust Kriging* dan Penerapannya pada data Geostatistika” dengan studi kasus yaitu kandungan timah dilokasi penambangan, Negara (2016) meneliti tentang “Penerapan Interpolasi *Robust Kriging* Pada Data Spasial Berpencilan Menggunakan *Semivariogram* isotropik” dengan studi kasus data curah hujan Kabupaten Malang. Hal yang membedakan antara penelitian ini dengan beberapa penelitian sebelumnya adalah *Semivariogram* yang digunakan yaitu *Semivariogram Robust* serta studi kasus yaitu curah hujan wilayah Sulawesi Selatan pada tahun 2021. Oleh karena itu penulis memilih judul untuk penelitian ini yaitu **“Penerapan Metode *Robust Kriging* Pada Data Curah Hujan Wilayah Sulawesi Selatan Untuk Mengestimasi Adanya *Outlier* Yang Disebabkan Oleh Data Hilang”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah, maka dapat dirumuskan sebagai berikut:

1. Mencari nilai data hilang dengan menggunakan *Median Imputation* pada data curah hujan wilayah Sulawesi Selatan pada tahun 2021?
2. Bagaimana model *Semivariogram* terbaik yang didapatkan dengan metode *robust kriging* untuk data curah hujan wilayah Sulawesi Selatan pada tahun 2021?

1.3 Batasan Masalah

1. Penelitian hanya berfokus pada data curah hujan wilayah Sulawesi Selatan pada tahun 2021.
2. Model semivariogram yang digunakan dalam penelitian ini adalah model *spherical*, model *exponential*, dan model *gaussian*.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini yaitu sebagaimana yang telah dibahas pada rumusan masalah diatas maka tujuan penelitian adalah:

1. Mendapatkan nilai median untuk dilakukan *imputation* kedalam data yang hilang agar data tersebut menjadi data yang kompleks.
2. Mendapatkan model *Semivariogram* terbaik dengan menggunakan metode *robust kriging* untuk data curah hujan wilayah Sulawesi Selatan pada tahun 2021.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini yaitu:

1. Mengetahui bagaimana cara mengisi kembali data yang telah hilang dengan melakukan median *imputation* kedalam data agar data tersebut menjadi lengkap.
2. Mendapatkan model *Semivariogram* terbaik dari metode *robust kriging* berdasarkan kriteria *RMSE*.

BAB II TINJAU PUSTAKA

2.1 Data Spasial

Data spasial merupakan data yang menunjukkan lokasi letak data tersebut di permukaan bumi. Data spasial juga biasa diartikan sebagai data yang meninjau pada geografis dari suatu objek, berkaitan dengan lokasi, bentuk dan hubungan diantaranya dalam ruang bumi. Bentuk penyajian dalam data spasial mempunyai tiga cara yaitu titik (*point*), garis (*line*), dan area (*polygon*). Data dapat dibedakan menjadi 3 tipe berdasarkan jenis datanya, yaitu data geostatistik (*geostatistical data*), data area (*lattice data*), dan pola titik (*point pattern*). 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).

Geostatistika muncul pada awal tahun 1980-an yang merupakan gabungan dari ilmu teknik pertambangan, geologi, matematika dan statistika. Kelebihannya jika dibandingkan dengan pendeteksian klasik yaitu bahwa geostatistika mampu memodelkan baik kecenderungan spasial (*spatial trend*) maupun korelasi spasial (*spatial correlation*).

Data area berhubungan dengan daerah spasial karena merupakan kumpulan data atribut diskrit yang merupakan hasil dari perhitungan atau jumlah pada wilayah tertentu. Data area dapat didefinisikan sebagai sebuah konsep dari garis tepi dan *neighbour* (tetangga sebelah).

Pola titik akan timbul ketika sebuah variabel yang akan dianalisis pada lokasi kejadian. Dalam hal ini, apakah lokasi kejadian yang akan diteliti menunjukkan suatu keacakan spasial, *cluster*, atau keteraturan (*regularity*) sehingga dapat dilihat apakah ada pola titik yang terbentuk memiliki ketergantungan titik atau tidak.

2.2 Imputasi

Acuna & Rodriguez (2003) metode imputasi adalah proses pengisian atau penggantian nilai-nilai yang hilang pada data dengan nilai-nilai yang ada berdasarkan informasi yang didapatkan pada data. Terdapat beberapa metode

teknik imputasi diantaranya adalah *Median Imputation*, *Mode Imputation*, dan *Forward Imputation*.

a) *Median Imputation*

Median imputation adalah metode yang dimana data hilang untuk peubah digantikan oleh median dari semua nilai yang terdapat pada peubah yang didasarkan pada semua nilai yang diketahui/valid. *Metode imputasi* ini diterapkan secara terpisah pada masing-masing peubah yang mengandung data hilang (Acuna & Rodriguez, 2003).

b) *Mode Imputation*

Mode Imputation adalah metode dimana nilai-nilai yang hilang untuk peubah digantikan oleh modus dari semua nilai yang terdapat pada peubah yang didasarkan pada semua nilai yang diketahui/valid. Metode imputasi ini diterapkan secara terpisah pada masing-masing peubah yang mengandung nilai-nilai yang hilang.

