



TESIS-TE142599

VISUALISASI VARIABEL CUACA MENGGUNAKAN METODE KRIGING BERBASIS BIG DATA

BAYU SETYATMOKO
NRP 2213 206 712

Dosen Pembimbing :
Mochammad Hariadi, ST., M.Sc, Ph.D.
Dr. I Ketut Eddy Purnama, ST., MT.

PROGRAM MAGISTER
BIDANG KEAHLIAN TELEMATIKA
KONSENTRASI CHIEF INFORMATION OFFICER (CIO)
JURUSAN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI INDUSTRI
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016



THESIS-TE142599

VISUALIZATION OF WEATHER VARIABLE USING KRIGING METHOD BASE ON BIG DATA

BAYU SETYATMOKO
NRP 2213 206 712

SUPERVISORS

Mochammad Hariadi, ST., M.Sc, Ph.D.
Dr. I Ketut Eddy Purnama, ST., MT.

MASTER PROGRAM
TELEMATICS
CHIEF INFORMATION OFFICER (CIO)
DEPARTMENT OF ELECTRICAL ENGINEERING
FACULTY OF INDUSTRIAL TECHNOLOGY
INSTITUTE TECHNOLOGY OF SEPULUH NOPEMBER
SURABAYA
2016

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Teknik (MT)

di
Institut Teknologi Sepuluh Nopember

oleh :

BAYU SETYATMOKO
NRP. 2213206712

Tanggal Ujian : 7 Januari 2016
Periode Wisuda : Maret 2016

Disetujui oleh :


1. **Mochammad Hariadi, S.T., M.Sc, Ph.D.** (Pembimbing I)
NIP. 196912091997031002


2. **Dr. I Ketut Eddy Purnama, ST., MT.** (Pembimbing II)
NIP. 196907031995121001


3. **Dr. Eko Mulyanto Yuniarno, ST., MT.** (Penguji)
NIP. 196806011995121009


4. **Dr. Supeno Mardi Susiki Nugroho, ST., MT.** (Penguji)
NIP. 197003131995121001



Direktur Program Pascasarjana,


Prof. Ir. Djauhar Manfaat, M.Sc, Ph.D.
NIP. 196012021987011001

VISUALISASI VARIABEL CUACA MENGGUNAKAN METODE KRIGING BERBASIS BIG DATA

Nama Mahasiswa : Bayu Setyatmoko
NRP : 2213206712
Dosen Pembimbing : 1. Mochammad Hariadi, ST., M.Sc, Ph.D.
2. Dr. I Ketut Eddy Purnama, ST., MT.

ABSTRAK

Indonesia adalah salah satu negara di dunia yang rentan mengalami bencana alam. Bencana banjir bahkan terjadi hampir diseluruh wilayah Indonesia. Salah satu upaya "mitigasi bencana" yang perlu dilakukan adalah peningkatan kualitas pemantauan kondisi cuaca menggunakan visualisasi geospasial variabel cuaca yang dihasilkan oleh peralatan observasi cuaca. Peralatan observasi cuaca yang sering digunakan di bandara udara dan pelabuhan laut di seluruh dunia adalah *Automatic Weather Station* (AWS).

Beberapa permasalahan pengolahan data AWS adalah lokasi data tersebar diseluruh wilayah Indonesia yang luas (107 lokasi), periode update data setiap 15 menit (*velocity*) dan bentuk data semi terstruktur (*variety*). Sehingga proses pemindahan seluruh data ke server datawarehouse sulit untuk dilaksanakan karena terbatasnya kapasitas bandwidth jaringan komputer dan kapasitas penyimpanan data.

Proses pengolahan data semi terstruktur AWS menggunakan aplikasi *Big Data* yang berjalan pada sistem data terdistribusi *Hadoop Distributed File System* (HDFS) tanpa perlu memindahkan semua data AWS tapi hanya data yang dibutuhkan. Sedangkan visualisasi variabel cuaca menggunakan analisa data spasial dengan metode interpolasi *kriging* yang dapat digunakan sebagai alat bantu pemantauan kondisi cuaca saat ini.

Analisa *cross validation* menunjukkan nilai error yang kecil yaitu *Mean Error* (ME) sebesar -0,0281 dan *Mean Squared Prediction Error* (MSPE) sebesar 1,337 pada hasil visualisasi geospasial yang baik.

Kata kunci : *big data, hadoop, kriging, visualisasi geospasial.*

VISUALIZATION OF WEATHER VARIABLE USING KRIGING METHOD BASE ON BIG DATA

Student Name : Bayu Setyatmoko

Student Identify Number : 2213206712

Supervisor : 1. Moch. Hariadi, ST., M.Sc, Ph.D.

2. Dr. I Ketut Eddy Purnama, ST., MT.

ABSTRACT

Indonesia is one country in the world that is prone to natural disasters. Flood disaster even occurs almost throughout the territory of Indonesia. One effort "*disaster mitigation*" that needs to be done is to increase the quality monitoring of geospatial visualization of weather conditions using weather variables generated by the weather observation equipment. Weather observation equipment that are often used in airports and seaports around the world is the Automatic Weather Station (AWS).

Some of the problems of AWS data processing is the location of data scattered throughout the vast territory of Indonesia (107 location), the period of data updates every 15 minutes (velocity) and semi-structured data form (variety). So that the process of moving all the data to the server datawarehouse difficult to implement because of the limited capacity of the computer network bandwidth and data storage capacity.

Semi structured AWS data processing using Big Data applications running on distributed data system Hadoop Distributed File System (HDFS) without the need to move all the data AWS but only the required data. While the visualization of weather variables using spatial data analysis with kriging interpolation method that can be used as a tool for monitoring current weather conditions.

Cross validation analysis shows that a small error values are Mean Error (ME) of -0,0281 and Mean Squared Prediction Error (MSPE) of 1,337 on the outcome of good geospatial visualization.

Kata kunci : *big data, geospatial visualization, hadoop, kriging.*

KATA PENGANTAR

Rasa syukur dan terima kasih kami panjatkan kepada Alloh SWT karena berkat kasihNYA penulis diberikan kekuatan dan kemudahan menyelesaikan Tesis yang berjudul “Visualisasi Variable Cuaca Menggunakan Metode *Kriging* Berbasis *Big Data*”.

Terima kasih kami sampaikan kepada Kementrian Kominfo atas bantuan beasiswa yang telah diberikan hingga terselesaikannya Studi S2 kami di Jurusan Teknik Elektro ITS dengan bidang konsentrasi CIO Telematika.

Terima kasih kami sampaikan kepada Pemerintah Kabupaten Pati dan Kepala BKD Pemerintah Kabupaten Pati yang telah memberikan kesempatan pada penulis untuk melanjutkan pendidikan pada Program Magister Jurusan Teknik Elektro ITS

Terima kasih kami sampaikan kepada Bapak Mochammad Hariadi, Bapak I Ketut Eddy Purnama, Bapak Eko Mulyanto Yuniarno, Bapak Supeno Mardi Susiki Nugroho, Bapak Achmad Zaini dan seluruh bapak dosen di lingkungan Jurusan Teknik Elektro ITS atas bimbingan dan tambahan ilmu pengetahuan selama kami menempuh kuliah. Terima kasih kami sampaikan kepada Bapak Djoko Purwanto (ketua program studi pasca sarjana jurusan teknik elektro) dan Bapak Gamantyo Hendratoro (dosen wali) atas segala bantuan selama penulis menempuh pendidikan. Terima kasih kami sampaikan kepada segenap Civitas Akademika Jurusan Teknik Elektro ITS atas segala bantuan selama penulis menempuh pendidikan.

Terima kasih kami sampaikan kepada Bapak Lalu Muhamad Jaelani atas bimbingan dan tambahan ilmu mengenai *Geomatics Engineering*.

Terima kasih setulusnya kami sampaikan kepada almarhum ayahanda tercinta dan ibunda tercinta yang telah mendidik penulis dari kecil hingga kami dewasa. Terima kasih kami sampaikan kepada seluruh keluarga besar di Surabaya yang banyak memberikan bantuan dan dukungan kepada penulis selama kami tinggal di Surabaya. Terima kasih kami sampaikan kepada ayah mertua, ibu

mertua dan seluruh keluarga di Pati yang telah banyak membantu menjaga anak-anak selama penulis menyelesaikan pendidikan di Surabaya.

Tidak lupa terima kasih sebesar-besarnya kami sampaikan kepada istri tercinta (Nur Kartikasari) dan kedua putraku tersayang (Syarifa Nurrusyifa dan Arrisa Maya Syafiya) atas kesabaran dan pengertiannya tidak bisa mendampingi secara penuh saat penulis harus menyelesaikan pendidikan di Surabaya.

Terima kasih tak terhingga kami panjatkan kepada Allah SWT dan Nabi Muhammad SAW. Jika tanpa kasihNYA dan cinta Rasulullah sesungguhnya kami adalah orang merugi (tanpa ilmu pengetahuan). Dalam setiap kesulitan memahami sebuah ilmu pengetahuan baru akan membuka peluang kemudahan-kemudahan memahami ilmu pengetahuan yang lebih luas lagi.

Penulis menyadari bahwa Tesis kami ini masih jauh dari kesempurnaan dan masih ada kekurangan dalam banyak hal. Besar harapan penulis mendapatkan masukan dan saran dari pembaca untuk menyempurnakan penelitian kami.

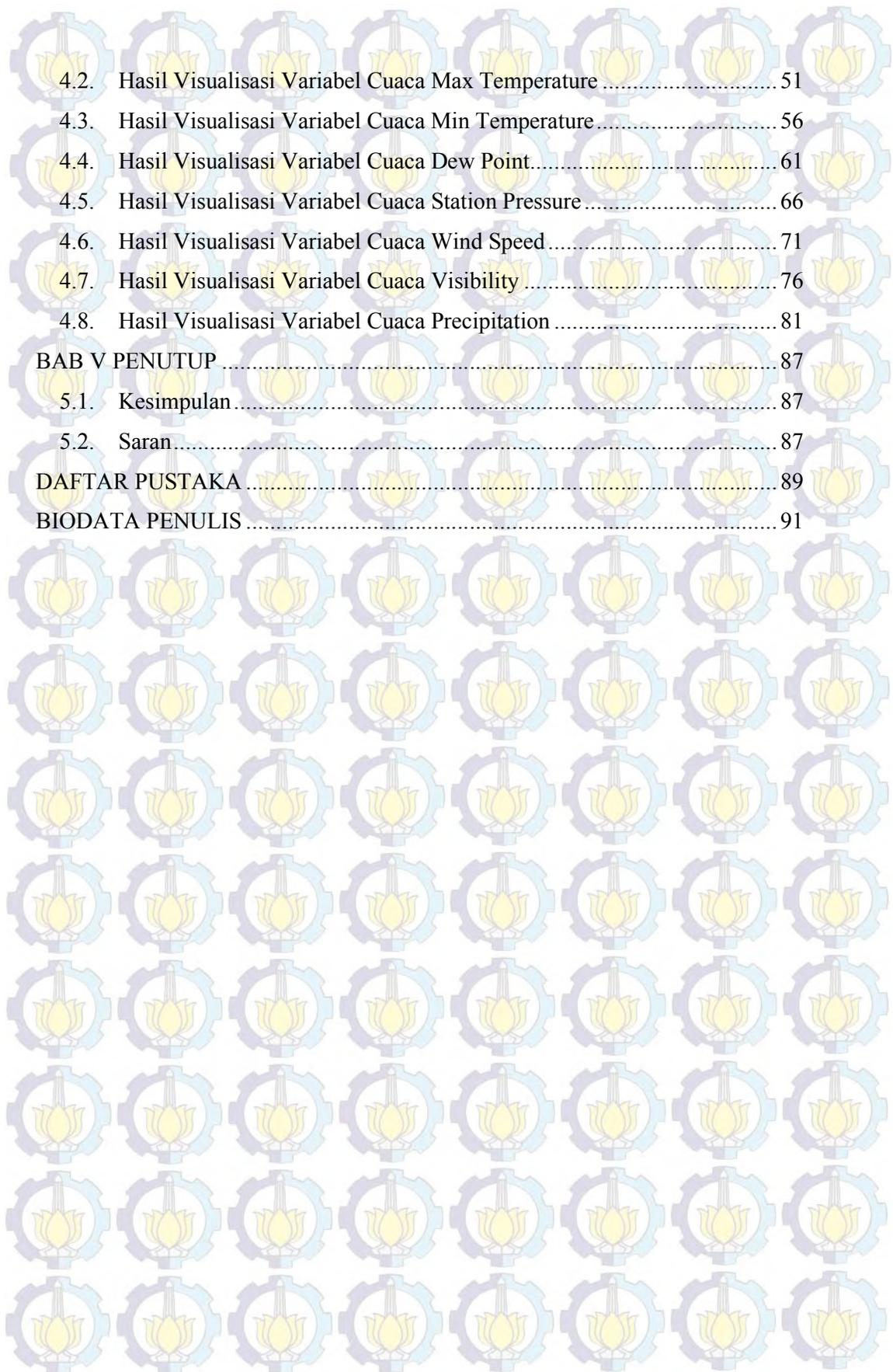
Penulis mohon maaf apabila terdapat kesalahan dan kekurangan dalam penyusunan tesis kami. Harapan penulis, semoga tesis kami bermanfaat bagi semua pihak dan dapat memberikan kontribusi bagi penelitian di bidang cuaca, penelitian di bidang visualisasi geospasial dan penelitian di bidang *Big Data*.

Surabaya, Januari 2016

Penulis

DAFTAR ISI

SURAT PERNYATAAN KEASLIAN TESIS	i
LEMBAR PENGESAHAN	iii
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan dan Manfaat Penelitian	3
1.4.1 Tujuan Penelitian	3
1.4.2 Manfaat Penelitian	3
1.5 Sistematika Penulisan	4
BAB II KAJIAN PUSTAKA	7
2.1 Konsep Cuaca	7
2.2 Konsep Analisa Data Spasial (Analisa Geostatistik)	16
2.3 Konsep Interpolasi Kriging	19
2.4 Konsep Bigdata dan Mapreduce	22
BAB III METODOLOGI PENELITIAN	31
3.1. Gambaran Umum Penelitan	31
3.2. Tahap Persiapan	32
3.3. Tahap Visualisasi Variabel Cuacadengan Metode Kriging	34
3.4. Tahap Evaluasi Cross Validation	38
BAB IV HASIL DAN PEMBAHASAN	41
4.1. Hasil Penelitian Tahap Persiapan	41



4.2.	Hasil Visualisasi Variabel Cuaca Max Temperature	51
4.3.	Hasil Visualisasi Variabel Cuaca Min Temperature	56
4.4.	Hasil Visualisasi Variabel Cuaca Dew Point.....	61
4.5.	Hasil Visualisasi Variabel Cuaca Station Pressure	66
4.6.	Hasil Visualisasi Variabel Cuaca Wind Speed	71
4.7.	Hasil Visualisasi Variabel Cuaca Visibility	76
4.8.	Hasil Visualisasi Variabel Cuaca Precipitation	81
BAB V PENUTUP		87
5.1.	Kesimpulan	87
5.2.	Saran	87
DAFTAR PUSTAKA		89
BIODATA PENULIS		91

DAFTAR GAMBAR

Gambar 2.1. Ketinggian Lapisan Atmosfir Bumi.....	9
Gambar 2.2. Perputaran Bumi Mengelilingi Matahari	10
Gambar 2.3. AlurPerubahan dan Perpindahan Air.....	12
Gambar 2.4. Peralatan Pengamatan Cuaca	13
Gambar 2.5. Peralatan Automatic Weather Station Portable.....	14
Gambar 2.6. Satelit Remote Sensing untuk Pengamatan Cuaca	15
Gambar 2.7. Data Geostatistik Ketinggian Tanah di Wisconsin USA.....	17
Gambar 2.8. Data Lattice Perumahan di Wisconsin USA	18
Gambar 2.9. Data Point Pattern Kejahatan di Los Angeles USA	18
Gambar 2.10. Grafik Variogram.....	20
Gambar 2.11. Parameter Model Fit Variogram.....	20
Gambar 2.12. Karakteristik Dasar Big Data	23
Gambar 2.13. Konsep Apache Hadoop Stack	25
Gambar 2.14. Alur Proses Model Pemrograman Mapreduce	26
Gambar 2.15. Arsitektur Hadoop Master Slave Pada Sistem Terdistribusi.....	27
Gambar 2.16. Komponen Apache Spark Stack.....	27
Gambar 2.17. Model Integrasi Apache Spark dengan YARN	28
Gambar 2.18. Alternatif Aplikasi Bigdata Berbasis Open Source	29
Gambar 3.1. Blok Diagram Tahapan Proses Penelitian.....	31
Gambar 3.2. Flowchart Visualisasi Variabel Cuaca dengan Metode Kriging	34
Gambar 3.3. Data Grid Seluruh Wilayah Indonesia	35
Gambar 3.4. Visualisasi Geospasial dengan Metode Ordinary Kriging.....	37
Gambar 4.1. Peta Lokasi AWS di Seluruh Wilayah Indonesia.....	45
Gambar 4.2. Histogram Variabel MAX, MIN dan DEWP	46
Gambar 4.3. Histogram Variable STP, MXSPD, SIB dan PRCP	47
Gambar 4.4. Grafik Model Fit Variogram Max Temperature.....	52
Gambar 4.5. Grafik Scatterplot3d Variabel Max Temperature.....	53
Gambar 4.6. Visualisasi Data Observasi Max temperature	54
Gambar 4.7. Visualisasi Variabel Cuaca Max Temperature.....	54
Gambar 4.8. Visualisasi Nilai Residual Max Temperature.....	55

Gambar 4.9. Grafik Model Fit Variogram Min Temperature	57
Gambar 4.10. Grafik Scatterplot3d Variabel Min Temperature.....	58
Gambar 4.11. Visualisasi Data Observasi Min temperature	59
Gambar 4.12. Visualisasi Variabel Cuaca Min Temperature.....	59
Gambar 4.13. Visualisasi Nilai Residual Min Temperature	60
Gambar 4.14. Grafik Model Fit Variogram Dew Point.....	62
Gambar 4.15. Grafik Scatterplot3d Variabel Dew Point	63
Gambar 4.16. Visualisasi Data Observasi Dew Point.....	64
Gambar 4.17. Visualisasi Variabel Cuaca Dew Point	64
Gambar 4.18. Visualisasi Nilai Residual Dew Point.....	65
Gambar 4.19. Grafik Model Fit Variogram Station Pressure.....	67
Gambar 4.20. Grafik Scatterplot3d Variabel Station Pressure.....	68
Gambar 4.21. Visualisasi Data Observasi Station Pressure	69
Gambar 4.22. Visualisasi Variabel Cuaca Station Pressure.....	69
Gambar 4.23. Visualisasi Nilai Residual Station Pressure	70
Gambar 4.24. Grafik Model Fit Variogram Wind Speed.....	72
Gambar 4.25. Grafik Scatterplot3d Variabel Wind Speed.....	73
Gambar 4.26. Visualisasi Data Observasi Wind Speed.....	74
Gambar 4.27. Visualisasi Variabel Cuaca Wind Speed.....	74
Gambar 4.28. Visualisasi Nilai Residual Wind Speed	75
Gambar 4.29. Grafik Model Fit Variogram Visibility.....	77
Gambar 4.30. Grafik Scatterplot3d Variabel Visibility.....	78
Gambar 4.31. Visualisasi Data Observasi Visibility	79
Gambar 4.32. Visualisasi Variabel Cuaca Visibility	79
Gambar 4.33. Visualisasi Nilai Residual Visibility.....	80
Gambar 4.34. Grafik Model Fit Variogram Precipitation.....	82
Gambar 4.35. Grafik Scatterplot3d Variabel Precipitation.....	83
Gambar 4.36. Visualisasi Data Observasi Precipitation	84
Gambar 4.37. Visualisasi Variabel Cuaca Precipitation.....	84
Gambar 4.38. Visualisasi Nilai Residual Precipitation.....	85

DAFTAR TABEL

Tabel 2.1. Nama Lapisan Atmosfir Bumi	8
Tabel 2.2. Daftar Nama Bentuk Awan.....	11
Tabel 4.1. Nama dan Lokasi AWS di Indonesia	42
Tabel 4.2. Rangkuman Data AWS Seluruh Variabel Cuaca.....	45
Tabel 4.3. Data Observasi AWS Indonesia Tanggal 15-08-2015	48
Tabel 4.4. Nilai Variogram Variabel Max Temperature.....	51
Tabel 4.5. Rangkuman Error dan Korelasi Visualisasi Max Temperature	55
Tabel 4.6. Nilai Variogram Variabel Min Temperature	56
Tabel 4.7. Rangkuman Error dan Korelasi Visualisasi Min Temperature.....	60
Tabel 4.8. Nilai Variogram Variabel Dew Point.....	61
Tabel 4.9. Rangkuman Error dan Korelasi Visualisasi Dew Point	65
Tabel 4.10. Nilai Variogram Variabel Station Pressure	66
Tabel 4.11. Rangkuman Error dan Korelasi Visualisasi Station Pressure	70
Tabel 4.12. Nilai Variogram Variabel Wind Speed	71
Tabel 4.13. Rangkuman Error dan Korelasi Visualisasi Wind Speed.....	75
Tabel 4.14. Nilai Variogram Variabel Visibility.....	76
Tabel 4.15. Rangkuman Error dan Korelasi Visualisasi Visibility	80
Tabel 4.16. Nilai Variogram Variabel Precipitation.....	81
Tabel 4.17. Rangkuman Error dan Korelasi Visualisasi Precipitation	85

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kondisi negara Indonesia yang berpotensi mengalami bencana alam harus diikuti dengan kesiapan sumber daya manusia dalam melakukan upaya mitigasi bencana. Mitigasi bencana adalah serangkaian upaya untuk mengurangi resiko bencana, baik melalui pembangunan fisik maupun penyadaran dan peningkatan kemampuan menghadapi ancaman bencana (Undang-undang No 24 Tahun 2007 dan PP No 21 Tahun 2008).

Dalam rangka melakukan upaya mitigasi bencana pemerintah Indonesia banyak mengeluarkan regulasi dan mendirikan lembaga yang menjalankan regulasi tersebut. Salah satu regulasi tersebut adalah Undang-Undang No 21 Tahun 2013 Tentang Keantariksaan yang menyatakan bahwa LAPAN (Lembaga Penerbangan dan Antariksa) berkewajiban menyediakan data pengideraan jarak jauh dengan tutupan awan minimal dan bebas awan setiap tahunnya untuk seluruh wilayah Indonesia kedalam Sistem Jaringan Data Spasial Nasional (pasal 20), dan berkewajiban pula mengupayakan terjadinya alih teknologi keantariksaan (pasal 25).

Sedangkan institusi yang diberi tugas untuk mengelola informasi geospasial dan mengelola Jaringan Data Spasial Nasional (JDSN) adalah Badan Informasi Geospasial (BIG) sebagai pengganti Badan Koordinasi Survei dan Pemetaan Nasional (Bakorsurtanal). Bahkan dalam Perpres No.85 Tahun 2007 yang diperkuat oleh Undang-Undang No.4 Tahun 2011 mewajibkan Pemerintah Daerah membangun JDSN dan menggunakan informasi geospasial yang akurat dalam pengambilan keputusan. Yang tidak kalah pentingnya adalah lembaga khusus yang ditugaskan mengelola bidang meteorologi, klimatologi, kualitas udara dan geofisika yang dikenal dengan nama Badan Meteorologi Klimatologi dan Geofisika (BMKG).

Sayangnya belum banyak instansi pemerintah daerah yang tertarik memanfaatkan informasi yang telah dihasilkan oleh lembaga seperti LAPAN, BIG

dan BMKG. Padahal tidak sedikit anggaran pemerintah yang harus dikeluarkan bagi lembaga negara tersebut untuk menghasilkan informasi cuaca di Indonesia. Selama ini hanya beberapa institusi pemerintah saja seperti bandara udara dan pelabuhan laut yang memanfaatkan informasi cuaca tersebut. Kurangnya antusias dari pemerintah daerah memanfaatkan informasi cuaca dikarenakan kurangnya sosialisasi dan proses alih teknologi ke pemerintah daerah. Sebagai contoh adalah pemanfaatan data cuaca BMKG pada proses visualisasi variabel cuaca sebagai alat bantu pemantauan kondisi cuaca.

Beberapa permasalahan pengolahan data *Automatic Weather Station* (AWS) adalah lokasi data tersebar di seluruh wilayah Indonesia yang luas, data berbentuk *text compressed* (txt.gz) atau disebut data semi terstruktur (*variety*) dan periode akuisisi data setiap 15 menit sekali (*velocity*). Sehingga pendekatan memindahkan data ke server datawarehouse untuk menganalisa data sudah tidak dapat dilakukan lagi. Sebab kapasitas storage dan bandwidth jaringan yang terbatas sudah tidak sanggup lagi untuk menampung data tersebut.

Jadi diperlukan sebuah pendekatan baru, bukan lagi data yang datang kelokasi server komputasi tapi justru proses komputasi yang mendatangi lokasi data. Sebagai solusinya digunakan aplikasi Big Data Apache Spark yang berjalan diatas sistem data terdistribusi *Hadoop Distributed File System* (HDFS) untuk memfilter data yang dibutuhkan saja tanpa harus memindahkan seluruh datanya. Sedangkan visualisasi variabel cuaca digunakan aplikasi statistik Rstudio untuk melakukan proses analisa data spasial dengan metode interpolasi kriging. Hasil akhir berupa visualisasi variabel cuaca yang mudah dipahami diharapkan dapat membantu proses pemantauan kondisi cuaca saat ini.

1.2 Perumusan Masalah

Pada penelitian ini memiliki dua permasalahan yaitu :

1. Presentasi informasi cuaca tidak mudah dipahami oleh semua orang untuk memantau kondisi cuaca saat ini.
2. Data AWS berbentuk *file* semi terstruktur (*text compressed file*).

1.3 Batasan Masalah

Untuk tidak meluasnya pokok pembahasan maka penelitian ini dititikberatkan pada :

1. Data observasi cuaca yang digunakan adalah data AWS seluruh dunia yang dikoordinir oleh lembaga World Meteorological Organization (WMO). Dataset observasi cuaca dikelola dan dipelihara oleh National Oceanic and Atmospheric Administration (NOAA) dengan alamat website di <http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD>
2. Metode yang dipergunakan untuk melakukan visualisasi geospasial adalah metode interpolasi kriging. Adapun variabel cuaca yang divisualisasikan yaitu *Maximum Temperature (MAX)*, *Minimum Temperature (MIN)*, *Dew Point (DEWP)*, *Station Pressure (STP)*, *Maximum Wind Speed (MXSPD)*, *Visibility (SIB)* dan *Precipitation (PRCP)*. Lokasi AWS sebagai obyek penelitian adalah AWS di seluruh Indonesia.
3. Framework komputasi Bigdata yang dipergunakan adalah Apache HadoopStack, yang meliputi HDFS (Hadoop Distributed File System), Apache Spark dan RStudio.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Penelitian ini bertujuan untuk menghasilkan visualisasi geospasial variabel cuaca yang mudah dipahami untuk memantau kondisi cuaca yang terjadi saat ini.

1.4.2 Manfaat Penelitian

Hasil penelitian dapat dimanfaatkan oleh instansi pemerintah maupun swasta untuk memantau kondisi cuaca saat ini berdasarkan hasil observasi peralatan Automatic Weather Station (AWS).

1.5 Sistematika Penulisan

Sistematika penulisan penelitian ini terdiri atas 5 bab, setiap bab saling berhubungan satu dengan lainnya sesuai dengan urutan permasalahan yang akan dibahas. Garis besar susunan penulisannya adalah sebagai berikut:

BAB I PENDAHULUAN

Berisi latar belakang, perumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, serta sistematika pembahasan penelitian.

BAB II KAJIAN PUSTAKA DAN DASAR TEORI

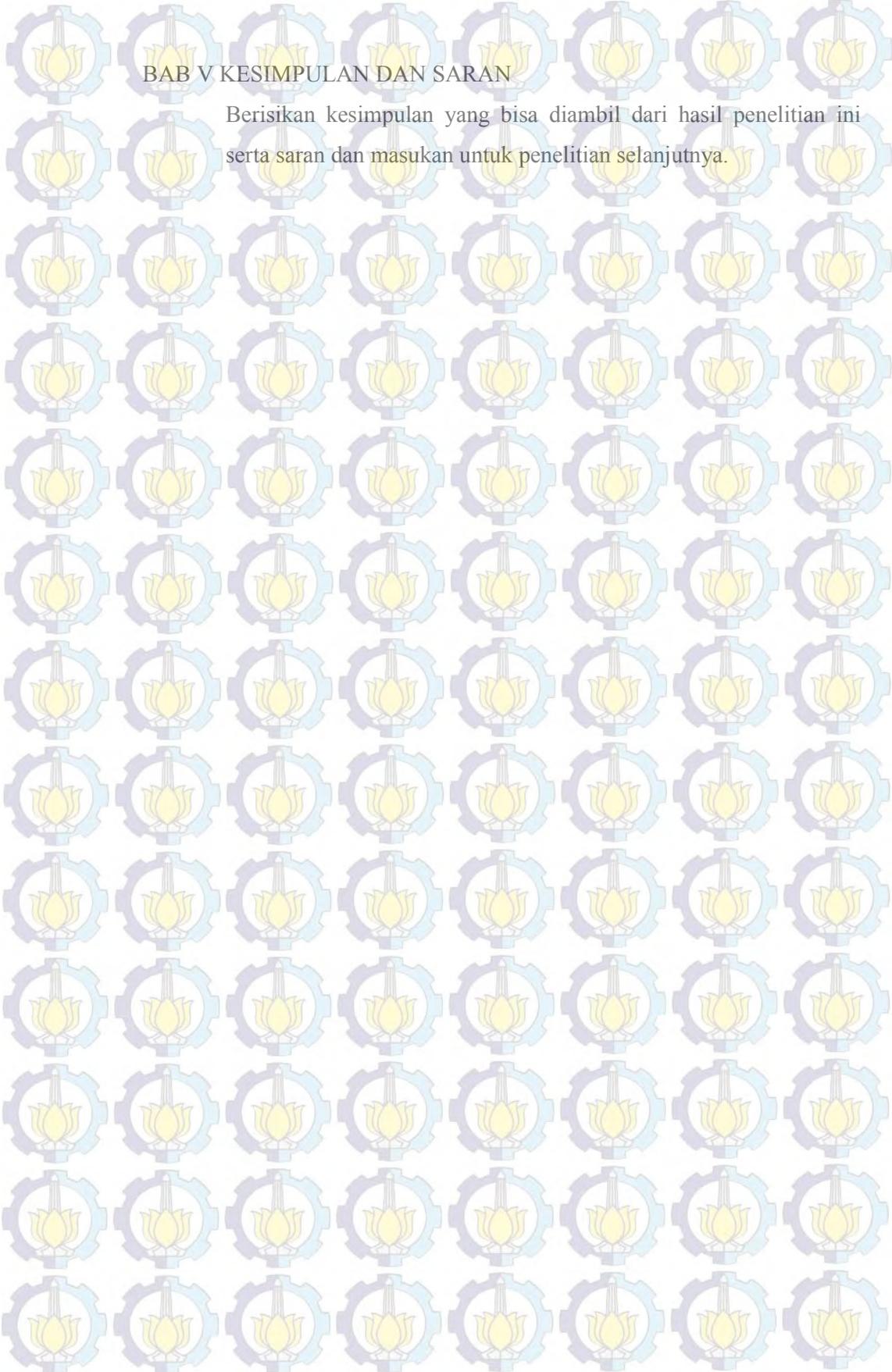
Berisi tentang kajian teoritis mengenai konsep cuaca, konsep geostatistik, konsep interpolasi *kriging*, konsep *Big Data* dan Mapreduce. Disamping itu melakukan studi terhadap hasil-hasil penelitian sebelumnya serta literatur pendukung lainnya.

