

SKRIPSI

**PEMODELAN *STATISTICAL DOWNSCALING*
MENGUNAKAN REGRESI KOMPONEN UTAMA
DENGAN METODE *MINIMUM VECTOR VARIANCE*
UNTUK PENDUGAAN CURAH HUJAN**
(Studi kasus : Data Curah Hujan Kabupaten Pangkep, Sulawesi Selatan)

Disusun dan diajukan oleh:

AYU RISKI RAMADANI

H 121 16 014



PROGRAM STUDI STATISTIKA

DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

**PEMODELAN *STATISTICAL DOWNSCALING* MENGGUNAKAN
REGRESI KOMPONEN UTAMA DENGAN METODE *MINIMUM
VECTOR VARIANCE* UNTUK PENDUGAAN CURAH HUJAN
(Studi kasus : Data Curah Hujan Kabupaten Pangkep, Sulawesi Selatan)**

SKRIPSI

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

AYU RISKI RAMADANI

H 121 16 014

PROGRAM STUDI STATISTIKA

DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

LEMBAR PENGESAHAN SKRIPSI

**Pemodelan *Statistical Downscaling* Menggunakan Regresi
Komponen Utama Dengan Metode *Minimum Vector Variance*
Untuk Pendugaan Curah Hujan
(Studi kasus : Data Curah Hujan Kabupaten Pangkep, Sulawesi Selatan)**

Disusun dan diajukan oleh:

**AYU RISKI RAMADANI
H12116014**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Studi Program Sarjana Program Studi Statistika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 11 Februari 2021
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama

Sitti Sahrman, S.Si., M.Si.
NIP. 19881018 201504 2 002

Pembimbing Pertama

Dr. Dr. Georgina Maria Tinungki, M.Si
NIP. 19620926 198702 2 001

Ketua Program Studi

Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2 002

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Ayu Riski Ramadani
NIM : H12116014
Program Studi : Statistika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

**Pemodelan *Statistical Downscaling* Menggunakan Regresi
Komponen Utama Dengan Metode *Minimum Vector Variance*
Untuk Pendugaan Curah Hujan
(Studi kasus : Data Curah Hujan Kabupaten Pangkep, Sulawesi Selatan)**

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan Skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 11 Februari 2021

Menyatakan



Ayu Riski Ramadani

KATA PENGANTAR

Segala puji saya bagi Allah SWT. Tuhan pencipta alam ,yang Maha Kekal sebelum sesuatunya ada dan akan tetap kekal setelah segala sesuatunya tiada. Shalawat dan salam senantiasa semoga terus tercurahkan kepada Nabiullah Muhammad SAW dan para sahabat-sahabat beliau yang senantiasa kita rindukan. Alhamdulillahirobbil'alamiin, berkat rahmat dan kemudahan yang diberikan oleh Allah SWT, penulis dapat menyelesaikan skripsi yang berjudul “Pemodelan *Statistical Downscaling* Menggunakan Regresi Komponen Utama Dengan Metode *Minimum Vector Variance* Untuk Pendugaan Curah Hujan (Studi kasus : Data Curah Hujan Kabupaten Pangkep, Sulawesi Selatan)” sebagai salah satu syarat memperoleh gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis tidak akan sampai pada titik ini, jikalau tanpa dukungan dan bantuan dari pihak yang selalu ada, peduli dan menyayangi penulis. Oleh karena itu, penulis haturkan rasa terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya untuk orang tua penulis: Ibunda tercinta dan terkasih **Fatima** dan Ayahanda tercinta dan tersayang **Acho Pande** (almarhum) yang telah memberikan dukungan penuh, pengorbanan, dan kesabaran hati memberikan doa, cinta dan kasih sayang, serta semangat selama penulis menempuh pendidikan. Untuk Kakak-Kakakku **Asdjarmad, Juanas, Jusniati, Aluddin, Amanul** dan **Sri Asriani** terima kasih telah menjadi kakak yang sangat baik, selalu ada dan selalu memberikan dorongan dukungan baik batin dan juga raga untuk penulis. Serta untuk kakak-kakak iparku **Wirdah Yuni, Nuryanti, Noor Azizah** dan **Ibrahim** terima kasih sudah menjadi ipar yang baik dan selalu memberi motivasi.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada :

1. Direktorat Pendidikan Tinggi yang telah memberikan beasiswa **Bidikmisi**.
2. **Ibu Prof. Dr. Dwia Aries Tina Palubuhu, MA**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.

3. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
4. **Ibu Dr. Nurtiti Sunusi, S.Si, M.Si**, selaku Ketua Departemen Statistika, segenap Dosen Pengajar dan Staf Departemen Matematika yang telah membekali ilmu dan kemudahan-kemudahan kepada Penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
5. **Ibu Sitti Sahrinan, S.Si.,M.Si** selaku Pembimbing Utama dan **Ibu Dr. Dr. Georgina Maria Tinungki, M.Si** selaku Pembimbing Pendamping yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya untuk memberikan arahan, dorongan, dan motivasi kepada penulis mulai dari awal hingga selesainya penulisan skripsi ini.
6. **Ibu Sitti Sahrinan, S.Si, M.Si** selaku Ketua Tim Penguji, **Ibu Dr. Dr. Georgina Maria Tinungki, M.Si** selaku Sekretaris Tim Penguji, serta **Ibu Nasrah Sirajang, S.Si, M.Si** dan **Ibu Sri Astuti Thamrin, S.Si.,M.Stat.,Ph.D**, selaku Anggota Tim Penguji. Terima kasih telah memberikan saran dan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini serta waktu yang telah diberikan kepada Penulis.
7. **Ibu Nasrah Sirajang, S.Si, M.Si**, selaku Penasehat Akademik. Terima kasih atas segala masukan bantuan, nasehat serta motivasi yang selalu diberikan kepada Penulis selama menjalani pendidikan di Departemen Statistika.
8. Kak **Fadil** dan Kak **Fitri**, terima kasih atas bantuan dan arahnya, terima kasih selalu menjadi tempat diskusi penulis.
9. Sahabat tersayang **Hayati Hasan** yang selalu menemani, memberi semangat, dan sampai saat ini masih terus berbagi tawa, cerita , suka maupun duka. Walaupun sudah fokus dengan kesibukan tapi pertemanan kita masih terkenang hingga masa tua kelak.
10. Sahabat terkasih “**Sahabat until Jannah**” **Sisca Tiara Hanandita** dan **Dewi Sartika** terima kasih atas nasehat dan dukungan luar biasa kepada penulis. *Jazakumullahu khairan katsiran.*