Permasalahan data hilang dalam atribut bertipe kategori dapat digunakan metode *Mode Imputation*. dalam *Mode Imputation* struktur berhubungan antara data tidak dianggap. Menurut Acuna & Rodriguez (2003), keberadaan relasi yang tinggi antara atribut yang memiliki data hilang dengan suatu atribut yang lain dapat menyebabkan teknik *Mode Imputation* tidak berguna atau bahkan berbahaya.

c) *Forward Imputation*

Ferrari *et al.* (2011) mengusulkan alternatif penanganan data hilang dengan tujuan mengubah suatu dataset tidak lengkap menjadi data lengkap dengan mempertimbangkan pola hubungan diseluruh peubah. Usulan ini didasarkan pada algoritma iterative yang menggunakan hasil analisis komponen utama Nonlinier pada data hilang dan imputasi yang dilakukan berurutan dengan metode *nearest neighbor*. Proses yang berurutan ini dimulai dari objek yang memiliki data hilang dengan jumlah terendah dan berakhir pada objek yang memiliki data hilang dengan jumlah tertinggi. Metode *Forward Imputation* menggunakan seluruh informasi yang terkandung dalam data, memperhitungkan nilai yang diamati pada peubah lain (bertentangan dengan metode *median imputation* dan *mode imputation*).

2.3 Spasial Outlier

Spasial *outlier* adalah suatu nilai yang menyimpang atau ekstrim pada suatu data, di mana nilai atribut non spasial secara signifikan berbeda dari nilai lainnya dalam kedekatan spasial atau tetangga spasialnya (Shekhar, Lu, dan Zhang, 2003). Dalam hal ini kedekatan spasial ditentukan berdasarkan metode k tetangga terdekat. Salah satu metode yang dapat digunakan untuk mendeteksi adanya *outlier* adalah spatial statistic Z test yang didefinisikan sebagai berikut:

$$Z_{hit} = \left| \frac{x_i - \mu_x}{\sigma_x} \right| > Z_{\alpha/2} \quad i = 1, 2, \dots, n$$

- x_i : Selisih antara nilai amatan dari lokasi ke- i dengan rata-rata nilai amatan lokasi yang dekat dengan lokasi ke- i
- μ_x : Nilai mean dari x_i
- σ_x : Standar deviasi dari x_i
- $Z_{\alpha/2}$: Nilai Z tabel untuk tingkat signifikan tertentu

Jika $Z_{hitung} > Z_{\alpha/2}$, maka nilai amatan dari lokasi ke- i dideteksi sebagai *outlier*.

2.4 Kriging

Metode *kriging* digunakan oleh G. Matheron pada tahun 1960-an, untuk menonjolkan metode khusus dalam *moving average* terbobot (*weighted moving average*) yang meminimalkan varians dari hasil estimasi. *Kriging* merupakan analisis data *geostatistika* yang digunakan untuk mengestimasi besarnya nilai yang mewakili suatu titik yang tidak tersampel berdasarkan titik-titik tersampel yang berada di sekitarnya dengan mempertimbangkan *korelasi spasial* yang ada dalam data tersebut. Metode *kriging* dibagi menjadi tiga jenis, yaitu *Simple Kriging*, *Ordinary Kriging*, dan *Kriging With a Trend*.

- Simple Kriging* Pada diasumsikan bahwa mean atau rata-rata sudah diketahui dan mempunyai nilai yang konstan. Metode ini dapat dikembangkan lagi, dimana data *spasial* yang akan diestimasi akan dipartisi menjadi beberapa bagian yang disebut metode *sequential kriging*.
- Ordinary Kriging* merupakan metode *kriging* yang paling sederhana pada *geostatistika*. Pada metode ini mempunyai asumsi bahwa rata-rata (mean) dari populasi tidak diketahui tetapi konstan. Data *spasial* yang digunakan dalam

metode *ordinary kriging* harus tidak mengandung trend. Selain itu, data yang digunakan juga tidak mengandung *outlier*.

- c. *Kriging with a Trend* atau yang biasa disebut dengan universal *kriging* merupakan metode *kriging* yang mempunyai kecenderungan trend dan merupakan bentuk umum dari simple *kriging*. Metode *kriging with a trend* ini digunakan untuk menangani masalah *kenonstasioneran* dari data yang diambil.