BAB III METODOLOGI PENELITIAN

Membahas metodologi penelitian yang digunakan, diawali dengan tahapan *pre-processing* yaitu melakukan query data tanpa harus mengunduh datanya terlebih dahulu. Tahapan berikutnya adalah proses visualisasi spasial pada variabel cuaca (tujuh variabel cuaca) menggunakan metode *kriging*. Tahapan terakhir penelitian adalah melakukan evaluasi *cross validation* pada tiap variabel cuaca.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Dalam bab ini dijelaskan hasil penelitian dan analisa dari hasil penelitian. Pembahasan diawali dengan data yang diperoleh pada tahapan *pre-processing* dan bagaimana proses mendapatkan data. Pembahasan berikutnya adalah menampilkan visualisasi spasial pada variabel cuaca menggunakan metode *kriging* beserta hasil evaluasi *cross validation* pada tiap variabel cuaca.



BAB V KESIMPULAN DAN SARAN

Berisikan kesimpulan yang bisa diambil dari hasil penelitian ini serta saran dan masukan untuk penelitian selanjutnya.

BAB II

KAJIAN PUSTAKA

2.1 Konsep Cuaca

Definisi cuaca adalah kondisi atmosfer pada waktu tertentu dan lokasi tertentu. Sedangkan definisi iklim adalah cuaca rata-rata yang khas pada daerah tertentu (biasanya ditentukan berdasarkan pada data cuaca dalam kurun waktu 30 tahun atau lebih). Beberapa parameter penting dari cuaca sebagai data observasi dalam penelitian cuaca adalah *Temperature, Precipitation, Station Pressure, Dew Point Temperature, Wind Speed, Visibility*, dan sebagainya. Sedangkan beberapa kondisi cuaca yang seringkali menjadi obyek penelitian adalah *Rain, Snow, Hail, Mist, Thunder, Tornado*, dan sebagainya.

Topik mengenai cuaca sangat dekat dengan kehidupan kita sehari-hari. Banyak kejadian kecelakaan transportasi (kapal laut dan pesawat), bencana alam (banjir dan tsunami) dan kegagalan produksi tanaman (padi dan palawija) terkait dengan kondisi cuaca yang buruk. Maka diperlukan kewaspadaan dan kesiagaan untuk menghadapi kondisi cuaca saat ini yang cenderung tidak menentu akibat terjadi perubahan iklim diseluruh dunia (*global warming*). Salah satu upaya pemantauan terhadap kondisi cuaca adalah memanfaatkan visualisasi spasial kondisi cuaca berdasarkan hasil pengamatan peralatan observasi cuaca.

Kondisi cuaca di bumi sangat dipengaruhi oleh matahari. Hampir seluruh energi yang ada di bumi adalah berasal dari matahari yang disalurkan ke bumi dalam bentuk radiasi cahaya matahari. Sekitar 46% cahaya matahari dipantulkan oleh benda-benda sehingga kita dapat melihat bentuk benda dengan jelas. Sekitar 40% adalah infra red yang dapat kita rasakan dalam bentuk panas. Sedangkan sisanya adalah ultra violet yang dapat membuat kulit kita terbakar. Dalam matahari sendiri sering terjadi badai matahari yang sebagian energi panasnya juga terkirim ke seluruh alam semesta termasuk ke bumi kita. Namun berbagai bahaya yang ada di alam semesta tidak sampai memasuki bumi, hal ini disebabkan adanya perlindungan atmosfer yang menyelimuti bumi.

Separuh energi radiasi matahari diserap oleh daratan dan lautan, sebagian diserap oleh atmosfer dan sistem awan, sedangkan sisanya ada yang dipantulkan kembali ke alam semesta. Pada dasarnya keseluruhan energi di bumi ini adalah konstan, sebab apa yang diserap bumi di siang hari akan dipantulkan kembali pada malam harinya. Atmosfer bumilah yang mengatur siklus perputaran energi di bumi. Sayangnya akibat ulah manusia keteraturan sistem atmosfer menjadi terganggu hingga terjadi *global warming* dan perubahan iklim hampir di seluruh dunia.

Begitu strategis fungsi atmosfer bumi kita, namun kita tidak menyadari sebenarnya ada berapa lapisan atmosfer yang melindungi bumi. Nama lapisan atmosfer bumi beserta isinya dapat dilihat pada Tabel 2.1.

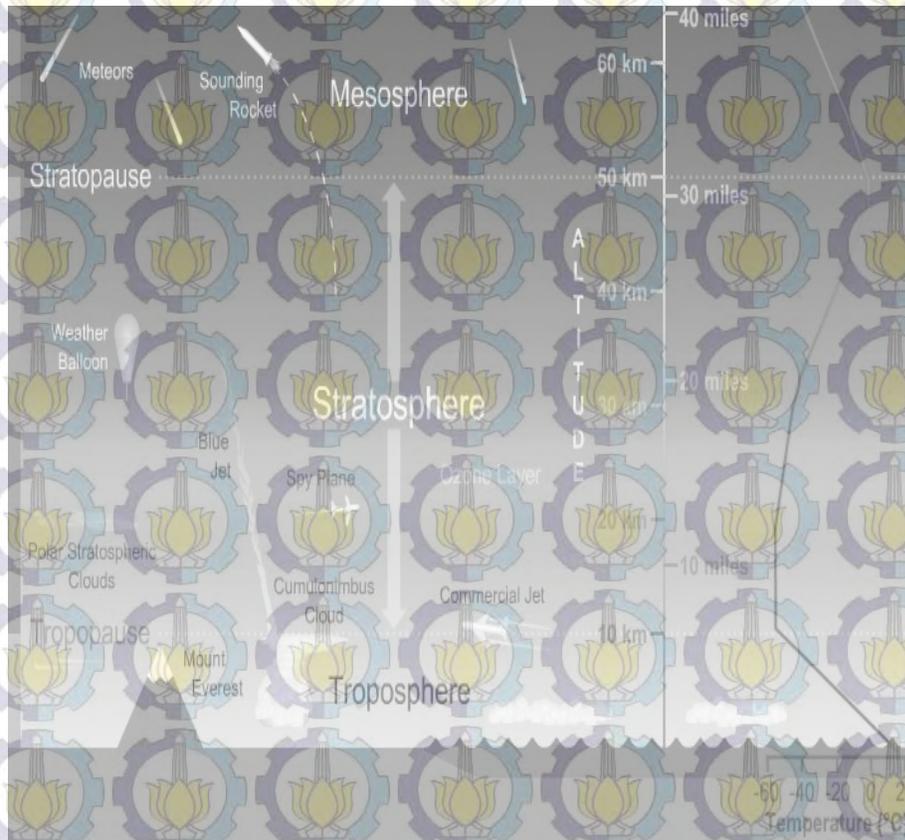
Tabel 2.1. Nama Lapisan Atmosfir Bumi

Nama Lapisan	Ketinggian	Isi Lapisan
Exosphere	Over 300 Miles	Satelites
Thermosphere	50 – 300 Miles	Auroras
Mesosphere	30 – 50 Miles	Spacecraft & Meteors
Stratosphere	6 – 30 Miles	Ozone Layer
Troposphere	0 – 80 Miles	Most Our Weather

Sumber : Yorke, 2010

Manusia hanya bisa bertahan hidup dan bernapas pada lapisan atmosfer setinggi sekitar 2 mil saja, dimana pada lapisan inilah proses pergantian cuaca terjadi. Diatas lapisan troposphere ada lapisan tropopause yang ketinggiannya bervariasi di beberapa daerah permukaan bumi. Ketinggian tropopause di daerah khatulistiwa sekitar 12 mil sedangkan di daerah kutub ketinggiannya sekitar 4 mil. Dibawah lapisan tropopause terjadi penurunan suhu seiring dengan kenaikan ketinggian, namun kondisi berbalik ketika ketinggian sudah melampaui diatas lapisan tropopause. Hal ini terjadi akibat penyerapan radiasi ultraviolet oleh lapisan ozon. Sebagian besar awan terbentuk pada ketinggian dibawah 20000 ft, tapi pada kondisi tertentu kelompok awan cumulonimbus dapat menembus hingga

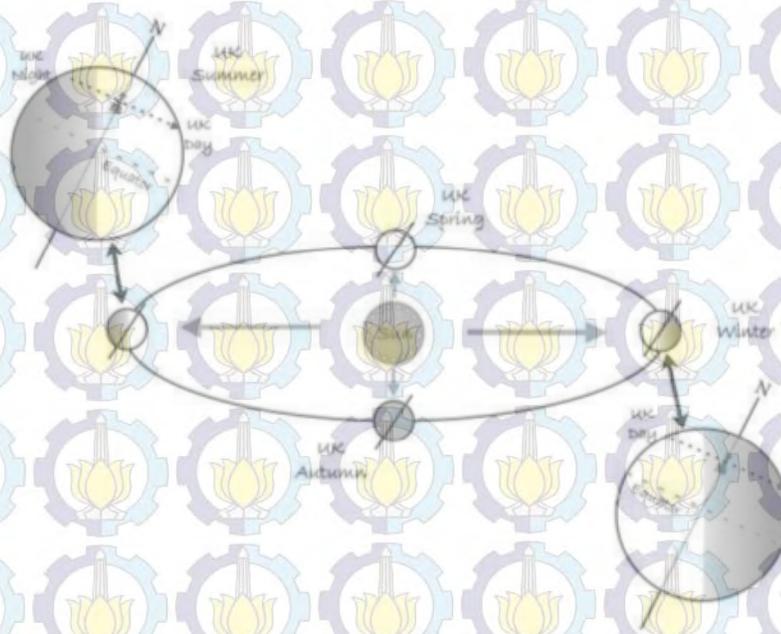
lapisan tropopause. Ketinggian setiap lapisan atmosfer tempat terbentuknya kumpulan awan dapat digambarkan pada Gambar 2.1.



Gambar 2.1. Ketinggian Lapisan Atmosfir Bumi
 Sumber : The University Corporation for Atmospheric Research (UCAR)

Selain energi radiasi matahari yang sampai ke bumi dapat mempengaruhi kondisi cuaca di bumi, pergerakan rotasi perputaran bumi mengelilingi matahari ikut memberikan kontribusi kondisi cuaca di bumi. Bumi mengelilingi matahari dalam bentuk lintasan elips yang membutuhkan waktu 365 hari (1 tahun) untuk menyelesaikan satu putaran penuh. Bumi juga berputar pada porosnya selama 24 jam (1 hari) sehingga kita dapat merasakan siang dan malam. Namun sumbu bumi dimiringkan sebesar 23,5° selama rotasi mengelilingi bumi, hal ini menempatkan matahari diatas lintang utara selama musim panas, sedangkan pada daerah lintang

selatan akan mengalami musim dingin. Peristiwa perputaran bumi mengelilingi matahari inilah yang menyebabkan terjadinya pergantian musim di bumi. Hal ini dapat dilihat pada Gambar 2.2.



Gambar 2.2. Perputaran Bumi Mengelilingi Matahari
Sumber : Yorke, 2010

Faktor penting selain matahari yang mempengaruhi kondisi cuaca adalah lautan. Laut juga memiliki aliran seperti aliran udara tapi karena sifat air lebih padat daripada udara, maka membutuhkan waktu yang cukup lama untuk memanaskan dan mendinginkan (pergerakan air laut juga lebih lambat dibanding dengan pergerakan udara). Pergerakan air laut yang terbesar dan mempengaruhi kondisi cuaca di bumi adalah *North Atlantic Gulf Stream*, dengan kecepatan pergerakan air sebesar 30 milyar galon per detik.

Uap air yang terbentuk dari air laut yang dingin hanya bisa terjadi saat kondisi udara diatas air laut adalah kering. Sedangkan pada permukaan air laut yang lebih hangat akan menghasilkan uap air yang lebih banyak, yang nantinya akan menghasilkan awan dan hujan pada atmosfer bumi. Uap air yang dihasilkan

adalah air murni yang tidak mengandung garam, namun karena faktor polusi yang dibuat oleh manusia menjadikan uap air menjadi lebih asam. Sekitar 3% air di bumi berbentuk es murni, gletser dan salju. Sekitar 7% air tersimpan di dalam perut bumi dalam bentuk air tanah.

Saat iklim udara panas, lautan seluas 2 mil persegi dapat merubahsekitar 2 juta galon air menjadi uap air. Uap air adalah bahan dasar terbentuknya awan dan hujan, uap air sendiri adalah bentuk gas dari air (terjadi saat suhu dan tekanan udara tertentu). Kelembaban (*humidity*) adalah jumlah uap air (*water vapour*) yang terkandung dalam udara. Sedangkan batas atas suhu saat uap air berubah menjadi kembali menjadi tetesan air (*water droplets*) disebut titik embun (*dew point*). Berapa jumlah uap air yang dapat ditampung oleh udara dan berapa titik embun sangat ditentukan oleh suhu udara (*temperature*).

Awan (*cloud*) terbentuk dari sekumpulan uap air saat suhu udara lebih dingin dibandingkan kumpulan uap air yang cenderung lebih hangat. Setelah cukup banyak uap air yang terkumpul maka pada titik tertentu akan berubah menjadi tetesan air. Jika jumlah tetesan air semakin banyak maka akan terjadi hujan (*rain*). Penamaan awan sesuai dengan bentuknya, sebagai contoh awan *cumulonimbus*, terdiri dari dua kata yakni kata *cumulus* yang artinya berbentuk bulat yang saling menumpuk dan kata *nimbus* yang artinya merujuk pada hujan. Adapun daftar nama bentuk awan dapat dilihat pada Tabel 2.2.

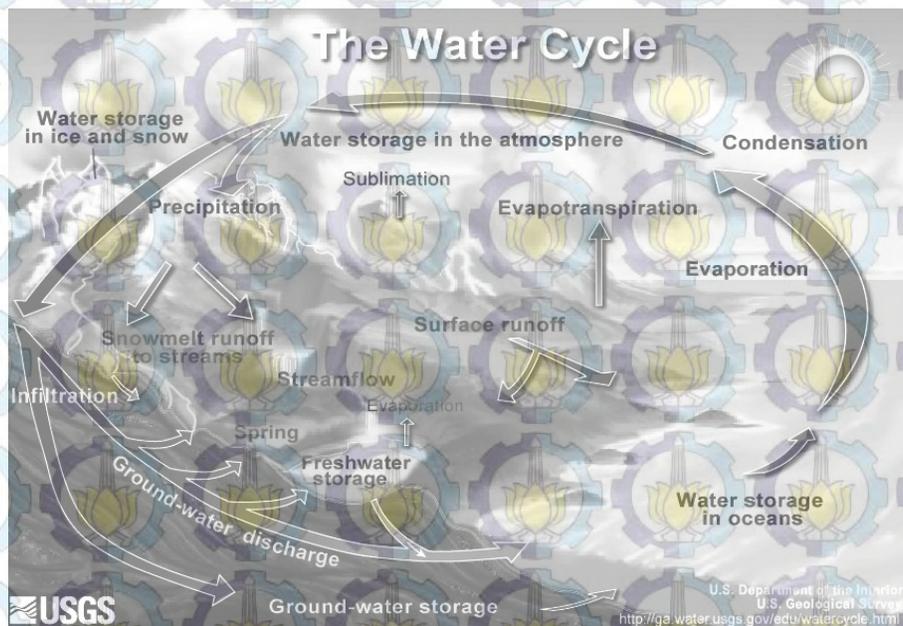
Tabel 2.2. Daftar Nama Bentuk Awan

High Clouds	Medium Clouds	Low Clouds
0 Cirrus	3 Altocumulus	6 Stratocumulus
1 Cirrocumulus	4 Altostratus	7 Stratus
2 Cirrostratus	5 Nimbostratus	8 Cumulus
		9 Cumulonimbus

Sumber : Yorke, 2010

Variable penting lainnya dalam penelitian cuaca adalah *Precipitation* (curah hujan). Dalam meteorologi definisi *precipitation* adalah hasil dari kondensasi uap air yang jatuh ke tanah akibat gaya gravitasi bumi. Beberapa bentuk dari *Precipitation* adalah hujan gerimis kecil (*drizzle*), hujan deras (*rain*), hujan salju (*sleet*), salju (*snow*) dan hujan es (*hail*). *Precipitation* terjadi ketika atmosfer menjadi jenuh dengan uap air yang kemudian uap air tersebut mengembun dan jatuh ke tanah. Maka kabut tipis (*mist*) dan kabut tebal (*fog*) bukan termasuk *precipitation* melainkan hanya kumpulan uap air yang pekat namun belum mencapai proses pengembunan.

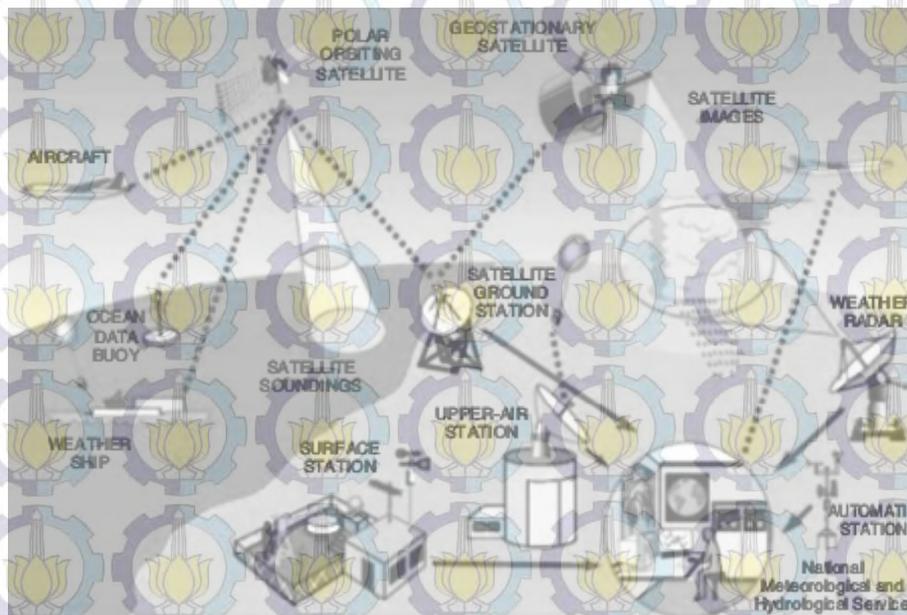
Sesungguhnya jumlah air di bumi ini relatif konstan, hanya berubah dari satu bentuk ke bentuk lainnya, dan berpindah dari satu tempat ke tempat lainnya. Rangkuman perubahan dan perpindahan air (*water cycle*) dapat dilihat pada Gambar 2.3.



Gambar 2.3. Alur Perubahan dan Perpindahan Air

Sumber : U.S. Geological Survey (USGS)

Proses penelitian cuaca tidak bisa dilepaskan dari sistem observasi yang baik. Peralatan observasi cuaca yang baik akan menghasilkan data-data cuaca yang akurat yang sangat dibutuhkan pada proses penelitian cuaca. Para peneliti cuaca akan saling berbagi pakai data observasi cuaca dengan para peneliti cuaca diseluruh dunia melalui organisasi World Meteorological Organization (WMO). Semakin banyak jenis peralatan survey yang digunakan dengan tingkat keakuratan yang tinggi akan menghasilkan penelitian cuaca yang baik pula. Beberapa jenis peralatan observasi cuaca yang masih dipergunakan saat ini dapat dilihat pada gambar 2.4.



Gambar 2.4. Peralatan Pengamatan Cuaca

Sumber : Inness, 2013

Pada sistem transportasi laut (kapal) dan transportasi udara (pesawat) pasti dilengkapi dengan berbagai peralatan pengamatan cuaca. Peralatan pengamatan cuaca selain dipasang di bandara udara / pelabuhan laut, juga dipasang pada alat transportasi (kapal atau pesawat). Hal ini diperlukan untuk mengantisipasi kondisi cuaca yang dapat berubah sewaktu-waktu selama perjalanan. Selain itu di

bandara udara / pelabuhan laut juga dipasang peralatan *Automatic Weather Station* (AWS) yang sangat bermanfaat untuk menentukan apakah aman untuk berlayar bagi kapal laut / apakah aman untuk mengudara bagi pesawat terbang. Selain itu hampir seluruh bandara udara / pelabuhan laut di dunia saling berbagi pakai data-data AWS melalui organisasi dunia WMO. Saat ini bahkan sudah ada peralatan AWS portable yang dapat digunakan untuk keperluan pribadi atau hobby.

Peralatan pengamatan cuaca AWS termasuk peralatan yang sudah ada sejak tahun 1930 yang umum digunakan oleh para peneliti cuaca. Seiring dengan perkembangan teknologi peralatan AWS semakin canggih dan bentuknya semakin kecil. Salah satu model AWS yang ada saat ini dapat dilihat pada gambar 2.5.

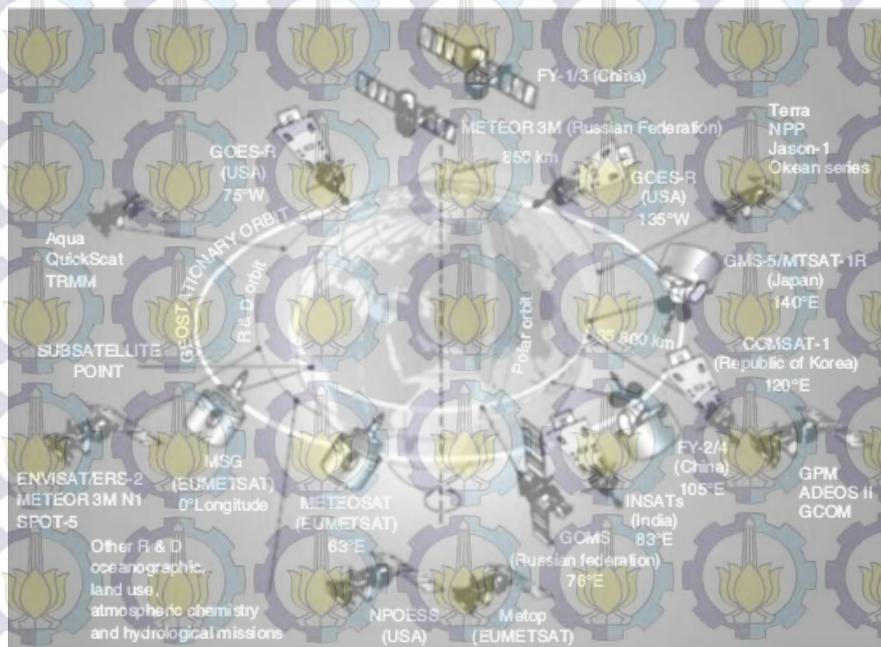


Gambar 2.5. Peralatan Automatic Weather Station Portable
Sumber : Wikipedia

Seiring dengan kemajuan teknologi satelit maka mulai banyak digunakan satelit pengideraan jarak jauh (*remote sensing*) sebagai alternatif peralatan untuk pengamatan cuaca. Teknologi *remote sensing* menghasilkan jumlah data yang luar

biasa besar (*big data*) yang membutuhkan pemrosesan data yang efisien serta pengarsipan data yang baik. Pada tiap satelit memiliki bermacam sensor yang hasil pengukurannya harus diterjemahkan dengan algoritma tertentu agar data mentah satelit (data biner) dapat dirubah menjadi data representatif spasial yang lebih bermakna.

Saat pertama kali *US National Oceanic and Atmospheric Administration* (NOAA) meluncurkan satelit TIROS pada tahun 1960, mulailah pusat pengolahan *Numerical Weather Prediction* (NWP) seperti *European Centre for Medium-Range Weather Forecasts* (ECMWF) menggunakan data satelit *remote sensing* untuk melakukan pengolahan model matematika. Sejak saat itu data satelit mulai digunakan lembaga riset di seluruh dunia. Mengingat biaya riset satelit sangat mahal, maka perlu kolaborasi antar negara untuk pengembangannya. Beberapa satelit *remote sensing* yang sudah diluncurkan hingga saat ini dapat dilihat pada gambar 2.6.



Gambar 2.6. Satelit Remote Sensing untuk Pengamatan Cuaca
Sumber : Inness, 2013

2.2 Konsep Analisa Data Spasial (Analisa Geostatistik)

Pengertian data spasial (sering disebut data geospasial) adalah informasi mengenai sebuah obyek fisik yang direpresentasikan dengan nilai numerik dalam sebuah koordinat sistem geografi. Data spasial sangat sering kita jumpai dalam kehidupan kita sehari-hari seperti informasi kondisi cuaca dalam berita koran, televisi dan media online. Secara umum data spasial merepresentasikan lokasi, ukuran, dan bentuk obyek pada planet bumi seperti gedung, danau, gunung atau lingkungan kota. Informasi yang terkandung dalam data spasial akan diproses lebih lanjut oleh software *Geographic Information System (GIS)* untuk keperluan visualisasi, manipulasi dan analisa data.

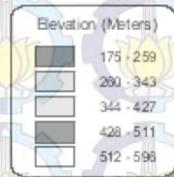
Sedangkan data *spatio temporal* adalah sekumpulan data spasial yang terjadi dalam periode waktu tertentu. Pada proses analisa sebuah data *spatio temporal* selain memperhatikan lokasi sebuah data juga memperhatikan kapan kegiatan observasi dilakukan. Hal inilah yang menyebabkan analisa data *spatio temporal* tidak mudah untuk dilakukan.

Salah satu contoh visualisasi data spasial adalah pembuatan sebuah peta yang berisi gambar geografi lokasi wilayah administratif seperti negara, propinsi, kota, kabupaten, dan sebagainya. Selain peta dasar yang umum kita gunakan sebagai petunjuk lokasi sebuah wilayah dikenal juga peta tematik. Pengertian peta tematik adalah peta khusus yang didesain untuk menggambarkan tema tertentu yang terkait dengan lokasi geografis. Peta tematik dapat menggambarkan secara fisik kondisi sosial, politik, ekonomi, pertanian, atau semua aspek yang terkait dengan wilayah kota, propinsi, negara, kepulauan atau benua. Sebagai contoh peta tematik adalah visualisasi variabel cuaca di Indonesia.

Kondisi cuaca pada lokasi tertentu dipermukaan bumi dapat dipantau melalui data kuantitatif variabel cuaca seperti data *temperature, dew point, pressure, wind speed, visibility, precipitation* dan sebagainya. Metode presentasi data dalam bentuk tabulasi data numerik variabel cuaca akan sulit dibaca dan sulit dipahami. Sedangkan metode presentasi data dalam bentuk visualisasi geospasial akan lebih mudah dibaca dan dipahami.

Ada beberapa tipe data yang digunakan dalam proses analisa data spasial (disebut juga analisa geostatistik) adalah sebagai berikut :

1. **Geostatistical data.** Sekumpulan data numerik mengenai obyek tertentu yang terkait dengan koordinat geografi di permukaan bumi yang digambarkan secara kontinyu dalam bentuk gradasi warna (menggunakan interpolasi geospasial).
2. **Lattice data.** Sekumpulan data numerik mengenai obyek tertentu yang terkait dengan lokasi wilayah administratif sebuah negara yang digambarkan dalam bentuk petak-petak (*data grid*).
3. **Point pattern data.** Sekumpulan data mengenai kejadian tertentu yang terkait dengan koordinat geografi di permukaan bumi yang digambarkan dengan sebuah penanda (*point marker*).



Gambar 2.7. Data Geostatistik Ketinggian Tanah di Wisconsin USA

Pada Gambar 2.7 menggambarkan data geostatistik ketinggian tanah dari permukaan air laut di *Wisconsin*. Data lokasi disekitarnya digambarkan dalam bentuk gradasi warna sesuai ketinggian tanah dari permukaan air laut. Sumber data berasal dari data satelit Remote Sensing.

2.3 Konsep Interpolasi Kriging

Pada analisa geostatistik dikenal *spatial prediction* yaitu proses estimasi data pada lokasi yang tidak diobservasi berdasarkan data pada lokasi yang telah diobservasi. Perbedaan *spatial prediction* (disebut juga *spatial interpolation*) dengan pendekatan model statistik biasa yaitu pada keterkaitan antara data dengan lokasi geografis. Pada teknik interpolasi spasial dilakukan perhitungan estimasi data pada lokasi tertentu berdasarkan *weighthed average* dari data-data terdekat. Pembobotan (*weighthed*) berdasarkan kriteria statistik atau deterministik.

Ada beberapa teknik interpolasi yang umum digunakan yaitu metode *Inverse Distance Weigthing* (IDW) dan metode *kriging* (*simple kriging*, *ordinary kriging*, *universal kriging*, dan sebagainya). Pada IDW proses pembobotan hanya bergantung kepada jarak lokasi data yang akan diprediksi, sedangkan pada metode kriging selain mengukur jarak lokasi diantara data juga juga memperhitungkan *spatial autocorelation* antar data pada keseluruhan lokasi.

Pada *simple kriging* diasumsikan korelasi spasial disemua lokasi adalah mempunyai nilai rata-rata yang sama (*stationary randomize function*) sedangkan pada *ordinary kriging* nilai rata-rata korelasi spasial yang berbeda-beda disemua lokasi (*non-stationary randomize function*). Pada *ordinary kriging* dapat dibatasi distance yang akan dihitung serta nilai maksimum atau minimum dari data. Terlihat metode *ordinary kriging* lebih realistik dibanding metode simple kriging.

