

Performa Model *Statistical Downscaling* dengan Peubah *Dummy* Berdasarkan *K-Means* dan *Average Linkage*

Fitri Annisa^{1*}, Raupong², Sitti Sahriman³

¹²³Departemen Statistika Fakultas MIPA Universitas Hasanuddin, 90245,
Indonesia.

*Corresponding author, email: fitri.annisa1@gmail.com

Abstract

Climate change that occurs is often used to predict future climate conditions. For future climate predictions it is only possible to use climate models. One of the climate models used to predict climate conditions is the Global Circulation Models (GCM). GCM represents global climatic conditions but not on a regional or local scale. The approach that has been widely used to bridge the difference in scale is statistical downscaling. Large-scale GCM data allows for multicollinearity. estimation liu regression and principal component regression is used to solve the multicollinearity problem. In addition, dummy variables based on k-means and average linkage are used in the model to overcome the heterogeneous variance of residue. There are 4 dummy variables in the cluster technique. In this paper, Liu k-means regression model parameter estimation method is the best model.

Keywords: *Average Linkage, Estimastion Liu Regression, Global Circulation Model, K-Means, Dummy Variabel, Principal Component Regression, Statistical Downscaling.*

Abstrak

Perubahan iklim yang terjadi sering digunakan untuk meramalkan keadaan iklim yang akan datang. Untuk pendugaan iklim yang akan datang hanya mungkin dilakukan dengan menggunakan model iklim. Salah satu model iklim yang digunakan untuk meramalkan keadaan iklim adalah *Global Circulation Models* (GCM). GCM merepresentasikan keadaan iklim global tapi tidak dalam skala regional atau skala lokal. Pendekatan yang telah banyak dilakukan untuk menjembatani perbedaan skala itu adalah dengan *statistical downscaling*. Data GCM yang berskala besar memungkinkan adanya multikolinearitas. Metode Estimasi Parameter Model Regresi LIU dan Regresi Komponen Utama digunakan untuk mengatasi masalah multikolinearitas. Selain itu, peubah *dummy* berdasarkan *k-means* dan *Average Linkage* digunakan dalam model untuk mengatasi ragam sisaan yang heterogen. Terdapat 4 peubah *dummy* dalam teknik cluster. Pada makalah ini metode estimasi parameter model regresi Liu *k-means* merupakan model terbaik.

Kata Kunci: *Average Linkage, Estimasi Parameter Model Regresi Liu, Global Circulation Model, K-Means, Multikolinearitas, Peubah dummy, Regresi Komponen Utama, Statistical Downscaling.*

1. Pendahuluan

Curah hujan merupakan peubah yang paling sering digunakan dalam penelitian mengenai dampak perubahan iklim. Ketersediaan informasi ramalan tentang curah hujan sangat berguna bagi berbagai sektor tercakup di dalamnya, mulai dari masyarakat pesisirnya, nelayan, pulau-pulau kecil, perikanan, sampai sumber daya kelautan lainnya termasuk salah satunya garam. Produksi garam merupakan salah satu isu nasional yang

menjadi perhatian pemerintah saat ini. Pola Pertanian Garam masyarakat Indonesia masih menggunakan cara tradisional. Kendala yang dihadapi oleh petambak garam yaitu penurunan produksi yang diakibatkan oleh faktor cuaca yang tidak menentu[1].

Salah satu model iklim yang digunakan untuk meramalkan keadaan iklim adalah *Global Circulation Models* (GCM). GCM adalah alat terpenting dalam upaya memahami sistem iklim. GCM menghasilkan data dalam bentuk *grid* atau petak wilayah dengan resolusi rendah ($2,5^\circ$ atau $\pm 300 \text{ km}^2$) yang merepresentasikan keadaan iklim global tapi tidak dalam skala regional atau skala lokal. Untuk menggunakan data GCM sebagai penduga bagi penelitian perubahan iklim lokal, diperlukan pendekatan yang mampu mengatasi perubahan iklim tersebut. Salah satu pendekatan untuk mengatasi masalah tersebut digunakan Teknik *Downscaling*[2]. . Salah satu jenis teknik *downscaling* adalah *statistical downscaling* (SD), di mana merupakan model statistik yang menggambarkan hubungan antara data pada grid-grid berskala besar (GCM) dengan data pada grid berskala lebih kecil. Data GCM yang berskala besar memungkinkan adanya multikolinearitas[3].