2.5 Robust Kriging

Pada data spasial seringkali ditemukan data yang mengandung pencilan (*outlier*) sehingga data tersebut nantinya tidak digunakan dalam analisis, ini dapat mengurangi informasi pada data tersebut. Oleh karena itu untuk mengatasi hal tersebut dibutuhkan suatu metode untuk mengatasi data yang mengandung pencilan (*outlier*).

Metode *Robust kriging* merupakan pengembangan dari metode *ordinary kriging* yang mentransformasi bobot *semivariogram* klasik menjadi *semivariogram* yang *Robust* terhadap *outlier* dalam data. Berbeda dengan perhitungan *semivariogram* klasik, untuk mengakomodir adanya *outlier* dalam data spasial maka digunakan *variogram Robust* yang didefinisikan sebagai berikut:

$$2\gamma(h) = \frac{\left\{ \frac{1}{N(h)} \sum_{i=1}^{N(h)} |Z(s_i) - Z(s_i + h)|^{1/2} \right\}^4}{\left(0.457 + \frac{0.494}{N(h)} \right)}$$

Sehingga model *semivariogram Robust* dapat dirumuskan sebagai berikut:

$$\gamma(h) = \frac{\frac{1}{2} \left\{ \frac{1}{N(h)} \sum_{i=1}^{N(h)} |Z(s_i) - Z(s_i + h)|^2 \right\}^4}{\left(0.457 + \frac{0.494}{N(h)} \right)}$$

- $2\gamma(h)$: nilai *variogram* dengan jarak h
- $\gamma(h)$: nilai *semivariogram* dengan jarak h
- s_i : lokasi ke-i
- $Z(s_i)$: nilai pengamatan pada lokasi ke-i
- $s_i + h$: lokasi ke-i dengan menambahkan jarak h
- $Z(s_i + h)$: nilai pengamatan pada lokasi ke-i dengan penambahan jarak h
- h : rentang jarak antara dua lokasi
- $N(h)$: banyak pasangan data yang memiliki jarak h

2.6 Pemodelan *Semivariogram*

2.6.1. *Semivariogram* Eksperimental

Semivariogram Eksperimental atau biasa disebut *semivariogram cloud* merupakan *semivariogram* yang dihitung dari data pengukuran dan kemudian diplotkan sebagai fungsi jarak (Gaetan dan Guyon 2010). Misalkan $Z(s_i)$ merupakan hasil pada pengukuran lokasi ke- i , sedangkan $s_i = (x_i, y_i)$ merupakan vektor koordinat x, y sehingga *semivariogram* eksperimental dapat ditulis sebagai berikut:

$$\gamma_{ij} = \frac{1}{2} E[Z(s_i) - Z(s_i + h)]^2$$

Semua pasangan jarak yang mungkin $(s_i, s_i + h)$ untuk $i = 1, 2, \dots, n$ kemudian diplotkan sebagai fungsi jarak sebagai berikut:

$$|h| = |s_i - (s_i + h)| = |(x_i - (x_i + h))^2 + (y_i - (y_i + h))^2|^{\frac{1}{2}}$$

Kelemahan dari *semivariogram* eksperimental adalah sulit untuk melihat pola apabila perhitungan melibatkan ribuan titik, sehingga untuk mengatasi hal tersebut dilakukan pengelompokkan berdasarkan kesamaan jarak yang disebut proses *binning*. *Semivariogram* eksperimental dirumuskan sebagai berikut:

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

$\gamma(h)$: nilai *semivriogram* dengan jarak h

$Z(s_i)$: nilai pengamatan di titik s_i

$Z(s_i + h)$: nilai pengamatan di titik $s_i + h$

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

2.6.2. *Semivariogram Robust*

Apabila data spasial terdapat pencilan maka akan menghasilkan nilai prediksi yang kurang tepat, oleh karena itu dibutuhkan model *semivariogram* yang bisa mengakomodasi adanya pencilan. *Semivariogram robust* adalah *semivariogram* yang dapat mengakomodir adanya pencilan pada data spasial. *Semivariogram robust* dapat ditulis sebagai berikut:

$$\gamma(h) = \frac{\frac{1}{2} \left\{ \frac{1}{N(h)} \sum_{i=1}^{N(h)} |Z(s_i) - Z(s_i + h)|^2 \right\}}{\left(0,457 + \frac{0,494}{N(h)} \right)}$$

- $\gamma(h)$: nilai semivriogram dengan jarak h
- $Z(s_i)$: nilai pengamatan di titik s_i
- $Z(s_i + h)$: nilai pengamatan di titik $s_i + h$
- $N(h)$: banyak pasangan data yang memiliki jarak h