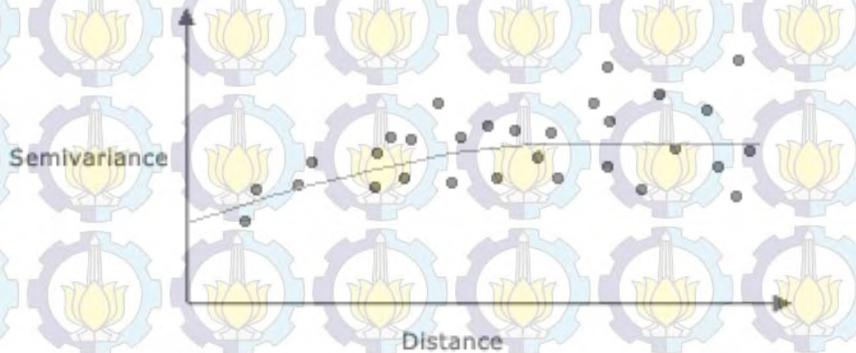
Proses *spatial autocorelation* dalam analisa geostatistik menggunakan metode *variogram* (disebut juga *semi variogram*) yaitu melakukan proses estimasi nilai *sample data* pada seluruh lokasi data. Berdasarkan metode *Matheron* (Oliver, 2015) untuk menghitung nilai variogram digunakan Persamaan 2.1.

$$\gamma(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{z(x_i) - z(x_i - h)\}^2 \dots \dots \dots (2.1)$$

dengan :

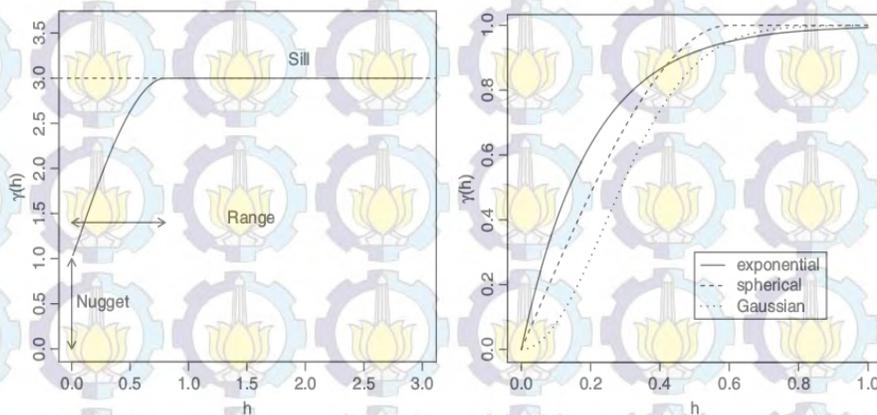
$z(x_i)$ dan $z(x_i - h)$ adalah nilai observasi dari semua lokasi
 h adalah *lag* atau *distance*

Hubungan antara distance dan variance setiap titik data dapat digambarkan menggunakan grafik variogram pada Gambar 2.10.



Gambar 2.10. Grafik Variogram

Proses *model fit semivariogram* (pembuatan kurva garis yang mewakili semua titik data) yang baik terkait dengan pemilihan parameter model fit (misal model, sill, range, nugget, dll) yang tepat. Kualitas pembuatan model fit akan sangat menentukan keakuratan hasil prediksi metode kriging. Penentuan parameter model fit variogram dapat dilakukan secara manual atau secara otomatis dengan metode autofit variogram yang akan menguji semua kemungkinan parameter dengan nilai *Sum Squared Error* (SSE) terendah. Variabel *model fit variogram* dapat dilihat pada Gambar 2.11.



Gambar 2.11. Parameter Model Fit Variogram

Setelah parameter *model fit variogram* yang terbaik ditentukan maka langkah selanjutnya adalah membuat *output data grid* yang meliputi seluruh data yang telah diobservasi. Proses estimasi data menggunakan metode kriging hanya akan mengolah data yang ada di dalam output data grid. Tiap titik dalam grid akan ditandai mana lokasi data hasil observasi dan lokasi data yang belum di observasi yang akan diprediksi nilai datanya. Hasil *model fit variogram*, output data grid dan input data point kemudian akan diolah menggunakan metode kriging untuk mendapatkan nilai pada lokasi data yang belum dilakukan observasi.

Metode *Ordinary Kriging* adalah metode *spatial interpolation* untuk mengestimasi nilai pada lokasi yang belum diobservasi berdasarkan nilai *weighed average* data *input* hasil observasi. Metode *Ordinary Kriging* menggunakan asumsi bahwa variasi nilai adalah acak dan tergantung kepada lokasi geospasial (Oliver, 2015), yang dapat dihitung menggunakan Persamaan 2.2.

$$Z(x_0) = \sum_{i=1}^N \lambda_i z(x_i) \dots \dots \dots (2.2)$$

dengan :

x_i adalah lokasi *data sample*

λ adalah nilai pembobotan (*weight*) pada setiap lokasi *data sample*

Untuk memastikan hasil estimasi tidak mengalami bias (overfitting), maka nilai total dari *weight sum* adalah 1, yang dapat dihitung dengan Persamaan 2.3.

$$\sum_{i=1}^N \lambda_i = 1 \dots \dots \dots (2.3)$$

Nilai pembobotan (*weight*) dihitung menggunakan Persamaan 2.4.

$$[C] = [A]^{-1}[B] \dots \dots \dots (2.4)$$

dengan :

A adalah matrix semivariogram diantara *data point*

B adalah vektor *semivariogram* diantara *data point* yang diprediksi

C adalah nilai pembobotan (*weight*)

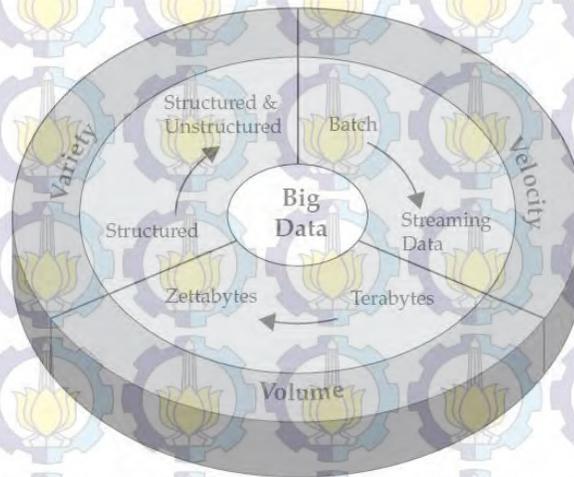
2.4 Konsep Bigdata dan Mapreduce

Bigdata adalah sebuah istilah yang mewakili kondisi *dataset* yang begitu besar dan kompleks sehingga menjadi sulit untuk melakukan proses analisa terhadap data jika dilakukan dengan pendekatan proses analisa datawarehouse dan database tradisional yang ada saat ini. Namun *Bigdata* bukan sekedar ukuran data yang besar, tetapi timbulnya permasalahan analisa data (*Big Problem*) pada data berukuran besar dan peluang besar (*Big Opportunity*) bagi kemajuan teknologi *Predictive Analytics*.

Kemajuan teknologi *Big Data* sangat bermanfaat bagi para *Data Scientist* dalam menyelesaikan permasalahan analisa data berukuran besar. Kemajuan teknologi *Big Data* sangat dipengaruhi oleh kemajuan ilmu pengetahuan di bidang matematika, statistik dan pemrograman komputer.

Beberapa karakteristik penting sebuah *Big Data* yaitu :

1. **Volume.** Ukuran data *Big Data* yang relatif besar. Berapa ukuran minimal sebuah *Big Data* akan berubah seiring dengan perkembangan teknologi storage terbaru. Saat ini satuan terkecil sebuah Bigdata sudah mencapai satuan Tera. Bukan ukuran datanya yang penting tapi proses analisa pada data berukuran besar membutuhkan teknologi analisa data yang lebih maju dibandingkan teknologi analisa data yang tersedia saat ini.
2. **Variety.** Jenis data yang bervariasi mulai data terstruktur (data pada sebuah database), data semi terstruktur (data text, file excel, xml, dsb) dan data tidak terstruktur (data gambar, data video, dsb). Semua jenis data tersebut harus dapat diolah secara langsung pada sebuah sistem berbasis Bigdata.
3. **Velocity.** Kecepatan proses akuisisi data yang begitu cepat mengakibatkan perubahan ukuran data yang semakin besar pula. Beberapa sistem tidak bisa memakai proses batch (menyimpan data dahulu baru dianalisa) melainkan sudah harus segera diproses secara realtime (saat data baru diterima saat itu pula dianalisa). Perkembangan teknologi *social media* dan teknologi satelit *remote sensing* membutuhkan proses analisa data besar secara cepat dan akurat.



Gambar 2.12. Karakteristik Dasar Big Data

Pada Gambar 2.12 digambarkan karakteristik dasar *Big Data* yang dimiliki oleh sekumpulan data yang cenderung mengalami ledakan data akibat kemajuan teknologi informasi dan komunikasi. Kondisi ledakan data yang terjadi saat ini adalah adanya fenomena dimana penyumbang data terbesar adalah berasal dari user bukan dari vendor penyedia layanannya. Sebagai contohnya adalah fenomena media sosial yang menjamur seperti facebook, twitter, dsb. Maka pendekatan untuk memindahkan data yang begitu besar ke server datawarehouse (untuk dianalisa) sudah tidak bisa dilakukan lagi, karena adanya keterbatasan kapasitas storage dan bandwidth jaringan komputer. Jadi diperlukan pendekatan baru yaitu menjalankan proses komputasi di lokasi server tempat data disimpan. Maka munculah beberapa framework komputasi *Big Data*, yang paling banyak digunakan saat ini adalah framework komputasi Apache Hadoop.

Apache hadoop adalah sebuah opensource framework untuk menyimpandan memproses dataset berukuran besar pada komputer biasa (commodity hardware) atau pada sebuah computer cluster yang besar (sebuah cluster terbesar milik google saat ini sudah memiliki sekitar 4000 komputer). Jadi Apache Hadoop memiliki beberapa keunggulan dalam hal *scalability*, *distributed*, *data intensive computing* dan *fault tolerant*.

Apache Hadoop memiliki beberapa komponen dasar sebagai berikut :

1. **Hadoop Common.** Komponen dasar yang berisi *library* dan *utility* penting yang sangat dibutuhkan oleh hadoop. Berisi perintah dasar sistem operasi seperti *ls (list file and directory)*, *rm -r (hapus file or directory)*, dsb.

Hadoop mengadopsi banyak perintah-perintah sistem operasi linux, perbedaannya linux dibuat dengan bahasa pemrograman c sedangkan hadoop dibuat menggunakan bahasa pemrograman java.

2. **Hadoop Distributed File System (HDFS).** Komponen *file system* (sama dengan *file system* pada *operating system* seperti *ntfs*, *ext2*, dsb) yang bertanggung jawab pada proses baca dan tulis file. Perbedaan dibandingkan dengan *file system* sistem operasi adalah kemampuan untuk melakukan proses duplikasi file ke beberapa sistem operasi HDFS lainnya (sangat diperlukan pada sistem server yang terdistribusi). Adanya server HDFS Master (berfungsi seperti DNS pada sistem internet) akan memberitahukan lokasi server HDFS slave (berfungsi sebagai lokasi redundancy file dan sebagai *worker* proses komputasi). Perbedaan pokok lainnya terletak pada ukuran blok terkecilnya bisa dikonfigurasi berukuran besar misal 100MB (pada sistem operasi windows biasanya blok terkecilnya sebesar 512kb). Sehingga proses pengolahan data pada data-data besar dapat dilakukan dengan cepat. Fasilitas HDFS inilah yang menyebabkan hadoop sering disebut sebagai sistem operasi ketimbang sebagai sistem database.

3. **Hadoop YARN.** Komponen *resource management* yang bertanggung jawab mengelola tugas komputasi (*task tracker*) serta memantau berjalannya proses komputasi (*job tracker*). Selain itu juga bertanggung jawab pada proses alokasi penggunaan *resource* komputer (cpu, storage atau memory).

Adanya YARN memungkinkan untuk saling berbagi proses komputasi pada jaringan server HDFS yang lainnya didalam sistem server terdistribusi.

4. **Hadoop Mapreduce.** Komponen *programming model* yang didukung oleh hadoop agar dapat berjalan dengan baik diatas HDFS dan pada sistem data terdistribusi. Model pemrograman Mapreduce berbeda dengan model pemrograman lainnya, namun bahasa pemrograman yang digunakan adalah

bahasa pemrograman java. Meskipun demikian tidak mudah melakukan proses analisa data menggunakan model pemrograman Mapreduce. Saat ini banyak alternatif model pemrograman lain yang dapat berjalan dengan baik diatas HDFS, salah satu contohnya adalah konsep pemrograman dengan menggunakan Apache Spark. Contoh konsep *hadoop stack* lengkap dapat digambarkan pada Gambar 2.13.



Gambar 2.13. Konsep Apache Hadoop Stack

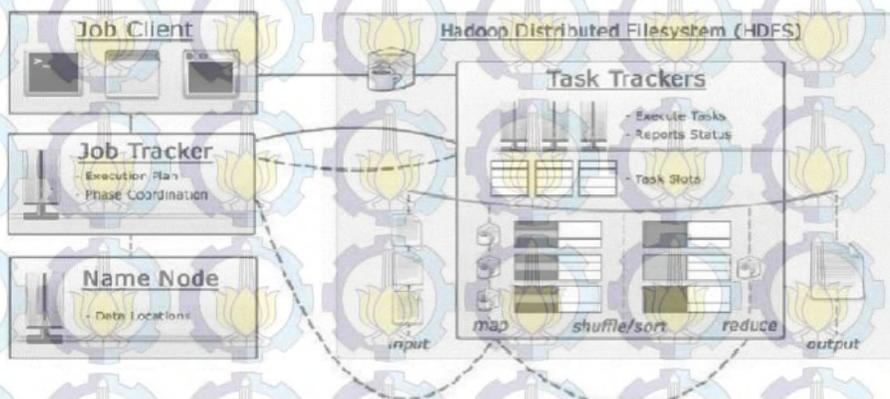
Mapreduce adalah sebuah model pemrograman yang dikhususkan untuk menangani pemrosesan dataset besar secara paralel dan juga dikhususkan untuk menangani algoritma pada data yang terdistribusi dalam sebuah cluster. Meski bahasa pemrograman yang digunakan adalah bahasa java, namun memiliki tahapan proses yang tidak dimiliki model pemrograman lain.

Tahapan proses mapreduce saat mengolah data adalah sebagai berikut :

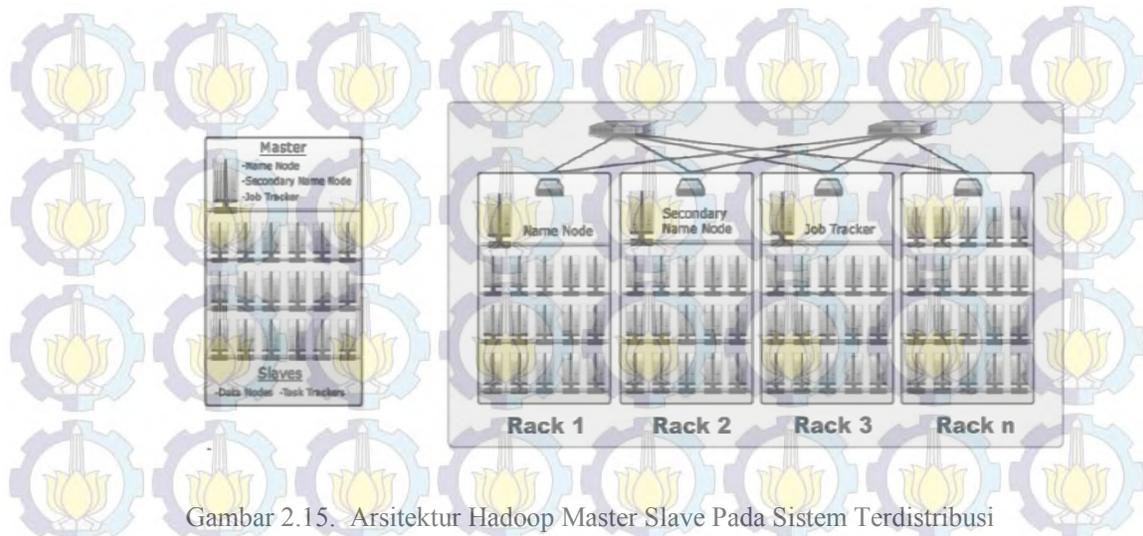
1. **Menyiapkan data map() input.** Pada tahap ini disiapkan sebuah map processor yang berisi data input yang diasosiasikan dengan sebuah K1 key value.
2. **Menjalankan user map() code.** Pada tahap ini dijalankan code map() yang dibuat user pada tiap K1 key value, dan masing-masing akan menghasilkan K2 key value.

3. **Melakukan prose shuffle data map output untuk diberikan pada reduce processor.** Pada tahap ini disiapkan sebuah *reduce processor* yang telah diasosiasikan dengan K2 key value.
4. **Menjalankan user reduce() code.** Pada tahap ini dijalankan program `reduce()` yang dibuat user pada tiap K2 key value hasil proses map pada tahap sebelumnya.
5. **Menghasilkan final output.** Pada tahap ini system mengumpulkan semua hasil dari proses reduce dan diurutkan berdasarkan K2 key value.

Karena tahapan proses mapreduce bersifat berurutan, maka tahap berikutnya baru akan dijalankan jika tahap sebelumnya telah diselesaikan seluruhnya. Jadi menempatkan data input yang sudah terdistribusi ke dalam satu cluster (shard) dapat meringankan proses map karena tinggal melakukan proses asosiasi datanya saja. Konsep mapreduce yang kompleks inilah yang menyebabkan pembuatan program analisa *Big Data* menggunakan framework mapreduce tidak mudah untuk dilakukan. Saat ini beberapa aplikasi *Big Data* (misal Apache Spark) membuat interface tambahan untuk menyederhanakan proses pembuatan analisa Big Data. Pada aplikasi Apache Spark program dapat ditulis menggunakan bahasa java, scala, python dan R, namun saat program dieksekusi tetap akan menjalankan alur proses MapReduce yang berjalan diatas framework Hadoop. Adapun alur proses mapreduce dapat digambarkan pada Gambar 2.14.



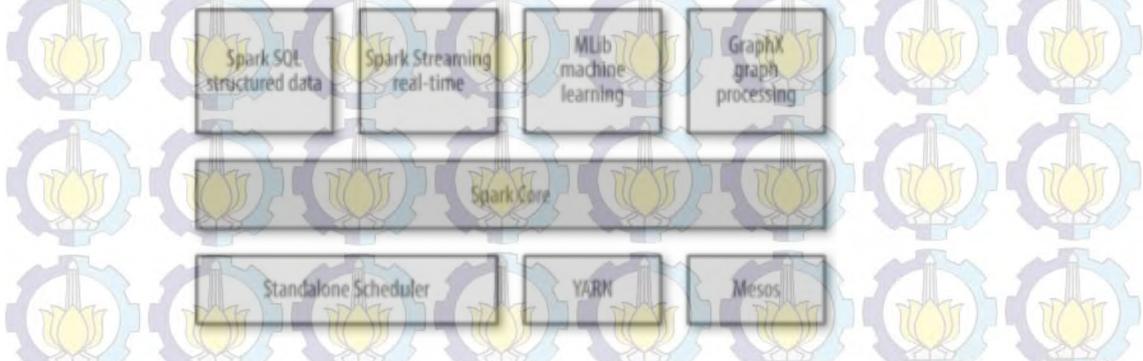
Gambar 2.14. Alur Proses Model Pemrograman Mapreduce



Gambar 2.15. Arsitektur Hadoop Master Slave Pada Sistem Terdistribusi

Pada Gambar 2.15 digambarkan arsitektur *Hadoop Master Slave*, dengan *Hadoop Master* berfungsi seperti DNS yang menunjukkan dimana lokasi *Hadoop Node* lainnya, sedangkan *Hadoop Slave* berfungsi sebagai server komputasi yang melakukan proses analisa data.

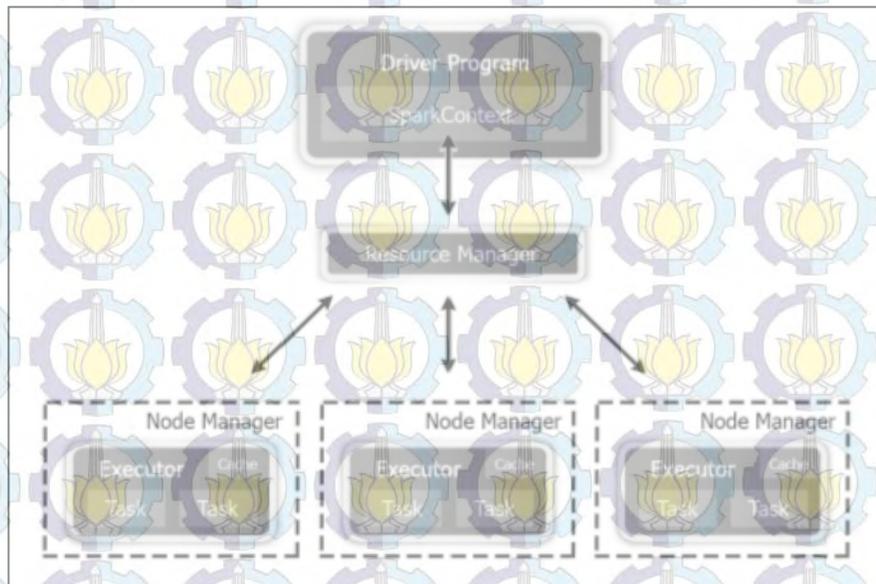
Sejak Apache Hadoop menjadi populer maka bermunculan pula aplikasi Bigdata yang berbasis open source. Salah satu aplikasi yang begitu populer di kalangan *data scientist* di dunia sejak mulai di open source-kan pada tahun 2013 adalah Apache Spark. Salah satu keunggulan Apache Spark adalah kemampuan mengoptimasikan *in-memory processing* ketimbang *io-storage processing* pada sistem data terdistribusi. Optimasi pemakaian memory sangat berguna saat harus melakukan komputasi *machine learning* yang membutuhkan proses komputasi iteratif yang berulang-ulang untuk mendapatkan model yang terbaik. Adapun komponen *Apache park Stack* dapat dilihat pada Gambar 2.16.



Gambar 2.16. Komponen Apache Spark Stack

Pada *Spark Stack* terdiri dari beberapa modul diantaranya yaitu Spark GraphX (pengolahan data tree atau graph), Spark MLlib (pengolahan machine learning), SparkSQL (pengolahan data terstruktur), Spark Streaming (pengolahan data streaming secara real time) dan SparkR (integrasi aplikasi statistik R kedalam sistem Spark Core).

Keunggulan Apache Spark berikutnya adalah dukungan terhadap berbagai Framework Komputasi *Big Data*, seperti YARN (dikembangkan oleh Apache Hadoop), Mesos dan Model Standalone aplikasi. Dengan dukungan terhadap berbagai framework komputasi *Big Data* menguntungkan pemakai Spark agar tidak tergantung pada satu jenis framework komputasi saja. Keunggulan inilah yang menyebabkan banyak *vendor Big Data* mulai menggunakan Spark sebagai solusi layanan analisa *Big Data* bagi pengguna mereka. Model integrasi aplikasi *Big Data* Apache Spark dengan YARN yang dikembangkan oleh Apache Hadoop dapat dilihat pada Gambar 2.17.



Gambar 2.17. Model Integrasi Apache Spark dengan YARN

Kompatibilitas aplikasi Bigdata Apache Spark dengan banyak aplikasi open source semakin memperkaya kemampuan Apache Spark melakukan proses analisa data berukuran besar pada sistem data terdistribusi. Beberapa aplikasi open

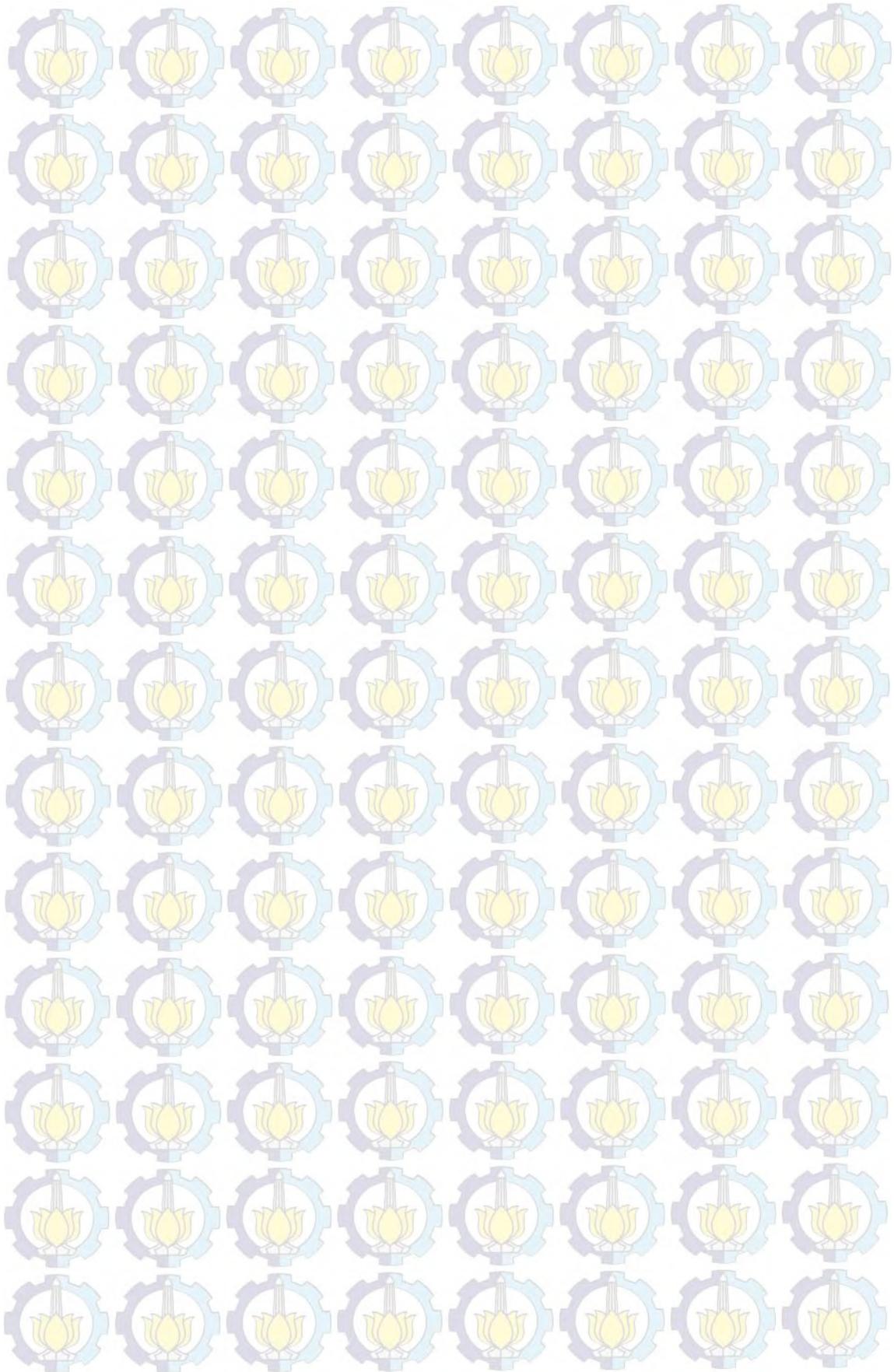
source yang telah diadopsi oleh Apache Spark diantaranya adalah Kafka (realtime data analysis), Flume (log data analysis), H2O (machine learning analysis), Hive (SQL structure data analysis), Itan (data tree analysis), Hbase (NoSql database) dan Cassandra (NoSql database).

Perkembangan aplikasi Apache Spark tidak lepas dari dukungan para programmer dan data scientist dari seluruh dunia yang ikut mengembangkan aplikasi open source ini. Selain itu sudah banyak dukungan dari perusahaan besar dunia yang memberikan solusi Bigdata seperti perusahaan Cloudera, Revolution Analytics, HortaWork, Amazon Web Service, dsb. Dukungan layanan aplikasi Spark oleh Amazon Web Service juga ikut mempopulerkan Apache Spark.

Spark mendukung berbagai bahasa pemrograman seperti java, scala, python, dan R. Sehingga banyak programmer yang memilih spark karena alternatif solusi yang diberikan Spark. Beberapa alternatif solusi aplikasi Bigdata berbasis open source yang ada saat ini dapat dilihat pada Gambar 2.18.



Gambar 2.18. Alternatif Aplikasi Bigdata Berbasis Open Source



BAB III

METODOLOGI PENELITIAN

3.1. Gambaran Umum Penelitian

Ada beberapa tahapan pada proses pembuatan visualisasi variabel cuaca yakni tahapan persiapan, tahapan visualisasi spasial variabel cuaca menggunakan metode *kriging* dan tahapan terakhir adalah proses evaluasi *cross validation*.

Tahapan penelitian secara blok diagram dapat digambarkan pada Gambar 3.1.



Gambar 3.1. Blok Diagram Tahapan Proses Penelitian

Pada penelitian ini data yang digunakan adalah data *Automatic Weather System* (AWS) seluruh Indonesia berdasarkan data *World Meteorological Organization* (WMO) yang pengelolaan datanya dilakukan oleh *National Oceanic and Atmospheric Administration* (NOAA). Lokasi data cuaca secara online dapat diakses di <http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD>.

Beberapa perangkat lunak yang digunakan pada penelitian ini adalah aplikasi *Big Data* Apache Spark dan aplikasi Rstudio. Data observasi AWS yang berupa *file* semi struktur (*file txt.gz*) akan diolah menggunakan aplikasi *Big Data* Apache Spark, sedangkan proses visualisasi spasial variabel cuaca menggunakan aplikasi Rstudio. Sedangkan proses pada tiap tahapan akan dijelaskan lebih lanjut.

3.2. Tahap Persiapan

Data observasi aws yang berupa file semi struktur (file txt.gz) pada server NOAA akan membutuhkan waktu yang cukup lama jika data harus di download dahulu baru dilakukan proses analisa data. Jadi diperlukan aplikasi *Big Data* Apache Spark - Java untuk melakukan proses query data secara langsung tanpa harus mendownload datanya terlebih dahulu. Hasil query data tersebut dapat langsung dilakukan proses analisa lebih lanjut menggunakan Apache Spark R. Kemampuan bahasa pemrograman R dapat melakukan proses analisa pada framework hadoop menggunakan aplikasi *Big Data* Spark karena dukungan Spark dalam pengolahan data dengan tipe *Dataframe*.