11. Sahabat di program studi Statistika “**Waknda 4ever**” **Rusydah, Rosdiana, Desa, Jum, Widya, Zhaza, Fissa, Amalah, Dera, Ray, Bunga, Atik, Reski, Isna, Affan, Jay, Agung, Samsul, Suritman** dan **Riski** terima kasih sudah berbagi keceriaan dan kesedihan, selalu membantu apalagi dalam hal perkuliahan di statistika, teman bukan sekedar berada di radius kampus tapi juga diluar kampus untuk melepas kepenatan tapi tetap stay positif.
12. Sahabat seperjuangan di program studi **Statistika 2016**, terima kasih atas segala kebersamaan, kebahagiaan, dan dukungannya selama ini
13. Sahabat “**BFF SQUAD**” **Mutia, Sylva, Chya, Tri, Kiki, Angel** dan **Madi** terima kasih atas motivasi yang selalu diberikan kepada penulis.
14. Keluarga tak sedarah **A16ORITMA 2016**, semoga perjuangan kita akan selalu dikenang dalam memori dan akan tetap selalu menjaga silaturahmi, serta terima kasih atas kebahagiaan dan kebersamaannya untuk menjadi “**Seperti Seharusnya**”.
15. Kepada semua pihak yang tidak dapat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi **Allah SWT**.

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

Wassalamu’alaikum Warahmatullahi Wabarakatuh

Makassar, 11 Februari 2021

Ayu Riski Ramadani

ABSTRAK

Curah hujan merupakan salah satu unsur iklim yang sangat penting dan sangat berdampak besar bagi kehidupan manusia, namun keberadaannya masih sulit diprediksi. *Global circulation model* (GCM) merupakan salah satu model untuk meramalkan data curah hujan. Kelemahan dari data GCM adalah masih bersifat global sehingga akan sulit untuk menjelaskan keragaman dalam skala lokal yang lebih rinci. *Statistical Downscaling* (SD) hadir untuk menangani permasalahan tersebut. SD menghubungkan antara data luaran GCM dan curah hujan untuk menduga perubahan pada skala lokal dengan menggunakan metode regresi. Pemodelan SD dapat dipengaruhi oleh adanya data pencilan dan multikolinearitas. Keberadaan data tersebut perlu diperhatikan sehingga diperoleh model terbaik antara lain dengan cara mengoreksi atau dengan metode pendugaan model yang *robust* untuk mengurangi pengaruh pencilan serta mengatasi multikolinearitas. Regresi komponen utama (RKU) dan *minimum vector variance* (MVV) digunakan dalam model SD untuk mengatasi masalah tersebut. Terdapat 5 komponen utama (KU) yang optimal untuk digunakan pada model SD dengan regresi komponen utama. Selain itu penambahan 3 peubah *dummy* berdasarkan *k-means cluster* sebagai peubah prediktor untuk mengatasi masalah heterogenan sisaan data. Hasil penelitian curah hujan lokal yang diperoleh dari regresi komponen utama *dummy* (k-means) merupakan model yang terbaik berdasarkan nilai koefisien determinasi (R^2) sebesar 92% dengan *root mean square error* (RMSE) yaitu 84.33, serta validasi modelnya menghasilkan *root mean square error prediction* (RMSEP) yang lebih rendah yaitu 94.74 dengan korelasi curah hujan aktual dan dugaannya sebesar 0.9570.

Kata kunci : *Statistical Downscaling, Global Circulation Model, Regresi Komponen Utama, Minimum Vector Variance, Multikolinearitas, Pencilan, Peubah Dummy.*

ABSTRACT

Rainfall is one of the most important climate elements and has a huge impact on human life, but its existence is still difficult to predict. The global circulation model (GCM) is a model for forecasting rainfall data. The weakness of the GCM data is that it is still global in nature so it will be difficult to explain the diversity at a local scale in more detail. Statistical Downscaling (SD) exists to address this problem. SD relates the GCM output data and rainfall to predict changes at a local scale using a regression method. SD modeling can be influenced by the presence of outliers and multicollinearity data. The existence of this data needs to be considered so that the best model is obtained, among others, by means of correcting or by using a robust estimation method to reduce the effect of outliers and overcome multicollinearity. Principal component regression (RKU) and minimum vector variance (MVV) were used in the SD model to solve this problem. There are 5 main components (KU) that are optimal for use in the SD model with main component regression. In addition, the addition of 3 dummy variables based on k-means clusters as predictor variables to overcome the heterogeneity problem of remaining data. The results of local rainfall research obtained from the main component dummy regression (k-means) are the best model based on the coefficient of determination (R^2) of 92% with a root mean square error (RMSE) of 84.33, and the validation of the model produces a root mean. The lower square error prediction (RMSEP) is 94.74 with a correlation of actual and estimated rainfall of 0.9570.

Key Words : Statistical Downscaling, Global Circulation Model, Principal Component Regression, Minimum Vector Variance, Multicollinearity, Outlier, Dummy Variables.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN JUDUL	ii
LEMBAR PENGESAHAN SKRIPSI	iii
PERNYATAAN KEASLIAN	iv
KATA PENGANTAR	v
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat	4
BAB II TINJAUAN PUSTAKA	5
2.1 <i>Global Circulation Model</i>	5
2.2 <i>Statistical Downscaling</i>	6
2.2.1 Pendekatan <i>Statistical Downscaling</i>	7
2.2.2 Asumsi dan Permasalahan dalam Teknik <i>Statistical Downscaling</i>	7
2.3 Multikolinearitas	8
2.4 Metode <i>Minimum Vector Variance</i>	9
2.5 Analisis Komponen Utama	10
2.6 Regresi Komponen Utama	12
2.7 Teknik <i>Cluster</i>	13
2.8 Peubah <i>Dummy</i> berbasis <i>K-Means</i>	14

