



PENGEMBANGAN PREDIKSI REGRESI GEROMBOL SEBARAN GAMMA PADA *STATISTICAL DOWNSCALING* UNTUK PENDUGAAN CURAH HUJAN HARIAN

**(Aplikasi pada Curah Hujan Harian di Beberapa Stasiun di Jawa Barat
periode 2010 – 2019)**

MUTHIA NADHIRA FALADIBA



**STATISTIKA TERAPAN
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2021**

PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis berjudul Pengembangan Prediksi Regresi Gerombol Sebaran Gamma pada *Statistical Downscaling* Untuk Pendugaan Curah Hujan Harian (Aplikasi pada Curah Hujan Harian di Beberapa Stasiun di Jawa Barat periode 2010 – 2019) adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Februari 2021

Muthia Nadhira Faladiba
G152180401



RINGKASAN

MUTHIA NADHIRA FALADIBA. Pengembangan Prediksi Regresi Gerombol Sebaran Gamma pada *Statistical Downscaling* Untuk Pendugaan Curah Hujan Harian (Aplikasi pada Curah Hujan Harian di Beberapa Stasiun di Jawa Barat periode 2010 – 2019). Dibimbing oleh AGUS MOHAMAD SOLEH dan ANIK DJURAIDAH.

Letak geografis Indonesia dilalui hembusan angin muson (monsoon) menjadikan Indonesia memiliki dua musim yaitu musim kemarau dan musim penghujan. Iklim tropis dan kelembapan udara tinggi membuat musim penghujan dapat berlangsung lebih lama dibanding musim kemarau. Informasi mengenai curah hujan sangat penting untuk diketahui terutama dalam bidang pertanian yang produktivitasnya bergantung pada cuaca dan ketersediaan air. Metode *Statistical Downscaling* (SD) merupakan salah satu metode yang dapat digunakan untuk pendugaan curah hujan. Metode ini membuat Informasi data berskala global GCM sebagai peubah prediktor diproyeksikan terhadap informasi skala lokal sebagai peubah respon. Peubah prediktor mempunyai korelasi tinggi antar peubah dan menyebabkan adanya multikolinearitas. Teknik pereduksian dimensi atau seleksi peubah dapat digunakan untuk mengatasinya.

Penelitian mengenai pendugaan curah hujan menggunakan metode SD dengan regresi gerombol telah dilakukan sebelumnya. Metode yang digunakan antara lain regresi gerombol dengan peubah *dummy*, regresi gerombol yang dibandingkan dengan metode Regresi Komponen Utama (RKU) dan Regresi Kuadrat Terkecil Parsial (RKTP) dan regresi gerombol sebaran Gamma. Hasil dari penelitian tersebut regresi gerombol menghasilkan *Root Mean Square Error of Prediction* (RMSEP) yang lebih kecil jika dibandingkan dengan RKU dan RKTP. Selain itu, pada regresi gerombol sebaran Gamma terbukti mampu memodelkan data curah hujan dengan baik, namun belum mampu menghasilkan nilai dugaan curah hujan.

Penelitian yang dilakukan merupakan kajian lanjutan regresi gerombol sebaran Gamma untuk dapat melakukan dugaan pada data curah hujan. Penelitian ini terbagi kedalam dua kajian. Kajian pertama yaitu simulasi regresi gerombol dengan sebaran Gamma dan sebaran Normal. Kajian kedua yaitu penerapan regresi gerombol pada data curah hujan harian di stasiun Bandung, Bogor, Citeko dan Jatiwangi. Analisis komponen utama terlebih dahulu dilakukan untuk pereduksian data prediktor (GCM) sebagai solusi untuk multikolinearitas.

Simulasi regresi gerombol dibagi kedalam simulasi dua kelompok dan tiga kelompok. Hal ini bertujuan untuk melihat kemampuan regresi gerombol mengelompokkan data sesuai sebaran sebenarnya. Simulasi dua kelompok memiliki (3) skenario dengan model Gamma-Gamma (GG), Normal-Normal (NN) dan Gamma-Normal (GN). Simulasi tiga kelompok memiliki (4) skenario dengan model Gamma-Gamma-Gamma (GGG), Normal-Normal-Normal (NNN), Gamma- Normal-Normal (GNN) dan Gamma-Gamma-Normal (GNN). Sebaran Gamma dengan parameter bentuk (ξ) yaitu $\xi = 0,5$, $\xi = 0,85$ dan $\xi = 15$. Hasil simulasi menunjukkan regresi gerombol untuk skenario sebaran Gamma, normal dan campuran tersebut mampu membentuk gerombol dengan tepat dan sesuai dengan karakteristik data sebenarnya.