Pada penelitian ini data yang digunakan adalah hasil observasi AWS seluruh Indonesia pada tanggal 15 Agustus 2015. Pemilihan data observasi AWS pada tanggal tersebut dikarenakan kualitas distribusi datanya dan sedikitnya data kosong (NA). Adapun variabel cuaca yang akan dilakukan proses visualisasi spasial meliputi 7 variabel cuaca yaitu :

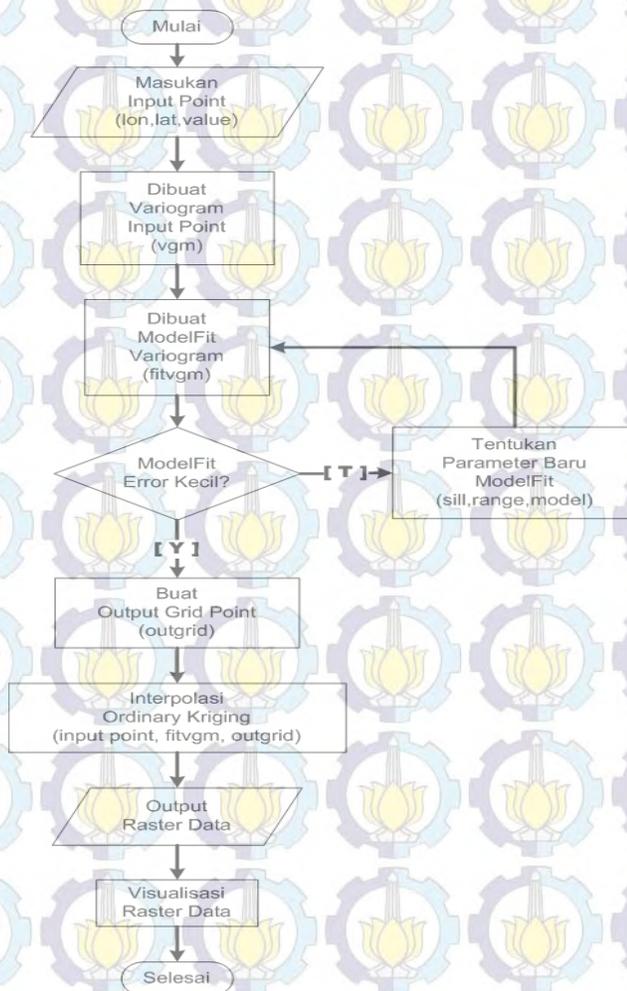
1. **Maximum Temperature (MAX).** Variabel cuaca yang berisi data suhu (satuan celcius) tertinggi dalam sehari pada lokasi observasi AWS.
2. **Minimum Temperature (MIN).** Variabel cuaca yang berisi data suhu (satuan celcius) terendah dalam sehari pada lokasi observasi AWS.
3. **Dew Point (DEWP).** Variabel cuaca yang berisi data suhu titik embun (satuan celcius) dalam sehari pada lokasi observasi AWS.
4. **Station Pressure (STP).** Variabel cuaca yang berisi data tekanan udara (satuan milibar) dalam sehari pada lokasi observasi AWS.
5. **Max Wind Speed (MXSPD).** Variabel cuaca yang berisi data kecepatan angin maksimum (satuan knot) dalam sehari pada lokasi observasi AWS.
6. **Visibility (SIB).** Variabel cuaca yang berisi data jarak pandang (satuan *mile*) dalam sehari pada lokasi observasi AWS.
7. **Precipitation (PRCP).** Variabel cuaca yang berisi data curah hujan (satuan jumlah per inch) dalam sehari pada lokasi observasi AWS.

Setelah proses akuisisi semua data yang dibutuhkan berhasil, maka perlu dilakukan langkah-langkah pre-processing sebagai berikut :

1. **Validasi Lokasi AWS.** Perlu dilakukan pengecekan lokasi AWS dengan cara melakukan plotting data lokasi AWS satu per satu di atas peta Indonesia. Jika lokasi AWS berada di lokasi lautan atau diluar wilayah Indonesia maka data AWS tidak perlu dimasukkan. Selain itu perlu dicek apakah ada lokasi AWS yang sama (disebabkan ada pergantian peralatan AWS yang baru) pilih data AWS yang baru.
2. **Konversi Data Satuan AWS.** Untuk memudahkan pemahaman terhadap hasil visualisasi data maka perlu dilakukan konversi data satuan pada variabel cuaca. Sebagai contoh adalah data suhu yang ada di AWS adalah dalam satuan Kelvin, maka perlu dikonversi menjadi Celcius.
3. **Konversi Data Kosong Menjadi NA.** Kadangkala AWS gagal melakukan proses akuisisi data karena kualitas peralatan atau usia peralatan, maka secara otomatis oleh AWS diberi data “99999” (artinya data kosong). Maka perlu dikonversi menjadi data NA (Not Available) yang dimengerti oleh bahasa pemrograman R, supaya tidak terjadi kesalahan analisa data.
4. **Pembuatan Data Summary.** Rangkuman kondisi dataset seperti nilai maksimum, nilai minimum, nilai rata-rata, dan sebagainya. Hal ini perlu dilakukan untuk mengetahui secara sekilas kondisi *dataset*.
5. **Pembuatan Grafik Histogram dan Density.** Grafik histogram dan density sebuah dataset dapat menunjukkan distribusi data sebuah dataset. Informasi ini sangat membantu dalam proses analisa data.
6. **Pembuatan Tabel Lokasi AWS.** Sebagai gambaran lokasi AWS secara gespasial yang berisi data koordinat lattitude dan longitudinal. Selain itu perlu juga informasi nama bandara udara atau pelabuhan laut lokasi dari AWS yang digunakan sebagai peralatan observasi.
7. **Pembuatan Tabel Data Observasi.** Sebagai gambaran atau informasi mengenai data observasi ketujuh variabel cuaca pada sebuah AWS. Hal ini dapat memberikan gambaran kondisi cuaca pada lokasi AWS tersebut.

3.3. Tahap Visualisasi Variabel Cuacadengan Metode Kriging

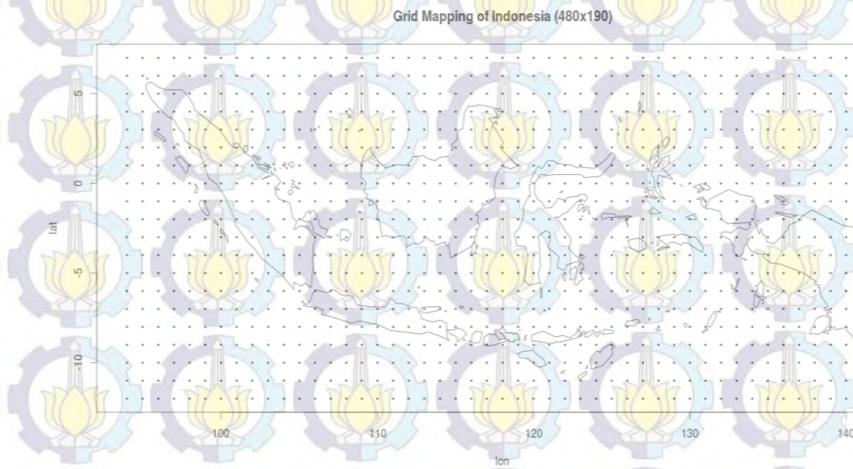
Flowchart langkah-langkah yang dilakukan pada tahap visualisasi variabel cuaca dengan metode *kriging* dapat dilihat pada Gambar 3.2.



Gambar 3.2. Flowchart Visualisasi Variabel Cuaca dengan Metode Kriging

Tahapan awal pada proses visualisasi spasial adalah pembuatan *variogram* seluruh data *input* yang siap diolah. *Variogram* bermanfaat untuk menunjukkan *spatial corelation* diantara seluruh data input. Setelah itu ditentukan parameter *model fit variogram* dengan metode auto fit variogram yang akan menghasilkan parameter optimal pada proses *model fit variogram*. Hasil *model fit variogram* akan digunakan pada tahapan estimasi data menggunakan metode *kriging*.

Proses selanjutnya adalah pembuatan *output data grid* pada seluruh lokasi data AWS seluruh Indonesia. *Data grid* harus meliputi semua lokasi AWS karena metode *kriging* hanya mengolah data yang ada di dalam *data grid*. Adapun contoh sebuah *data grid* yang meliputi seluruh wilayah Indonesia dapat dilihat pada Gambar 3.3.



Gambar 3.3. Data Grid Seluruh Wilayah Indonesia

Langkah selanjutnya adalah melakukan proses interpolasi kriging terhadap seluruh data masukan, *model fit variogram* dan *output data grid*. Adapun langkah proses interpolasi *kriging* adalah sebagai berikut :

1. Dicari data *input point* yang valid. Jika data tidak memiliki koordinat atau data memiliki nilai NA maka data diabaikan. Jika dua data mempunyai lokasi data yang sama maka diambil data pertama yang ditemukan, tidak dibuat nilai rata-ratanya.
2. Proses output pixel ke-n. Nilai data *output pixel* ditentukan oleh jarak dan nilai *input point* lainnya. Jika ada *input point* yang jaraknya melebihi batasan jarak yang telah ditentukan maka data diabaikan. Jika ada *input point* didalam batasan jarak yang nilainya dibawah nilai *nr* minimum maka *output pixel* diisi dengan *undefined value*. Jika terdapat banyak *input point* masih didalam batasan jarak tertentu, maka digunakan nilai *input point* terdekat.

3. Ditentukan jarak untuk semua *input point* terhadap *output pixel* (langkah 2) dan cari nilai *semivariogram* pada semua jarak yang ditemukan. Masukkan nilai *semivariogram* kedalam matrix C. Untuk semua proses *kriging* menggunakan Persamaan 3.1.

$$[C] * [W] = [D] \dots \dots \dots (3.1)$$

dengan :

[C] adalah matrix C yang berisi data *semivariogram input point* terpilih

[W] adalah vektor W yang berisi *weight factor* input point i

[D] adalah matrix D yang berisi data *semivariogram output pixel* sesuai dengan *input point* terpilih

4. Ditentukan jarak *output pixel* (pada langkah 2) terhadap semua *input pixel* dan cari nilai *semivariogram* pada semua jarak yang ditemukan. Masukkan nilai *semivariogram* kedalam matrix D.

5. Dihitung *weight facttor* (vektor W) hanya pada *output pixel* (langkah 2), dengan Persamaan 3.2.

$$\sum_{i=1}^N W_i = 1 \dots \dots \dots (3.2)$$

dengan :

W_i adalah *weight vector* untuk *input point* i

6. Dihitung prediksi nilai pada *output pixel* (langkah 2) dengan Persamaan 3.3.

$$Z = S_i(w_i * Z_i) \dots \dots \dots (3.3)$$

dengan :

Z adalah prediksi nilai pada *output pixel*

w_i adalah *weight factor* untuk *input point* i

Z_i adalah nilai data *input point* i

7. Dihitung *Error Variance* dan *Standart Error* pada *output pixel* dengan menggunakan Persamaan 3.4 dan Persamaan 3.5.

$$s^2 = S_i (w_i * g(h_{pi})) + I \dots \dots \dots (3.4)$$

$$s = \sqrt{S_i (w_i * g(h_{pi})) + I} \dots \dots \dots (3.5)$$

dengan :

s^2 adalah *Error Variance*

s adalah *Standart Error*

h_{pi} adalah jarak antara *output pixel* p dengan *input point* i

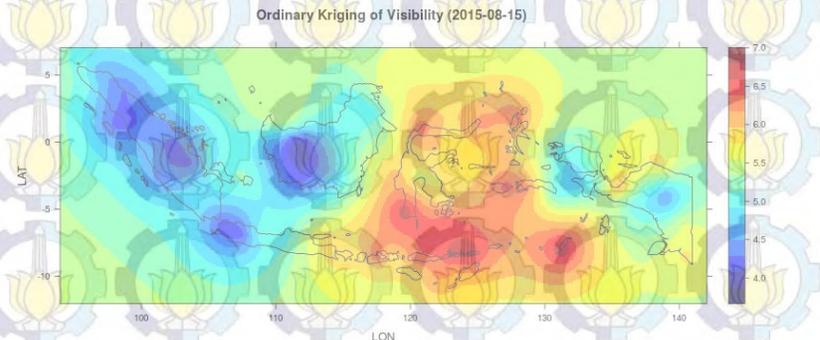
$g(h_{pi})$ adalah nilai *semivariogram* untuk jarak antara *output pixel* p dengan *input point* i

w_i adalah *weight factor* untuk *input point* i

I adalah *langrange multiplier* (untuk meminimalisir estimasi error)

8. Proses dilanjutkan untuk *ouput pixel* selanjutnya, diulangi mulai langkah ke 2 hingga langkah ke 8, sampai semua *output pixel* selesai diproses.

Demikianlah algoritma metode interpolasi *kriging* yang akan menghasilkan data *ouput raster pixel*. Setelah itu seluruh data *output raster pixel* siap di tampilkan kedalam bentuk visualisasi geospasial seperti pada contoh gambar 3.4.



Gambar 3.4. Visualisasi Geospasial dengan Metode Ordinary Kriging

3.4. Tahap Evaluasi Cross Validation

Proses *Cross Validation* adalah teknik untuk menguji seberapa akurat sebuah *Predictive Model*. Teknik *Cross Validating* berbeda dengan teknik validasi umumnya yang akan membagi *dataset* menjadi dua bagian yaitu 70% data sebagai *training dataset* dan 30% data sebagai *test dataset* dengan tujuan akhir untuk menghitung *Root Mean Square Error* (RMSE). Pada teknik *Cross Validation* digunakan nilai rata-rata dari *prediction error* untuk mengoreksi nilai *training error* yang terlalu optimis, agar didapatkan hasil yang lebih akurat saat mengukur kinerja sebuah *Predictive Model*.

Teknik visualisasi spasial data variabel cuaca perlu diukur keakuratannya pada ketujuh variabel cuaca yang diteliti. Kinerja sebuah model visualisasi spasial tidak selalu baik pada semua jenis sample data yang dimiliki. Banyak faktor yang menyebabkan terjadinya penurunan kinerja diantaranya adalah kurangnya jumlah lokasi data yang diobservasi atau banyaknya data kosong (NA). Dengan teknik *Cross Validation* dapat diukur kinerja model visualisasi spasial yang diteliti.

Ada tiga kriteria pada *Cross Validation* (Oliver, 2015) yang digunakan untuk mengukur kinerja model visualisasi spasial adalah :

1. Mean Error (ME) dihitung menggunakan Persamaan 3.6.

$$ME = \frac{1}{N} \sum_{i=1}^N \{z(x_i) - Z(x_i)\} \dots \dots \dots (3.6)$$

dengan :

N adalah jumlah data yang diobservasi

$z(x_i)$ adalah nilai data hasil observasi

$Z(x_i)$ adalah nilai data hasil prediksi

Nilai ideal ME adalah 0

2. Mean Square Prediction Error (MSPE) dihitung dengan Persamaan 3.7.

$$MSPE = \frac{1}{N} \sum_{i=1}^N \{z(x_i) - Z(x_i)\}^2 \dots \dots \dots (3.7)$$

Nilai ideal MSPE adalah bernilai kecil

3. Mean Squared Normalized Error (MSNE) dihitung dengan Persamaan 3.8.

$$MSNE = \frac{1}{N} \sum_{i=1}^N \frac{\{z(x_i) - Z(x_i)\}^2}{\sigma^2(x_i)} \dots \dots \dots (3.8)$$

dengan :

$\sigma^2(x_i)$ adalah kriging variance

Nilai ideal MSNE adalah mendekati 0

Selain ketiga kriteria diatas perlu juga diukur kriteria lainnya seperti korelasi antara data hasil observasi dengan data hasil prediksi (COP) dan korelasi antara data hasil prediksi dengan data residual (CPR). Adapun nilai ideal COP adalah 1 sedangkan nilai ideal CPR adalah 0.

Satu kriteria lagi yang penting untuk dievaluasi adalah Sum of Squared Error of Prediction (SSE) yang dihitung menggunakan Persamaan 3.9.

$$SSE = \sum_{i=1}^n (y_i - f(x_i))^2 \dots \dots \dots (3.9)$$

dengan :

y_i adalah nilai data hasil observasi

$f(x_i)$ adalah nilai data hasil prediksi

Nilai SSE kecil menunjukkan kecocokan antara model dengan data observasi

BAB IV

HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian Tahap Persiapan

Proses akuisisi data semi terstruktur (*file txt.gz*) *Daily Global Weather Measurements* (GSOD) pada lokasi server NOAA menggunakan Aplikasi Bigdata Apache Spark Java (berjalan diatas framework HDFS) membutuhkan waktu sekitar tujuh jam (tanpa kendala komputer berhenti bekerja). Saat dicoba secara manual (diunduh dahulu datanya) menggunakan Aplikasi Rstudio, membutuhkan waktu tujuh hari (dengan mengalami kendala komputer berhenti bekerja).

Langkah-langkah akuisisi data dengan aplikasi Bigdata Apache Spark Java adalah sebagai berikut :

1. Dilakukan proses inialisasi *library* aplikasi Big Data Spark (JavaRDD, JavaSparkContext dan FlatMapFunction).
2. Dilakukan proses penyaringan data AWS seluruh Indonesia.
3. Dilakukan proses *parsing* data semi terstruktur (*file txt.gz*) AWS.
4. Dilakukan proses inialisasi data hasil proses *parsing* menjadi data bertipe *Dataframe* yang siap untuk dilakukan proses analisa lebih lanjut oleh aplikasi statistik Rstudio (aplikasi Big Data Apache Spark mendukung juga bahasa pemrograman python, java dan scala).

Hasil akuisisi data AS|W pada periode mulai tahun 1929 hingga tahun 2014 adalah sebagai berikut :

1. File GSOD_DESC.txt (deskripsi data GSOD)
2. File isd-history.txt (berisi informasi nama dan lokasi AWS)
3. File coords.csv (berisi lokasi AWS di seluruh dunia)
4. File aws_data_all.csv (data AWS sedunia dengan ukuran file 12,7 GB)

Jumlah AWS diseluruh dunia (yang ikut berbagi data dibawah koordinasi WMO) adalah sebanyak 29.352. Sedangkan jumlah AWS seluruh Indonesia adalah sebanyak 107 dan jumlah AWS di pulau Jawa sebanyak 38.

Tabel 4.1. Nama dan Lokasi AWS di Indonesia

NO	AWSID	NAMA	LON	LAT
1	960010	MAIMUN SALEH	95.34	5.874
2	960011	SRI AMAN / SIMANGGAN	111.45	1.217
3	960090	MALIKUS SALEH	96.95	5.227
4	960110	SULTAN ISKANDARMUDA	95.42	5.524
5	960150	MEULABOH/CUT NYAK	96.117	4.25
6	960330	MEDAN/BELAWAN	98.7	3.8
7	960350	POLONIA	98.672	3.558
8	960730	PINANG SORE	98.889	1.556
9	960750	BINAKA	97.705	1.166
10	961090	SULTAN SYARIF KASIM II	101.445	0.461
11	961630	TABING	100.352	-0.875
12	961710	JAPURA	102.335	-0.353
13	961950	SULTAN THAHA	103.644	-1.638
14	962070	KERINCI/DEPATI PARB	101.367	-2.767
15	962210	SULTAN MAHMUD	104.701	-2.898
16	962370	PANGKALPINANG	106.13	-2.17
17	962490	H AS HANANDJOEDDIN	107.755	-2.746
18	962530	FATMAWATI SOEKARNO	102.339	-3.864
19	962950	RADIN INTEN II	105.179	-5.242
20	964130	KUCHING INTL	110.347	1.485
21	964810	TAWAU	118.122	4.313
22	965050	LONGBAWAN/JUVAI	115.683	3.733
23	965090	JUWATA	117.566	3.327
24	965250	TANJUNG SELOR	117.333	2.85
25	965290	KALIMARU	117.432	2.155
26	965350	PALOH	109.3	1.767
27	965570	NANGAPINOH	111.783	-0.35
28	965590	SUSILO	111.473	0.064
29	965650	PUTUSSIBAU/PAGSUMA	112.933	0.883
30	965810	SUPADIO	109.404	-0.151
31	965950	MUARATEWE/BERINGIN	114.9	-0.95
32	966070	TEMINDUNG	117.157	-0.485
33	966150	RAHADI USMAN	109.963	-1.817
34	966330	SEPINGGAN	116.894	-1.268
35	966450	ISKANDAR	111.673	-2.705
36	966550	TJILIK RIWUT	113.943	-2.225
37	966850	SYAMSUDIN NOOR	114.763	-3.442
38	966950	KOTABARU/SETAGEN	116.217	-3.4
39	967370	SERANG	106.133	-6.117
40	967390	BUDIARTO	106.57	-6.293

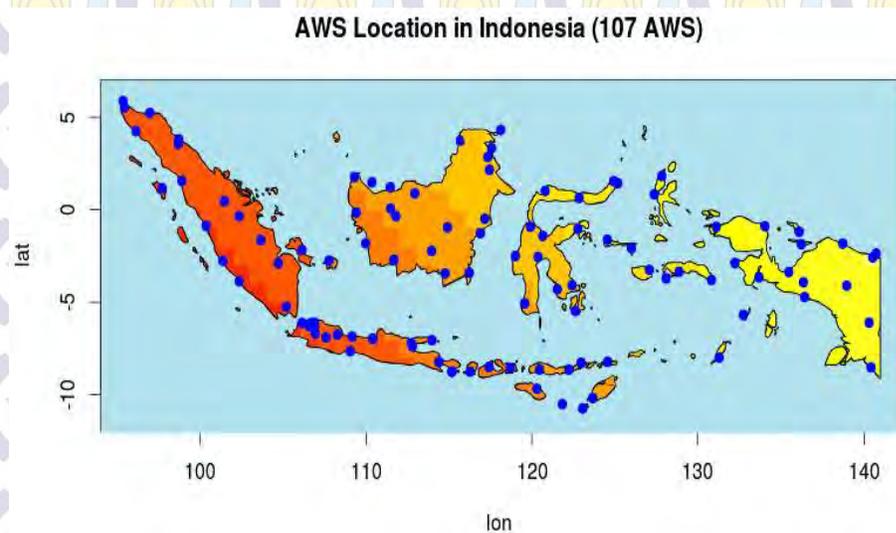
Lanjutan Tabel 4.1. Nama dan Lokasi AWS di Indonesia

NO	AWSID	NAMA	LON	LAT
41	967410	JAKARTA/TANJUNG PRIOK	106.867	-6.1
42	967450	JAKARTA/OBSERVATORY	106.833	-6.183
43	967490	SOEKARNO HATTA INTL	106.656	-6.126
44	967495	HALIM PERDANAKUSUMA	106.891	-6.267
45	967510	BOGOR/CITEKO	106.933	-6.7
46	967810	HUSEIN SASTRANEGARA	107.576	-6.901
47	967910	CIREBON/JATIWANGI	108.267	-6.75
48	967970	TEGAL	109.15	-6.85
49	968050	TUNGGUL WULUNG	109.034	-7.645
50	968370	SEMARANG	110.417	-6.967
51	968390	ACHMAD YANI	110.375	-6.973
52	969330	SURABAYA/PERAK	112.717	-7.217
53	969350	SURABAYA JUANDA	112.787	-7.38
54	969730	KALIANGET (MADURA IS)	113.967	-7.05
55	969870	BANYUWANGI	114.383	-8.217
56	970140	SAM RATULANGI	124.926	1.549
57	970160	BITUNG	125.183	1.433
58	970280	TOLI-TOLI/LALOS	120.8	1.017
59	970480	JALALUDDIN	122.85	0.637
60	970720	MUTIARA	119.91	-0.919
61	970860	BUBUNG	122.772	-1.039
62	970960	KASIGUNCU	120.658	-1.417
63	971200	MAJENE	119	-2.5
64	971260	MASAMBA/ANDI JEMMA	120.367	-2.55
65	971420	KOLAKA/POMALA	121.533	-4.3
66	971460	WOLTER MONGINSIDI	122.417	-4.082
67	971800	HASANUDDIN	119.554	-5.062
68	971820	UJANG PANDANG/PAOTERE	119.55	-5.067
69	971920	BAU-BAU/BETO AMBIRI	122.617	-5.467
70	971925	BALI INTL	115.167	-8.748
71	972300	DENPASAR NGURAH RAI	115.167	-8.749
72	972400	MATARAM-LOMBOK INTL	116.267	-8.75
73	972600	SUMBAWA BESAR	117.412	-8.489
74	972700	MUHAMMAD SALAHUDDIN	118.687	-8.54
75	972840	RUTENG/SATAR TACIK	120.45	-8.633
76	973000	WAI OTI	122.237	-8.641
77	973100	LARANTUKA/GEWAYENTA	122.967	-8.267
78	973200	ALOR/MALI/KALABAHI	124.567	-8.217
79	973400	MAU HAU	120.302	-9.669
80	973720	EL TARI	123.671	-10.172

Lanjutan Tabel 4.1. Nama dan Lokasi AWS di Indonesia

NO	AWSID	NAMA	LON	LAT
81	973780	ROTE/BAA/LEKUNIK	123.067	-10.733
82	973800	SABU/TARDAMU	121.833	-10.5
83	974060	GALELA/GAMARMALAMU	127.833	1.817
84	974300	BABULLAH	127.381	0.831
85	974600	LABUHA/TALIABU	124.55	-1.617
86	975020	JEFMAN	131.121	-0.926
87	975300	RENDANI	134.049	-0.892
88	975600	FRANS KAISIEPO	136.108	-1.19
89	975700	SERUI/YENDOSA	136.233	-1.867
90	975800	SARMI/ORAI	138.717	-1.833
91	976000	SANANA	126	-2.083
92	976300	FAK-FAK/TOREA	132.25	-2.883
93	976820	NABIRE	135.496	-3.368
94	976860	WAMENA	138.957	-4.103
95	976900	SENTANI	140.516	-2.577
96	976980	JAYAPURA/DOK II	140.717	-2.367
97	977000	NAMLEA/BURU ISLAND	127.083	-3.25
98	977220	AMAHAI	128.883	-3.35
99	977240	PATTIMURA	128.089	-3.71
100	977480	GESER	130.833	-3.8
101	977600	KAIMANA	133.696	-3.645
102	977800	ENAROTALI	136.367	-3.917
103	977960	KOKONAO/TIMUKA	136.433	-4.717
104	978100	TUAL/DUMATUBUN	132.75	-5.683
105	978760	TANAH MERAH	140.3	-6.1
106	979000	SAUMLAKI	131.3	-7.983
107	979800	MOPAH	140.418	-8.52

Pada Tabel 4.1 total AWS seluruh Indonesia berjumlah 107 buah adalah hasil penyaringan data AWS berdasarkan lokasi wilayah Indonesia (dengan koordinat Longitude diantara 94 hingga 142 dan Lattitude diantara -12 hingga 7) yang berjumlah 318 buah. Namun setelah dilakukan pengecekan banyak koordinat lokasi AWS yang sama (ada pergantian peralatan AWS, namun peralatan AWS lama tetap hidup) dan lokasi AWS yang ada di tengah lautan (AWS di pulau kecil yang terpencil). Setelah dilakukan penyaringan ulang data AWS akhirnya didapatkan data AWS di Indonesia sebanyak 107.



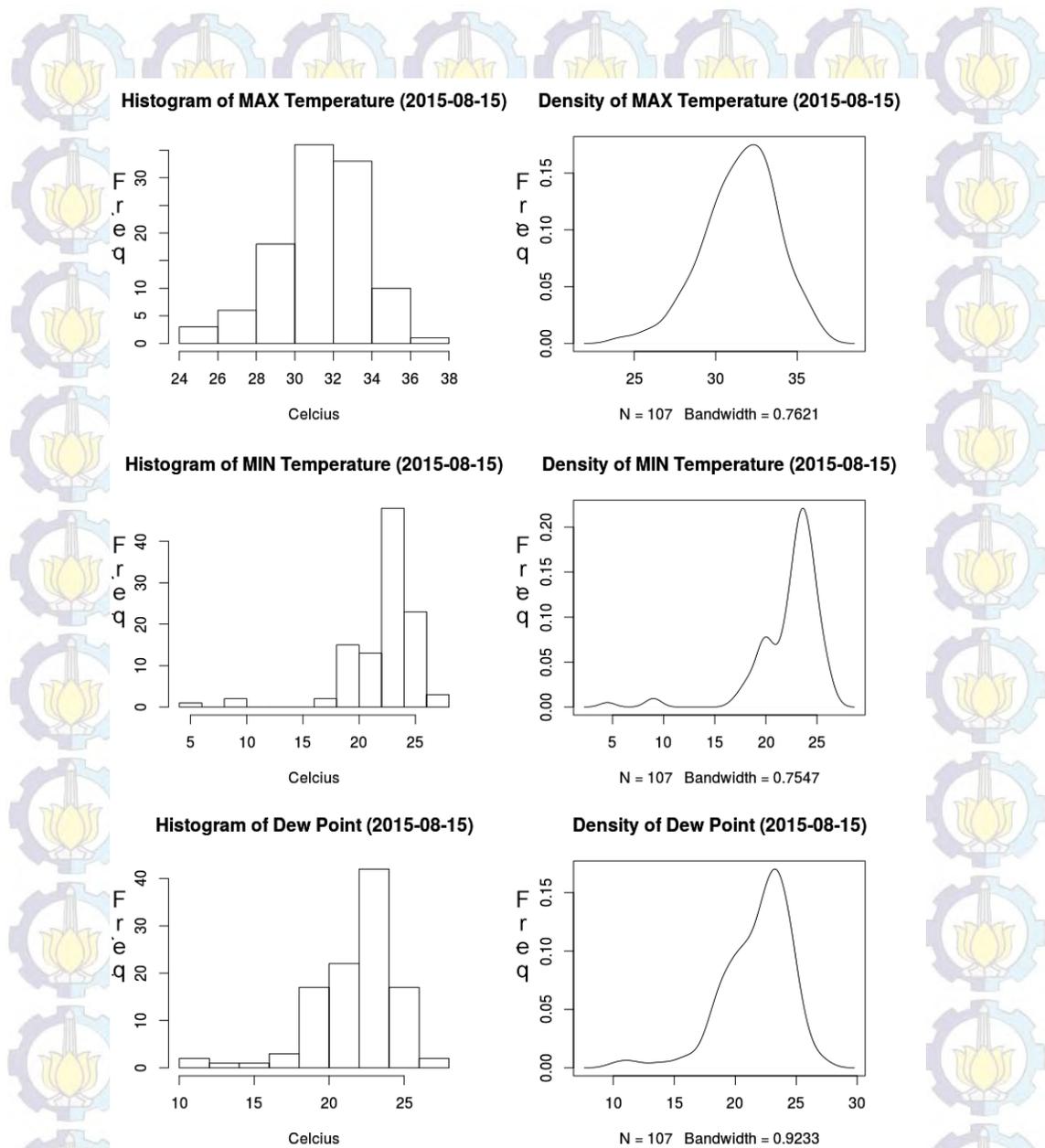
Gambar 4.1. Peta Lokasi AWS di Seluruh Wilayah Indonesia

Pada Gambar 4.1 terlihat bahwa lokasi AWS terdistribusi cukup merata diseluruh wilayah Indonesia. Meratanya penyebaran lokasi data AWS diharapkan dapat mewakili kondisi cuaca seluruh Indonesia.