Ada beberapa metode yang dapat dilakukan dalam mengatasi masalah multikolinearitas salah satunya adalah Regresi Komponen Utama (RKU), Estimasi Parameter Model Regresi LIU. Dalam hal ini metode yang digunakan Estimasi Parameter Model Regresi LIU. Pada tahun 1993 Liu memperkenalkan suatu metode yang dapat mengatasi multikolinearitas, metode tersebut disebut metode estimasi liu. Seperti halnya regresi ridge, metode estimasi liu juga memiliki nilai konstanta taksiran yang digunakan untuk mengatasi multikolinearitas.

Berdasarkan Uraian tersebut maka tujuan penelitian ini adalah mendapatkan nilai peramalan regresi metode estimasi parameter model regresi liu dan regresi komponen utama dengan peubah dummy berdasarkan *k-means* dan *average linkage*.

2. Material dan Metode

Data yang digunakan pada penelitian ini adalah data presipitasi GCM *Climate Model Intercomparison Project* (CMIP5) dalam satuan mm/bulan dan data curah hujan di sekitar wilayah Kabupaten Pangkep pada periode 1996-2017 yang diperoleh dari Skripsi Melly Amelia. Data curah hujan yang digunakan dalam penelitian ini merupakan rata-rata curah hujan pada 3 pos hujan di Kabupaten Pangkep yaitu Kecamatan Bunggoro, Kecamatan Labakkang, Kecamatan Ma'rang. Data GCM CMIP5 diperoleh dari situs web <http://www.climatexp.knmi.nl/> yang dikeluarkan oleh KNMI Belanda. Domain GCM yang digunakan dalam penelitian ini adalah sejumlah grid berbentuk persegi berukuran 8×8 grid ($2.5^\circ \times 2.5^\circ$) untuk setiap grid pada 109.54° BT s.d 129.54° BT dan -14.75° LS s.d 5.25° LU . Variabel penelitian terdiri dari satu variabel respon dan 64 variabel prediktor sesuai ukuran domain GCM. variabel respon (y) yaitu intensitas curah hujan (mm/bulan) dan variabel prediktor (x) yang akan digunakan dalam penelitian ini yaitu data curah hujan GCM.

Data dianalisis menggunakan metode estimasi parameter model regresi liu dan regresi komponen utama. Sebelum dianalisis, tahap pertama penentuan peubah *dummy* dengan mengelompokkan data curah hujan yang ditentukan berdasarkan teknik *k-means* dan *average linkage*.

Dalam mengatasi masalah multikolinearitas, Kejian Liu (1993) mengusulkan metode *Liu Estimator* yaitu suatu metode yang dapat mengatasi masalah multikolinearitas, dengan menambahkan augmentasi $d\hat{\beta} = \beta + \tilde{\epsilon}$ ke Persamaan $y = X\beta + \epsilon$ dan dengan menggunakan metode MKT maka akan diperoleh estimasi LIU yang dirumuskan sebagai berikut (Liu, 1993):

$$\hat{\beta}_{LE} = (X'X + I)^{-1}(X'X + d_L I)\hat{\beta}_{MKT} \quad (1)$$

atau

$$\hat{\alpha}_{LE} = (\Lambda + I)^{-1}(\Lambda + d_L I_P)\hat{\alpha}_{MKT}$$

Dimana d adalah konstanta *LIU* dan dirumuskan sebagai berikut (Liu, 1993):

$$d_L = \frac{\sum_{j=1}^p \left[\frac{\hat{\alpha}_{MKTj}^2 - \hat{\sigma}_{MKT}^2}{(\lambda_j + 1)^2} \right]}{\sum_{j=1}^p \left[\frac{\lambda_j \hat{\alpha}_{MKTj}^2 + \hat{\sigma}_{MKT}^2}{\lambda_j (\lambda_j + 1)} \right]} \quad (2)$$

dengan $\hat{\alpha} = Q'\hat{\beta}_{MKT}$ dan $\hat{\sigma}_{MKT}^2 = \frac{(y-\hat{y})'(y-\hat{y})}{n-p}$ adalah estimasi MKT dari α dan variansi eror dari MKT, dan Q adalah matriks orthogonal yang berukuran $p \times p$ yang elemennya adalah nilai eigen vektor dari $X'X$, dengan nilai $0 < d < 1$.

