

PEMODELAN CLUSTERWISE REGRESSION PADA STATISTICAL DOWNSCALING UNTUK PENDUGAAN CURAH HUJAN BULANAN*

Victor P. Butar-butur¹, Agus M. Soleh^{2‡}, and Aji H. Wigena³

¹Department of Statistics, IPB University, Indonesia, vctrbr@gmail.com

²Department of Statistics, IPB University, Indonesia, agusms@apps.ipb.ac.id

³Department of Statistics, IPB University, Indonesia, ajiwigena@ymail.com

[‡]corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 3 No 3 (2019), 236 - 246

Copyright © 2019 Victor P. Butar-butur, Agus M. Soleh, and Aji H. Wigena. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Statistical downscaling (SDS) is one of the developing models for rainfall estimation. The SDS model is a regression model used to analyze the relation of global (GCM output) and local data (rainfall). Rainfall has large variance so that clustering is needed to minimize the variance. One of the analytical methods that can be used in clustering rainfall estimation is cluster wise regression. There are three methods for clusterwise regression namely linear regression, finite mixture method (FMM) and cluster-weighted method (CWM). This study used GCM outputs data namely CFRsv2 as a covariate. The response variable is rainfall data in four stations such as Bandung, Bogor, Citeko and Jatiwangi from BMKG. The purpose of this study is to increase the accuracy of rainfall estimation using the three methods and compare the clusterwise regression with PCR and PLS models. Based on the value of RMSEP, the clusterwise regression with FMM was the best method to estimate rainfall in four stations.

Keywords: cluster-weighted method, clusterwise regression, finite mixture method, statistical downscaling.

1. Pendahuluan

Perubahan iklim dapat menjadi salah satu masalah bagi manusia, khususnya pada bidang pertanian, peternakan dan beberapa bidang lainnya. Salah satu faktor iklim adalah curah hujan dan salah satu anomali iklim adalah El Nino atau La Nina. EL Nino adalah penurunan curah hujan yang dapat menyebabkan kekeringan, dan La Nina

* Received Jun 2019; Accepted Oct 2019; Published online on Oct 2019

adalah peningkatan curah hujan yang dapat menyebabkan banjir. Peningkatan atau penurunan curah hujan yang besar (ekstrem) sering menjadi salah satu masalah, seperti gagal panen pada bidang pertanian. Oleh karena itu pendugaan curah hujan sangat penting dilakukan untuk mengantisipasi peningkatan atau penurunan curah hujan yang ekstrem (Irawan, 2016).

General Circulation Model (GCM) merupakan model matematik yang dapat digunakan sebagai pendugaan curah hujan. GCM menghasilkan informasi yang berskala global, tetapi mempunyai resolusi sangat rendah untuk menduga curah hujan yang terjadi pada skala lokal. Meskipun demikian data GCM masih dapat digunakan untuk mendapatkan informasi di stasiun berskala lokal dengan menggunakan teknik *downscaling* (Wigena, 2011).

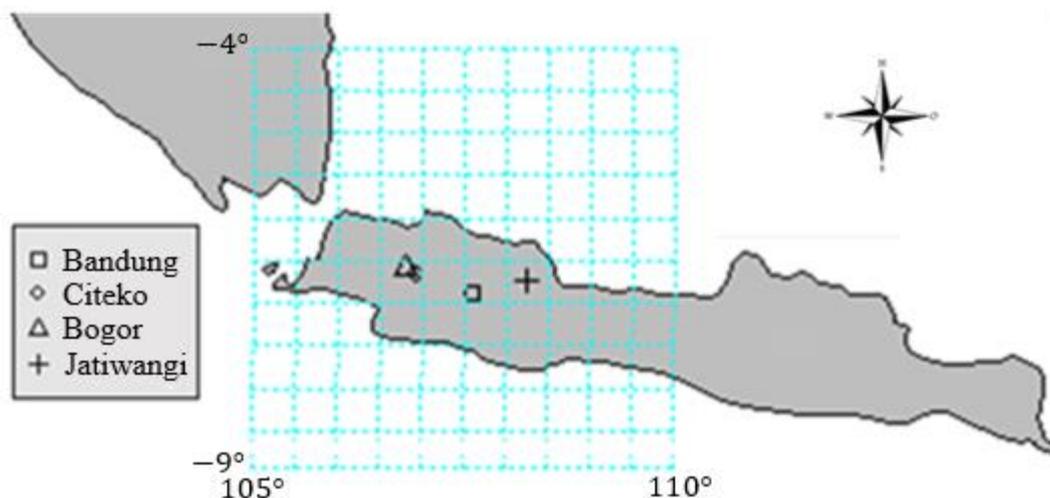
Model *Statistical Downscaling* (SDS) adalah salah satu model untuk pendugaan curah hujan. Model SDS adalah model statistik yang digunakan untuk menganalisis hubungan antara data berskala global dengan data berskala lokal. SDS didasarkan pada hubungan empiris antara data GCM sebagai data global dengan data stasiun sebagai data lokal. Model SDS tersusun atas kovariat dengan skala besar dan multikolinearitas (Soleh *et al.*, 2015).