Tabel 4.2. Rangkuman Data AWS Seluruh Variabel Cuaca

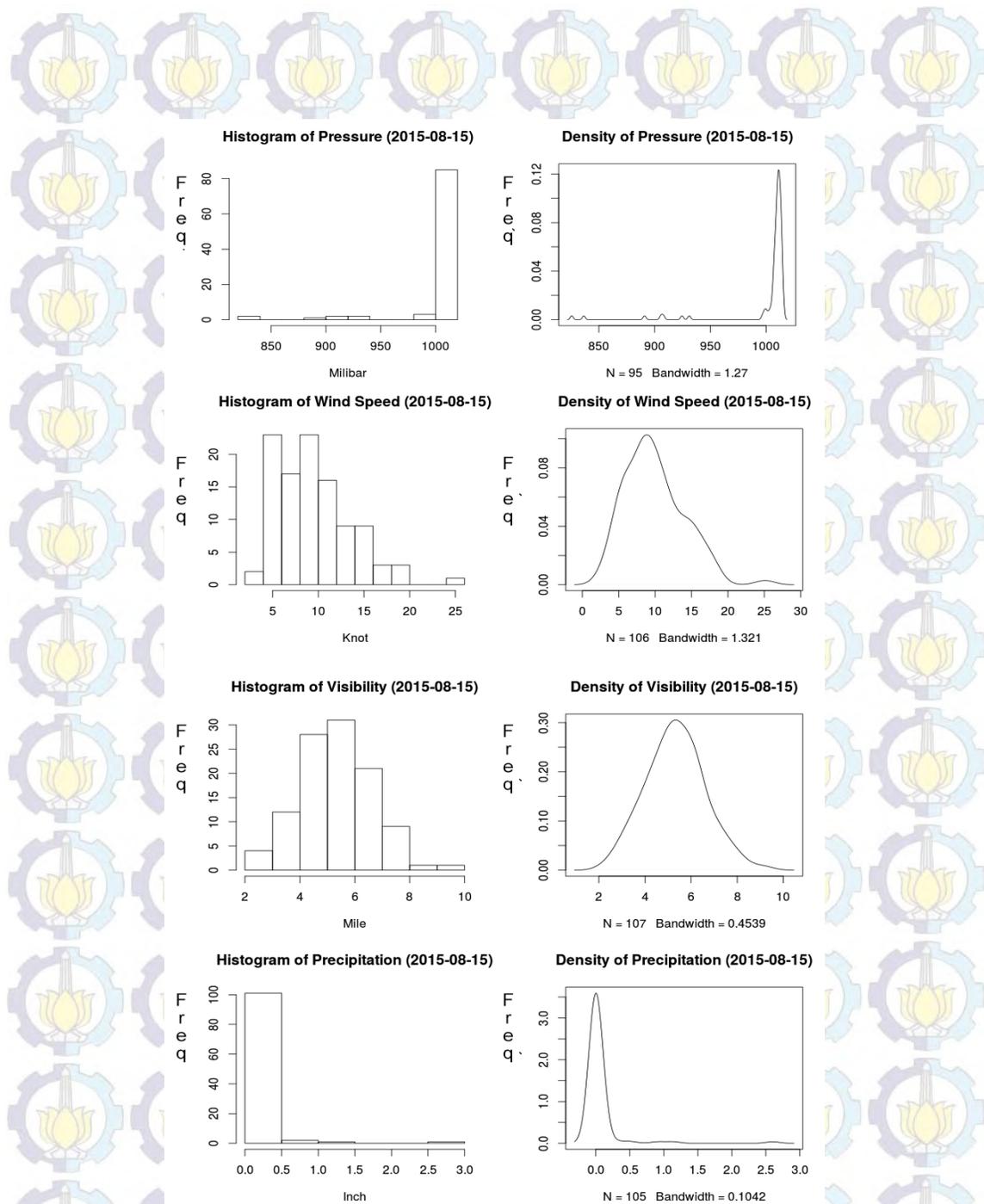
No	Nama Variabel	Nilai Minimal	Rata-Rata	Nilai Maksimal	Data Kosong
1	MAX	24,22	31,52	36,22	0
2	MIN	4,50	22,33	26,39	0
3	DEWP	10,39	21,76	27,00	0
4	STP	825,6	1001,1	1014,8	12
5	MXSPD	2,900	9,925	25,100	1
6	SIB	2,300	5,322	9,100	0
7	PRCP	0	0,05676	2,60000	2

Pada Tabel 4.2 terlihat ada tiga variabel cuaca yang memiliki data kosong (NA) yaitu variabel cuaca *Station Pressure* (12 data kosong), *Maximum Wind Speed* (1 data kosong) dan *Precipitation* (1 data kosong).



Gambar 4.2. Histogram Variabel MAX, MIN dan DEWP

Pada Gambar 4.2 pola distribusi datanya mengikuti pola distribusi data normal yang artinya frekuensi kemunculan nilai tidak didominasi pada nilai tertentu saja tapi terdistribusi merata dibandingkan nilai-nilai yang lainnya. Pola distribusi data normal berpeluang besar menghasilkan hasil prediksi *raster data* yang bersifat kontinyu.



Gambar 4.3. Histogram Variable STP, MXSPD, SIB dan PRCP

Pada Gambar 4.3 terlihat pola variasi variabel cuaca Station Pressure (STP) dan Precipitation (PRCP) tidak mengikuti pola distribusi data normal.

Tabel 4.3. Data Observasi AWS Indonesia Tanggal 15-08-2015

NO	AWSID	MAX	MIN	DWP	STP	WSP	SIB	PRC
1	960010	31.39	23.11	24.94	996.8	8	4.5	0.52
2	960011	35	23	24.33	NA	4.1	6.2	0
3	960090	32.61	23.22	23.72	1006.4	7	3.7	1.14
4	960110	33	23.39	24	1007.6	8	5.6	0.04
5	960150	32	23.39	25.06	1010.6	8	5.4	0.08
6	960330	31	25.5	25.28	1010.3	6	3.8	0.01
7	960350	33	24.39	26.61	NA	8	4.2	0
8	960730	33	21	23.39	1008.8	NA	3.6	0.03
9	960750	32.22	22.22	25.17	1009.2	6	3.8	NA
10	961090	33.89	24	24.11	NA	6	4.1	0.91
11	961630	31	23.28	24.11	1009.6	7	5.1	0
12	961710	33	24	24	1008	7	3.6	0
13	961950	33	23.78	23.72	1007.9	8.9	3.4	0.12
14	962070	29.5	18.22	19.67	924.5	12	3.7	0
15	962210	33.61	23.5	23.17	NA	13	5.2	0
16	962370	32	25	23.06	1007.3	11.1	5.2	0
17	962490	33.89	23.39	22.44	1005.6	9.9	5.7	0
18	962530	31.5	24.61	24.5	1007.7	11.1	5	0
19	962950	33.22	23.78	23.17	1002.3	14	4.7	0
20	964130	33.11	23	23.78	NA	11.1	6.2	0.34
21	964810	32.39	23.22	24.33	1009.5	8.9	6.4	0
22	965050	29.22	18.28	20.72	907.8	4.1	4.7	0
23	965090	32.78	25	24.61	1010.7	5.1	5.1	0
24	965250	35.5	24.61	24.39	1011	9.9	5.3	0
25	965290	35.22	22	23.44	1010.6	5.1	4.3	0
26	965350	30.72	22.22	23.72	1011.1	2.9	5.6	0.08
27	965570	33.39	23.89	23.67	1006.6	6	3.2	0
28	965590	34.78	23.22	24.83	1007.7	5.1	2.3	0.08
29	965650	34	23.39	27	1006.8	5.1	4.8	0
30	965810	32	23	24.17	1011.3	15	4.1	2.6
31	965950	33.22	25.22	23.89	1007.2	5.1	4.8	0
32	966070	31.61	25.39	23.39	1011.1	6	5.3	0
33	966150	33.5	24.11	23.33	1011.1	9.9	4.5	0
34	966330	31	24	23.44	NA	11.1	6	0
35	966450	32.39	23	24.5	1009.1	8	3.4	0
36	966550	34.22	23.5	23	1010.1	8	4.1	0.01
37	966850	36	23	21.11	1010	17.1	5	0
38	966950	31.78	24	22.33	1012.2	8	5.2	0
39	967370	32	23.61	23.5	1008	8	3.1	0
40	967390	33.78	22.78	22.17	1007.1	5.1	3.4	0

Lanjutan Tabel 4.3. Data Observasi AWS Indonesia Tanggal 15-08-2015

NO	AWSID	MAX	MIN	DWP	STP	WSP	SIB	PRC
41	967410	32	26.39	23.06	1011.6	12	4.9	0
42	967450	33	26.22	22.94	1011.7	5.1	4.3	0
43	967490	33	24.39	22.67	NA	15.9	5.1	0
44	967495	33	24	22.56	NA	17.1	3	0
45	967510	28.22	17.61	18.28	905.6	4.1	4.6	0
46	967810	29	20.11	19.28	931	8.9	2.8	0
47	967910	33.61	22.39	22.33	1007.2	7	5.4	0
48	967970	31.22	24.78	23.11	1012.3	15.9	6.1	0
49	968050	28.78	25	23.44	1012.8	9.9	4.7	0
50	968370	34.61	25.39	22.56	1012.3	9.9	5	0
51	968390	35.39	25.39	21.67	1011.9	14	4.2	0
52	969330	31.78	25.61	22.11	1013	18.1	6.2	0
53	969350	30.5	24.22	22.44	1013	15	5.4	0
54	969730	30	25.61	22.61	1013.5	18.1	6.2	0
55	969870	30.39	22.61	22	1008.6	6	6	0
56	970140	33	17	16.06	NA	17.1	6.2	0
57	970160	30	23.89	20.28	1013.1	8	6.8	0
58	970280	31.78	22.39	22.33	1011	11.1	7.4	0
59	970480	31.78	19.5	19.28	1009.8	14	5.4	0
60	970720	36.22	22.39	20.11	1003.6	14	5.5	0
61	970860	29.61	24.5	19.67	1012.6	13	6.3	NA
62	970960	34.61	22.22	17.94	1011.9	9.9	6.2	0
63	971200	31	24.39	20.89	1010	9.9	6.8	0
64	971260	31.61	22	20.72	1009.9	2.9	4.7	0
65	971420	33	22	18.61	1013.1	6	5.7	0
66	971460	29	18.22	20.11	1008.9	9.9	5.3	0
67	971800	35	21.89	13.83	NA	11.1	5.6	0
68	971820	32.39	24.5	19.11	1013	8.9	9.1	0
69	971920	32	20	18.61	1010.6	8	7	0
70	971925	30	24	22.56	NA	15.9	6.1	0
71	972300	30	22.89	22.56	1013.7	11.1	6	0
72	972400	31.39	19.78	20.5	NA	14	4.8	0
73	972600	33.39	19.78	17.5	1013.5	8.9	7.9	0
74	972700	32.78	20	21.5	1014.8	9.9	5	0
75	972840	24.22	8.778	15.83	890.9	6	7.1	0
76	973000	30.22	18.89	18.94	1013	6	6.2	0
77	973100	28.78	22.22	20.06	1014.1	8	7.1	0
78	973200	29.89	18.78	19.56	1014.5	15	7.6	0
79	973400	32.22	19.78	18.44	1014.8	12	5.3	0
80	973720	32.22	20.22	18.33	1004.3	15	5.6	0

Lanjutan Tabel 4.3. Data Observasi AWS Indonesia Tanggal 15-08-2015

NO	AWSID	MAX	MIN	DWP	STP	WSP	SIB	PRC
81	973780	29.39	21.39	18.56	999.5	15	6.2	0
82	973800	30.22	24	20.94	1013.9	11.1	6.6	0
83	974060	32	19.78	20.44	1008.5	12	7.5	0
84	974300	31	24	20.5	1009.7	13	5.6	0
85	974600	29.61	20.22	19.89	1011.3	6	5.6	0
86	975020	27.78	23.22	22.33	1012.5	8	4.4	0
87	975300	31.39	23.72	23.61	1010.6	6	5.2	0
88	975600	30.78	24.78	25	1009.9	8.9	6	0
89	975700	29.89	23.61	23.67	1009.8	11.1	5.5	0
90	975800	30.78	24.5	24.89	1010.7	4.1	6.8	0
91	976000	30.78	20.78	20.44	1013.7	8	6	0
92	976300	26	19.61	20.67	998.5	5.1	4.2	0
93	976820	33.11	22.89	24	1009.9	8.9	8	0
94	976860	27	4.5	10.39	836.4	25.1	2.8	0
95	976900	33.72	24.28	23.06	1000.3	8.9	5.5	0
96	976980	30.61	26.22	25.17	1010.6	12	4.7	0
97	977000	30.61	19.72	18.94	1011.6	15.9	7.2	0
98	977220	27.72	21.28	21.28	1013.7	11.1	7.1	0
99	977240	28	24	21.06	1012.6	9.9	6.2	0
100	977480	27.78	23.39	22.17	1013.7	8.9	4	0
101	977600	27.5	23	21.78	1012.6	8.9	4.3	0
102	977800	25.78	9.222	11.67	825.6	8.9	5	0
103	977960	30.22	21	21.44	1010.4	9.9	4.3	0
104	978100	29.22	23	22.61	1011	12	6.4	0
105	978760	31.89	20.72	18.44	1011.8	8.9	6.1	0
106	979000	28.72	23.61	21.39	1012.7	14	8.2	0
107	979800	30.5	19.72	19.83	1014.7	18.1	6.2	0

Pada Tabel 4.3 terlihat bahwa pada variabel cuaca *Precipitation* (PRC) hanya 13 AWS yang memiliki nilai selain nilai nol, sisanya memiliki nilai data yang sama yakni data nol. Pada variabel cuaca *Station Pressure* (STP) hanya 22 AWS yang memiliki nilai dibawah 1000 milibar, sisanya memiliki nilai data diatas 1000 milibar. Selain itu ada tiga variabel cuaca yang memiliki data kosong (NA) yaitu variabel *Station Pressure* (STP), *Wind Speed* (WSP), *Precipitation* (PRC). Perlu diingat bahwa pada variabel cuaca yang memiliki data kosong (NA) sebaiknya dilakukan penyaringan data terlebih dahulu.

4.2. Hasil Visualisasi Variabel Cuaca Max Temperature

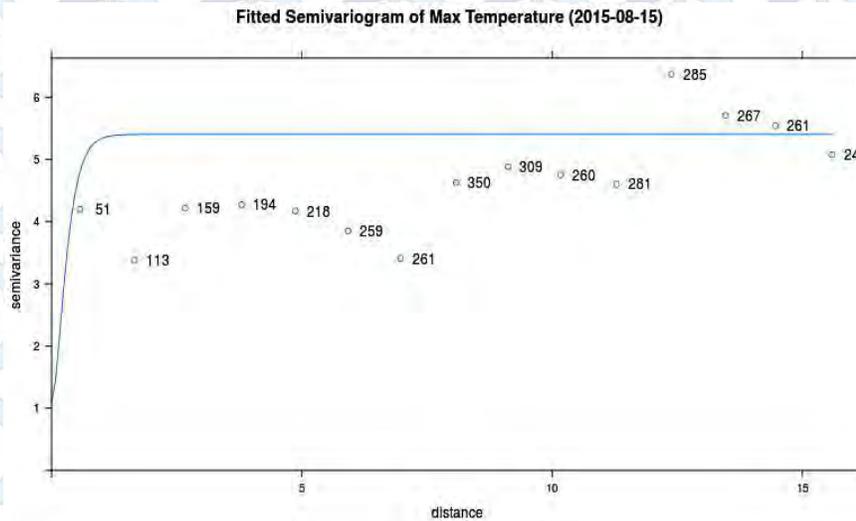
Proses visualisasi variabel cuaca *Maximum Temperature* diawali dengan proses penyaringan data yang akan digunakan dipastikan tidak memiliki data kosong (NA) yang kemudian disimpan ke dalam sebuah *Dataframe Input Point*. Setelah itu dilakukan analisa *Spatial Auto Correlation* menggunakan teknik variogram yang memetakan jarak antar semua titik terhadap nilai data.

Tabel 4.4. Nilai Variogram Variabel Max Temperature

No	NP	Distance	Gamma
1	51	0,568614327	4,197107238
2	113	1,65749184	3,37820933
3	159	2,66892615	4,217874059
4	194	3,79613696	4,271286751
5	218	4,870572606	4,170751218
6	259	5,925177198	3,850392059
7	261	6,971594147	3,409541885
8	350	8,081070753	4,627632275
9	309	9,119211367	4,882191658
10	260	10,17306417	4,756629867
11	281	11,28916286	4,602582268
12	285	12,37990014	6,374496426
13	267	13,46569205	5,708639663
14	261	14,46359164	5,539319332
15	243	15,59193618	5,077268455

Berdasarkan Tabel 4.4 secara manual (hasil pengamatan mata) dapat diduga nilai variabel *model fit variogram* yaitu $range = 0,569$ dan $sill = 4,197$ (data nomor 1) namun tetap harus diuji nilai dugaan tersebut apakah menghasilkan nilai *Sum Squared Error* (SSE) terkecil.

Penentuan nilai variabel *model fit variogram* selain menggunakan hasil pengamatan langsung dapat juga menggunakan teknik *Auto Fit Variogram* yang secara otomatis dihasilkan nilai variabel model yaitu model = Ste, psill = 4,40265, range = 0,39723, nugget = 0 dan kappa = 1,6.

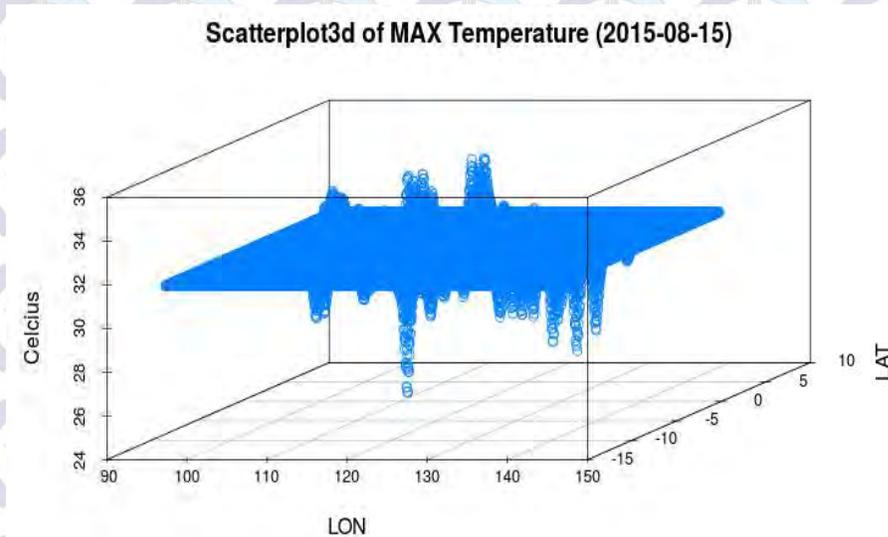


Gambar 4.4. Grafik Model Fit Variogram Max Temperature

Pada Gambar 4.4 menggambarkan hasil akhir *model fit variogram* yaitu variabel psill (garis mendatar) memiliki nilai 4,30474, variabel nugget (titik awal grafik garis) memiliki nilai 1,09781, variabel range (nilai jarak sejak titik awal hingga garis mendatar) memiliki nilai 0,40047 dan variabel kappa memiliki nilai 1,6. Hasil akhir memiliki nilai mendekati nilai hasil proses *Auto Fit Variogram*. Sedangkan nilai *Sum Squared Error* (SSE) adalah 337,44521.

Setelah didapatkan nilai *model fit variogram* dapat dilanjutkan dengan proses pembuatan *Output Grid* seluruh wilayah Indonesia dengan koordinat *Longitudinal* antara 94 hingga 142 dan koordinat *Lattitude* antara -12 hingga 7. Sedangkan resolusi data grid yang digunakan adalah 0,1 x 0,1. Pembuatan data grid bertujuan untuk memberikan batasan lokasi yang akan diolah menggunakan metode interpolasi kriging.

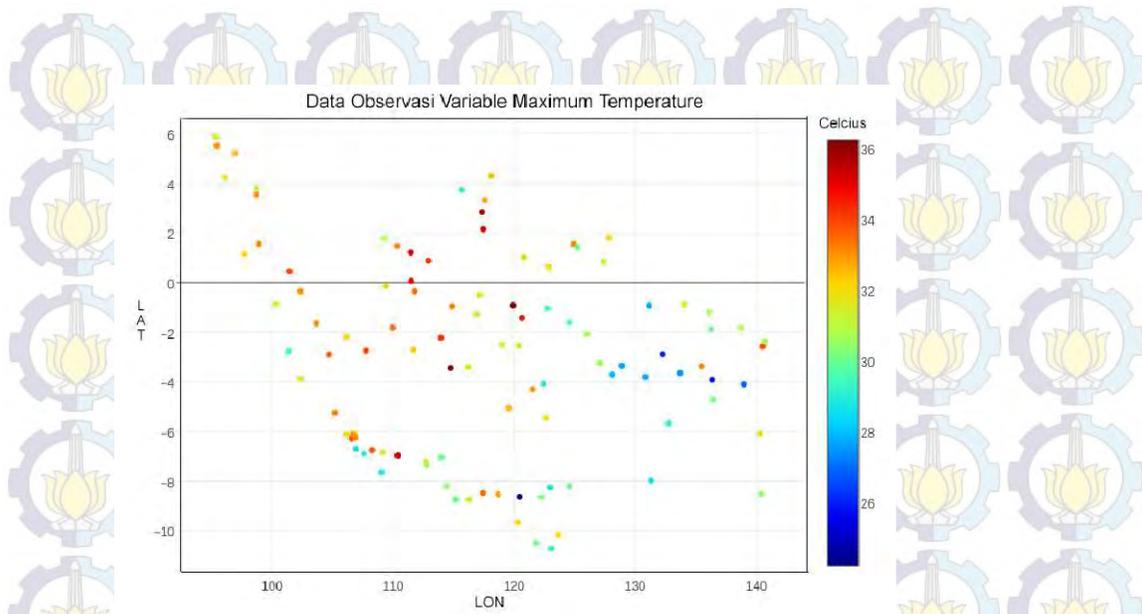
Proses selanjutnya adalah analisa metode interpolasi kriging berdasarkan data *Input Point*, *Model Fit Variogram* dan *Output Grid* yang telah didefinisikan sebelumnya. Hasil prediksi keseluruhan grid (seluruh wilayah Indonesia) dapat digambarkan dengan grafik scatterplot3d *Max Temperatur* pada Gambar 4.5.



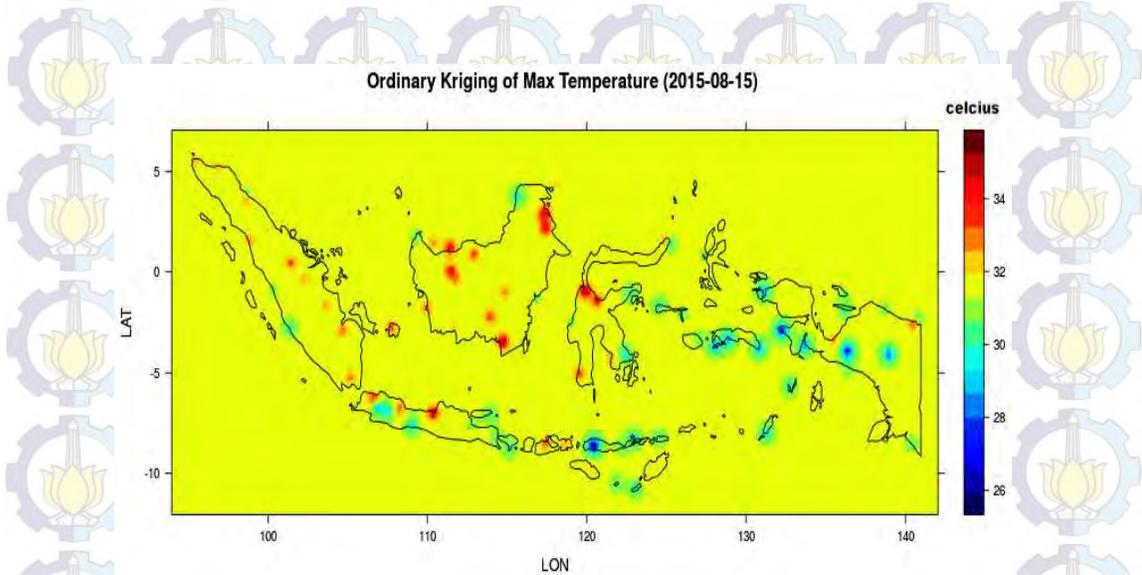
Gambar 4.5. Grafik Scatterplot3d Variabel Max Temperature

Berdasarkan Gambar 4.5 terlihat hasil prediksi variabel cuaca *Maximum Temperature* mempunyai nilai rata-rata 32 derajat celcius (terlihat dari luasan bidang mendatar yang cukup luas). Juga terlihat nilai terkecil dengan nilai berkisar 24 derajat celcius dan nilai terbesar dengan nilai berkisar 36 derajat celcius. Maka dapat diduga hasil *plotting raster data* bersifat tidak kontinyu.

Tingkat akurasi hasil prediksi dapat diukur dengan metode *Cross Validation* namun secara visual hasil *plotting raster data* hasil prediksi dapat dibandingkan dengan hasil *plotting data* hasil observasi. Pemakaian skala gradasi warna (*color pallete*) yang sama dapat memberikan gambaran perbandingan penyebaran warna pada grafik data hasil prediksi (Gambar 4.6) dengan grafik data observasi (Gambar 4.7).



Gambar 4.6. Visualisasi Data Observasi Max temperature

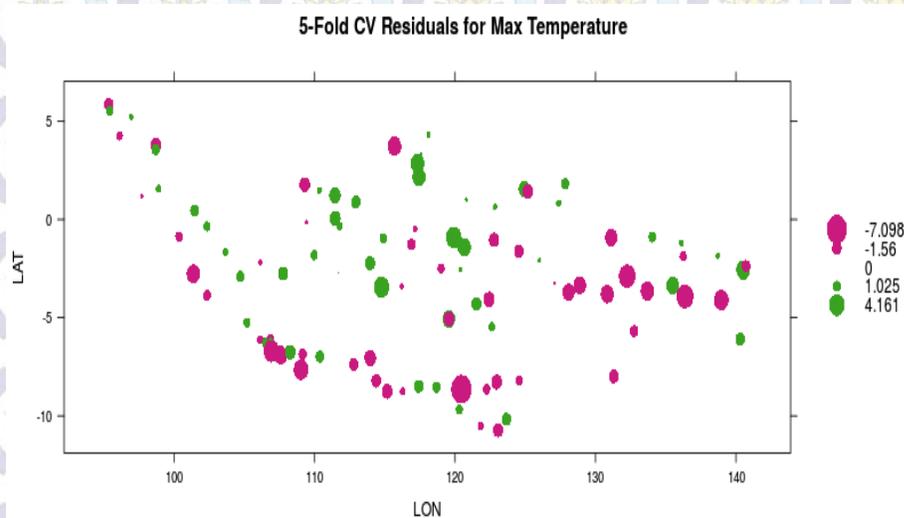


Gambar 4.7. Visualisasi Variabel Cuaca Max Temperature

Pada Gambar 4.6 dan Gambar 4.7 mempunyai kesamaan distribusi warna biru disekitar daerah Pulau Jawa dan Pulau Papua. Pada Gambar 4.7 *data raster* hasil prediksi terlihat tidak kontinyu.

Pengukuran keakuratan hasil visualisasi umumnya menggunakan metode *cross validation* dengan memanfaatkan rangkuman hasil nilai residual (beda nilai

prediksi dengan nilai hasil observasi). Berdasarkan nilai residual didapatkan nilai *error* untuk mengukur tingkat keakuratan hasil prediksi sebuah *predictive model*.



Gambar 4.8. Visualisasi Nilai Residual Max Temperature

Pada Gambar 4.8 terlihat lokasi Pulau Jawa dan Pulau Papua memiliki nilai residual tinggi yang artinya hasil prediksinya kurang akurat dibandingkan dengan daerah lainnya.

Tabel 4.5. Rangkuman Error dan Korelasi Visualisasi Max Temperature

No	Nama Parameter Kinerja	Nilai
1	Mean Error (ME) Nilai Residual	-0,2633828957143
2	Mean Squared Prediction Error (MSPE) Nilai Residual	3,94653510527702
3	Mean Squared Normalized Error (MSNE)	0,769679947599303
4	Korelasi Observasi &Prediksi (COP)	0,501067560994295
5	Korelasi Prediksi& Nilai Residual (CPR)	0,109897013884282
6	Sum Squared Error (SSE) Fit Variogram	337,445217710806

Pada Tabel 4.5 terlihat rata-rata nilai *error* cukup rendah kecuali nilai MSPE dan nilai SSE yang agak tinggi.

4.3. Hasil Visualisasi Variabel Cuaca Min Temperature

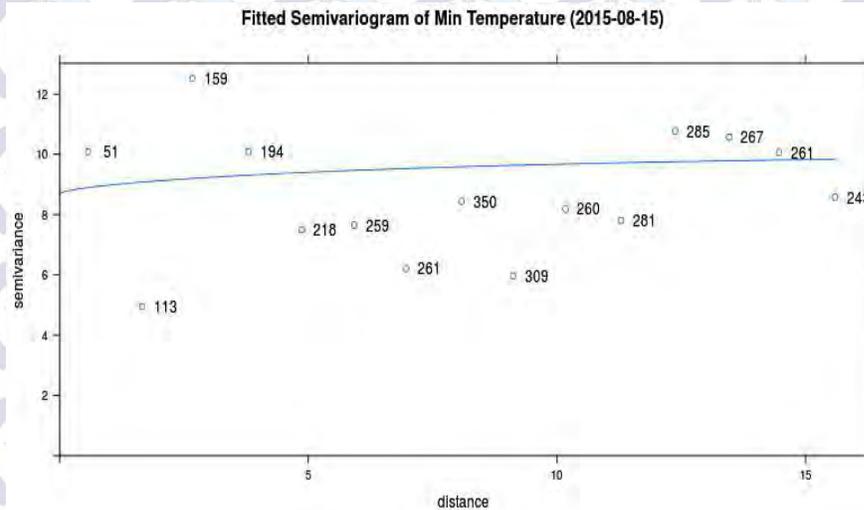
Proses visualisasi variabel cuaca *Minimum Temperature* diawali dengan proses penyaringan data yang akan digunakan dipastikan tidak memiliki data kosong (NA) yang kemudian disimpan ke dalam sebuah *Dataframe Input Point*. Setelah itu dilakukan analisa *Spatial Auto Correlation* menggunakan teknik variogram yang memetakan jarak antar semua titik terhadap nilai data.

Tabel 4.6. Nilai Variogram Variabel Min Temperature

No	NP	Distance	Gamma
1	51	0,568614327	10.08632898
2	113	1,65749184	4,95446848
3	159	2,66892615	12,52135259
4	194	3,79613696	10,08443108
5	218	4,870572606	7,492722845
6	259	5,925177198	7,653826445
7	261	6,971594147	6,204602431
8	350	8,081070753	8,439620811
9	309	9,119211367	5,96584462
10	260	10,17306417	8,189707977
11	281	11,28916286	7,809257063
12	285	12,37990014	10,77296946
13	267	13,46569205	10,57161187
14	261	14,46359164	10,06619838
15	243	15,59193618	8,577713001

Berdasarkan Tabel 4.6 secara manual (hasil pengamatan mata) dapat diduga nilai variabel *model fit variogram* yaitu $range = 0,569$ dan $sill = 10.086$ (data nomor 1) namun tetap harus diuji nilai dugaan tersebut apakah menghasilkan nilai *Sum Squared Error (SSE)* terkecil.

Penentuan nilai variabel *model fit variogram* selain menggunakan hasil pengamatan langsung dapat juga menggunakan teknik *Auto Fit Variogram* yang secara otomatis dihasilkan nilai variabel model yaitu model = Ste, psill = 9,38014, range = 0,82623, nugget = 0 dan kappa = 0,3.