2.9	Validasi Model.....	15
BAB III METODOLOGI PENELITIAN		16
3.1	Sumber Data.....	16
3.2	Deskripsi Peubah	16
3.3	Struktur Data Penelitian	16
3.4	Metode Analisis	17
BAB IV		18
4.1	Deskripsi Data Curah Hujan	18
4.2	Mendeteksi Multikolinearitas	19
4.3	Menghitung Penduga Vektor Varians Minimum.....	20
4.4	Mereduksi Dimensi Peubah Menggunakan Analisis Komponen Utama dengan metode <i>Minimum Vector Variance</i>	22
4.5	Pemodelan <i>Statistical Downscaling</i> dengan Regresi Komponen Utama	26
4.6	Peubah <i>Dummy</i> berbasis <i>K-Means</i>	28
4.7	Pemodelan <i>Statistical Downscaling</i> dengan Regresi Komponen Utama <i>Dummy</i>	29
4.8	Validasi Model.....	31
BAB V PENUTUP		33
5.1	Kesimpulan	33
5.2	Saran	33
DAFTAR PUSTAKA		34
LAMPIRAN.....		36

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Pemodelan GCM (Sumber: Sutikno, 2008)..... 5

Gambar 4.1 Diagram Kotak Garis Curah Hujan Bulanan Kabupaten Pangkep 1997-
2018..... 19

Gambar 4.2 Plot sisaan model RKU_1 27

Gambar 4.3 Plot Curah Hujan Kabupaten Pangkep..... 28

Gambar 4.4 Plot sisaan model RKU_1 dummy 30

Gambar 4.5 Plot curah hujan aktual, RKU_1 dan RKU_1 dummy 32

Gambar 4.6 Plot curah hujan aktual, RKU_2 dan RKU_2 dummy 32

Gambar 4.7 Plot curah hujan aktual, RKU_3 dan RKU_3 dummy 32

Gambar 4.8 Plot curah hujan aktual, RKU_4 dan RKU_4 dummy 32

Gambar 4.9 Plot curah hujan aktual, RKU_5 dan RKU_5 dummy 32

DAFTAR TABEL

Tabel 3.1 Struktur Data Penelitian	16
Tabel 4.1 <i>Deskripsi Curah Hujan Bulanan Tahun 1997 – 2018</i>	18
Tabel 4.2 Hasil Iterasi Data MVV	20
Tabel 4.3 Mean Berdasarkan Metode MVV	20
Tabel 4.4 Matriks Variansi Kovariansi Berdasarkan Metode MVV	22
Tabel 4.5 Nilai Eigen	23
Tabel 4.6 Proporsi Keragaman dan Keragaman Kumulatif	25
Tabel 4.7 Nilai R ² dan RMSE model regresi komponen utama	27
Tabel 4.8 Nilai R ² dan RMSE model Regresi Komponen Utama Dummy	29
Tabel 4.9 Nilai korelasi dan RMSEP model regresi komponen utama dummy	31

DAFTAR LAMPIRAN

Lampiran 1. Tabulasi data CH dan presipitasi GCM periode 1996-2017.....	37
Lampiran 2. Nilai VIF dengan presipitasi GCM-lag.....	38
Lampiran 3. Skor Komponen Utama	39
Lampiran 4. Peubah Dummy	40

BAB I PENDAHULUAN

1.1 Latar Belakang

Curah hujan merupakan peubah yang paling sering digunakan dalam penelitian mengenai dampak perubahan iklim. Peningkatan dan penurunan curah hujan merupakan salah satu fenomena perubahan iklim yang akan mengakibatkan pergeseran awal musim, baik musim hujan maupun kemarau. Perubahan ini memberi dampak serius terhadap berbagai sektor, misalnya kesehatan, pertanian, perekonomian dan industri (Sahrinan, 2014).

Kabupaten Pangkep merupakan salah satu daerah Sulawesi Selatan yang mempunyai iklim tropis basa dan musim kemarau. Hal tersebut dipengaruhi oleh keadaan geografi dan perputaran/pertemuan arus udara, menyebabkan wilayah Kabupaten Pangkep memiliki jumlah curah hujan beragam, yang berpengaruh terhadap produktivitas industrinya (BMKG, 2018). Dampak kerugian tersebut sering terjadi karena kekurangan informasi tentang curah hujan yang akurat, cepat, dan bersifat spesifik lokasi dan tingkat kemampuan peramalan yang masih belum baik (Berliana & Sutikno, 2007).

Curah hujan merupakan kejadian kompleks yang melibatkan topografi dan interaksi antara laut, darat dan atmosfer sehingga mempersulit prediksi curah hujan itu sendiri. Hal ini menjadikan studi tentang perubahan iklim sangat diperlukan untuk meminimumkan kerugian yang mungkin terjadi. Oleh karena itu diperlukan model peramalan curah hujan yang akurat pada skala lokal dengan mempertimbangkan informasi tentang sirkulasi atmosfer global yang dapat diperoleh dari data luaran *Global Circulation Models* (GCM) (Wigena, 2006).

GCM adalah metode numerik yang dibangun berdasarkan Persamaan hukum klasik fisika. Menurut Wilby dkk pada tahun 2009, GCM merupakan model yang dapat mensimulasikan perubahan iklim. Simulasi unsur iklim untuk masa mendatang suhu, curah hujan dan yang terkait dengan peubah hidrologi. GCM menghasilkan data dalam bentuk grid atau petak wilayah dengan resolusi rendah ($\pm 2,5^\circ$ atau $\pm 300 \text{ km}^2$). Data GCM berupa grid yang menunjukkan bahwa data GCM merupakan salah satu bentuk data spasial yang berkaitan dengan keruangan. GCM mempresentasikan perubahan iklim global tidak secara lokal, untuk itu

diperlukan teknik untuk menduga peubah perubahan iklim dengan skala lokal (Wigena, 2006).

Pendekatan yang telah banyak dilakukan untuk menjembatani perbedaan skala itu adalah dengan *Statistical Downscaling* (SD). SD adalah suatu proses transformasi data dari suatu grid dengan unit skala besar menjadi data pada grid-grid dengan unit skala yang lebih kecil. Menurut Zorita dan Storch (1999), metode SD menduga nilai peubah meteorologis dalam selang waktu tertentu berdasarkan karakteristik sirkulasi atmosfer berskala besar atau transformasi data GCM ke peubah yang berskala lebih kecil. Pendekatan ini menggunakan model regresi untuk menentukan hubungan fungsional antara peubah iklim skala global GCM sebagai prediktor dengan peubah iklim lokal sebagai respon.