Pemodelan SDS telah banyak dilakukan di Indonesia. Permatasari *et al.* (2017) menggunakan SDS dengan sebaran gamma dan regularisasi *elastic net* untuk pendugaan curah hujan dengan data *Global Precipitation Climatology Project* (GPCP). Penelitian ini menghasilkan nilai RMSEP yang lebih kecil jika ditambahkan peubah *dummy* gerombol curah hujan. Irvan *et al.* (2017) menggunakan regresi linear dengan persentil L_1 dan persentil L_2 untuk pendugaan curah hujan bulanan dengan data *Couple Model Intercomparison Project* (CMIP5). Penambahan peubah *dummy* pada persentil L_1 menghasilkan prediksi curah hujan yang lebih baik seperti dalam Ali *et al.* (2017) yang menggunakan sebaran gamma dengan persentil lasso dan gulud untuk pendugaan curah hujan bulanan dengan data CMIP5. Model gamma GLM persentil L_1 lebih konsisten dibandingkan dengan gamma GLM persentil L_2 dalam memprediksi curah hujan. Nadya (2018) menggunakan regresi linear gerombol dan pemodelan dua tahap dengan data GPCP. Metode regresi gerombol menghasilkan nilai galat prediksi yang lebih kecil dibandingkan dengan penambahan peubah *dummy* menggunakan metode *Classification and Regression Trees* (CART), Regresi Komponen Utama (RKU) dan RKU dengan peubah *dummy* hasil dari regresi logistik. Penelitian tersebut menggunakan keluaran GCM yang tidak dapat meramal periode ke depan kecuali CMIP5 yang memperoleh data simulasi.

Climate Forecast System Reanalysis Volume 2 (CFSRv2) adalah jenis GCM yang pada saat ini dapat menduga curah hujan beberapa waktu ke depan. (Khairunnisa *et al.*, 2019) memodelkan curah hujan dengan memanfaatkan data CFSRv2 sebagai data global dan data curah hujan pada empat stasiun di Provinsi Jawa Barat yaitu stasiun Bandung, Bogor, Citeko dan Jatiwangi dengan Regresi Kuadrat Terkecil Parsial (RKTP) dan RKU. Hasil pendugaan curah hujan menggunakan kedua metode tersebut masih menghasilkan galat prediksi atau *Root Mean Square Error of Prediction* (RMSEP) yang besar. Pitri *et al.*, (2018) menggunakan data CFSRv2 dalam penelitian SDS menggunakan penggerombolan K-means yang menghasilkan nilai

RMSEP yang lebih kecil. Penggerombolan dapat memberikan hasil RMSEP yang lebih kecil dibandingkan dengan tanpa penggerombolan.

Data curah hujan pada stasiun pengamatan beragam besar sehingga penggerombolan dibutuhkan untuk meminimalkan nilai ragam. Salah satu metode yang dapat digunakan untuk meminimalkan nilai ragam dalam memodelkan curah hujan adalah regresi gerombol sebagaimana Nadya (2018) yang menghasilkan nilai RMSEP yang lebih baik. Penelitian ini akan menerapkan regresi gerombol dengan data luaran CFSRv2 sebagai kovariat. Penerapan regresi gerombol diduga dapat memperkecil nilai RMSEP. *Software R* menyediakan beberapa paket untuk menerapkan regresi gerombol yaitu *Flexible Procedures of Clustering* (fpc) dengan regresi linear, *Flexible Mixture Modeling* (flexmix) dengan *Finite Mixture Model* (FMM), dan *Flexible Cluster-Weighted Modeling* (flexCWM) dengan *Cluster Weighted Model* (CWM). Penelitian ini bertujuan melakukan pemodelan SDS untuk pendugaan curah hujan menggunakan tiga metode pada empat stasiun di Provinsi Jawa Barat yaitu stasiun Bandung, Bogor, Citeko dan Jatiwangi.



Gambar 1: Letak stasiun curah hujan penelitian.