2.6.3 Semivariogram Teoritis

Variogram merupakan karakteristik dari *korelasi* spasial, yaitu korelasi antara dua buah data spasial yang disebabkan oleh bertambahnya jarak dari data yang akan diambil sehingga menyebabkan *korelasi* antara dua buah data spasial menjadi berkurang ataupun tidak memiliki korelasi. Simbol dari *variogram* adalah yang didefinisikan sebagai berikut:

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

- h : Jarak antara dua lokasi
- $Z(s)$: Pengamatan pada lokasi
- $Z(s + h)$: Nilai pengamatan pada lokasi dengan penambahan jarak h

Semivariogram merupakan suatu perangkat statistik untuk menggambarkan, memodelkan, dan menjelaskan korelasi spasial antara variabel random $Z(s)$ dan $Z(s + h)$ sehingga *Semivariogram* didefinisikan sebagai berikut:

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

Dalam menghitung *semivariogram teoritis* diperlukan parameter-parameter berikut:

1. Range

Range adalah jarak *semivariogram* mencapai sebuah masa stabil atau sill. *Range* menunjukkan jarak terjauh saat tidak ada lagi korelasi spasial.

2. Sill ($c_0 + c$)

Sill adalah saat dimana nilai semivariogram cenderung mencapai nilai stabil, sehingga *sill* merupakan nilai tertinggi dari *semivariogram*. *Sill* terdiri dari dua bagian yaitu, *nugget effect* dan *Partial sill* (*sill* yang dikurangi *nugget effect*).

3. Nugget Effect (c_0)

Nugget effect adalah lompatan vertikal dari nilai 0 pada titik asal menuju nilai dari *semivariogram* pada jarak pemisah yang terkecil.

Setelah diperoleh komponen-komponen semivariogram dari perhitungan nilai *semivariogram eksperimental*, dilakukan analisis lebih lanjut untuk mengganti *semivariogram eksperimental* menjadi *semivariogram teoritis*. Ada beberapa model *semivariogram teoritis* yang dapat digunakan sebagai pembanding dengan *semivariogram eksperimental* yaitu sebagai berikut:

1. Model Spherical

$$\gamma(h) = \begin{cases} [c_0 + c] \left[\left(\frac{3h}{2a} \right) - 0.5 \left(\frac{h}{a} \right)^3 \right] & \text{untuk } h \leq a \\ c_0 + c & \text{untuk } h > a \end{cases}$$

2. Model Eksponensial

$$\gamma(h) = [c_0 + c] \left[1 - \exp\left(-\frac{3h}{a}\right) \right]$$

3. Model Gaussian

$$\gamma(h) = [c_0 + c] \left[1 - \exp\left(-\frac{3h^2}{a^2}\right) \right]$$

2.7 Root Mean Square Error

Kegunaan dari *Root Mean Square Error (RMSE)* adalah untuk membandingkan akurasi antara dua atau lebih model. Semakin kecil nilai *RMSE* yang dihasilkan suatu model maka semakin akurat model tersebut. Rumus *RMSE* sebagai berikut:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Z}(s_i) - Z(s_i))^2}{n}}$$

$Z(s_i)$: nilai aktual pada peubah s pada lokasi ke- i

$\hat{Z}(s_i)$: nilai prediksi pada peubah s pada lokasi ke- i

n : banyak pengamatan

Metode *kriging* dikatakan memiliki akurasi pendugaan yang baik apabila memiliki nilai *RMSE* yang kecil (Oliver dan Webster, 2015).

2.8 Bobot Kriging

Hasil dari analisis struktural dengan nilai *RMSE* terkecil akan digunakan untuk menentukan bobot dari lokasi-lokasi yang tersampel terhadap lokasi yang akan diestimasi. Besarnya nilai bobot pada lokasi tersampel dapat dituliskan dalam notasi matriks sebagai berikut:

$$C \cdot W = D$$

$$\begin{bmatrix} C_{Z_{ij}} & \dots & C_{Z_{in}} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{Z_{nj}} & \dots & C_{Z_{nn}} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} W_1 \\ \vdots \\ W_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{Z_{ko}} \\ \vdots \\ C_{Z_{no}} \\ 1 \end{bmatrix}$$

Matriks C merupakan matriks *nonsingular*, yaitu sebuah matriks bujur sangkar dan $\det(C) \neq 0$ sehingga diperoleh persamaan sebagai berikut:

$$w = C^{-1} \cdot D$$

Estimasi *kriging* $\hat{Z}(s)$ merupakan kombinasi linear, yaitu penjumlahan dari hasil kali anggota himpunan pasangan berurutan dari variable $Z(s_i)$ yang tersampel dalam data *spasial* yang dirumuskan sebagai berikut:

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

$\hat{Z}(s)$: nilai estimasi pada suatu lokasi

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

$Z(s_i)$: nilai pengamatan pada lokasi ke- i

n : banyak sampel yang digunakan dalam mengestimasi