Penerapan regresi gerombol pada curah hujan harian stasiun hujan Bogor, Bandung, Citeko dan Jatiwangi dilakukan pemodelan tanpa gerombol (regresi Gamma) dan dengan penggerombolan sebaran campuran Gamma-Normal. Prediksi data dilakukan dengan pendekatan jarak *centroid* dan jarak antar amatan. Regresi gerombol memberikan nilai RMSEP lebih kecil jika dibandingkan dengan tanpa penggerombolan. Pendugaan pada regresi gerombol dilakukan sebanyak dua dan tiga gerombol. Hasil menunjukkan tiga gerombol menghasilkan RMSEP lebih baik dari dua gerombol. Model terbaik bagi seluruh stasiun hujan adalah model Gamma-Normal-Normal (GNN) yaitu model dengan satu sebaran Gamma dan dua kelompok Normal. Metode prediksi dengan pendekatan jarak *centroid* dan jarak antar amatan mempunyai kemampuan yang sama dalam melakukan prediksi data.

Kata kunci: Analisis Komponen Utama, Prediksi Regresi gerombol, RMSE, Sebaran Gamma, *Statistical downscaling*

@Hak cipta milik IPB University

IPB University





SUMMARY

MUTHIA NADHIRA FALADIBA. Development of Clusterwise Regression Prediction Model using Gamma Distribution for Statistical Downscaling to Predict Daily Rainfall (Application on Daily Rainfall at Several Stations in West Java for the period 2010 - 2019). Supervised by AGUS MOHAMAD SOLEH and ANIK DJURAI DAH.

The geographical location of Indonesia is passed by the monsoon winds, making Indonesia has two seasons, the dry season and the rainy season. The tropical climate and high humidity make the rainy season could be lasts longer than the dry season. Information about rainfall is very important to know especially in agriculture, whose productivity depends on the weather and water availability. The statistical downscaling (SD) is a method that can be used to estimate rainfall. This method makes global scale GCM data information (as predictor variables) projected against local scale information (as response variables). Predictor variables have a high correlation between variables and it cause multicollinearity. Dimensional reduction techniques or variable selection can be used to overcome this.

Research on rainfall estimation using the SD method with clusterwise regression has been done before. The SD method used includes clusterwise regression with dummy variables, clusterwise regression compared to the Principal Component Regression (RKU) method and Partial Least Squares Regression (RKTP) and clusterwise regression using Gamma distribution. The clusterwise regression research resulted in lower Root Mean Square Error of Prediction (RMSEP) when compared to RKU and RKTP. In addition, the Gamma distribution cluster regression is proven to be able to model rainfall data well but has not been able to produce an estimated rainfall value.

This study is a further study of clusterwise regression using Gamma distribution to be able to predict rainfall data. This research is divided into two studies. The first study is a clusterwise regression simulation with the Gamma distribution and the Normal distribution. The second study is the application of clusterwise regression to daily rainfall data at Bandung, Bogor, Citeko and Jatiwangi stations. First, principal component analysis (PCA) was performed for predictor data reduction (GCM) as a solution for multicollinearity.

Two-clusters and three-clusters simulations aim to see the ability of clusterwise regression to classify data according to their actual distribution. The two-clusters simulation has (3) scenarios with the Gamma-Gamma (GG), Normal-Normal (NN) and Gamma-Normal (GN) models. The three-clusters simulation has (4) scenarios with the Gamma-Gamma-Gamma (GGG), Normal-Normal-Normal (NNN), Gamma-Normal-Normal (GNN) and Gamma-Gamma-Normal (GNN) models. The Gamma distribution with the shape parameter (ξ) consists of $\xi=0,5$, $\xi=0,85$ and $\xi=15$. The simulation results show clusterwise regression for the Gamma, Normal and mixed distribution response data scenario is able to form clusters appropriately according to the characteristics of the actual data (generated data).

The application of clusterwise regression to the daily rainfall of the Bogor, Bandung, Citeko and Jatiwangi rain stations begins with no-cluster modeling (Gamma regression) and with the clustering of the Gamma-Normal mix distribution.

Hak Cipta Dilindungi Undang-undang

IPB University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Prediction data using the centroid distance and distance between observations. Estimates in clusterwise regression were carried out in two and three clusters. The results showed that three-clusters produced better RMSEP than two-clusters. The best model for all rain stations is the Gamma-Normal-Normal (GNN) model, which is a model with one Gamma and two Normal distribution. The prediction method using the centroid distance approach and the distance between observations have the same ability to predict data.