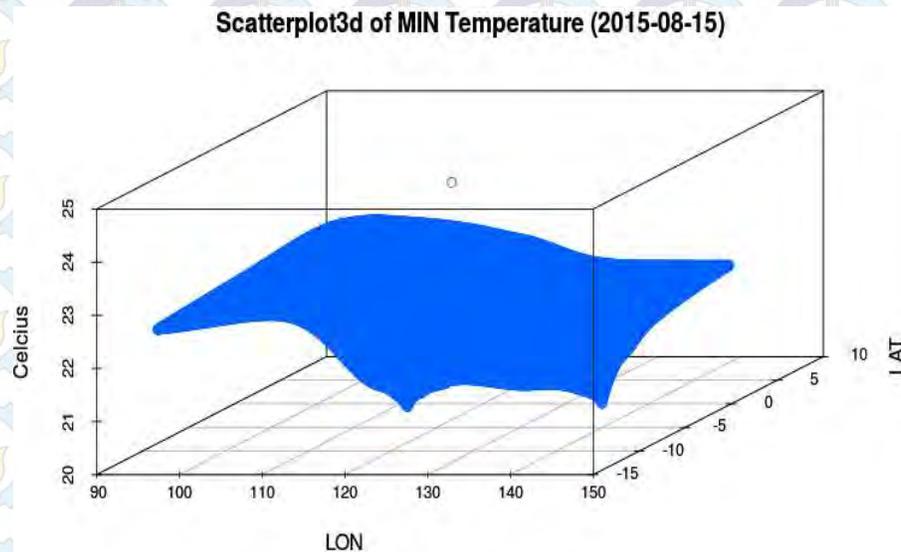


Gambar 4.9. Grafik Model Fit Variogram Min Temperature

Pada Gambar 4.9 menggambarkan hasil akhir *model fit variogram* yaitu variabel psill (garis mendatar) memiliki nilai 1,3961, variabel nugget (titik awal grafik garis) memiliki nilai 8,68961, variabel range (nilai jarak sejak titik awal hingga garis mendatar) memiliki nilai 13,56011 dan variabel kappa memiliki nilai 0,3. Hasil akhir memiliki nilai mendekati nilai hasil proses *Auto Fit Variogram*. Sedangkan nilai *Sum Squared Error (SSE)* adalah 1369.1492.

Setelah didapatkan nilai *model fit variogram* dapat dilanjutkan dengan proses pembuatan *Output Grid* seluruh wilayah Indonesia dengan koordinat *Longitudinal* antara 94 hingga 142 dan koordinat *Lattitude* antara -12 hingga 7. Sedangkan resolusi data grid yang digunakan adalah 0,1 x 0,1. Pembuatan data grid bertujuan untuk memberikan batasan lokasi yang akan diolah menggunakan metode interpolasi kriging.

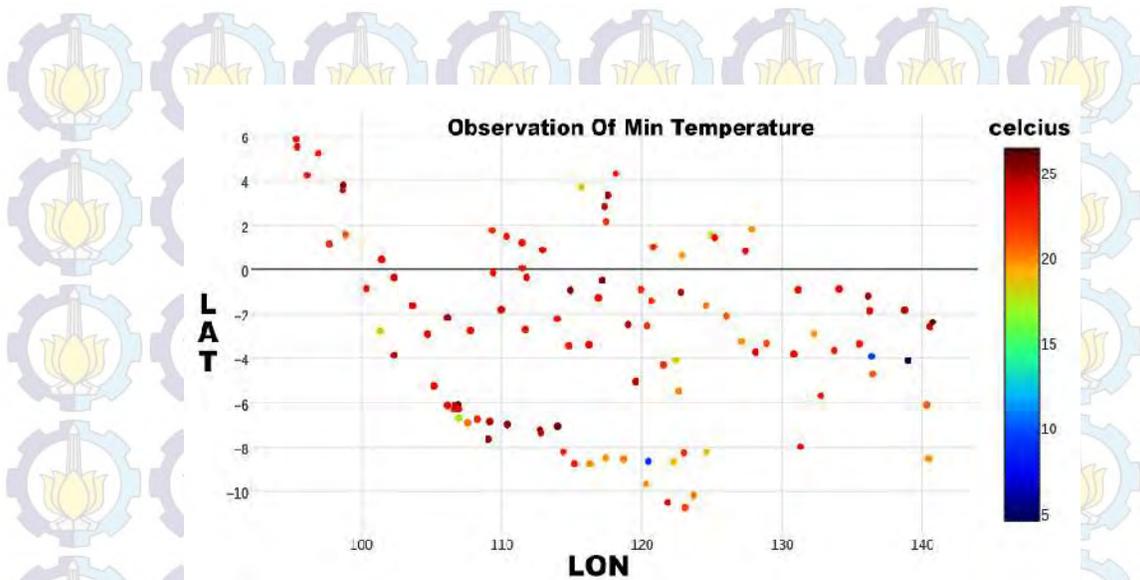
Proses selanjutnya adalah analisa metode interpolasi kriging berdasarkan data *Input Point*, *Model Fit Variogram* dan *Output Grid* yang telah didefinisikan sebelumnya. Hasil prediksi keseluruhan grid (seluruh wilayah Indonesia) dapat digambarkan dengan grafik *scatterplot3d Min Temperatur* pada Gambar 4.10.



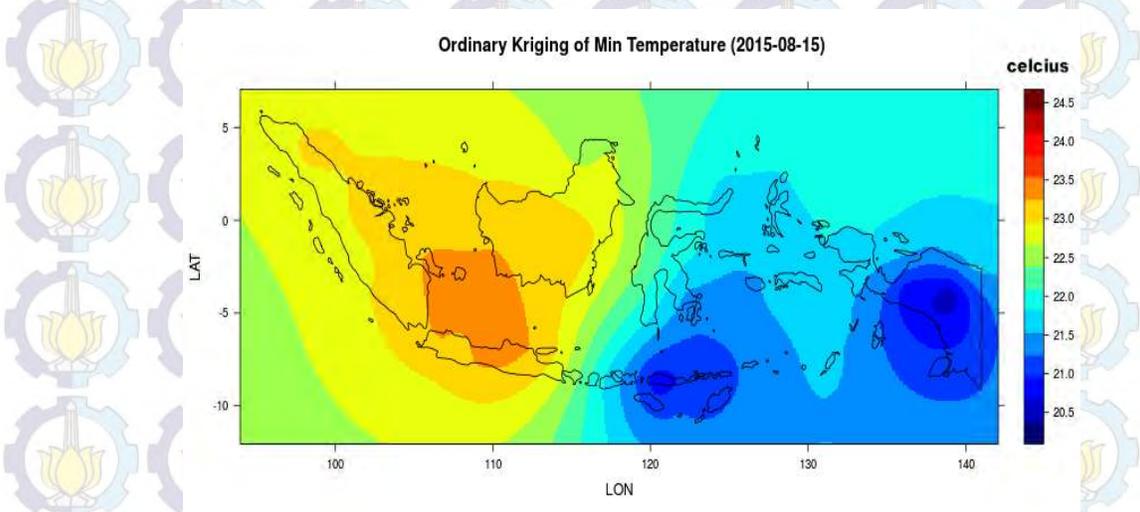
Gambar 4.10. Grafik Scatterplot3d Variabel Min Temperature

Berdasarkan Gambar 4.10 terlihat hasil prediksi variabel cuaca *Minimum Temperature* mempunyai nilai rata-rata 29 derajat celcius (terlihat dari luasan bidang mendatar yang tidak luas). Juga terlihat nilai terkecil dengan nilai berkisar 21 derajat celcius dan nilai terbesar dengan nilai berkisar 25 derajat celcius. Maka dapat diduga hasil *plotting raster data* bersifat kontinyu.

Tingkat akurasi hasil prediksi dapat diukur dengan metode *Cross Validation* namun secara visual hasil *plotting raster data* hasil prediksi dapat dibandingkan dengan hasil *plotting data* hasil observasi. Pemakaian skala gradasi warna (*color pallette*) yang sama dapat memberikan gambaran perbandingan penyebaran warna pada grafik data hasil prediksi (Gambar 4.11) dengan grafik data observasi (Gambar 4.12).



Gambar 4.11. Visualisasi Data Observasi Min temperature

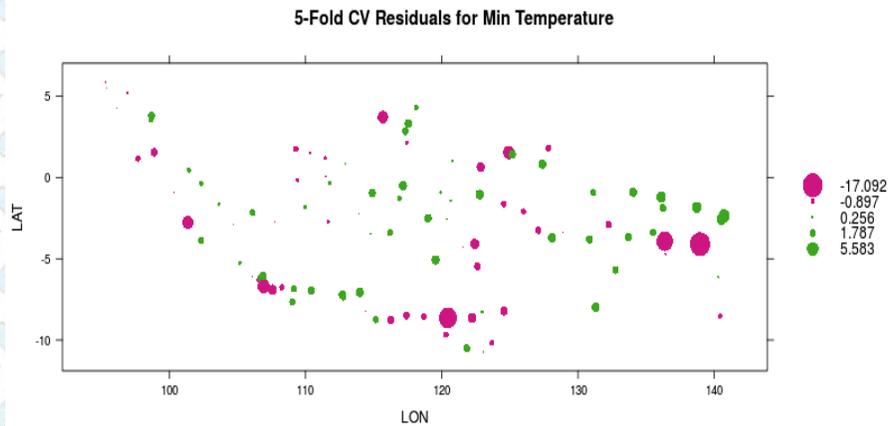


Gambar 4.12. Visualisasi Variabel Cuaca Min Temperature

Pada Gambar 4.11 dan Gambar 4.12 mempunyai kesamaan distribusi warna oranye disekitar daerah Kalimantan dan Sumatera. Pada Gambar 4.12 *raster data* hasil prediksi terlihat kontinyu.

Pengukuran keakuratan hasil visualisasi umumnya menggunakan metode *cross validation* dengan memanfaatkan rangkuman hasil nilai residual (beda nilai

prediksi dengan nilai hasil observasi). Berdasarkan nilai residual didapatkan nilai *error* untuk mengukur tingkat keakuratan hasil prediksi sebuah *predictive model*.



Gambar 4.13. Visualisasi Nilai Residual Min Temperature

Pada Gambar 4.13 terlihat lokasi Pulau Jawa dan Pulau Papua memiliki nilai residual tinggi yang artinya hasil prediksinya kurang akurat dibandingkan dengan daerah lainnya.

Tabel 4.7. Rangkuman Error dan Korelasi Visualisasi Min Temperature

No	Nama Parameter Kinerja	Nilai
1	Mean Error (ME) Nilai Residual	-0.07130256592272
2	Mean Squared Prediction Error (MSPE) Nilai Residual	9.76604837093814
3	Mean Squared Normalized Error (MSNE)	1.00779818348819
4	Korelasi Observasi &Prediksi (COP)	0.294452088091784
5	Korelasi Prediksi& Nilai Residual (CPR)	0.001362269834185
6	Sum Squared Error (SSE) Fit Variogram	1369.14928903451

Pada Tabel 4.7 terlihat rata-rata nilai *error* cukup rendah kecuali nilai MSPE dan nilai SSE yang agak tinggi.

4.4. Hasil Visualisasi Variabel Cuaca Dew Point

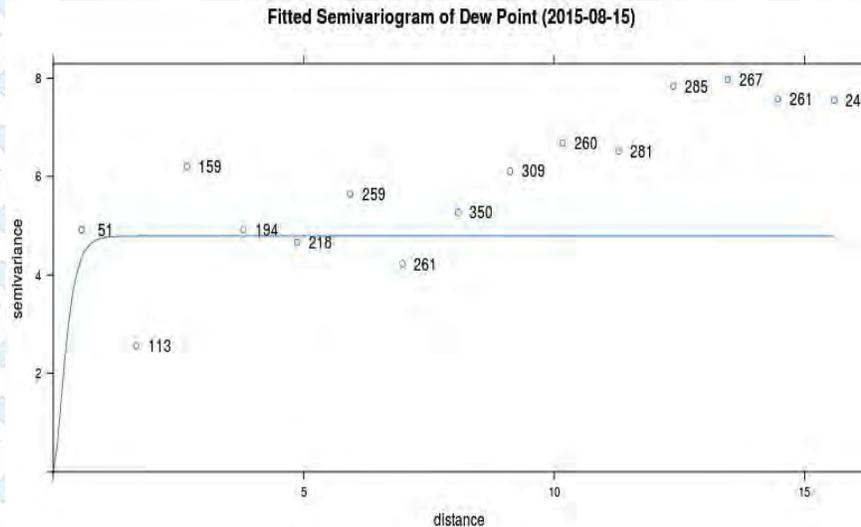
Proses visualisasi variabel cuaca *Dew Point* diawali dengan proses penyaringan data yang akan digunakan dipastikan tidak memiliki data kosong (NA) yang kemudian disimpan ke dalam sebuah *Dataframe Input Point*. Setelah itu dilakukan analisa *Spatial Auto Correlation* menggunakan teknik variogram yang memetakan jarak antar semua titik terhadap nilai data.

Tabel 4.8. Nilai Variogram Variabel Dew Point

No	NP	Distance	Gamma
1	51	0,568614327	4,917816509
2	113	1,65749184	2,563271605
3	159	2,66892615	6,206576598
4	194	3,79613696	4,923690658
5	218	4,870572606	4,664217352
6	259	5,925177198	5,648684399
7	261	6,971594147	4,229341091
8	350	8,081070753	5,270410053
9	309	9,119211367	6,10875884
10	260	10,17306417	6,679837369
11	281	11,28916286	6,519754185
12	285	12,37990014	7,840632445
13	267	13,46569205	7,973707634
14	261	14,46359164	7,583694007
15	243	15,59193618	7,553967891

Berdasarkan Tabel 4.8 secara manual (hasil pengamatan mata) dapat diduga nilai variabel *model fit variogram* yaitu $\text{range} = 0,569$ dan $\text{sill} = 4,917$ (data nomor 1) namun tetap harus diuji nilai dugaan tersebut apakah menghasilkan nilai *Sum Squared Error* (SSE) terkecil.

Penentuan nilai variabel *model fit variogram* selain menggunakan hasil pengamatan langsung dapat juga menggunakan teknik *Auto Fit Variogram* yang secara otomatis dihasilkan nilai variabel model yaitu model = Ste, psill = 4,76326, range = 0,34871, nugget = 0.02692 dan kappa = 1,6.

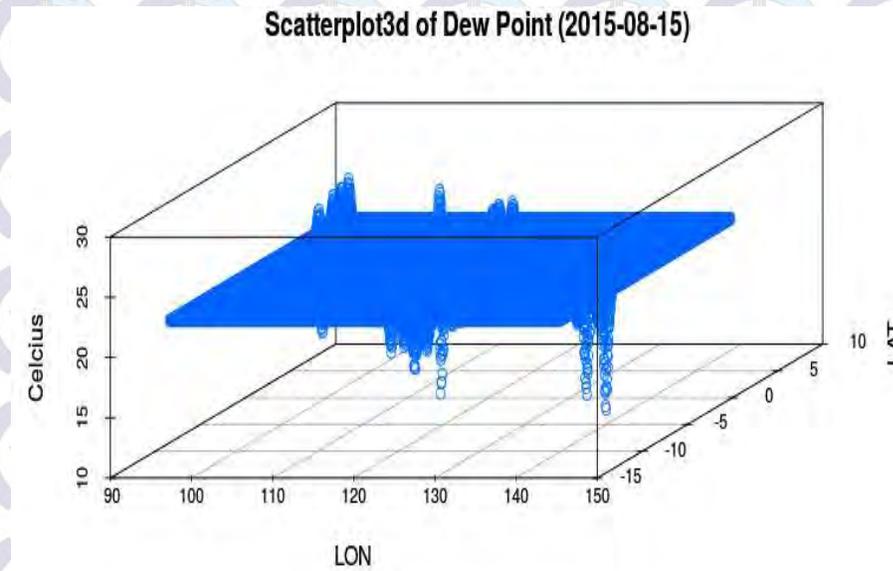


Gambar 4.14. Grafik Model Fit Variogram Dew Point

Pada Gambar 4.14 menggambarkan hasil akhir *model fit variogram* yaitu variabel psill (garis mendatar) memiliki nilai 4.75487, variabel nugget (titik awal grafik garis) memiliki nilai 0,03531, variabel range (nilai jarak sejak titik awal hingga garis mendatar) memiliki nilai 0.34890 dan variabel kappa memiliki nilai 0,3. Hasil akhir memiliki nilai mendekati nilai hasil proses *Auto Fit Variogram*. Sedangkan nilai *Sum Squared Error* (SSE) adalah 378,98336.

Setelah didapatkan nilai *model fit variogram* dapat dilanjutkan dengan proses pembuatan *Output Grid* seluruh wilayah Indonesia dengan koordinat *Longitudinal* antara 94 hingga 142 dan koordinat *Lattitude* antara -12 hingga 7. Sedangkan resolusi data grid yang digunakan adalah 0,1 x 0,1. Pembuatan data grid bertujuan untuk memberikan batasan lokasi yang akan diolah menggunakan metode interpolasi kriging.

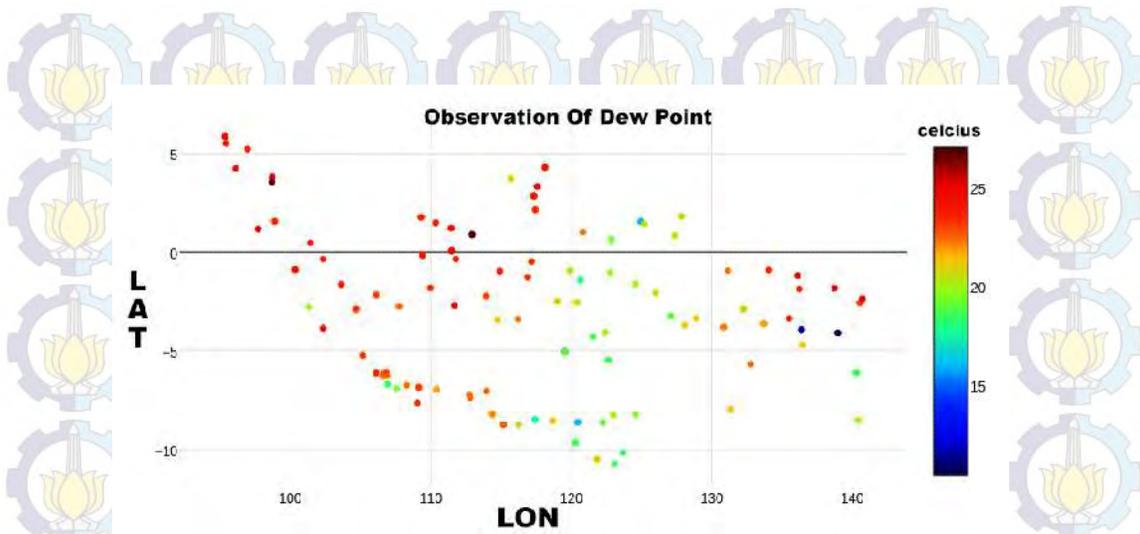
Proses selanjutnya adalah analisa metode interpolasi kriging berdasarkan data *Input Point*, *Model Fit Variogram* dan *Output Grid* yang telah didefinisikan sebelumnya. Hasil prediksi keseluruhan grid (seluruh wilayah Indonesia) dapat digambarkan dengan grafik scatterplot3d *Dew Point* pada Gambar 4.15.



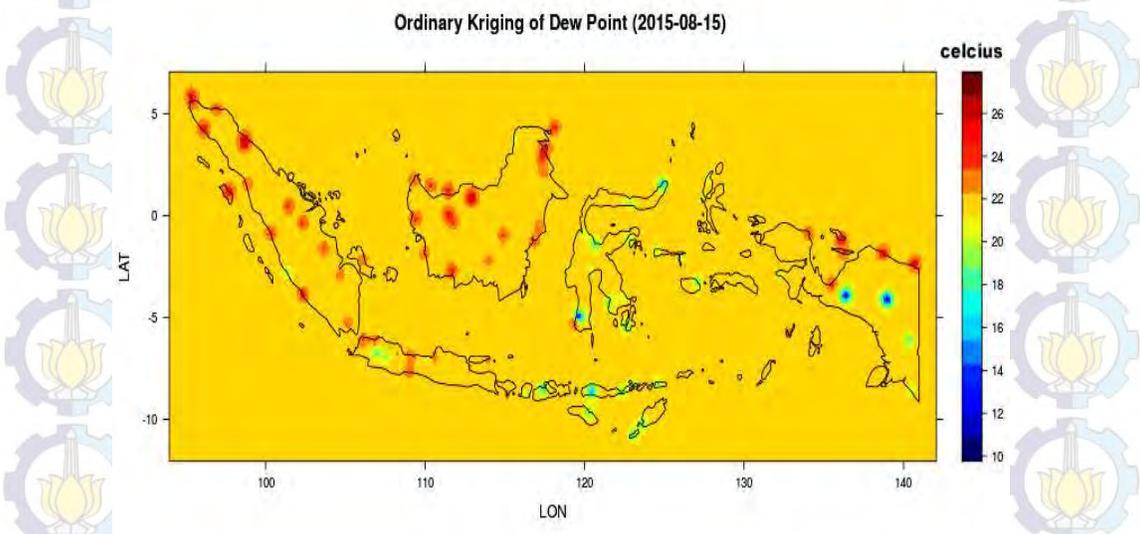
Gambar 4.15. Grafik Scatterplot3d Variabel Dew Point

Berdasarkan Gambar 4.15 terlihat hasil prediksi variabel cuaca *Dew Point* mempunyai nilai rata-rata 20 derajat celcius (terlihat dari luasan bidang mendatar yang cukup luas). Juga terlihat nilai terkecil dengan nilai berkisar 15 derajat celcius dan nilai terbesar dengan nilai berkisar 29 derajat celcius. Maka dapat diduga hasil *plotting raster data* bersifat tidak kontinyu.

Tingkat akurasi hasil prediksi dapat diukur dengan metode *Cross Validation* namun secara visual hasil *plotting raster data* hasil prediksi dapat dibandingkan dengan hasil *plotting data* hasil observasi. Pemakaian skala gradasi warna (*color pallette*) yang sama dapat memberikan gambaran perbandingan penyebaran warna pada grafik data hasil prediksi (Gambar 4.16) dengan grafik data observasi (Gambar 4.17).



Gambar 4.16. Visualisasi Data Observasi Dew Point

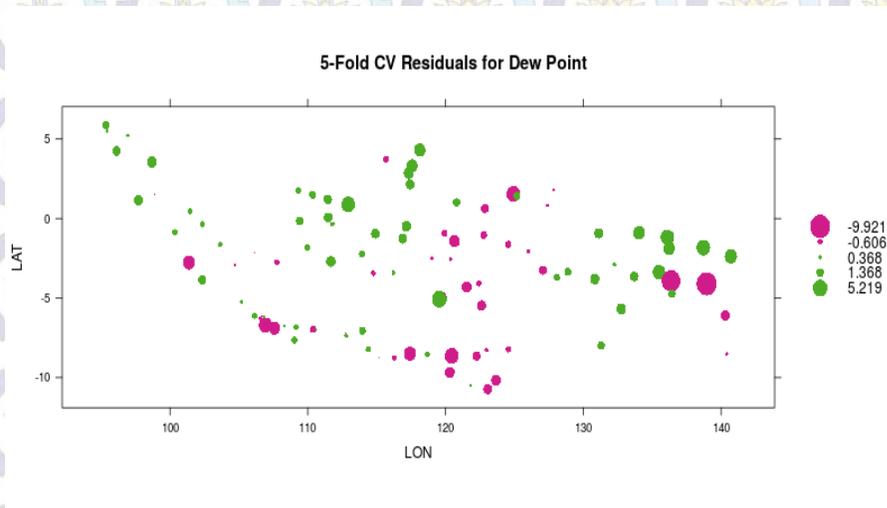


Gambar 4.17. Visualisasi Variabel Cuaca Dew Point

Pada Gambar 4.16 dan Gambar 4.17 mempunyai kesamaan distribusi warna merah disekitar daerah Kalimantan dan Sumatera. Pada Gambar 4.17 *raster data* hasil prediksi terlihat tidak kontinyu.

Pengukuran keakuratan hasil visualisasi umumnya menggunakan metode *cross validation* dengan memanfaatkan rangkuman hasil nilai residual (beda nilai

prediksi dengan nilai hasil observasi). Berdasarkan nilai residual didapatkan nilai *error* untuk mengukur tingkat keakuratan hasil prediksi sebuah *predictive model*.



Gambar 4.18. Visualisasi Nilai Residual Dew Point

Pada Gambar 4.18 terlihat lokasi Pulau Jawa dan Pulau Papua memiliki nilai residual tinggi yang artinya hasil prediksinya kurang akurat dibandingkan dengan daerah lainnya.

Tabel 4.9. Rangkuman Error dan Korelasi Visualisasi Dew Point

No	Nama Parameter Kinerja	Nilai
1	Mean Error (ME) Nilai Residual	0.17118565473626
2	Mean Squared Prediction Error (MSPE) Nilai Residual	5.54827322376675
3	Mean Squared Normalized Error (MSNE)	7.6435574530826
4	Korelasi Observasi &Prediksi (COP)	0.540802202795339
5	Korelasi Prediksi& Nilai Residual (CPR)	0.03180611643479
6	Sum Squared Error (SSE) Fit Variogram	378.98336027038

Pada Tabel 4.9 terlihat rata-rata nilai *error* cukup rendah kecuali nilai MSPE dan nilai SSE yang agak tinggi.

4.5. Hasil Visualisasi Variabel Cuaca Station Pressure

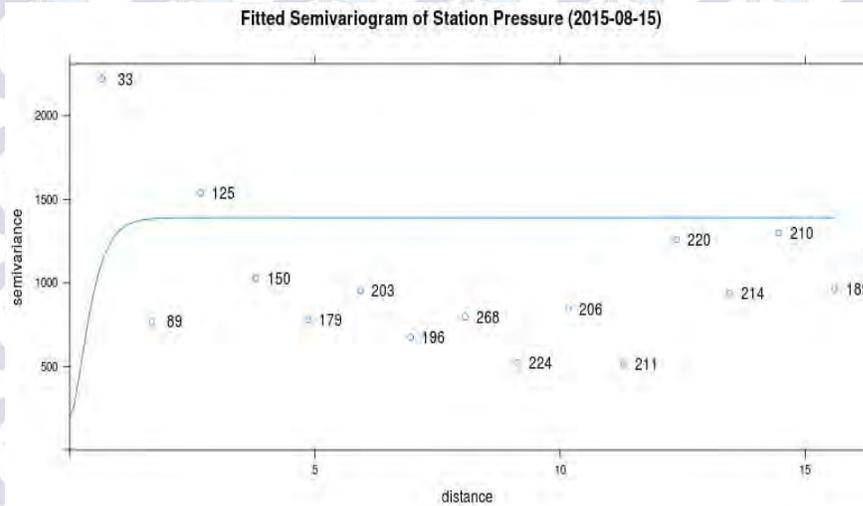
Proses visualisasi variabel cuaca *Station Pressure* diawali dengan proses penyaringan data yang akan digunakan dipastikan tidak memiliki data kosong (NA) yang kemudian disimpan ke dalam sebuah *Dataframe Input Point*. Setelah itu dilakukan analisa *Spatial Auto Correlation* menggunakan teknik variogram yang memetakan jarak antar semua titik terhadap nilai data.

Tabel 4.10. Nilai Variogram Variabel Station Pressure

No	NP	Distance	Gamma
1	33	0,664807478	2222,241061
2	89	1,676956439	769,0260112
3	125	2,668747406	1539,27808
4	150	3,797717203	1027,817467
5	179	4,860799383	780,5170391
6	203	5,932835258	956,3591379
7	196	6,952716116	674,6383929
8	268	8,066680333	798,4053545
9	224	9,136133275	521,4403795
10	206	10,16635527	845,1457524
11	211	11,29262183	511,8470379
12	220	12,37718328	1260,022886
13	214	13,46357276	936,7196262
14	210	14,45519849	1299,68181
15	189	15,59741244	965,6237302

Berdasarkan Tabel 4.10 secara manual (hasil pengamatan mata) dapat diduga nilai variabel *model fit variogram* yaitu $range = 0,6648$ dan $psill = 2222,241$ (data nomor 1) namun tetap harus diuji nilai dugaan tersebut apakah menghasilkan nilai *Sum Squared Error* (SSE) terkecil.

Penentuan nilai variabel *model fit variogram* selain menggunakan hasil pengamatan langsung dapat juga menggunakan teknik *Auto Fit Variogram* yang secara otomatis dihasilkan nilai variabel model yaitu model = Ste, psill = 1390,79, range = 0,5331, nugget = 0 dan kappa = 1,6.

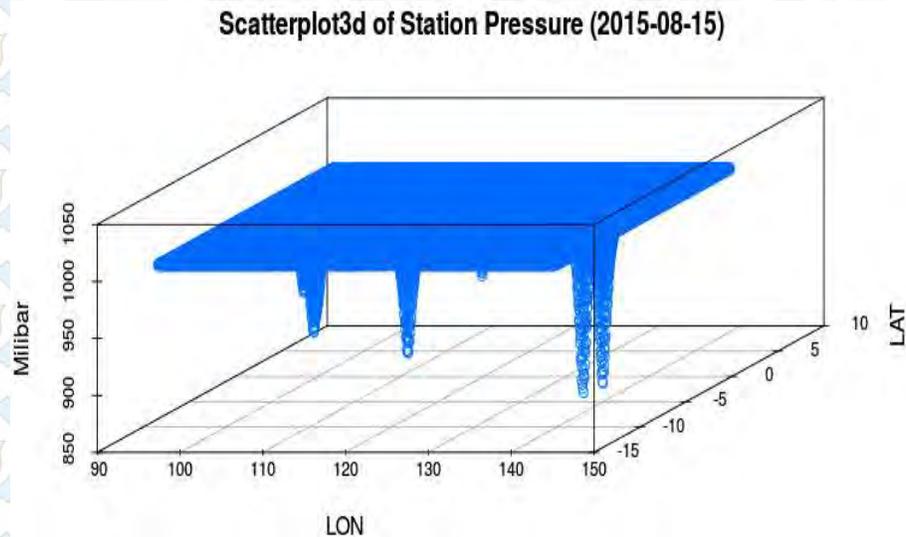


Gambar 4.19. Grafik Model Fit Variogram Station Pressure

Pada Gambar 4.19 menggambarkan hasil akhir *model fit variogram* yaitu variabel psill (garis mendatar) memiliki nilai 1185,75, variabel nugget (titik awal grafik garis) memiliki nilai 204.4291, variabel range (nilai jarak sejak titik awal hingga garis mendatar) memiliki nilai 0,5655 dan variabel kappa memiliki nilai 1,6. Hasil akhir memiliki nilai mendekati nilai hasil proses *Auto Fit Variogram*. Sedangkan nilai *Sum Squared Error* (SSE) adalah 114449683.41.