2. Metodologi

2.1 Data

Penelitian dilakukan menggunakan data luaran GCM yang dikeluarkan oleh *Climate Forecast System* (CFS) diambil dari laman <https://rda.ucar.edu/>. CFS adalah sebuah model yang menggambarkan sebuah interaksi global antara daratan, lautan dan udara bumi yang diukur setiap 6 jam (00:00, 06:00, 12:00, 18:00) atau 4 kali dalam sehari. CFS dihasilkan oleh beberapa peneliti yang dikeluarkan oleh dari *National Centers for Environmental Prediction* (NCEP). Peubah yang digunakan adalah jumlah curah hujan bulanan (*Precipitation Rate*) sebagai peubah kovariat (X) mulai dari Januari 2011 sampai Desember 2017 yang terletak pada 4° LS sampai dengan 9° LS dan 105° BT sampai dengan 110° BT dengan ukuran grid 0.5°x0.5°. Domain yang digunakan berdasarkan Khairunnisa et al., (2019) yaitu domain 9x9 grid untuk stasiun Bandung, domain 8x8 grid untuk stasiun Citeko dan domain 6x6 grid untuk stasiun

Bogor dan stasiun Jatiwangi. Gambar 1 (Khairunnisa *et al.*, 2019) memperlihatkan letak keempat stasiun.

Data peubah respon (y) adalah data curah hujan bulanan dari stasiun Bandung, Bogor, Citeko dan Jatiwangi dari Januari 2011 sampai dengan Desember 2017 yang dikeluarkan oleh <http://dataonline.bmkg.go.id/> Badan Meteorologi, Klimatologi dan Geofisika (BMKG). Stasiun Bandung terdiri dari 83 data curah hujan, stasiun Bogor terdiri dari 83 data curah hujan, stasiun Citeko terdiri dari 82 data curah hujan, dan stasiun Jatiwangi terdiri dari 82 data curah hujan.

2.2 Metode Penelitian

a. Regresi Linear Gerombol

Regresi gerombol menyatukan analisis gerombol kedalam kerangka regresi sehingga parameter dapat diduga secara bersama-sama. Gerombol diperoleh berdasarkan parameter regresi dari subpopulasi heterogen yang tidak teramati. Model regresi gerombol memungkinkan koefisien regresi bervariasi dengan amatan-amatan dari berbagai gerombol (Lau *et al.*, 1999).

Regresi gerombol melakukan pendugaan kemungkinan maksimum untuk regresi linear dengan galat normal, dengan fungsi sebagai berikut:

$$y = \mathbf{X}\beta_k$$

Dalam hal ini y adalah peubah respon, \mathbf{X} adalah kovariat tetap, k adalah gerombol dan β_k adalah koefisien regresi pada gerombol ke- k (DeSarbo & Cron, 1988).

b. Finite Mixture Model (FMM)

FMM adalah model peluang yang menggabungkan dua atau lebih fungsi kepekatan. Konsep dari FMM adalah data yang diobservasi merupakan gabungan dari dua atau lebih subpopulasi yang berbeda belum terobservasi.

Fungsi *finite mixtures* dengan komponen K adalah:

$$h(y|\mathbf{x}, \psi) = \sum_{k=1}^K \lambda_k f(y|\mathbf{x}, \theta_k) \quad \lambda_k \geq 0, \sum_{k=1}^K \lambda_k = 1$$

Dalam hal ini y adalah peubah respon dengan kepekatan peluang h , \mathbf{x} adalah sebuah vektor dari peubah bebas (kovariat tetap), λ_k adalah peluang prior (proporsi) dari komponen K , θ_k adalah vektor parameter komponen untuk fungsi kepekatan peluang f , dan $\psi = (\lambda_1, \dots, \lambda_K, \theta'_1, \dots, \theta'_K)'$ adalah vektor dari semua parameter (Grun & Leisch, 2007).

c. Cluster-Weighted Model (CWM)

Model regresi campuran dengan kovariat acak adalah *Cluster-Weighted Model* (CWM). CWM mengasumsikan hubungan fungsional $Y|X = \mathbf{x}$ memfaktorkan distribusi gabungan $p(\mathbf{x}, y)$ di antara distribusi bersyarat $Y|X = \mathbf{x}$ dan distribusi marginal dari \mathbf{X} .

Fungsi dari (\mathbf{X}, Y) pada CWM dapat ditulis sebagai berikut:

$$p(\mathbf{x}, y; \vartheta) = \sum_{k=1}^k \lambda_k p(y|\mathbf{x}; \beta_k, \gamma_k) p(\mathbf{x}; \alpha_k)$$

Dalam hal ini λ_k adalah proporsi gerombol, dengan $\lambda_k > 0$ dan $\sum_{k=1}^K \lambda_k = 1$, $p(y|x; \beta_k, \gamma_k)$ adalah peluang kepekatan peubah respon Y terhadap vektor \mathbf{X} , β_k adalah koefisien regresi, γ_k adalah parameter tambahan berdasarkan distribusi yang digunakan, $p(x; \alpha_k)$ adalah peluang kepekatan (terhadap α_k) dari \mathbf{X} (kovariat acak) dan ϑ mengandung semua parameter model (Mazza et al., 2018).