Domain GCM merupakan salah satu faktor dalam SD, yaitu lokasi dan luasan area permukaan atmosfer. Domain menentukan keakuratan penduga model sehingga penentuan domain merupakan langkah penting dalam SD. Dalam pemodelan SD domain ini dijadikan sebagai prediktor yang berdimensi banyak yang memungkinkan terjadi korelasi spasial antar grid dalam domain, dan multikolinearitas antar peubah. Semakin besar domain dan semakin banyak peubah akan menimbulkan permasalahan statistik dalam pemodelan, terutama bila data historisnya tidak lebih banyak dari besar domain dan jumlah peubah (Wilby, Charles, Timbal, Whetton, & Mearns, 2009).

Dalam beberapa tahun terakhir, beberapa teknik SD telah dikembangkan diantaranya adalah Regresi Komponen Utama (RKU). RKU dapat digunakan untuk mengatasi korelasi spasial atau multikolinearitas antar peubah prediktor. Melalui penggunaan analisis ini akan dihasilkan Peubah-Peubah baru yang merupakan kombinasi linier dari Peubah-Peubah bebas yang disebut komponen utama, dan selanjutnya diregresikan dengan Peubah tak bebas. Apabila seleksi Peubah diperbolehkan dan tidak mengubah teori yang ada maka cara yang paling mudah untuk mengatasi multikolinearitas adalah dengan mengeluarkan salah satu atau beberapa Peubah bebas tak penting dalam model sehingga akan diperoleh estimator dengan varian lebih kecil (Modiana, 2012).

Pemodelan SD dapat dipengaruhi oleh adanya data pencilan, terutama data pencilan yang berpengaruh terhadap pendugaan model. Data pencilan ini perlu

diperhatikan sehingga diperoleh model terbaik antara lain dengan cara mengoreksi atau dengan metode pendugaan model yang *robust* untuk mengurangi pengaruh pencilan (Wigena, 2006). Oleh karena itu diperlukan metode lain yang tidak mengeluarkan peubah bebas dalam model regresi yang dapat menghasilkan parameter dengan variansi lebih kecil. Metode alternatif yang akan digunakan disini adalah metode *Minimum Vector Variance* (MVV).

MVV merupakan salah satu cara yang bisa digunakan untuk mengetahui data pencilan yang merusak komposisi data. Rousseuw dan Van Driessen (1999) memperkenalkan metode *Fast Minimum Covariance Determinant* (FMCD) untuk mendeteksi *outlier* berdasarkan nilai determinan matriks varians-covarians yang minimum. Namun, metode FMCD mempunyai kelemahan ketika nilai determinan matriks varians-kovarians sama dengan nol. Herwindiati (2009), memodifikasi algoritma FMCD menjadi lebih efektif dan tingkat kompleksitas yang lebih rendah dengan menggunakan ukuran vektor varians yang minimum yang selanjutnya disebut dengan *Minimum Vector Variance* (MVV). Alasan yang mendasar menggunakan metode MVV dalam mendeteksi pencilan adalah karena metode ini *robust* (tegar) terhadap pencilan (Herdiani, 2017).

Oleh karena itu dalam penelitian tugas akhir ini akan dikaji SD menggunakan regresi komponen utama dengan metode *minimum vector variance* untuk pendugaan curah hujan di Kabupaten Pangkep.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, rumusan masalah yang akan dibahas adalah:

1. Berapa jumlah komponen utama pada AKU menggunakan matriks varians-kovarians berdasarkan metode MVV ?
2. Bagaimana mendapatkan hasil dugaan curah hujan menggunakan model SD dengan regresi komponen utama dari metode MVV?

1.3 Batasan Masalah

1. Penelitian ini difokuskan pada pencarian model prediksi curah hujan menggunakan SD dengan menggunakan data observasi dari stasiun cuaca di Kabupaten Pangkep.

2. Jumlah komponen utama yang digunakan sebagai acuan adalah komponen yang memiliki nilai akar ciri lebih besar dari satu.
3. Metode yang digunakan adalah regresi komponen utama dari matriks varians-kovarians berdasarkan metode MVV.

1.4 Tujuan Penelitian

1. Memperoleh jumlah komponen utama pada AKU menggunakan matriks varians-kovarians berdasarkan metode MVV.
2. Memperoleh hasil dugaan curah hujan menggunakan model SD dengan regresi komponen utama dari metode MVV.

1.5 Manfaat

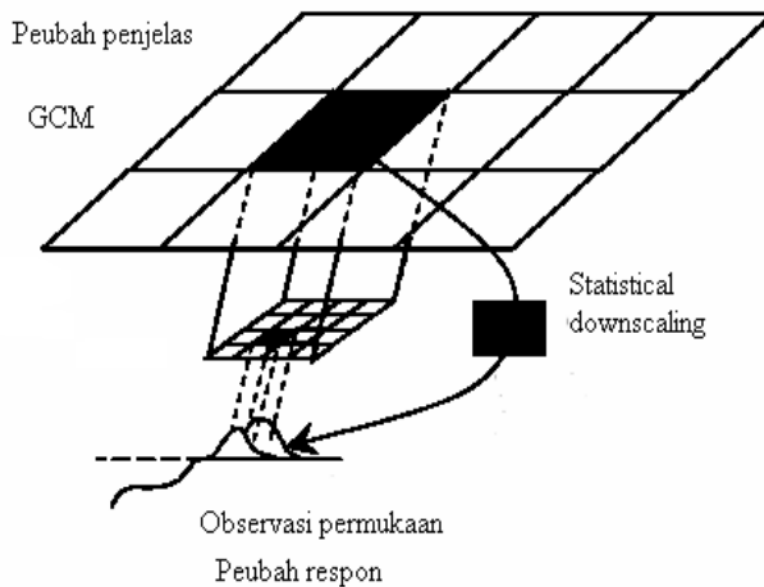
Penelitian ini bermanfaat sebagai referensi kepada pembaca tentang penggunaan regresi komponen utama dengan metode *minimum vector variance* dalam pemodelan peramalan curah hujan. Selain itu, penelitian ini bermanfaat sebagai sumber informasi tentang curah hujan di daerah Kabupaten Pangkep.

BAB II TINJAUAN PUSTAKA

2.1 *Global Circulation Model*

Global Circulation Model (GCM) atau model sirkulasi umum adalah model dinamik yang berdasarkan pemahaman yang mendalam mengenai sistem iklim saat ini untuk mensimulasi proses-proses fisik atmosfer dan lautan, yang dapat mengestimasi iklim global. GCM banyak digunakan untuk menilai dampak peningkatan gas rumah kaca terhadap kondisi atmosfer pada sistem iklim (Zorita & Storch, 1999).