Setelah didapatkan nilai *model fit variogram* dapat dilanjutkan dengan proses pembuatan *Output Grid* seluruh wilayah Indonesia dengan koordinat *Longitudinal* antara 94 hingga 142 dan koordinat *Lattitude* antara -12 hingga 7. Sedangkan resolusi data grid yang digunakan adalah 0,1 x 0,1. Pembuatan data grid bertujuan untuk memberikan batasan lokasi yang akan diolah menggunakan metode interpolasi kriging.

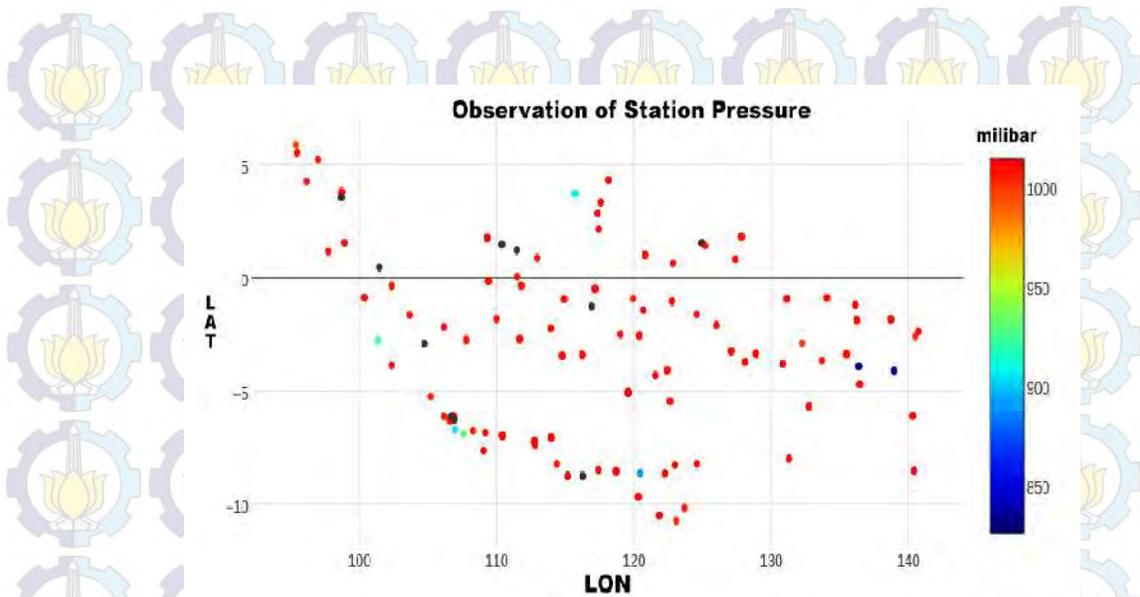
Proses selanjutnya adalah analisa metode interpolasi kriging berdasarkan data *Input Point*, *Model Fit Variogram* dan *Output Grid* yang telah didefinisikan sebelumnya. Hasil prediksi keseluruhan grid (seluruh wilayah Indonesia) dapat digambarkan dengan grafik *scatterplot3d Station Pressure* pada Gambar 4.20.



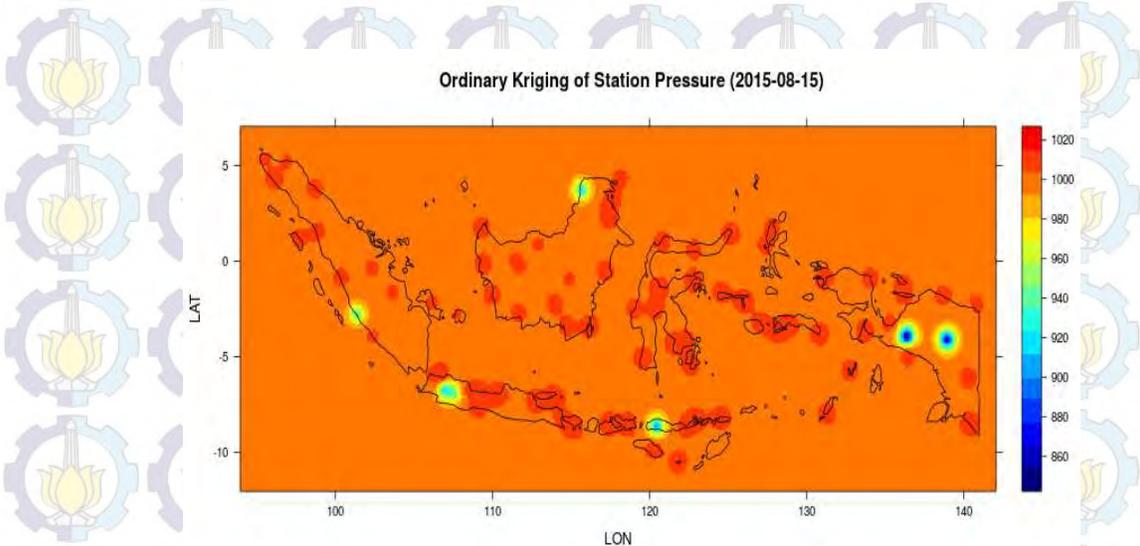
Gambar 4.20. Grafik Scatterplot3d Variabel Station Pressure

Berdasarkan Gambar 4.20 terlihat hasil prediksi variabel cuaca *Station Pressure* mempunyai nilai rata-rata 1000 milibar (terlihat dari luasan bidang mendatar yang cukup luas). Juga terlihat nilai terkecil dengan nilai berkisar 900 milibar dan nilai terbesar dengan nilai berkisar 1000 milibar. Maka dapat diduga hasil *plotting raster data* bersifat tidak kontinyu.

Tingkat akurasi hasil prediksi dapat diukur dengan metode *Cross Validation* namun secara visual hasil *plotting raster data* hasil prediksi dapat dibandingkan dengan hasil *plotting data* hasil observasi. Pemakaian skala gradasi warna (*color pallete*) yang sama dapat memberikan gambaran perbandingan penyebaran warna pada grafik data hasil prediksi (Gambar 4.21) dengan grafik data observasi (Gambar 4.22).



Gambar 4.21. Visualisasi Data Observasi Station Pressure

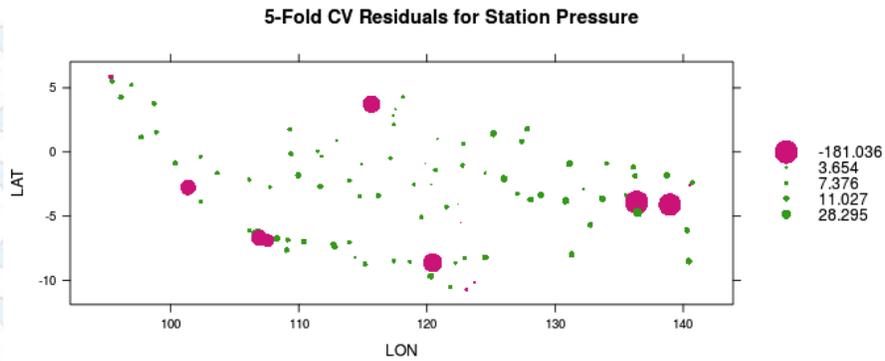


Gambar 4.22. Visualisasi Variabel Cuaca Station Pressure

Pada Gambar 4.21 dan Gambar 4.22 mempunyai kesamaan distribusi warna merah merata diseluruh pulau. Pada Gambar 4.22 *raster data* hasil prediksi terlihat tidak kontinyu.

Pengukuran keakuratan hasil visualisasi umumnya menggunakan metode *cross validation* dengan memanfaatkan rangkuman hasil nilai residual (beda nilai

prediksi dengan nilai hasil observasi). Berdasarkan nilai residual didapatkan nilai *error* untuk mengukur tingkat keakuratan hasil prediksi sebuah *predictive model*.



Gambar 4.23. Visualisasi Nilai Residual Station Pressure

Pada Gambar 4.23 terlihat lokasi Pulau Jawa dan Pulau Papua memiliki nilai residual tinggi yang artinya hasil prediksinya kurang akurat dibandingkan dengan daerah lainnya.

Tabel 4.11. Rangkuman Error dan Korelasi Visualisasi Station Pressure

No	Nama Parameter Kinerja	Nilai
1	Mean Error (ME) Nilai Residual	-0.9966286728421
2	Mean Squared Prediction Error (MSPE) Nilai Residual	1163.52139080968
3	Mean Squared Normalized Error (MSNE)	0.835819445649942
4	Korelasi Observasi &Prediksi (COP)	-0.04018904350328
5	Korelasi Prediksi& Nilai Residual (CPR)	-0.1861353578785
6	Sum Squared Error (SSE) Fit Variogram	114449683.409992

Pada Tabel 4.11 terlihat rata-rata nilai *error* cukup rendah kecuali nilai MSPE dan nilai SSE yang agak tinggi.

4.6. Hasil Visualisasi Variabel Cuaca Wind Speed

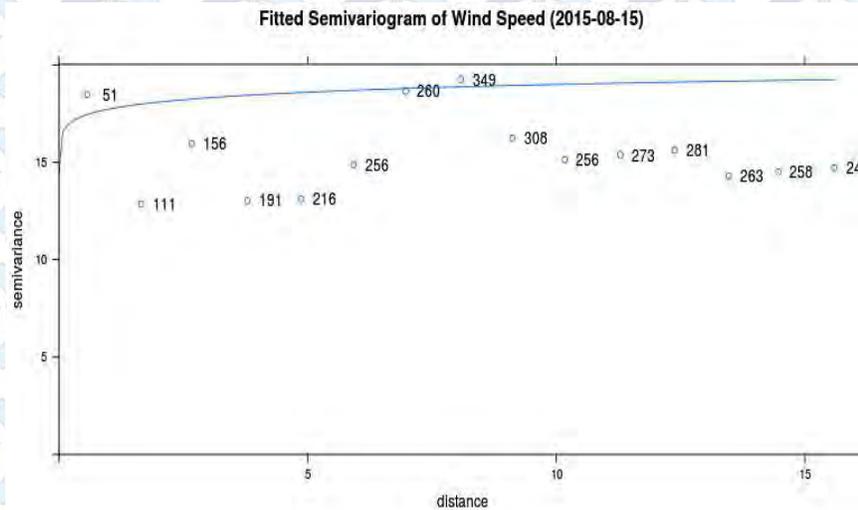
Proses visualisasi variabel cuaca *Wind Speed* diawali dengan proses penyaringan data yang akan digunakan dipastikan tidak memiliki data kosong (NA) yang kemudian disimpan ke dalam sebuah *Dataframe Input Point*. Setelah itu dilakukan analisa *Spatial Auto Correlation* menggunakan teknik variogram yang memetakan jarak antar semua titik terhadap nilai data.

Tabel 4.12. Nilai Variogram Variabel Wind Speed

No	NP	Distance	Gamma
1	51	0.568614327	18.48186275
2	111	1.65798445	12.85058559
3	156	2.669803597	15.94846154
4	191	3.793162661	13.03539267
5	216	4.868200921	13.10638889
6	256	5.925306874	14.86814453
7	260	6.970244916	18.65371154
8	349	8.080892125	19.25571633
9	308	9.118749124	16.23938312
10	256	10.17001569	15.12945313
11	273	11.29102099	15.38232601
12	281	12.378586	15.62199288
13	263	13.46718522	14.29627376
14	258	14.46625737	14.49736434
15	242	15.59222699	14.71491736

Berdasarkan Tabel 4.12 secara manual (hasil pengamatan mata) dapat diduga nilai variabel *model fit variogram* yaitu $range = 6.97024$ dan $psill = 18.6537$ (data nomor 7) namun tetap harus diuji nilai dugaan tersebut apakah menghasilkan nilai *Sum Squared Error* (SSE) terkecil.

Penentuan nilai variabel *model fit variogram* selain menggunakan hasil pengamatan langsung dapat juga menggunakan teknik *Auto Fit Variogram* yang secara otomatis dihasilkan nilai variabel model yaitu model = Ste, psill = 7.15185, range = 10.3476, nugget = 12.5704 dan kappa = 0.05.

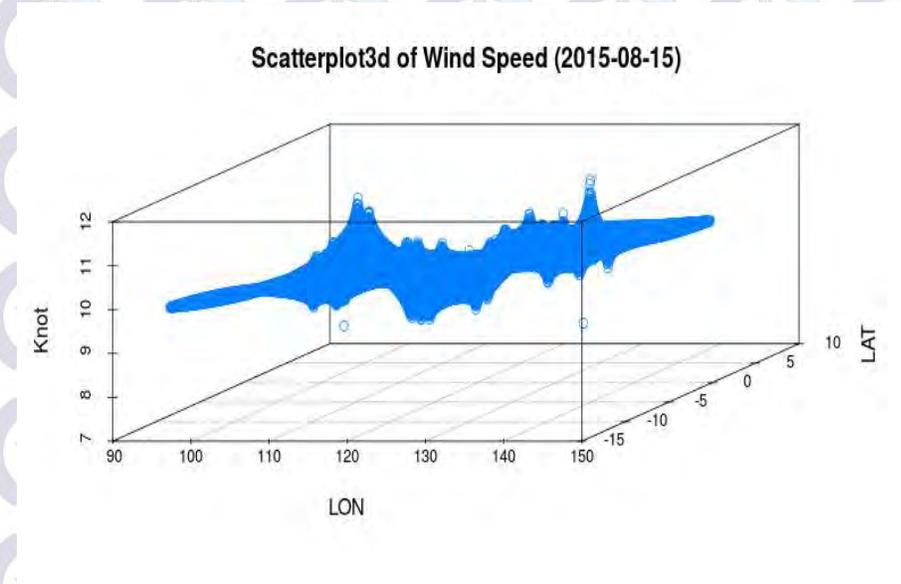


Gambar 4.24. Grafik Model Fit Variogram Wind Speed

Pada Gambar 4.24 menggambarkan hasil akhir *model fit variogram* yaitu variabel psill (garis mendatar) memiliki nilai 7.1922, variabel nugget (titik awal grafik garis) memiliki nilai 12.6911, variabel range (nilai jarak sejak titik awal hingga garis mendatar) memiliki nilai 13.9748 dan variabel kappa memiliki nilai 0.05. Hasil akhir memiliki nilai mendekati nilai hasil proses *Auto Fit Variogram*. Sedangkan nilai *Sum Squared Error* (SSE) adalah 2315.8338.

Setelah didapatkan nilai *model fit variogram* dapat dilanjutkan dengan proses pembuatan *Output Grid* seluruh wilayah Indonesia dengan koordinat *Longitudinal* antara 94 hingga 142 dan koordinat *Lattitude* antara -12 hingga 7. Sedangkan resolusi data grid yang digunakan adalah 0,1 x 0,1. Pembuatan data grid bertujuan untuk memberikan batasan lokasi yang akan diolah menggunakan metode interpolasi kriging.

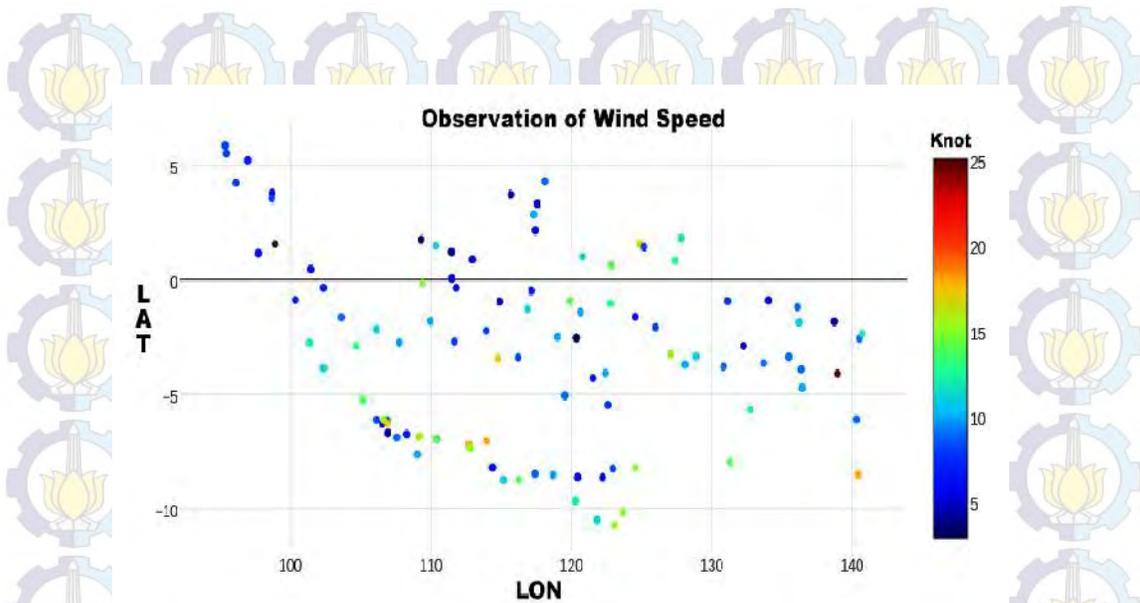
Proses selanjutnya adalah analisa metode interpolasi kriging berdasarkan data *Input Point*, *Model Fit Variogram* dan *Output Grid* yang telah didefinisikan sebelumnya. Hasil prediksi keseluruhan grid (seluruh wilayah Indonesia) dapat digambarkan dengan grafik scatterplot3d *Wind Speed* pada Gambar 4.24.



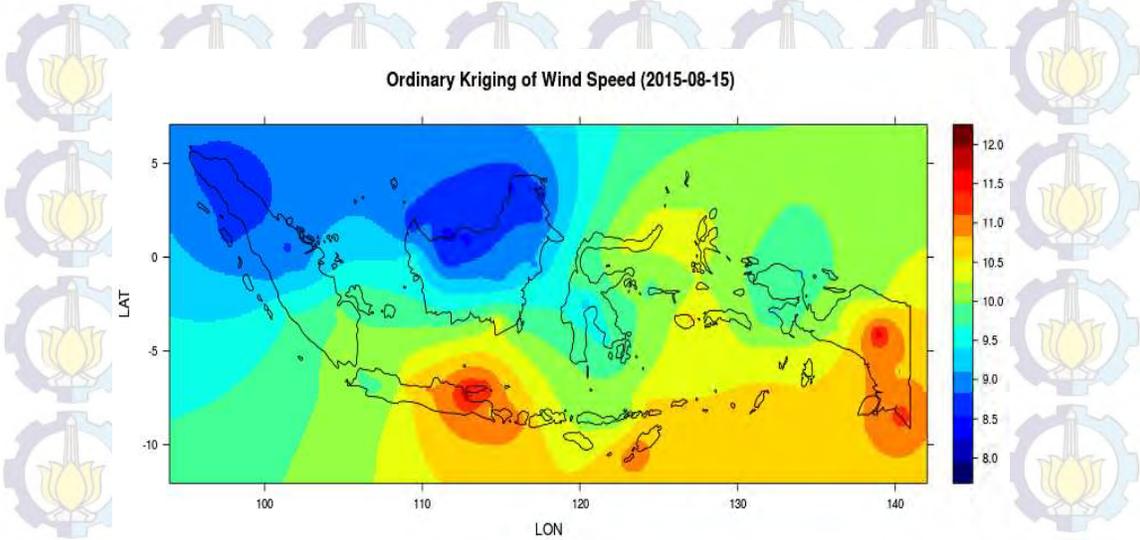
Gambar 4.25. Grafik Scatterplot3d Variabel Wind Speed

Berdasarkan Gambar 4.25 terlihat hasil prediksi variabel cuaca *Wind Speed* mempunyai nilai rata-rata 10 knot (terlihat dari luasan bidang mendatar yang tidak luas). Juga terlihat nilai terkecil dengan nilai berkisar 8 knot dan nilai terbesar dengan nilai berkisar 12 knot. Maka dapat diduga hasil *plotting raster data* bersifat kontinyu.

Tingkat akurasi hasil prediksi dapat diukur dengan metode *Cross Validation* namun secara visual hasil *plotting raster data* hasil prediksi dapat dibandingkan dengan hasil *plotting data* hasil observasi. Pemakaian skala gradasi warna (*color pallete*) yang sama dapat memberikan gambaran perbandingan penyebaran warna pada grafik data hasil prediksi (Gambar 4.26) dengan grafik data observasi (Gambar 4.27).



Gambar 4.26. Visualisasi Data Observasi Wind Speed

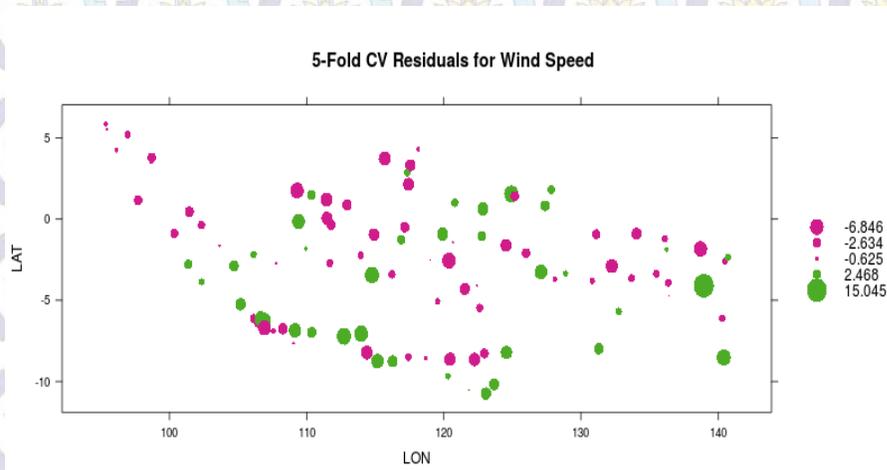


Gambar 4.27. Visualisasi Variabel Cuaca Wind Speed

Pada Gambar 4.26 dan Gambar 4.27 mempunyai kesamaan distribusi warna biru di pulau Sumatra dan Pulau Kalimantan. Pada Gambar 4.27 data *raster* hasil prediksi terlihat kontinyu.

Pengukuran keakuratan hasil visualisasi umumnya menggunakan metode *cross validation* dengan memanfaatkan rangkuman hasil nilai residual (beda nilai

prediksi dengan nilai hasil observasi). Berdasarkan nilai residual didapatkan nilai *error* untuk mengukur tingkat keakuratan hasil prediksi sebuah *predictive model*.



Gambar 4.28. Visualisasi Nilai Residual Wind Speed

Pada Gambar 4.28 terlihat hampir diseluruh lokasi di Indonesia memiliki nilai residual tinggi (kecuali Sumatra) yang artinya diluar Pulau Sumatra hasil prediksinya kurang akurat.

Tabel 4.13. Rangkuman Error dan Korelasi Visualisasi Wind Speed

No	Nama Parameter Kinerja	Nilai
1	Mean Error (ME) Nilai Residual	-0.0039039479838
2	Mean Squared Prediction Error (MSPE) Nilai Residual	15.3083876293672
3	Mean Squared Normalized Error (MSNE)	0.805165672276374
4	Korelasi Observasi &Prediksi (COP)	0.232699718655786
5	Korelasi Prediksi& Nilai Residual (CPR)	0.0591813965919
6	Sum Squared Error (SSE) Fit Variogram	2315.83383765353

Pada Tabel 4.13 terlihat rata-rata nilai *error* cukup rendah kecuali nilai MSPE dan nilai SSE yang agak tinggi.

4.7. Hasil Visualisasi Variabel Cuaca Visibility

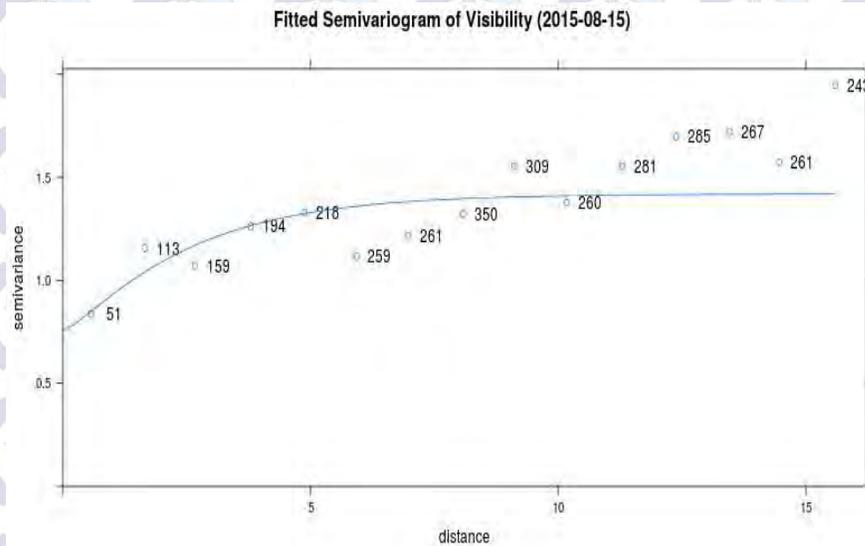
Proses visualisasi variabel cuaca *Visibility* diawali dengan proses penyaringan data yang akan digunakan dipastikan tidak memiliki data kosong (NA) yang kemudian disimpan ke dalam sebuah *Dataframe Input Point*. Setelah itu dilakukan analisa *Spatial Auto Correlation* menggunakan teknik variogram yang memetakan jarak antar semua titik terhadap nilai data.

Tabel 4.14. Nilai Variogram Variabel Visibility

No	NP	Distance	Gamma
1	51	0.568614327	0.836666667
2	113	1.65749184	1.156327434
3	159	2.66892615	1.072044025
4	194	3.79613696	1.262345361
5	218	4.870572606	1.328830275
6	259	5.925177198	1.115945946
7	261	6.971594147	1.214022989
8	350	8.081070753	1.322557143
9	309	9.119211367	1.554805825
10	260	10.17306417	1.377365385
11	281	11.28916286	1.55366548
12	285	12.37990014	1.697526316
13	267	13.46569205	1.718689139
14	261	14.46359164	1.572126437
15	243	15.59193618	1.948497942

Berdasarkan Tabel 4.14 secara manual (hasil pengamatan mata) dapat diduga nilai variabel *model fit variogram* yaitu $range = 11.2891$ dan $psill = 1.55366$ (data nomor 11) namun tetap harus diuji nilai dugaan tersebut apakah menghasilkan nilai *Sum Squared Error (SSE)* terkecil.

Penentuan nilai variabel *model fit variogram* selain menggunakan hasil pengamatan langsung dapat juga menggunakan teknik *Auto Fit Variogram* yang secara otomatis dihasilkan nilai variabel model yaitu model = Ste, psill = 0.90154, range = 11.1206, nugget = 0.8883 dan kappa = 0.7.

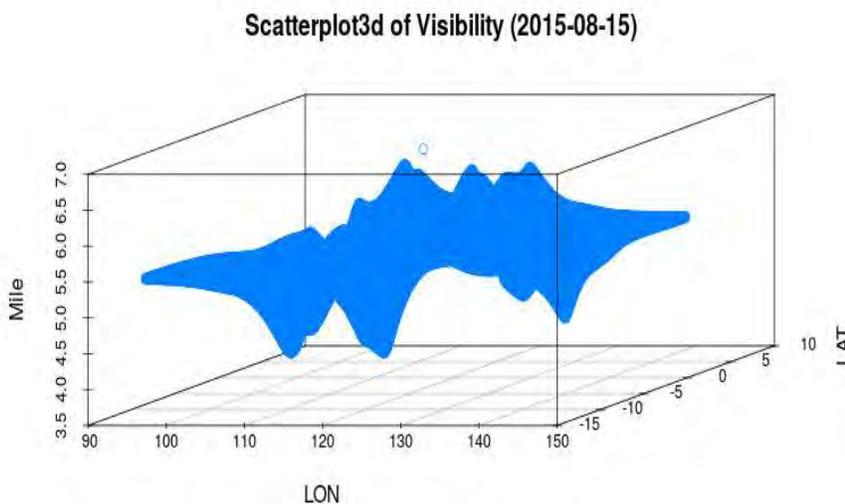


Gambar 4.29. Grafik Model Fit Variogram Visibility

Pada Gambar 4.29 menggambarkan hasil akhir *model fit variogram* yaitu variabel psill (garis mendatar) memiliki nilai 0.6635, variabel nugget (titik awal grafik garis) memiliki nilai 0.7553, variabel range (nilai jarak sejak titik awal hingga garis mendatar) memiliki nilai 3.5084 dan variabel kappa memiliki nilai 0.7. Hasil akhir memiliki nilai mendekati nilai hasil proses *Auto Fit Variogram*. Sedangkan nilai *Sum Squared Error* (SSE) adalah 2.1280.

Setelah didapatkan nilai *model fit variogram* dapat dilanjutkan dengan proses pembuatan *Output Grid* seluruh wilayah Indonesia dengan koordinat *Longitudinal* antara 94 hingga 142 dan koordinat *Lattitude* antara -12 hingga 7. Sedangkan resolusi data grid yang digunakan adalah 0,1 x 0,1. Pembuatan data grid bertujuan untuk memberikan batasan lokasi yang akan diolah menggunakan metode interpolasi *kriging*.

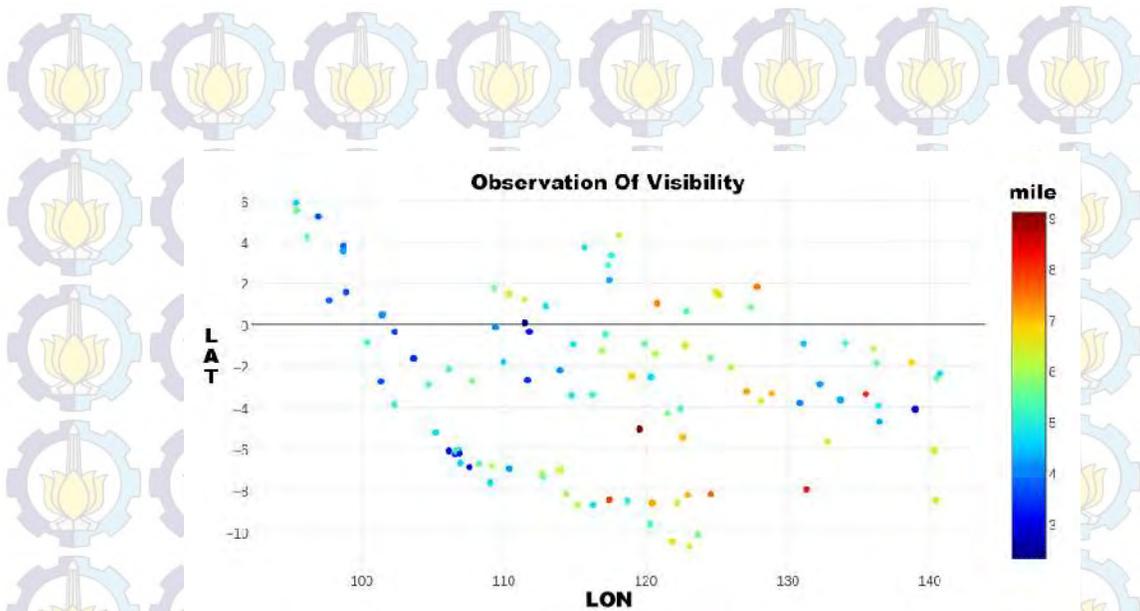
Proses selanjutnya adalah analisa metode interpolasi kriging berdasarkan data *Input Point*, *Model Fit Variogram* dan *Output Grid* yang telah didefinisikan sebelumnya. Hasil prediksi keseluruhan grid (seluruh wilayah Indonesia) dapat digambarkan dengan grafik *scatterplot3d Visibility* pada Gambar 4.29.