Keywords: cluster regression, Gamma distribution, General circulation model, principal component analysis, Statistical downscaling.

@Hak cipta milik IPB University

IPB University





© Hak Cipta Milik IPB, Tahun 2021 Hak Cipta Dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PENGEMBANGAN PREDIKSI REGRESI GEROMBOL SEBARAN GAMMA PADA *STATISTICAL DOWNSCALING* UNTUK PENDUGAAN CURAH HUJAN HARIAN

**(Aplikasi pada Curah Hujan Harian di Beberapa Stasiun di Jawa Barat
periode 2010 – 2019)**

MUTHIA NADHIRA FALADIBA

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister Sains pada
Program Studi Statistika Terapan

**STATISTIKA TERAPAN
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2021**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Tesis:

Dr. Ir. I Made Sumertajaya, M.Si.





Hak cipta milik IPB University

IPB University

Judul Tesis : Pengembangan Prediksi Regresi Gerombol Sebaran Gamma Pada *Statistical Downscaling* Untuk Pendugaan Curah Hujan Harian. (Aplikasi pada Curah Hujan Harian di Beberapa Stasiun di Jawa Barat periode 2010 – 2019)

Nama : Muthia Nadhira Faladiba
NIM : G152180401

Disetujui oleh

Pembimbing 1:
Dr. Agus Mohamad Soleh, S.Si., M.T.



Pembimbing 2:
Dr. Ir. Anik Djuraidah, MS.



Diketahui oleh

Ketua Program Studi:
Dr. Kusman Sadik, S.Si., M.Si
NIP 196909121997021001



Dekan Fakultas/Sekolah:
Prof. Dr. Anas Miftah Fauzi, M.Eng
NIP 196004191985031002



Tanggal Ujian:
21 Januari 2020

Tanggal Lulus: 23 FEB 2021

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PRAKATA

Alhamdulillah wasyukurillah segala puji syukur penulis panjatkan kehadirat Allah subhanahu wa ta'ala atas segala karunia dan ridho-Nya sehingga tesis dengan judul “Pengembangan Prediksi Regresi Gerombol Sebaran Gamma Pada *Statistical Downscaling* Untuk Pendugaan Curah Hujan Harian. (Aplikasi pada Curah Hujan Harian di Beberapa Stasiun di Jawa Barat periode 2010 – 2019)” ini dapat diselesaikan.

Terima kasih penulis ucapkan kepada Bapak Dr Agus M Soleh, M.T, dan ibu Dr. Anik Djuraidah MS, selaku pembimbing yang telah memberikan arahan, waktu, dan ilmu yang bermanfaat, juga kepada bapak Dr. Ir. I Made Sumertajaya, M.Si selaku dosen penguji tesis yang telah memberikan banyak saran dan arahan. Selain itu, penghargaan penulis sampaikan kepada seluruh Pimpinan, staf pengajar sekolah Pascasarjana, dan pegawai Departemen Statistika IPB yang telah banyak memberikan ilmu dan arahan selama perkuliahan sampai dengan penyusunan tesis ini.

Ucapan terima kasih yang tulus dan penghargaan yang tak terhingga juga penulis ucapkan kepada Ayahanda bapak Drs Erwanda, M.Pd dan Ibunda tercinta ibu Elyanti Nurlaeli S.Pd yang telah membesarkan dan mendidik penulis dengan penuh kasih sayang demi keberhasilan penulis selama menjalani proses pendidikan, dan keluarga besar atas doa dan semangatnya.

Semoga penelitian ini dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan dan teknologi serta dapat memperkaya wawasan bagi para pembaca.