Secara umum, ilustrasi GCM dapat di lihat pada gambar berikut :



Acti

Gambar 2.1 Ilustrasi Pemodelan GCM (Sumber: Sutikno, 2008)

Menurut Wilby dkk, Model GCM diyakini sebagai model penting dalam upaya memahami iklim masa lalu, sekarang, dan masa yang akan datang. GCM mampu menduga perubahan unsur-unsur cuaca dalam bentuk luaran grid-grid yang berukuran 100-500 km menurut lintang dan bujur (Zorita & Storch, 1999). Meskipun prediksi iklim berbasis GCM bisa akurat pada skala grid, namun prediksi-prediksi tersebut belum tentu sesuai bagi kondisi-kondisi lokal (Goldameir, 2015).

Model GCM mampu mensimulasi kondisi-kondisi iklim beresolusi rendah, namun tidak dirancang untuk menghasilkan informasi penting dengan resolusi yang lebih tinggi, misalnya untuk suhu dan curah hujan skala lokal. Akibat langsung resolusi spasial dari GCM adalah ketidakcocokan skala spasial antara prediksi iklim yang tersedia dan skala yang dibutuhkan oleh pengguna prediksi iklim (Berliana & Sutikno, 2007).

2.2 Statistical Downscaling

Downscaling didefinisikan sebagai upaya menghubungkan antara sirkulasi peubah skala global (peubah penjelas) dan peubah skala lokal (peubah respon) (Sutikno, 2008). Sedangkan menurut Wigena (2006), teknik *downscaling* adalah suatu proses transformasi data dari suatu grid dengan unit skala besar menjadi data pada grid-grid dengan unit skala yang lebih kecil.

Downscaling didasarkan pada asumsi bahwa iklim regional dipengaruhi oleh iklim skala global atau benua (Sutikno, 2008). Iklim regional merupakan hasil interaksi antara atmosfer, lautan, sirkulasi spesifik (lokal), seperti topografi, vegetasi, dan distribusi penggunaan lahan.

Statistical downscaling (SD) adalah pendekatan empiris mengenai hubungan secara statistika antara atmosfer global (GCM) dengan curah hujan. Ide dasar dari SD adalah menentukan parameter hubungan antara iklim skala global dengan iklim skala lokal dan menggunakan hubungan ini untuk proyeksi hasil simulasi GCM pada iklim masa lalu, sekarang, atau masa depan yang berskala lokal. SD menggunakan model statistik dalam menggambarkan hubungan antara data pada grid berskala global (prediktor) dengan data pada grid yang berskala lokal (respon) untuk menterjemahkan anomali-anomali skala global menjadi anomali dari beberapa peubah iklim lokal (Zorita & Storch, 1999). Persamaan umum model SD terdapat pada Persmaan (2.1) berikut:

$$\mathbf{y}_{n \times 1} = f(\mathbf{X}_{n \times p}) \quad (2.1)$$

Dengan $\mathbf{y}_{n \times 1}$ adalah peubah-peubah iklim lokal, $\mathbf{X}_{n \times p}$ adalah peubah GCM, n adalah banyaknya waktu, dan p adalah banyaknya grid domain GCM.

SD meliputi pengembangan kuantitatif hubungan antara peubah atmosfer (resolusi rendah : prediktor) dan peubah lokal permukaan (resolusi tinggi : respon) (Sutikno, 2008).

2.2.1 Pendekatan *Statistical Downscaling*

Pendekatan SD menggunakan data regional atau global untuk memperoleh hubungan fungsional antara skala lokal dengan skala global GCM, seperti model regresi (Wigena, 2006). SD luaran GCM diantaranya pernah digunakan dalam estimasi curah hujan di wilayah Iberian Peninsula (Zorita & Storch, 1999). Sementara itu untuk wilayah Indonesia pernah dilakukan oleh Sutikno pada tahun 2008 dalam pemanfaatan untuk peramalan produksi pertanian dengan menggunakan metode Regresi *Splines* Adaptif Berganda (RSAB) di daerah Indramayu, Subang dan Karawang. SD ini juga pernah dilakukan oleh Sahriman,dkk (2019) untuk peramalan produksi garam dengan menggunakan peubah *dummy* berdasarkan teknik *cluster hierarki* dan *nonhierarki* di daerah Pangkep, Sulawesi Selatan.

Secara umum prosedur SD adalah sebagai berikut (Sutikno, 2008): (1) pra-pemrosesan dan reduksi dimensi peubah penjelas (GCM) (2) menentukan domain peubah penjelas, (3) mengidentifikasi dan menentukan peubah penjelas yang berkorelasi tinggi terhadap peubah prediktan/respon, (4) validasi dan evaluasi model, dan (5) menerapkan data GCM.

Salah satu metode SD yaitu analisis regresi. Model regresi merupakan konsep sederhana yang menggambarkan bentuk hubungan linear atau non-linear antara peubah respon dengan peubah penjelas. Secara umum metode yang digunakan adalah regresi berganda, analisis korelasi kanonik, dan jaringan saraf tiruan yang serupa dengan regresi non-linear (Wilby, Charles, Timbal, Whetton , & Mearns, 2009).

2.2.2 Asumsi dan Permasalahan dalam Teknik *Statistical Downscaling*

Asumsi dalam teknik SD sangat diperlukan dalam penggunaan model statistik. Asumsi yang penting dalam menilai dampak iklim dengan pendekatan statistik adalah adanya hubungan antara sirkulasi atmosfer skala besar (prediktor) dengan iklim lokal (respon) yang tidak berubah dengan terjadinya perubahan iklim (Zorita & Storch, 1999).

SD didasarkan pada asumsi bahwa iklim regional dikendalikan oleh dua faktor yaitu iklim skala besar dan kondisi/karakteristik fisiografik regional/lokal seperti topografi distribusi daratan-lautan dan tataguna lahan (Storch, 1995). Dari prespektif ini, informasi iklim lokal/regional bisa didapatkan dari menentukan model statistik yang menghubungkan peubah iklim skala-besar (peubah penjelas) dengan peubah regional atau lokal (peubah respon).