Nilai *Mean Square Error* (MSE) estimasi *LIU* dapat dicari dengan menggunakan rumus berikut (Imdadullah *et al.* 2017):

$$MSE_{LIU} = \sigma^2 \sum_{i=1}^p \frac{(\lambda_i + d)^2}{\lambda_i (\lambda_i + 1)^2} + (d - 1)^2 \sum_{i=1}^p \frac{\hat{\alpha}^2}{(\lambda_i + 1)^2} \quad (3)$$

Regresi Komponen Utama (RKU) adalah suatu teknik analisis statistik untuk mentransformasi peubah-peubah asli yang masih saling berkorelasi satu dengan yang lain menjadi satu kumpulan peubah baru yang tidak berkorelasi lagi. Secara umum tujuan dari RKU adalah mereduksi dimensi data. RKU muncul sebagai solusi bagi proses pengumpulan data dimana data tersebut terdiri dari variabel-variabel yang jumlahnya sangat banyak sehingga diperoleh variabel-variabel baru yang jumlahnya lebih sedikit tetapi tetap mampu menjelaskan keragaman data (Johnson dan Wichern, 2007).

Misalkan suatu data terdiri dari n observasi dan p variabel dapat dinyatakan dengan matriks X sebagai berikut:

$$X_{t \times p} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t1} & x_{t2} & \dots & x_{tp} \end{bmatrix} \quad (4)$$

Persamaan (2.11) dapat dinyatakan dalam bentuk vektor yaitu $X = [X_1, X_2, \dots, X_p]$ dengan $X'_t = [X_{t1}, X_{t2}, \dots, X_{tp}]$ dengan $i = 1, 2, \dots, t$, matriks kovarian Σ

dan nilai eigen $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ dengan vektor eigen yang bersesuaian $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$.

Komponen utama yang terbentuk dapat dituliskan sebagai kombinasi linear dari variabel-variabel asalnya yaitu sebagai berikut:

$$\begin{aligned} K_1 &= \mathbf{v}'_1 \mathbf{X} = v_{11}X_1 + v_{12}X_2 + \dots + v_{1p}X_p \\ K_2 &= \mathbf{v}'_2 \mathbf{X} = v_{21}X_1 + v_{22}X_2 + \dots + v_{2p}X_p \\ &\vdots \\ K_p &= \mathbf{v}'_p \mathbf{X} = v_{p1}X_1 + v_{p2}X_2 + \dots + v_{pp}X_p \end{aligned} \tag{5}$$

atau

$$\mathbf{K} = \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_p \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1p} \\ v_{21} & v_{22} & \dots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \dots & v_{pp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

dengan

- $var(K_i) = var(\mathbf{v}'_i X) = \mathbf{v}'_i var(X) \mathbf{v}_i$
sehingga
 $var(K_i) = \mathbf{v}'_i \Sigma \mathbf{v}_i ; i = 1, 2, \dots, n$ (6)
dimana $\Sigma = var(X) = \Sigma_X$

- $cov(K_i, K_k) = cov(\mathbf{v}'_i X, \mathbf{v}'_k X) = \mathbf{v}'_i cov(X, X) \mathbf{v}_k = \mathbf{v}'_i \Sigma \mathbf{v}_k$
sehingga
 $cov(K_i, K_k) = \mathbf{v}'_i \Sigma \mathbf{v}_k ; i, k = 1, 2, \dots, n$ (7)
dimana $cov(X, X) = var(X) = \Sigma_X$

Langkah-langkah metode RCU adalah sebagai berikut:

- Menghitung matriks varian kovarians Σ dengan menggunakan rumus pada persamaan

$$\Sigma_{p \times p} = cov(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \tag{8}$$

- Menghitung nilai eigen dari matriks varian kovarians Σ dengan menggunakan rumus sebagai berikut:

$$|\Sigma - \lambda \mathbf{I}| = 0 \tag{9}$$

- Menghitung vektor eigen dari matriks varian kovarians Σ dengan menggunakan rumus sebagai berikut:

$$(\Sigma - \lambda \mathbf{I})\mathbf{x} = 0 \tag{10}$$

- Vektor eigen dinormalisasi dengan menggunakan rumus sebagai berikut:

$$\mathbf{e} = \frac{\mathbf{x}}{\sqrt{\mathbf{x}'\mathbf{x}}} \tag{11}$$

- Membentuk komponen utama ke-1, 2, ..., k dengan menggunakan persamaan (5)
- Menghitung proporsi varians dengan menggunakan rumus sebagai berikut:

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (12)$$

- Menghitung proporsi kumulatif yang dapat dijelaskan oleh komponen utama ke- k dengan menggunakan rumus persamaan sebagai berikut

$$\frac{\sum_{i=1}^k \lambda_j}{\sum_{i=1}^p \lambda_j} \quad (13)$$

Dengan adanya keheterogenen ragam sisaan model regresi liu dan regresi komponen utama. Selanjutnya, penambahan peubah *dummy*. Kemudian Memodelkan *statistical downscaling* dengan penambahan peubah *dummy*, langkah akhir yaitu melakukan validasi model dengan tujuan untuk mengukur sejauh mana model yang diperoleh mendekati ukuran sebenarnya menggunakan data periode 2017 menggunakan RMSEP dan menghitung korelasi antara y (data curah hujan aktual) dengan \hat{y} (prediksi curah hujan) model regresi liu dan regresi komponen utama dengan *dummy*.

3. Hasil dan Diskusi

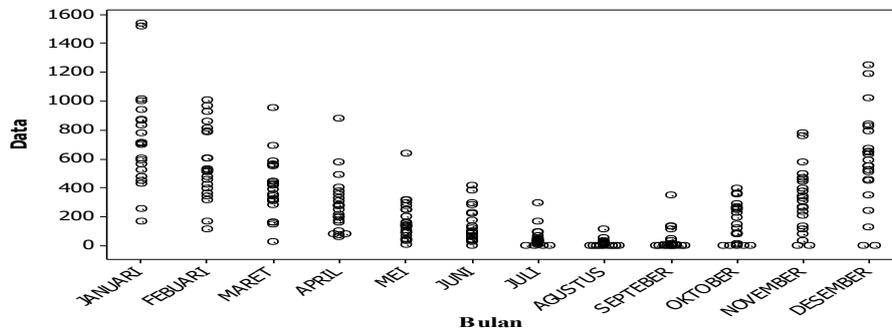
Deskripsi data curah hujan di Kabupaten Pangkep pada Tabel 1 menunjukkan deskripsi data curah hujan di Kabupaten Pangkep tahun 1996 – 2017 dengan rata-rata curah hujan tertinggi pada bulan Januari sebesar 739.32 mm/bulan. Sementara rata-rata curah hujan bulanan terendah terjadi pada bulan Agustus sebesar 16.25 mm/bulan. Curah hujan tertinggi terjadi pada bulan Januari dengan nilai maksimum 1540.50mm/bulan sedangkan curah hujan terendah 0 mm/bulan terjadi pada bulan Agustus, September, Oktober dan Desember. Pola lokal merupakan pola curah hujan yang membentuk U atau dengan kata lain memiliki satu puncak curah hujan. Pola lokal pada kabupaten Pangkep memiliki puncak tepatnya pada bulan Januari karena rata-rata curah hujannya paling tinggi dari bulan yang lain Simpangan baku terbesar berada pada bulan Januari, yaitu sebesar 338.86 mm/bulan. Simpangan baku terendah berada pada bulan Agustus, yaitu sebesar 26.71 mm/bulan yang menunjukkan bahwa curah hujan di kabupabeten Pangkep pada bulan Januari 1996-2017 sangat beragam.

Tabel 1. Deskripsi Curah Hujan Bulanan Tahun 1996 - 2017

Bulan	Rata-Rata (mm/bulan)	Simpangan Baku	Maksimum	Minumum
Januari	739.32	338.86	1540.50	166.50
Febuari	569.25	250.92	1006.50	116.00
Maret	404.36	203.14	958.00	27.50
April	280.73	192.11	884.00	62.00
Mei	177.64	139.81	641.50	6.00
Juni	138.25	118.76	419.00	4.00
Juli	51.75	67.38	296.50	0.00

Agustus	16.25	26.71	117.00	0.00
September	40.09	82.13	350.00	0.00
Oktober	156.11	136.26	398.00	0.00
November	322.34	220.19	779.50	0.00
Desember	586.84	331.03	1252.50	0.00