Bogor, Februari 2021

Muthia Nadhira Faladiba

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



DAFTAR ISI

DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penelitian	2
TINJAUAN PUSTAKA	3
2.1 <i>Statistical Downscaling</i>	3
2.2 Regresi Gerombol	4
2.3 Model Linear Terampat	4
III METODE	7
3.1 Data	7
3.2 Prosedur Analisis Data	8
3.2.1 Simulasi Sebaran Gamma, Normal dan Campuran	8
3.2.2 Penerapan Algoritma Regresi Gerombol pada Data Curah Hujan	11
IV HASIL DAN PEMBAHASAN	14
4.1 Simulasi Model 2 Gerombol Sebaran Gamma-Gamma	14
4.2 Simulasi Model 2 Gerombol Sebaran Normal-Normal	15
4.3 Simulasi Model 2 Gerombol Sebaran Gamma-Normal	16
4.4 Simulasi Model 3 Gerombol Sebaran Gamma-Gamma-Gamma	19
4.5 Simulasi Model 3 Gerombol Sebaran Normal-Normal-Normal	20
4.6 Simulasi Model 3 Gerombol Sebaran Gamma-Normal-Normal	21
4.7 Simulasi Model 3 Gerombol Sebaran Gamma-Gamma-Normal	22
4.8 Aplikasi Regresi Gerombol dengan Sebaran Gamma	26
4.8.1 Eksplorasi Data	26
4.8.2 Analisis Komponen Utama	28
4.8.3 Pemodelan Regresi Gerombol	30
V SIMPULAN DAN SARAN	37
5.1 Simpulan	37
DAFTAR PUSTAKA	38
LAMPIRAN	40
RIWAYAT HIDUP	50

Hak cipta milik IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR TABEL

1	Simulasi model regresi dua kelompok	9
2	Simulasi model regresi tiga kelompok	10
3	Perhitungan akurasi atau ketepatan klasifikasi	11
4	Rangkuman akurasi model dua kelompok pada data simulasi	18
5	Akurasi dan proporsi gerombol terbaik pada model simulasi dua kelompok	18
6	Rangkuman akurasi model tiga kelompok pada data simulasi	24
7	Akurasi dan proporsi gerombol terbaik pada model simulasi tiga kelompok	25
8	Statistik deskriptif curah hujan pada 4 stasiun hujan	26
9	Proporsi kumulatif keragaman komponen utama pada 4 stasiun hujan	28
10	Nilai RMSE regresi gerombol dari 4 stasiun hujan	30
11	Nilai RMSEP regresi gerombol dari 4 stasiun hujan	30
12	Proporsi gerombol terbaik pada 4 stasiun hujan	33

DAFTAR GAMBAR

1	Ilustrasi <i>statistical downscaling</i>	3
2	Peta sebaran 4 stasiun hujan terpilih di Jawa Barat (peta tidak berskala)	7
3	Diagram alur regresi gerombol	8
4	Diagram sebaran data simulasi Gamma 2 kelompok	14
5	Diagram sebaran data simulasi Gamma 2 kelompok (2)	15
6	Diagram sebaran data simulasi normal 2 kelompok	16
7	Diagram sebaran data simulasi GN 2 kelompok	17
8	Diagram sebaran data simulasi GN 2 kelompok (2)	17
9	Diagram sebaran data simulasi Gamma 3 kelompok	19
10	Diagram sebaran data simulasi normal 3 kelompok	20
11	Diagram sebaran data simulasi GNN 3 kelompok	21
12	Diagram sebaran data simulasi GNN 3 kelompok (2)	22
13	Diagram sebaran data simulasi GGN 3 kelompok	23
14	Diagram sebaran data simulasi GGN 3 kelompok (2)	24
15	Diagram kotak-garis pada 4 stasiun hujan	26
16	Histogram curah hujan pada 4 stasiun hujan	27
17	Diagram <i>scree</i> pada 4 stasiun hujan	29
18	Nilai RMSEP pada 50 ulangan untuk tiga gerombol stasiun Bandung	32
19	Nilai RMSEP pada 50 ulangan untuk tiga gerombol stasiun Bogor	32
20	Nilai RMSEP pada 50 ulangan untuk tiga gerombol stasiun Citeko	33
21	Nilai RMSEP pada 50 ulangan untuk tiga gerombol stasiun Jatiwangi	33
22	Diagram kotak-garis dugaan curah hujan pada gerombol terbaik stasiun Bandung	35
23	Diagram kotak-garis dugaan curah hujan pada gerombol terbaik stasiun Bogor	35
24	Diagram kotak-garis dugaan curah hujan pada gerombol terbaik stasiun	



	Citeko	36
25	Diagram kotak-garis dugaan curah hujan pada gerombol terbaik stasiun Jatiwangi	36

DAFTAR LAMPIRAN

	Diagram alur penelitian	40
	RMSEP simulasi dua kelompok stasiun Bandung	41
	RMSEP simulasi dua kelompok stasiun Bogor	42
	RMSEP simulasi dua kelompok stasiun Citeko	43
	RMSEP simulasi dua kelompok stasiun Jatiwangi	44
	RMSEP simulasi tiga kelompok stasiun Bandung	45
	RMSEP simulasi tiga kelompok stasiun Bogor	46
8	RMSEP simulasi tiga kelompok stasiun Citeko	47
9	RMSEP simulasi tiga kelompok stasiun Jatiwangi	48