2.3 Pengolahan dan Analisis data

Tahapan analisis data pada penelitian ini adalah sebagai berikut:

1. Mempersiapkan data CFS dan data stasiun lokal.
 - a. Mengkonversi data CFS dari format *ncdf4* menjadi format RData menggunakan *software* R dengan paket *ncdf4*.
 - b. Mengkonversi satuan pengukuran data luaran CFS dari $\frac{\text{kg}}{\text{m}^2\text{s}}$ menjadi $\frac{\text{mm}}{\text{day}}$ mengalikan data luaran GCM dengan 86400.
 - c. Mengunduh data harian curah hujan pada BMKG dengan format CSV.
 - d. Mengkonversi data harian curah hujan stasiun menjadi data jumlah curah hujan bulanan. Rata-rata data harian curah hujan setiap bulan dan dikali dengan jumlah hari pada bulan tersebut (dalam kondisi terdapat data curah hujan harian yang hilang).
 - e. Menyandingkan data CFS dengan data stasiun.
2. Mengeksplorasi data melalui diagram kotak garis untuk melihat sebaran tingkat curah hujan pada stasiun kajian.
3. Menganalisis data luaran CFS menggunakan analisis komponen utama (AKU) untuk mengatasi multikolinearitas dan menggunakan komponen utama yang terpilih untuk membangun model regresi gerombol.
4. Evaluasi model dilakukan dengan validasi silang *5-fold* yang diulang 2 kali.
5. Data *training* dari validasi silang digunakan untuk membuat model dengan regresi gerombol.
 - a. Menerapkan regresi gerombol menggunakan metode regresi linear dengan banyak gerombol 2, 3 atau 4.
 - b. Menerapkan regresi gerombol menggunakan metode FMM dengan banyak gerombol 2, 3 atau 4 menggunakan *gaussian* atau model linear.
 - c. Menerapkan regresi gerombol menggunakan metode CWM dengan banyak gerombol 2, 3 atau 4 menggunakan *gaussian* atau model linear.
6. Data *testing* diidentifikasi ke dalam gerombol menggunakan cara berikut:
 - a. Semua amatan data *testing* dimasukkan ke dalam setiap gerombol dan dihitung masing – masing nilai RMSEP, setelah nilai RMSEP masing – masing gerombol didapatkan maka dapat dihitung nilai rata – rata gabungan RMSEP.
 - b. Menduga amatan data *testing* ke dalam gerombol dengan menggunakan jarak terdekat atau euclid:

$$d_i = \sqrt{\sum_{j=1}^P (y_{ij} - \bar{x}_{ij})^2}$$

Dalam hal ini:

\bar{x}_{ij} = rata – rata pada data *training* peubah ke – j pada gerombol ke – i

y_{ij} = amatan pada data *testing* peubah ke – j pada gerombol ke – i

Dengan menggunakan nilai jarak euclid terkecil maka amatan data *testing* dapat digerombolkan dan dicari nilai RMSEP.

Data *testing* digunakan untuk memvalidasi kemampuan prediksi model dengan cara menghitung korelasi dan nilai RMSEP:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_{(-i)} - \hat{y}_{(-i)})^2}{n}}$$