Menurut Sutikno (2008), beberapa permasalahan yang muncul dalam SD adalah: (1) menentukan domain (grid) dan reduksi dimensi, (2) mendapatkan peubah penjelas yang mampu menjelaskan keragaman peubah lokal, dan (3) mendapatkan metode statistik yang sesuai karakteristik data, sehingga bisa menggambarkan hubungan antara peubah prediktan dan peubah penjelas, serta dapat mengakomodasi kejadian ekstrim.

2.3 Multikolinearitas

Istilah multikolinearitas mula-mula ditemukan oleh Ragnar Frisch pada tahun 1943. Pada mulanya multikolinearitas berarti adanya hubungan linear yang sempurna atau pasti, diantara beberapa atau semua peubah bebas dari model regresi ganda (Gujarati & Damodar, 1995). Ada beberapa cara untuk mengetahui keberadaan multikolinearitas dalam suatu model regresi, dan dalam penulisan ini menggunakan nilai *Tolerance* atau *Variance Inflation Factor* (VIF). Untuk menguji ada tidaknya multikolinearitas, dapat menggunakan bantuan *software* SPSS dengan melihat nilai *Tolerance* atau VIF pada tabel "*coefficients*". Jika nilai *Tolerance* kurang dari 0,1 atau nilai VIF melebihi 10 maka hal tersebut menunjukkan bahwa multikolinearitas adalah masalah yang pasti terjadi antar peubah bebas (Soemartini, 2008).

Mengidentifikasi multikolinearitas pada data presipitasi menggunakan *variance inflation factors* (VIF) dengan Persamaan (2.2) :

$$VIF_j = \frac{1}{1-R_j^2} \quad (2.2)$$

dengan R_j^2 adalah koefisien determinasi dari peubah penjelas yang diregresikan terhadap peubah penjelas lainnya. Menurut Wirnancy (2015), nilai VIF >10 mengindikasikan adanya multikolinieritas sehingga dilakukan reduksi dimensi.

2.4 Metode *Minimum Vector Variance*

Kriteria MVV pertama kali diperkenalkan oleh Herwindiati (2009) dengan mempertimbangkan himpunan data $\mathbf{x} = \{x_{1j}, x_{2j}, \dots, x_{ij}, \dots, x_{nj}\}$ dari satu observasi dengan peubah p dan $H \subseteq \mathbf{x}$. Misalkan \mathbf{T}_{MVV} dan \mathbf{S}_{MVV} adalah taksiran MVV untuk parameter lokasi dan matriks varians kovarians. Taksiran diperoleh berdasarkan himpunan H . Jumlah lokasi elemen dari H adalah $h = \frac{(n+p+1)}{2}$ data yang akan memberikan matriks varians kovarians \mathbf{S}_{MVV} dengan nilai $Tr(\mathbf{S}_{MVV}^2)$ minimum untuk semua kemungkinan himpunan yang mengandung h data. Oleh karena itu secara berturut-turut taksiran MVV untuk parameter lokasi dari matriks dapat ditunjukkan pada Persamaan (2.3) dan Persamaan (2.4) berikut (Herdiani, 2017) :

$$\mathbf{T}_{MVV} = \frac{1}{h} \sum_{i \in H} x_i \quad (2.3)$$

$$\mathbf{S}_{MVV} = \frac{1}{h-1} \sum_{i \in H} (x_i - \mathbf{T}_{MVV})(x_i - \mathbf{T}_{MVV})' . \quad (2.4)$$

Adapun algoritma untuk MVV adalah sebagai berikut (Herwindiati, 2009):

1. Ambillah himpunan data yang terdiri dari $h = \frac{(n+p+1)}{2}$ data, sebutlah himpunan data ini dengan \mathbf{H}_{old} . Dengan p merupakan peubah dan n merupakan banyaknya observasi.
2. Hitunglah vektor mean $\bar{\mathbf{x}}(\mathbf{H}_{old})$ dan matriks kovarians \mathbf{S}_{Hold} untuk semua data \mathbf{H}_{old} . Selanjutnya untuk $i = 1, 2, \dots, n$ hitunglah menggunakan Persamaan (2.5) berikut.

$$d_{H_{old}}^2 = d_{H_{old}}^2(\mathbf{x}_i, \bar{\mathbf{x}}_{Hold}) = (\mathbf{x}_i - \bar{\mathbf{x}}_{Hold})' \mathbf{S}_{Hold}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{Hold}) \quad (2.5)$$

3. Urutkan hasil perhitungan dari yang terkecil ke yang terbesar. Urutan ini akan memberikan permutasi indeks observasi π . Misalnya

$$d_{H_{old}}^2(\pi_1) \leq d_{H_{old}}^2(\pi_2) \dots \leq d_{H_{old}}^2(\pi_n)$$

4. Bentuklah suatu himpunan baru yang terdiri dari h observasi dengan indeks $\pi(1), \pi(2), \dots, \pi(h)$ dan berilah nama \mathbf{H}_{new}
5. Hitunglah $\bar{\mathbf{x}}_{H_{new}}$, $\mathbf{S}_{H_{new}}$ dan $(\mathbf{x}_i - \bar{\mathbf{x}}_{H_{new}})$ seperti pada tahap 2.

6. Jika $Tr(\mathbf{S}_{H_{new}}^2) = Tr(\mathbf{S}_{H_{old}}^2)$ maka proses dikerjakan. Jika $Tr(\mathbf{S}_{H_{new}}^2) < Tr(\mathbf{S}_{H_{old}}^2)$ maka proses dilanjutkan sampai iterasi ke-k mencapai $Tr(\mathbf{S}_{H_{new}}^2) = Tr(\mathbf{S}_{H_{old}}^2)$
7. Jika S_{H_i} adalah matriks kovarians dari iterasi ke-k. Pada akhir iterasi ke-k akan dimiliki $Tr(\mathbf{S}_{H_1}^2) \geq Tr(\mathbf{S}_{H_2}^2) \geq \dots \geq Tr(\mathbf{S}_{H_{k-1}}^2) = Tr(\mathbf{S}_{H_k}^2)$.

2.5 Analisis Komponen Utama

Variabel-variabel data GCM umumnya memiliki dimensi yang besar dan terdapat multikolinieritas atau korelasi yang tinggi antar grid data GCM. Untuk mereduksi dimensi peubah GCM dan mengatasi multikolinieritas digunakan Analisis Komponen Utama (AKU) sehingga akan terdapat peubah-peubah baru yang tidak berkorelasi sesamanya (Johnson & Wichern, 1998).