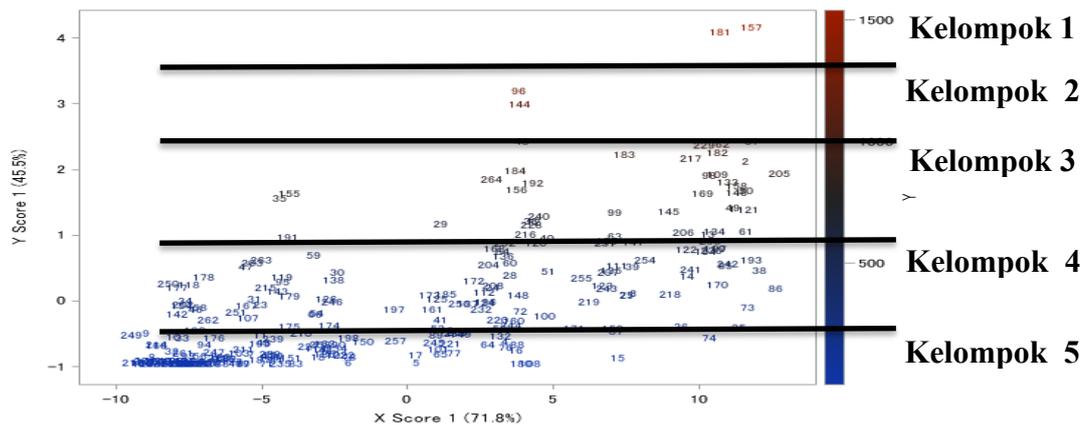
Sumber : Hasil olah data, 2019.



Gambar 1. Curah Hujan Bulanan Kabupaten Pangkep 1996-2017

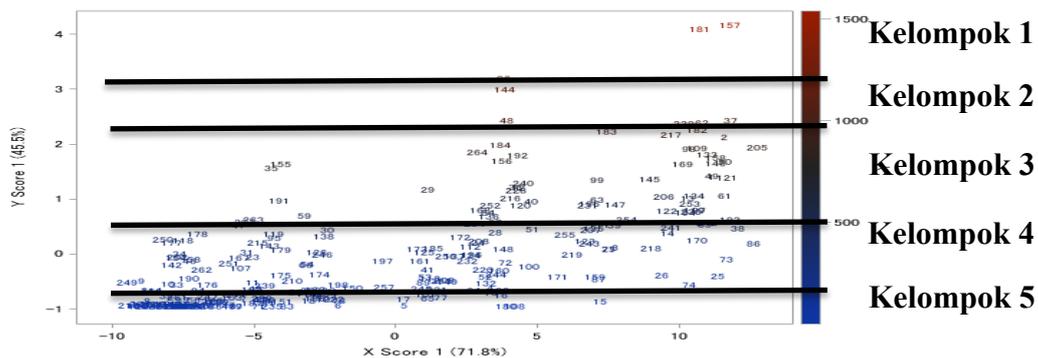
Untuk menguji ada tidaknya multikolinearitas, kita dapat menggunakan nilai VIF (Variance Inflation Factor). Jika nilai VIF melebihi 10 maka hal tersebut menunjukkan bahwa multikolinearitas adalah masalah yang pasti terjadi antar variabel bebas. Berdasarkan hasil perhitungan menunjukkan 20.57- 3399.57 , terlihat bahwa keseluruhan variabel yang memiliki nilai VIF lebih dari 10. Oleh karena itu, dapat dipastikan bahwa ada masalah multikolinearitas antar variabel bebas pada data.

Peubah *dummy* memiliki dua nilai yaitu 1 dan 0. Untuk *cluster* peubah *dummy* digunakan metode metode non-hirarki yaitu *K-means* dan metode hirarki yaitu *average linkage*. Pada metode non-hirarki, digunakan jarak *Euclidian*, untuk menetapkan nilai kedekatan antara objek. Hasil yang diperoleh *k-means* kelompok 1 yaitu curah hujan dengan intensitas 1519-1541 mm/bulan, kelompok 2 adalah curah hujan dengan intensitas 1190-1253 mm/bulan, kelompok 3 adalah curah hujan dengan intensitas 675-1019 mm/bulan, kelompok 4 adalah curah hujan dengan intensitas 344-652 mm/bulan dan kelompok 5 adalah curah hujan dengan intensitas 0-86 mm/bulan. Dengan demikian jumlah peubah *dummy* yang digunakan dalam pemodelan SD adalah 4 peubah yaitu D_1, D_2, D_3, D_4