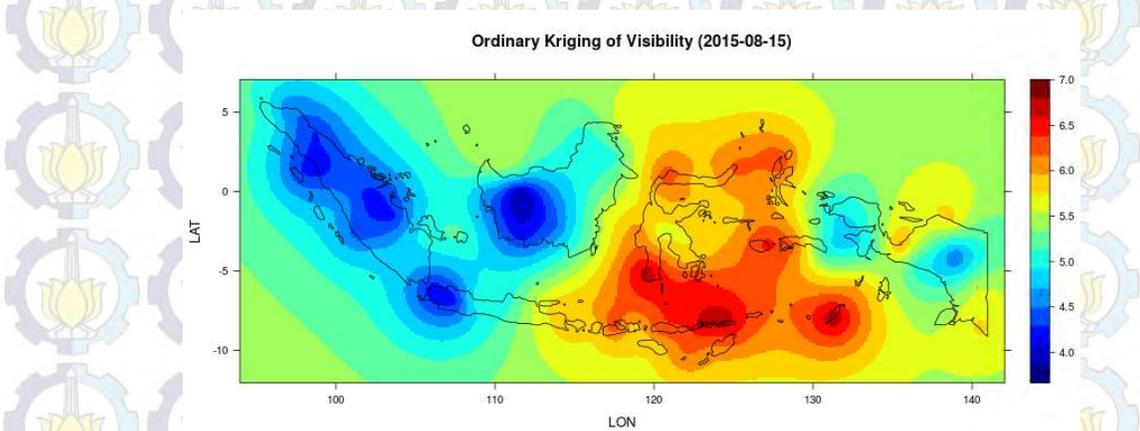
Gambar 4.30. Grafik Scatterplot3d Variabel Visibility

Berdasarkan Gambar 4.30 terlihat hasil prediksi variabel cuaca *Station Pressure* mempunyai nilai rata-rata 5,5 *mile* (terlihat dari luasan bidang mendatar yang tidak luas). Juga terlihat nilai terkecil dengan nilai berkisar 4 *mile* dan nilai terbesar dengan nilai berkisar 7 *mile*. Maka dapat diduga hasil *plotting raster data* bersifat kontinyu.

Tingkat akurasi hasil prediksi dapat diukur dengan metode *Cross Validation* namun secara visual hasil *plotting raster data* hasil prediksi dapat dibandingkan dengan hasil *plotting data* hasil observasi. Pemakaian skala gradasi warna (*color pallete*) yang sama dapat memberikan gambaran perbandingan penyebaran warna pada grafik data hasil prediksi (Gambar 4.31) dengan grafik data observasi (Gambar 4.32).



Gambar 4.31. Visualisasi Data Observasi Visibility

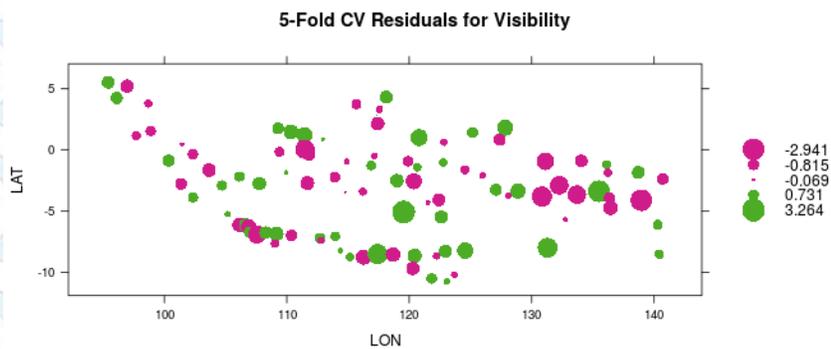


Gambar 4.32. Visualisasi Variabel Cuaca Visibility

Pada Gambar 4.31 dan Gambar 4.32 mempunyai kesamaan distribusi warna biru di pulau Sumatra dan Pulau Kalimantan. Pada Gambar 4.32 data *raster* hasil prediksi terlihat kontinyu.

Pengukuran keakuratan hasil visualisasi umumnya menggunakan metode *cross validation* dengan memanfaatkan rangkuman hasil nilai residual (beda nilai

prediksi dengan nilai hasil observasi). Berdasarkan nilai residual didapatkan nilai *error* untuk mengukur tingkat keakuratan hasil prediksi sebuah *predictive model*.



Gambar 4.33. Visualisasi Nilai Residual Visibility

Pada Gambar 4.33 terlihat lokasi Pulau Jawa dan Pulau Papua memiliki nilai residual tinggi yang artinya hasil prediksinya kurang akurat dibandingkan dengan daerah lainnya.

Tabel 4.15. Rangkuman Error dan Korelasi Visualisasi Visibility

No	Nama Parameter Kinerja	Nilai
1	Mean Error (ME) Nilai Residual	-0.0281488189523
2	Mean Squared Prediction Error (MSPE) Nilai Residual	1.33708557846553
3	Mean Squared Normalized Error (MSNE)	1.13494069164531
4	Korelasi Observasi &Prediksi (COP)	0.450910213263584
5	Korelasi Prediksi& Nilai Residual (CPR)	-0.16196139519729
6	Sum Squared Error (SSE) Fit Variogram	2.12803875215722

Pada Tabel 4.15 terlihat rata-rata nilai *error* cukup rendah kecuali nilai MSNE yang agak tinggi.

4.8. Hasil Visualisasi Variabel Cuaca Precipitation

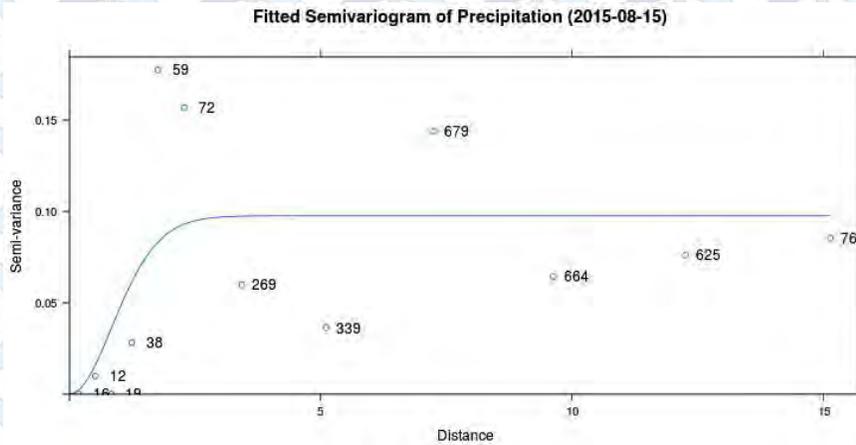
Proses visualisasi variabel cuaca *Precipitation* diawali dengan proses penyaringan data yang akan digunakan dipastikan tidak memiliki data kosong (NA) yang kemudian disimpan ke dalam sebuah *Dataframe Input Point*. Setelah itu dilakukan analisa *Spatial Auto Correlation* menggunakan teknik variogram yang memetakan jarak antar semua titik terhadap nilai data.

Tabel 4.16. Nilai Variogram Variabel Precipitation

No	NP	Distance	Gamma
1	51	0.568614327	0.002322549
2	109	1.654633642	0.135005963
3	153	2.662393809	0.080266667
4	185	3.805170104	0.064579459
5	210	4.867444404	0.023042619
6	253	5.927820306	0.055119565
7	252	6.974120124	0.223897619
8	341	8.077802642	0.116483724
9	302	9.117125578	0.046123013
10	255	10.17521092	0.093297647
11	261	11.28835767	0.129658429
12	277	12.37898864	0.030281047
13	256	13.46219658	0.057860352
14	255	14.46690758	0.055046275
15	236	15.59978313	0.125968856

Berdasarkan Tabel 4.16 secara manual (hasil pengamatan mata) dapat diduga nilai variabel *model fit variogram* yaitu $\text{range} = 0.5686$ dan $\text{psill} = 0.0023$ (data nomor 1) namun tetap harus diuji nilai dugaan tersebut apakah menghasilkan nilai *Sum Squared Error* (SSE) terkecil.

Penentuan nilai variabel *model fit variogram* selain menggunakan hasil pengamatan langsung dapat juga menggunakan teknik *Auto Fit Variogram* yang secara otomatis dihasilkan nilai variabel model yaitu model = Ste, psill = 0.0976, range = 1.2762, nugget = 0 dan kappa = 10.

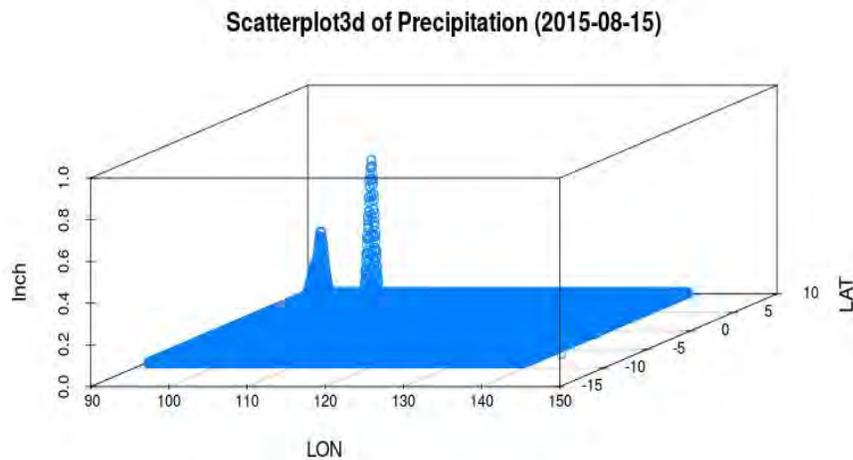


Gambar 4.34. Grafik Model Fit Variogram Precipitation

Pada Gambar 4.34 menggambarkan hasil akhir *model fit variogram* yaitu variabel psill (garis mendatar) memiliki nilai 0.0976, variabel nugget (titik awal grafik garis) memiliki nilai 0, variabel range (nilai jarak sejak titik awal hingga garis mendatar) memiliki nilai 1.2762 dan variabel kappa memiliki nilai 10. Hasil akhir memiliki nilai mendekati nilai hasil proses *Auto Fit Variogram*. Sedangkan nilai *Sum Squared Error* (SSE) adalah 0.4103.

Setelah didapatkan nilai *model fit variogram* dapat dilanjutkan dengan proses pembuatan *Output Grid* seluruh wilayah Indonesia dengan koordinat *Longitudinal* antara 94 hingga 142 dan koordinat *Lattitude* antara -12 hingga 7. Sedangkan resolusi data grid yang digunakan adalah 0,1 x 0,1. Pembuatan data grid bertujuan untuk memberikan batasan lokasi yang akan diolah menggunakan metode interpolasi kriging.

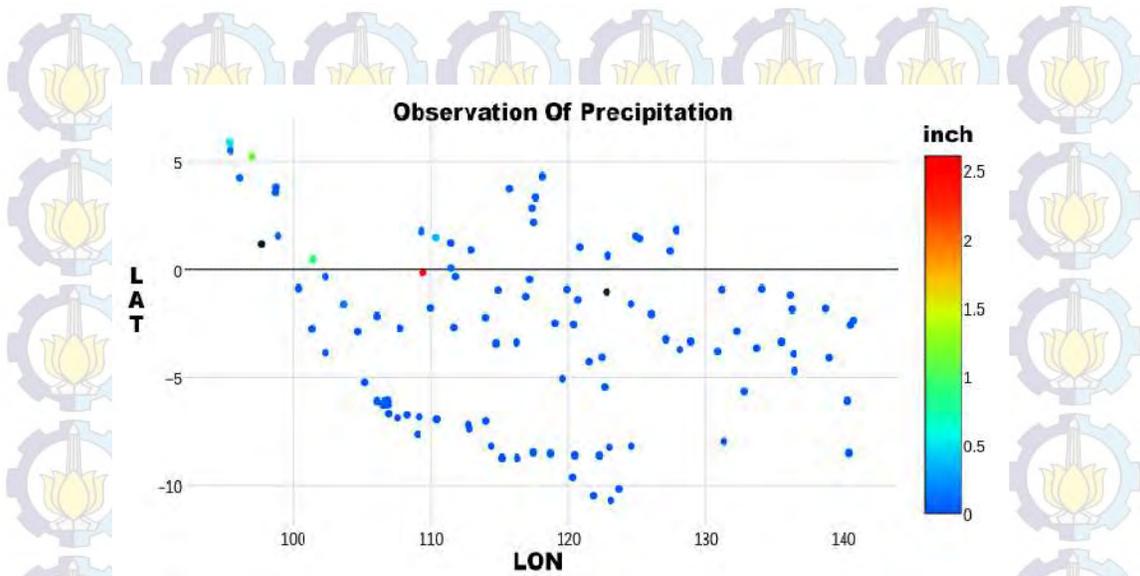
Proses selanjutnya adalah analisa metode interpolasi kriging berdasarkan data *Input Point*, *Model Fit Variogram* dan *Output Grid* yang telah didefinisikan sebelumnya. Hasil prediksi keseluruhan grid (seluruh wilayah Indonesia) dapat digambarkan dengan grafik scatterplot3d *Precipitation* pada Gambar 4.29.



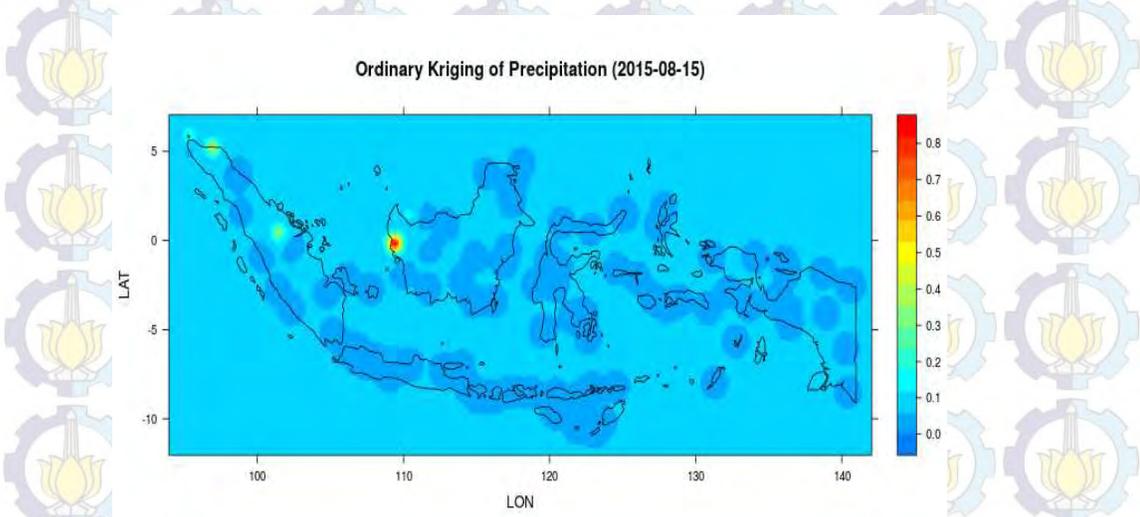
Gambar 4.35. Grafik Scatterplot3d Variabel Precipitation

Berdasarkan Gambar 4.35 terlihat hasil prediksi variabel cuaca *Precipitation* mempunyai nilai rata-rata 0 inch (terlihat dari luasan bidang mendatar yang tidak luas). Juga terlihat nilai terkecil dengan nilai 0 inch dan nilai terbesar dengan nilai 1 inch. Maka dapat diduga hasil *plotting raster data* bersifat tidak kontinyu.

Tingkat akurasi hasil prediksi dapat diukur dengan metode *Cross Validation* namun secara visual hasil *plotting raster data* hasil prediksi dapat dibandingkan dengan hasil *plotting data* hasil observasi. Pemakaian skala gradasi warna (*color pallete*) yang sama dapat memberikan gambaran perbandingan penyebaran warna pada grafik data hasil prediksi (Gambar 4.36) dengan grafik data observasi (Gambar 4.37).



Gambar 4.36. Visualisasi Data Observasi Precipitation

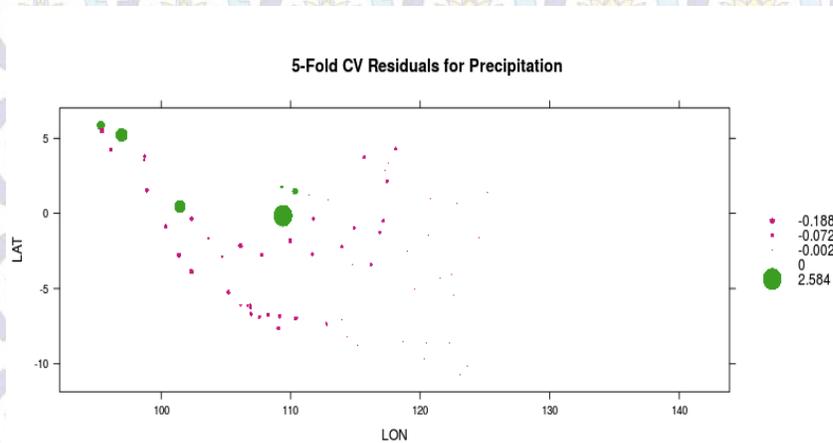


Gambar 4.37. Visualisasi Variabel Cuaca Precipitation

Pada Gambar 4.36 dan Gambar 4.37 mempunyai kesamaan distribusi warna biru di seluruh wilayah Indonesia. Pada Gambar 4.37 data *raster* hasil prediksi terlihat tidak kontinyu.

Pengukuran keakuratan hasil visualisasi umumnya menggunakan metode *cross validation* dengan memanfaatkan rangkuman hasil nilai residual (beda nilai

prediksi dengan nilai hasil observasi). Berdasarkan nilai residual didapatkan nilai *error* untuk mengukur tingkat keakuratan hasil prediksi sebuah *predictive model*.



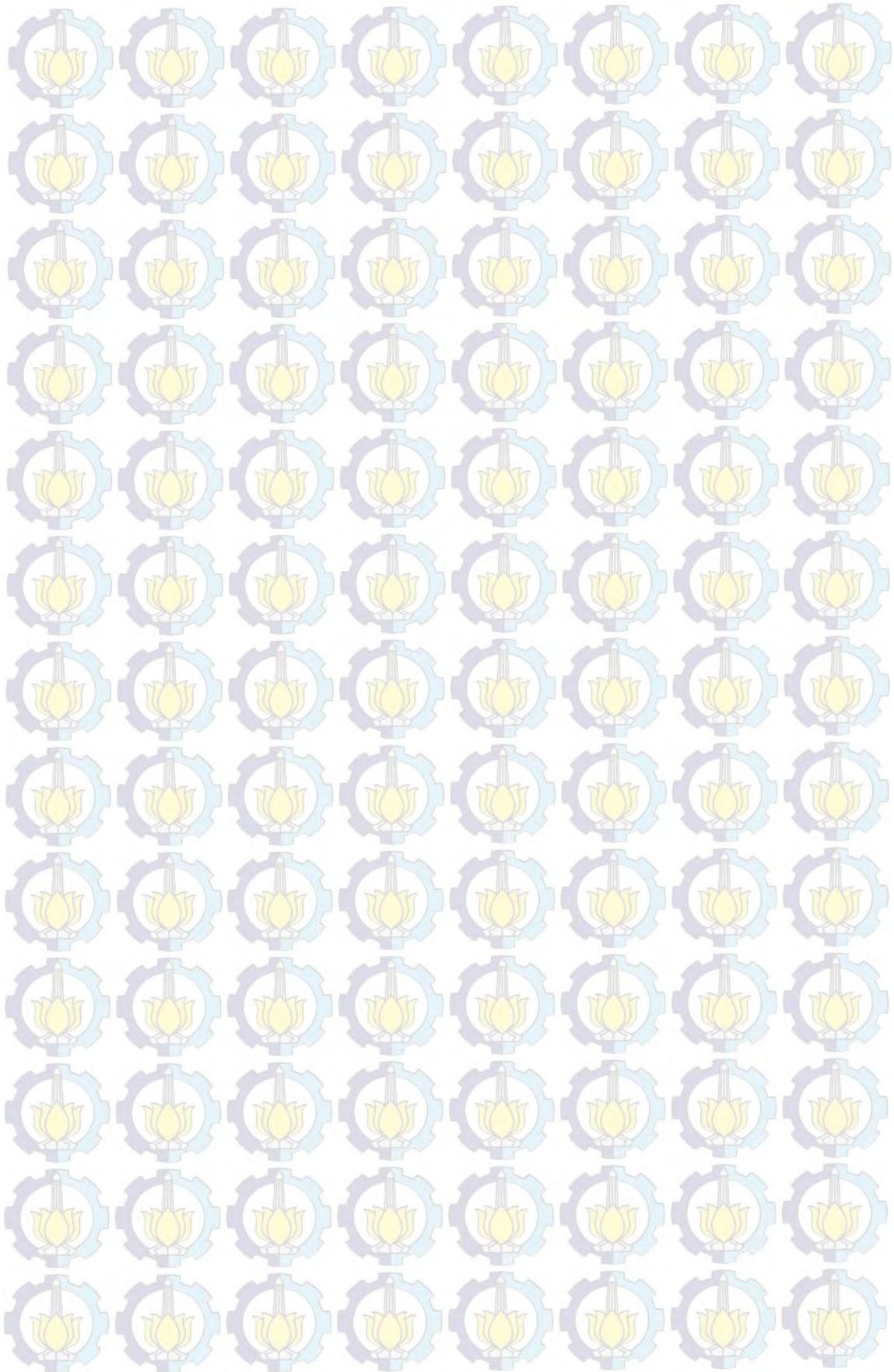
Gambar 4.38. Visualisasi Nilai Residual Precipitation

Pada Gambar 4.38 terlihat Pulau Sumatra dan Pulau Kalimantan memiliki nilai residual tinggi yang artinya hasil prediksi kurang akurat pada daerah tersebut.

Tabel 4.17. Rangkuman Error dan Korelasi Visualisasi Precipitation

No	Nama Parameter Kinerja	Nilai
1	Mean Error (ME) Nilai Residual	0.0170579305424
2	Mean Squared Prediction Error (MSPE) Nilai Residual	0.0863806120607
3	Mean Squared Normalized Error (MSNE)	0.0998323775587
4	Korelasi Observasi &Prediksi (COP)	0.06588958471497
5	Korelasi Prediksi& Nilai Residual (CPR)	-0.1086268176166
6	Sum Squared Error (SSE) Fit Variogram	0.4103

Pada Tabel 4.17 terlihat rata-rata nilai *error* cukup rendah kecuali nilai MSNE yang agak tinggi.



BAB V

PENUTUP

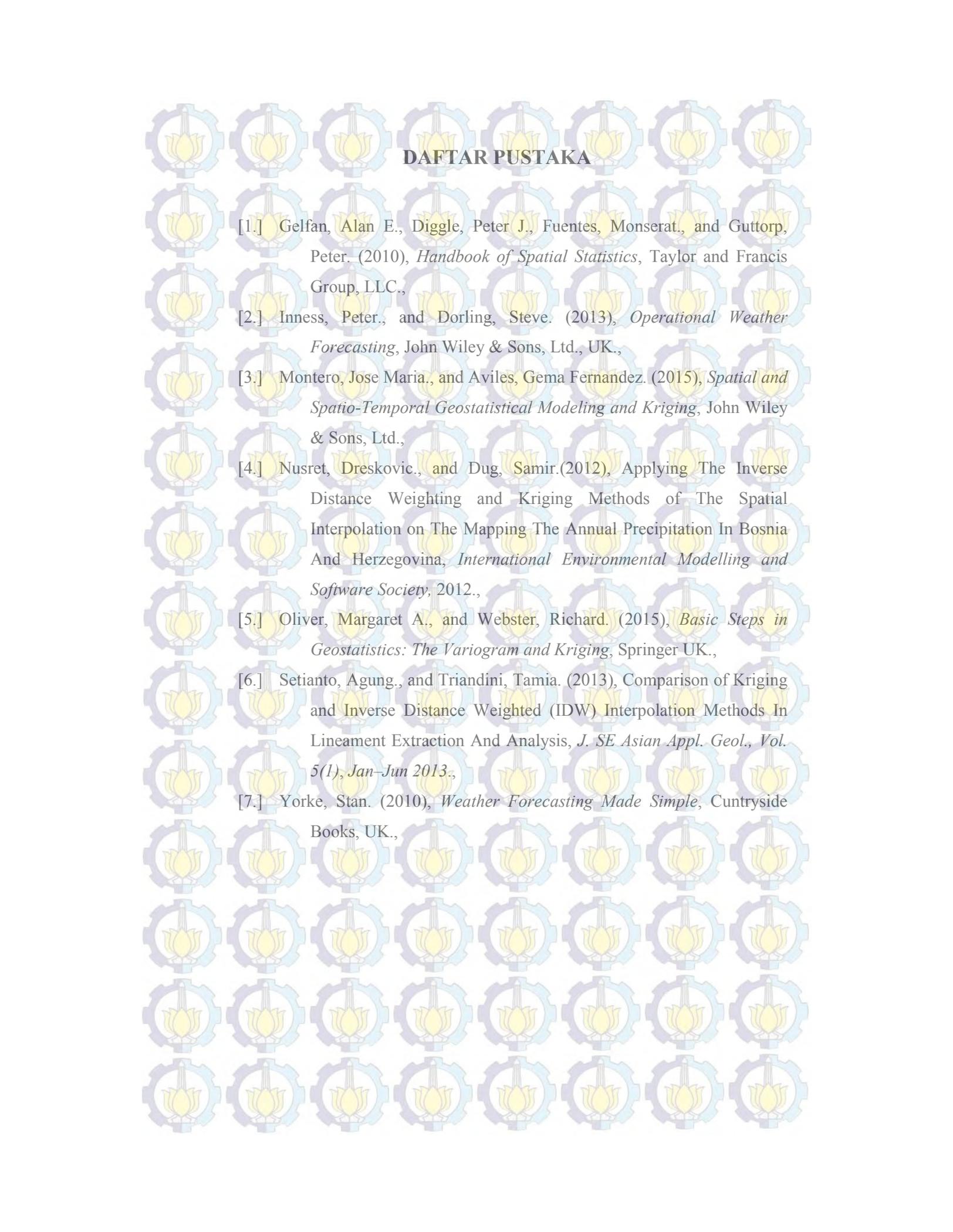
5.1. Kesimpulan

Presentasi visualisasi variabel cuaca yang baik selain berbentuk kontinyu juga memiliki nilai akurasi prediksi yang baik. Berdasarkan hasil evaluasi nilai *cross validation* pada model visualisasi variabel cuaca (tujuh variabel cuaca), maka hasil visualisasi spasial terbaik adalah visualisasi pada variabel *Visibility* (jarak pandang) sedangkan hasil visualisasi terburuk adalah visualisasi pada variabel *Station Pressure* (tekanan udara). Analisa *cross validation* hasil visualisasi pada variabel *Visibility* menunjukkan nilai error yang kecil yaitu Mean Error (ME) sebesar -0,0281 dan Mean Squared Prediction Error (MSPE) sebesar 1,337. Visualisasi geospasial yang baik pada variabel cuaca sangat bermanfaat untuk memantau kondisi cuaca bagi instansi pemerintah dan swasta, terutama untuk meningkatkan kualitas tanggap bencana.

5.2. Saran

Agar kualitas visualisasi spasial pada variabel cuaca menjadi lebih baik perlu diperbanyak jumlah lokasi observasi cuaca dan menggunakan beberapa jenis peralatan observasi cuaca yang berbeda seperti balon cuaca. Selain itu perlu digunakan metode probabilitas (misalkan metode bayesian) untuk melakukan prediksi data pada lokasi yang belum dilakukan proses observasi agar didapatkan hasil visualisasi yang lebih baik.

Selain itu dapat digunakan data observasi yang berasal dari Satelit Remote Sensing yang mempunyai jangkauan seluruh permukaan bumi. Semakin banyak data yang bisa diolah akan semakin akurat hasil prediksi yang dihasilkan. Namun hal ini akan membutuhkan proses komputasi yang berat pada data-data satelit yang berukuran besar. Pada pemrosesan data *raster* lebih tepat menggunakan aplikasi *Big Data* Geotrellis. Inilah saatnya teknologi *Big Data* dapat menjadi solusi bagi permasalahan pengolahan data-data berukuran besar.



DAFTAR PUSTAKA

- [1.] Gelfan, Alan E., Diggle, Peter J., Fuentes, Monserat., and Guttorp, Peter. (2010), *Handbook of Spatial Statistics*, Taylor and Francis Group, LLC.,
- [2.] Inness, Peter., and Dorling, Steve. (2013), *Operational Weather Forecasting*, John Wiley & Sons, Ltd., UK.,
- [3.] Montero, Jose Maria., and Aviles, Gema Fernandez. (2015), *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*, John Wiley & Sons, Ltd.,
- [4.] Nusret, Dreskovic., and Dug, Samir.(2012), Applying The Inverse Distance Weighting and Kriging Methods of The Spatial Interpolation on The Mapping The Annual Precipitation In Bosnia And Herzegovina, *International Environmental Modelling and Software Society*, 2012.,
- [5.] Oliver, Margaret A., and Webster, Richard. (2015), *Basic Steps in Geostatistics: The Variogram and Kriging*, Springer UK.,
- [6.] Setianto, Agung., and Triandini, Tamia. (2013), Comparison of Kriging and Inverse Distance Weighted (IDW) Interpolation Methods In Lineament Extraction And Analysis, *J. SE Asian Appl. Geol.*, Vol. 5(1), Jan-Jun 2013.,
- [7.] Yorke, Stan. (2010), *Weather Forecasting Made Simple*, Cuntrieside Books, UK.,

BIODATA PENULIS



Penulis lahir di Klaten, 15 Agustus 1974, putra pertama dari empat bersaudara dari Bapak (Alm) Kasno Waridi dan Ibu Titiek Maryati.

Menikah dengan dr. Nur Kartikasari dan dikaruniai dua orang anak yaitu Syarifa Nurussyifa dan Arissa Maya Syafiya.

Riwayat Pendidikan :

Tahun 1986 : Tamat SD PPSP IKIP Surabaya

Tahun 1989 : Tamat SMP Negeri 29 Surabaya

Tahun 1992 : Tamat SMA Negeri 1 Surabaya

Tahun 2001 : Tamat S1 Teknik Elektro ITS

Surabaya

Tahun 2016 : Tamat S2 Telematika (CIO) ITS

Surabaya

Riwayat Pekerjaan :

Tahun 2009 – saat ini : Staf Badan Kepegawaian Daerah (BKD) Pemerintah Kabupaten Pati Jawa Tengah

Email : bayusetyatmoko@gmail.com