(-i) = Observasi yang tidak digunakan dalam menduga model; i: 1, 2, ... , n

3. Hasil dan Pembahasan

3.1 Eksplorasi Data

a. Pola Curah Hujan di setiap Stasiun

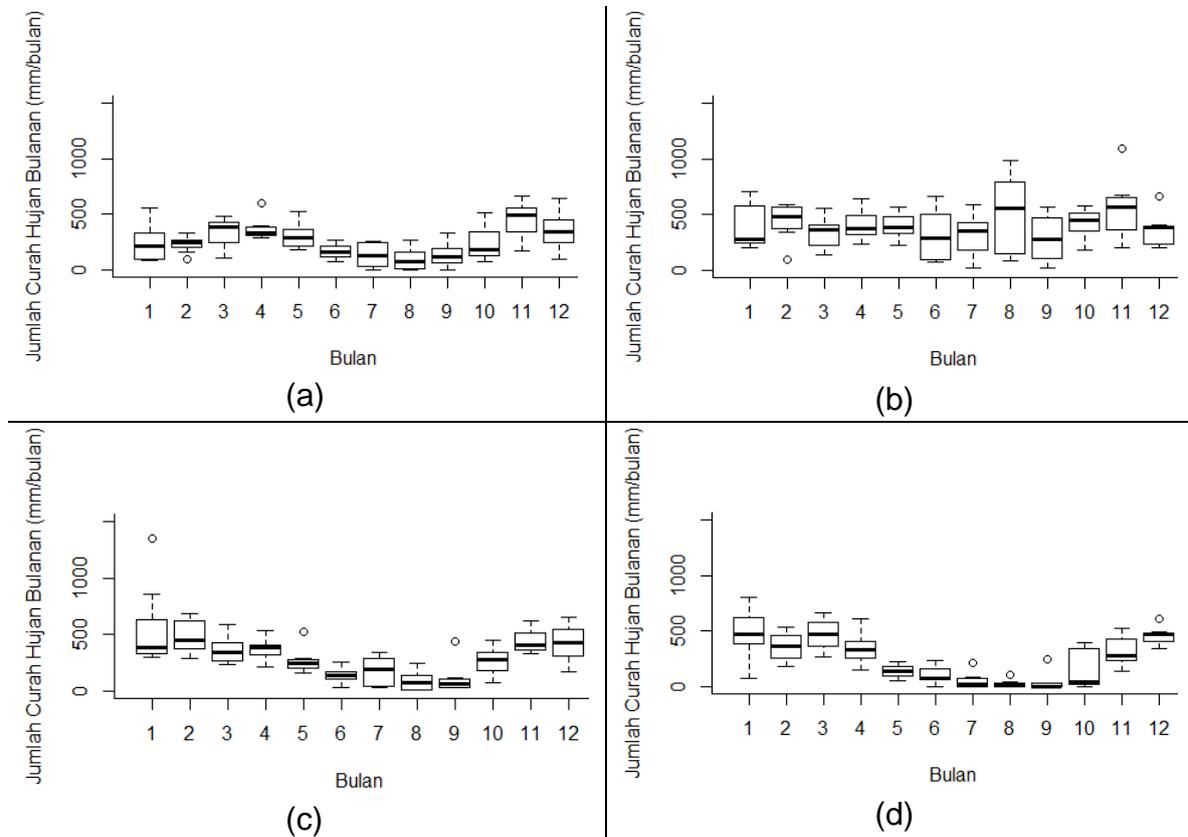
Gambar 2(a) memperlihatkan pola curah hujan untuk stasiun Bandung bulanan dari tahun 2011 – 2017. Di stasiun Bandung, curah hujan tertinggi terjadi pada bulan November dan terendah terjadi pada bulan Juli, Agustus dan September. Gambar 2(b) memperlihatkan pola curah hujan untuk stasiun bulanan Bogor. Di stasiun Bogor, curah hujan tertinggi terjadi pada bulan November dan curah hujan terendah terjadi pada bulan Juli dan September.

Gambar 2(c) memperlihatkan pola curah hujan untuk stasiun bulanan Citeko. Curah hujan tertinggi terjadi di stasiun Citeko pada bulan Januari sedangkan curah hujan terendah terjadi pada bulan Juni dan Agustus. Pola curah hujan pada stasiun Citeko menunjukkan pola monsunial, yaitu pola berbentuk U. Pola monsunial adalah daerah yang dimana memiliki perbedaan yang jelas antar musim hujan dan musim kemarau. Gambar 2(d) memperlihatkan pola curah hujan untuk stasiun bulanan Jatiwangi. Pola curah hujan pada stasiun Jatiwangi berbentuk U yang artinya menunjukkan pola monsunial. Curah hujan tertinggi terjadi pada bulan Januari dan curah hujan terendah terjadi antara bulan Juni sampai Oktober.

b. Pemilihan Komponen Utama pada Data CFSRv2

Data luaran GCM yang dikeluarkan oleh CFS (CFRSv2) merupakan rata-rata curah hujan bulanan yang terdapat pada domain terbaik berbentuk persegi (Khairunnisa *et al.*, 2019). Stasiun Bandung dengan 81 peubah kovariat, stasiun Bogor memiliki 36 peubah kovariat, stasiun Citeko dengan 64 peubah kovariat dan stasiun Jatiwangi memiliki 36 peubah kovariat. Banyaknya kovariat menyebabkan terjadi multikolinearitas, sehingga untuk mengatasi hal tersebut digunakan analisis komponen utama. Analisis komponen akan menghasilkan peubah – peubah yang baru atau komponen utama (KU) yang saling bebas. Penentuan KU berdasarkan *scree plot*. *Scree plot* adalah plot dari nilai akar ciri pada setiap KU berdasarkan persentase ragam

kumulatif yang dapat dijelaskan pada masing – masing KU. Dalam penentuan KU, nilai akar ciri pada KU lebih besar dari 1 dapat digunakan sebagai kovariat. Persentase ragam kumulatif yang digunakan untuk menentukan banyak KU adalah sebesar 70, 80, atau 90 (Mattjik & Sumertajaya, 2011).