Peubah baru yang dihasilkan disebut komponen utama (KU), yang memenuhi sifat: (1) merupakan kombinasi linier peubah-peubah asal dan (2) menghilangkan peubah-peubah asal yang mempunyai sumbangan informasi yang relatif kecil. KU dapat diperoleh dari pasangan nilai akar ciri dan vektor ciri matriks varians-kovarians atau matriks korelasi. Matrik varians-kovarians dari peubah \mathbf{X} digunakan apabila tidak terdapat perbedaan skala antar peubah prediktor. Sebaliknya, matriks korelasi dari peubah \mathbf{X} digunakan pada saat terdapat perbedaan skala antar peubah prediktor. Standardisasi data dilakukan terlebih dahulu jika satuan data antar peubah prediktor tidak sama (Sahrman, 2014).

Jika $\mathbf{X}' = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$ mempunyai matriks varians-kovarians $\mathbf{\Sigma}$ dengan nilai akar ciri $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, maka diperoleh \mathbf{w}_j yang merupakan kombinasi linier peubah asal (Johnson & Wichern, 1998), dapat dilihat pada Persamaan (2.6) berikut :

$$\begin{aligned}
 \mathbf{w}_1 &= \mathbf{e}_1' \mathbf{X} = e_{11}\mathbf{x}_1 + e_{12}\mathbf{x}_2 + \dots + e_{1p}\mathbf{x}_p \\
 \mathbf{w}_2 &= \mathbf{e}_2' \mathbf{X} = e_{21}\mathbf{x}_1 + e_{22}\mathbf{x}_2 + \dots + e_{2p}\mathbf{x}_p \\
 &\vdots \\
 \mathbf{w}_p &= \mathbf{e}_p' \mathbf{X} = e_{p1}\mathbf{x}_1 + e_{p2}\mathbf{x}_2 + \dots + e_{pp}\mathbf{x}_p
 \end{aligned} \tag{2.6}$$

dengan \mathbf{w}_j adalah komponen utama ($j = 1, 2, \dots, p$), \mathbf{e} adalah vektor eigen, dan \mathbf{X} adalah peubah asal.

Syarat untuk membentuk komponen utama yang merupakan kombinasi linier dari peubah asal X agar mempunyai variansi maksimum adalah dengan memilih vektor eigen yaitu: $e = (e_1, e_2, \dots, e_p)'$ sedemikian sehingga $var(w_j) = e_j' \Sigma e_j$ maksimum dengan batasan $e_j' e_j = 1$

- KU pertama adalah kombinasi linear $e_1' X$ yang memaksimumkan $Var(e_1' X)$ dengan fungsi kendala $e_1' e_1 = 1$.
- KU kedua adalah kombinasi linear $e_2' X$ yang memaksimumkan $Var(e_2' X)$ dengan fungsi kendala $e_2' e_2 = 1$ dan $Cov(e_2' X, e_1' X) = e_1' \Sigma e_2 = 0$.
- KU ke- j adalah kombinasi linear $e_j' X$ yang memaksimumkan $Var(e_j' X)$ dengan
- fungsi kendala $e_j' e_j = 1$ dan $Cov(e_j' X, e_1' X) = e_1' \Sigma e_j = 0$ untuk $j' < j$.

Langkah – langkah metode AKU adalah sebagai berikut:

1. Menghitung matriks varians-kovarians Σ dengan menggunakan Persamaan (2.7) sebagai berikut:

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' \quad (2.7)$$

2. Menghitung nilai eigen dari matriks varians-kovarians Σ dengan menggunakan Persamaan (2.8) sebagai berikut:

$$|\Sigma - \lambda I| = 0 \quad (2.8)$$

3. Menghitung vektor eigen dan matriks varians-kovarians Σ dengan menggunakan Persamaan (2.9) sebagai berikut :

$$|\Sigma - \lambda I|x = 0 \quad (2.9)$$

4. Menghitung proporsi varians dengan menggunakan Persamaan (2.10) sebagai berikut.

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (2.10)$$

5. Jika KU yang diambil sebanyak k komponen, dengan $r < p$, maka besarnya proporsi kumulatif dapat dijelaskan oleh KU ke- k dengan menggunakan rumus pada Persamaan (2.11) berikut :

$$\frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j} \times 100\% \quad (2.11)$$

6. Membentuk komponen utama ke-1,2,..., p menggunakan Persamaan (2.6)

7. Menghitung skor komponen utama berdasarkan pada poin 6.

Selanjutnya metode yang digunakan untuk menentukan jumlah KU yang akan digunakan dalam analisis adalah sebagai berikut (Fekedulegn, Colbert, Schuckers, & Hicks, 2002).

- Membuang komponen yang memiliki nilai akar ciri terkecil. Alasannya adalah bahwa KU dengan nilai akar ciri terkecil mengandung sedikit informasi. Menggunakan prosedur ini KU dieliminasi sampai komponen yang tersisa menjelaskan beberapa persentase yang dipilih sebelumnya dari total keragaman (misalnya 80% atau lebih).
- Beberapa peneliti menggunakan aturan memilih KU yang memiliki nilai akar ciri lebih besar dari satu.

2.6 Regresi Komponen Utama

Regresi komponen utama merupakan teknik analisis regresi yang dikombinasikan dengan analisis komponen utama, dimana analisis komponen utama dijadikan sebagai tahap analisis. Salah satu asumsi yang harus dipenuhi untuk uji hipotesis pada parameter analisis linier berganda adalah kebebasan multikolinieritas. Analisis komponen utama merupakan analisis yang memperkecil dimensi peubah tanpa kehilangan banyak informasi, dengan tujuan menyederhanakan peubah yang diamati dengan cara mereduksi dimensinya sehingga dihasilkan peubah-peubah baru yang merupakan kombinasi linier dari peubah-peubah prediktor asal dan antarpeubah baru ini bersifat saling bebas.