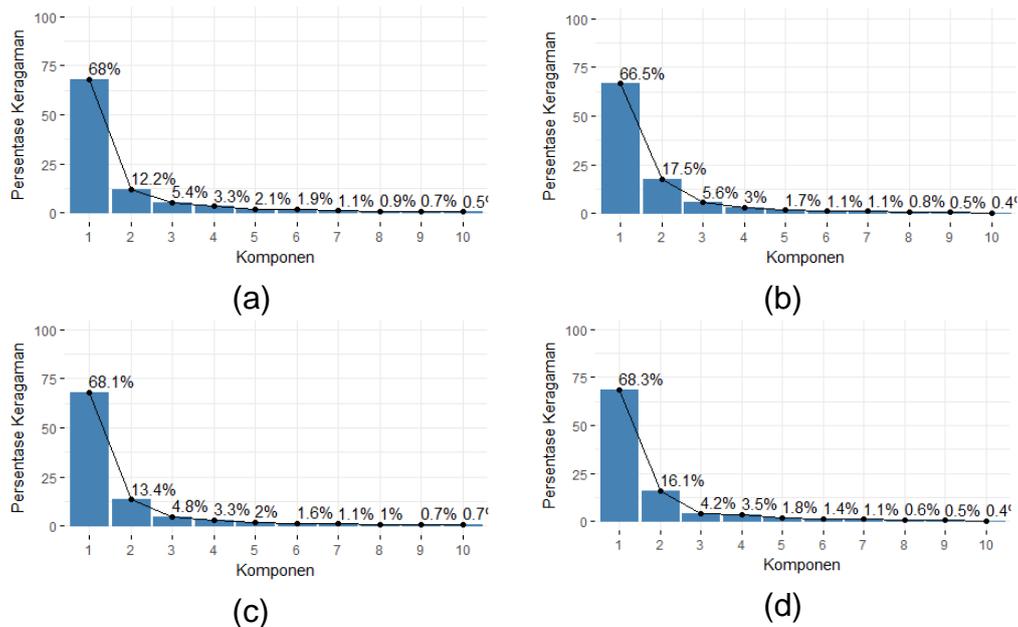


Gambar 2: *Boxplot* sebaran curah hujan bulanan pada stasiun: (a) Bandung (b) Bogor (c) Citeko (d) Jatiwangi.

Tabel 1: Nilai eigen dan ragam kumulatif.

Stasiun	Komponen Utama			
	Kriteria	KU 1	KU 2	KU 3
Bandung	Nilai Eigen	1728.16	3092.64	136.11
	Kumulatif	67.98	80.15	85.50
	Ragam (%)			
Bogor	Nilai Eigen	903.37	238.28	75.78
	Kumulatif	66.53	84.08	89.63
	Ragam (%)			
Citeko	Nilai Eigen	1460.90	288.05	102.21
	Kumulatif	68.06	81.48	86.23
	Ragam (%)			
Jatiwangi	Nilai Eigen	786.61	181.06	47.54
	Kumulatif	68.31	84.41	88.63
	Ragam (%)			

Nilai akar ciri dan ragam kumulatif pada empat stasiun diperlihatkan pada Tabel 1 dan *scree plot* untuk empat stasiun disajikan pada Gambar 3. KU yang digunakan sebagai kovariat untuk stasiun Bandung adalah sebanyak tiga. KU tiga mempunyai nilai akar ciri lebih besar dari satu. Persentase ragam yang terjadi setelah KU tiga tidak terdapat perubahan yang signifikan. Persentase ragam kumulatif pada KU tiga adalah 85.50. Demikian juga dengan stasiun Bogor, Citeko dan Jatiwangi KU yang di gunakan sebanyak tiga dengan persentase ragam kumulatif masing-masing sebesar 89.63, 86.23 dan 88.63.



Gambar 3: *Scree plot* untuk stasiun (a) Bandung (b) Bogor (c) Citeko (d) Jatiwangi.

3.2 Pemodelan dengan *Clusterwise Regression*

Clusterwise Regression atau regresi gerombol melakukan dua analisis secara bersama – sama yaitu analisis gerombol dan regresi. Terdapat tiga metode yang dapat melakukan regresi gerombol yaitu Regresi Linear (RL), Finite Mixture (FMM) dan *Cluster-Weighted* (CWM). Penelitian ini membandingkan nilai RMSEP yang dihasilkan analisis regresi gerombol dengan menggunakan 2, 3, dan 4 gerombol. Nilai RMSEP yang terbaik akan digunakan sebagai model dalam pendugaan curah hujan. Tabel 2 menunjukkan nilai RMSEP dari ketiga metode.

Analisis regresi gerombol dengan paket RL menghasilkan dua gerombol untuk stasiun Bandung. Dua gerombol tersebut menghasilkan pendugaan yang terbaik dengan nilai RMSEP terkecil dengan cara rata-rata gabungan. Dua gerombol untuk stasiun Bogor juga menghasilkan pendugaan curah hujan yang terbaik dengan cara jarak terdekat. Tiga gerombol untuk stasiun Citeko menghasilkan pendugaan curah hujan yang terbaik. Demikian juga untuk stasiun Jatiwangi, tiga gerombol menghasilkan pendugaan curah hujan yang terbaik.

Analisis regresi gerombol dengan paket CWM menghasilkan dua gerombol untuk stasiun Bandung. Dua gerombol menghasilkan pendugaan curah hujan yang terbaik dengan nilai RMSEP dengan cara rata-rata gabungan dan dengan cara jarak terdekat. Dua gerombol untuk stasiun Bogor dan Jatiwangi juga menghasilkan pendugaan curah hujan. Empat gerombol untuk stasiun Citeko menghasilkan pendugaan curah hujan yang terbaik.

Tabel 2: Nilai RMSEP RL, FMM dan CWM.

Stasiun	Gerombol	RMSEP				
		RL (a)*	RL (b)*	FMM	CWM (a)*	CWM (b)*
Bandung	2	136.68	146.77	75.69	138.03	109.81
	3	144.62	146.35	78.58	310.83	284.55
	4	155.70	165.84	71.83	236.50	166.99
Bogor	2	246.45	243.93	142.43	569.08	247.33
	3	295.74	312.78	124.92	461.39	293.55
	4	360.03	330.10	109.25	691.47	367.66
Citeko	2	171.61	183.07	125.01	252.05	193.91
	3	180.70	160.84	109.73	385.33	264.04
	4	193.33	207.76	126.67	276.38	142.04
Jatiwangi	2	135.66	131.95	106.94	143.31	120.77
	3	148.88	125.18	103.27	227.69	152.06
	4	179.45	159.31	86.47	506.98	218.33

* (a) Rata-rata gabungan, (b) Jarak terdekat

Regresi gerombol dengan paket FMM menghasilkan empat gerombol untuk stasiun Bandung. Pendugaan curah hujan yang diperoleh pada empat gerombol lebih baik dibandingkan dengan gerombol yang lain. Empat gerombol untuk stasiun Bogor dan Jatiwangi juga menghasilkan pendugaan curah hujan yang terbaik. Tiga gerombol untuk stasiun Citeko menghasilkan pendugaan curah hujan yang terbaik.

3.3 Perbandingan *Clusterwise Regression*, RKU dan RKTP

Tabel 3 menunjukkan perbandingan nilai RMSEP terbaik menggunakan model regresi gerombol dengan ketiga paket dan model RKU dan RKTP yang dihasilkan oleh Khairunnisa et al., (2019). Pendugaan curah hujan terbaik di stasiun Bandung adalah regresi gerombol dengan metode FMM. Nilai RMSEP yang dihasilkan regresi gerombol pada stasiun Bandung sebesar 71.83 jauh lebih baik dibandingkan model lainnya. Regresi gerombol juga menghasilkan pendugaan curah hujan yang terbaik di stasiun Bogor. Nilai RMSEP sebesar 109.25 sangat baik dibandingkan model yang lain.

Regresi gerombol untuk stasiun Citeko menghasilkan pendugaan curah hujan terbaik dengan nilai RMSEP sebesar 109.73 sedangkan regresi gerombol untuk

stasiun Jatiwangi menghasilkan pendugaan curah hujan yang terbaik dengan nilai RMSEP sebesar 86.47.

Tabel 3: Nilai RMSEP curah hujan terbaik.

Stasiun	RMSEP			RKU*	RKTP*
	<i>Clusterwise Regression</i>				
	RL	FMM	CWM		
Bandung	136.68	71.83	109.81	101.42	100.06
Bogor	243.93	109.25	247.33	200.42	194.3
Citeko	160.84	109.73	142.04	132.44	117.6
Jatiwangi	102.78	86.47	120.77	113.87	108.7

*Sumber Khairunnisa *et al.*, (2019).