Setelah dilakukan analisis komponen utama dan telah diperoleh skor komponen utama, langkah selanjutnya adalah meregresikan komponen-komponen utama terpilih (w_j) dengan peubah respon. Bentuk Persamaan RKU dapat dilihat pada Persamaan (2.12) berikut :

$$\hat{y}_i = \beta_0 + \sum_{j=1}^p \beta_j w_{ij} \quad (2.12)$$

dengan :

- \hat{y}_i = Penduga peubah iklim lokal pada pengamatan ke- i ; $i = 1, 2, \dots, n$,
- β_0 = nilai intersep,
- β_j = koefisien regresi bagi komponen utama ke - j ,

w_{ij} = Komponen Utama ke $-j$ pada pengamatan ke $-i$; $j = 1, 2, \dots, p$,
 p = banyaknya komponen utama.

2.7 Teknik Cluster

Teknik *cluster* adalah suatu teknik statistik yang bertujuan untuk mengelompokkan obyek ke dalam suatu kelompok sedemikian sehingga objek yang berada dalam satu kelompok akan memiliki kemiripan yang tinggi dibandingkan dengan objek yang berada di kelompok lain (Rahmawati, 2012). Kemiripan antar objek diukur dengan menggunakan ukuran jarak salah satunya yaitu jarak *Euclid*. Karakteristik objek-objek dalam suatu *cluster* memiliki tingkat kemiripan yang tinggi, sedangkan karakteristik antar objek pada suatu *cluster* dengan *cluster* lain memiliki tingkat kemiripan yang rendah. Keragaman dalam suatu *cluster* minimum sedangkan antar keragaman antar *cluster* maksimum (Mattjik & Sumertajaya, 2011).

Terdapat tiga metode yang umum digunakan dalam pengclusteran objek yaitu metode grafik, metode berhirarki dan metode tak berhirarki. Metode grafik, terdiri atas tiga jenis yaitu plot profil, plot Andrew dan plot Andrew termodifikasi. Pendekatan grafik yang paling sederhana adalah menggunakan plot profil dari setiap pengamatan. Plot ini hanya efektif untuk data yang tidak terlalu banyak pengamatannya, sehingga pembakuan data sangat membantu dalam proses ini. Plot lain yang biasa digunakan adalah plot Andrew dan plot Andrew termodifikasi. Kedua plot ini memberikan hasil yang lebih efektif dalam pengclusteran objek (Mattjik & Sumertajaya, 2011).

Metode berhirarki digunakan untuk mengclusterkan pengamatan secara terstruktur berdasarkan kemiripan sifatnya dan *cluster* yang diinginkan belum diketahui banyaknya. Ada dua cara untuk mendapatkan *cluster* dengan metode pengclusteran hirarki yaitu dengan cara penggabungan dan pemisahan *cluster*. Metode hirarki dengan cara penggabungan didapat dengan menggabungkan pengamatan atau *cluster* secara bertahap, sehingga pada akhirnya didapat hanya satu *cluster* saja. Sebaliknya, cara pemisahan pada metode hirarki dimulai dengan membentuk satu *cluster* besar beranggotakan seluruh pengamatan. *Cluster* besar

tersebut kemudian dipisah menjadi *cluster* yang lebih kecil, sampai satu *cluster* hanya beranggotakan satu pengamatan saja.

Secara umum, algoritma untuk metode pengelompokan hirarki adalah :

1. Tentukan matriks jarak antar data atau kelompok.
2. Gabungkan dua data atau kelompok terdekat ke dalam kelompok yang baru.
3. Tentukan kembali matrik jarak tersebut.
4. Lakukan langkah 2 dan 3 sampai semua data masuk dalam satu kelompok.

2.8 Peubah *Dummy* berbasis *K-Means*

Peubah *dummy* atau peubah indikator adalah peubah buatan yang dibuat untuk mewakili atribut dengan dua kategori atau kategori yang berbeda. Peubah *dummy* menetapkan angka “0” dan “1” untuk menunjukkan keanggotaan dalam kategori yang saling eksklusif dan menyeluruh. Jumlah peubah *dummy* yang diperlukan untuk mewakili peubah atribut tunggal sama dengan jumlah kategori kurang satu. Untuk peubah atribut tertentu, tidak ada peubah *dummy* yang dibangun dapat berulang. Artinya, satu peubah *dummy* tidak bisa menjadi banyak konstanta atau hubungan linier sederhana yang lain. peubah *dummy* adalah peubah bebas yang mengambil nilai “0” atau “1”. Dalam model regresi, peubah *dummy* dengan nilai 0 akan menyebabkan koefisiennya hilang dari Persamaan. Sebaliknya, nilai 1 menyebabkan koefisien berfungsi sebagai intercept tambahan, karena adanya properti identitas perkalian dengan 1. Jenis spesifikasi dalam model regresi linier ini berguna untuk menentukan himpunan bagian pengamatan yang memiliki kemiringan yang berbeda tanpa diciptakannya model terpisah (Parjiono, Faisal, & Hastiadi, 2018).

Penentuan peubah *dummy* berbasis *k-means* merupakan salah satu metode dalam pengclusteran tak berhirarki. MacQueen menyarankan penggunaan *kmeans* untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam *cluster* tertentu berdasarkan rata-rata terdekat. Dalam bentuk yang paling sederhana, proses ini terdiri dari tiga tahap (Mattjik & Sumertajaya, 2011) :

1. Bagi objek-objek tersebut ke dalam K *cluster* awal.

2. Masukkan tiap objek ke suatu *cluster* berdasarkan rata-rata terdekat. Jarak biasanya ditentukan dengan menggunakan Euclidean. Hitung kembali rata-rata untuk *cluster* yang mendapat objek dan yang kehilangan objek.
3. Ulangi langkah 2 sampai tidak ada lagi pemindahan objek antar *cluster*.

2.9 Validasi Model

Kriteria kebaikan model untuk validasi dapat menggunakan nilai *Root Mean Square Error Prediction* (RMSEP). RMSEP adalah nilai varians dari residual yang dapat menunjukkan keakuratan suatu model. RMSEP mempunyai nilai minimal 0, semakin kecil nilai RMSEP menunjukkan bahwa perbedaan antara nilai dugaan hasil pemodelan dengan data aktualnya semakin kecil pula. Sehingga model yang terbaik adalah model dengan nilai RMSEP terkecil. Nilai RMSEP diperoleh dengan menggunakan rumus pada Persamaan (2.13):

$$RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.13)$$

dengan :

y_i = nilai peubah iklim lokal pada data validasi ke i ; $i = 1, \dots, n$,

\hat{y}_i = nilai dugaan peubah iklim lokal pada data validasi ke i ,

n = banyaknya observasi data.