Tabel 4 menunjukkan regresi gerombol menggunakan metode FMM menghasilkan korelasi terbaik diantara model yang lain. Korelasi antar curah hujan aktual dan curah hujan prediksi yang dihasilkan regresi gerombol adalah kuat positif. Hal ini menjelaskan bahwa pendugaan curah hujan menggunakan model regresi gerombol dengan menggunakan regresi gerombol adalah sangat baik. Dugaan curah hujan yang dihasilkan regresi gerombol hampir sama dengan curah hujan aktual.

Tabel 4: Korelasi curah hujan terbaik.

Stasiun	Korelasi			RKU*	RKTP*
	<i>Clusterwise Regression</i>				
	RL	FMM	CWM		
Bandung	0.73	0.89	0.74	0.80	0.81
Bogor	0.12	0.85	0.14	0.41	0.38
Citeko	0.79	0.86	0.77	0.83	0.77
Jatiwangi	0.81	0.93	0.84	0.88	0.91

*Sumber Khairunnisa *et al.*, (2019).

4. Simpulan

Pendugaan curah hujan menggunakan regresi gerombol dengan metode FMM lebih baik dibandingkan RKU dan RKTP. Hal ini menunjukkan bahwa peningkatan akurasi pendugaan curah hujan dapat dilakukan dengan regresi gerombol khususnya dengan metode FMM.

Daftar Pustaka

Ali, I., Djuraidah, A., & Soleh, A. M. (2017). Gamma response regression with percentile lasso and ridge to estimate extreme rainfall. *Proceeding of ICSME: Contribution of Mathematics and Science Research for Sustainable Life in Facing Global Challenge*, 3, M-44-M-49. Semarang (ID): Faculty of Mathematics and Natural Sciences, Semarang State University, Indonesia.

- DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, **5**(2): 249–282.
- Grun, B., & Leisch, F. (2007). FlexMix: An R package for finite mixture modelling. *R News*, **7**(1): 8–13.
- Irawan, B. (2016). Fenomena anomali iklim El Niño dan La Niña: Kecenderungan jangka panjang dan pengaruhnya terhadap produksi pangan. *Forum Penelitian Agro Ekonomi*, **24**(1): 28–45.
- Irvan, M., Wigena, A. H., & Djuraidah, A. (2017). Linear regression with percentile lasso and ridge to predict rainfall. *Proceeding of ICSME: Contribution of Mathematics and Science Research for Sustainable Life in Facing Global Challenge*, **3**, M-56-M-62. Semarang (ID): Faculty of Mathematics and Natural Sciences, Semarang State University, Indonesia.
- Khairunnisa, K., Pitri, R., Butar-Butar, V. P., & Soleh, A. M. (2019). Pemanfaatan CFSRv2 untuk statistical downscaling menggunakan principal component regression dan partial least square. *Xplore: Journal of Statistics*, **8**(1): 37-44.
- Lau, K., Leung, P., & Tse, K. (1999). A mathematical programming approach to clusterwise regression model and its extensions. *European Journal of Operational Research*, **116**(3): 640–652.
- Mattjik, A. A., & Sumertajaya, I. M. (2011). *Sidik peubah ganda dengan menggunakan SAS*. Bogor (ID): IPB Press.
- Mazza, A., Punzo, A., & Ingrassia, S. (2018). FlexCWM: a flexible framework for cluster-weighted models. *J Stat Softw*, **86**(2): 1–30.
- Nadya, A. R. (2018). *Pemodelan Statistical Downscaling untuk Menduga Curah Hujan degan Regresi Linear Gerombol dan Pemodelan Dua Tahap* [Tesis]. Bogor (ID): IPB University.
- Permatasari, S. M., Djuraidah, A., & Soleh, A. M. (2017). Statistical downscaling with gamma distribution and elastic net regularization (case study: monthly rainfall 1981-2013 at Indramayu). *Proceeding The International Conference on Applied Statistics| Departemen Statistika FMIPA Universitas Padjadjaran*, **2**, 2. Bandung (ID): Departemen Statistika, FMIPA, Universitas Padjadjaran.
- Pitri, R., Soleh, A. M., & Djuraidah, A. (2018). Statistical downscaling modeling through k-means clustering. *International Journal of Scientific Research in Science, Engineering and Technology*, **4**(9): 220–227.
- Soleh, A. M., Wigena, A. H., Djuraidah, A., & Saefuddin, A. (2015). Statistical downscaling to predict monthly rainfall using linear regression with L1 regularization (LASSO). *Applied Mathematical Sciences*, **9**(108): 5361–5369.
- Wigena, A. H. (2011). Regresi kuadrat terkecil parsial multi respon untuk statistical downscaling (Multi response partial least square for statistical downscaling). *Forum Statistika dan Komputasi*, **16**(2):12-15.