Gambar 2. Plot Curah Hujan Kabupaten Pangkep Menggunakan *K-Means*

Pada metode *average linkage* teknik yang digunakan yang menggabungkan kluster-kluster menurut jarak rata-rata pasangan anggota masing-masing pada himpunan antara dua kluster. Hasilnya adalah data curah hujan dengan 5 *cluster* yaitu Kelompok 1 adalah curah hujan dengan intensitas 1190-1541 mm/bulan, kelompok 2 adalah curah hujan dengan intensitas 759-1019 mm/bulan, kelompok 3 adalah curah hujan dengan intensitas 443-715 mm/bulan, kelompok 4 adalah curah hujan dengan intensitas 195-430 mm/bulan dan kelompok 5 adalah curah hujan dengan intensitas 0-176 mm/bulan. Dengan demikian jumlah peubah *dummy* yang digunakan dalam pemodelan SD adalah 4 peubah yaitu D_1, D_2, D_3, D_4

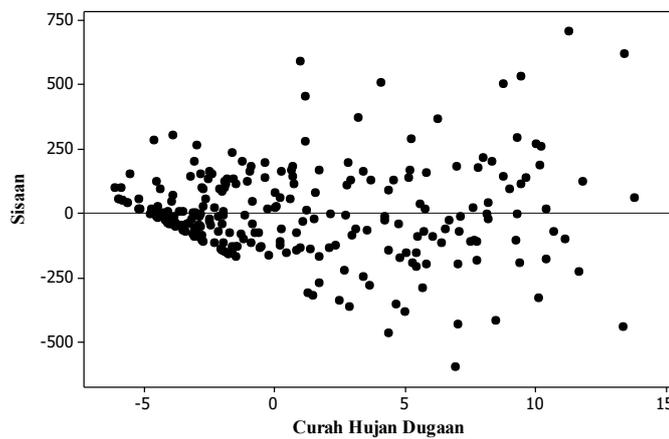


Gambar 3. Plot Curah Hujan Kabupaten Pangkep menggunakan Average Linkage

Untuk mengatasi masalah multikolinearitas dengan menggunakan metode estimasi parameter model regresi liu dilakukan yaitu dengan menentukan konstanta Regresi Estimasi *LIU* menggunakan persamaan (2) maka di dapatkan konstanta Regresi Estimasi *LIU* $d_L = 0.17$. Setelah menentukan konstanta parameter (d_L) maka dapat dihitung Regresi Estimasi *LIU*. Hasil yang diperoleh dari estimasi parameter model regresi liu menunjukkan bahwa kurang baik dalam menggambarkan keragaman data curah hujan berdasarkan nilai koefisien determinasi (R^2) berkisar 63.47% dan nilai RMSE yang

cukup besar, yakni berkisar 1.106. Salah satu asumsi yang harus dipenuhi dalam model regresi adalah sisaan memiliki ragam yang homogen (homoskedastisitas). Sisaan memiliki ragam yang heterogen apabila plot antara nilai dugaan dengan sisaannya membentuk pola tertentu. Tebaran sisaan dari estimasi parameter model regresi liu membentuk pola divergen yang semakin melebar seiring meningkatnya nilai dugaan curah hujan (Gambar 4). Oleh karena itu, peubah *dummy* digunakan untuk mengatasi ragam sisaan yang heterogen. Adapun persamaan model estimasi parameter model regresi liu yang terbentuk sebagai berikut:

$$\hat{y} = -75.55 + 89.47X_1 - 68.96X_2 + 73.52X_3 - 9.18X_4 + 98.00X_5 + 95.65X_6 - 72.29X_7 + 22.02X_8 + \dots - 42.35X_{57} + 95.21X_{58} - 51.39X_{59} + 56.25X_{60} + 105.92X_{61} - 141.99X_{62} + 33.18X_{63} - 24.90X_{64}$$



Gambar 4. Model Estimasi Parameter Model Regresi Liu

Model SD dengan estimasi parameter model regresi liu dengan penambahan peubah *dummy* mampu menghasilkan model yang bagus. Model yang ditentukan dengan teknik kluster *K-means* (Liu,D) yaitu nilai koefisien determinasi (R^2) berkisar 85.03% dan nilai *root mean square error* (RMSE) berkisar 77.83. Model regresi liu yang ditentukan dengan teknik kluster *avarange linkage* (Liu,D) yaitu nilai koefisien determinasi (R^2) berkisar 81.26% dan nilai *root mean square error* (RMSE) berkisar 104.42.

Secara umum tujuan dari RKU adalah mereduksi dimensi data atau mengatasi multikolinearitas dalam data. Tahap pertama yang dilakukan RKU yaitu menguji kehomogenan ragam menggunakan uji *levene*. Berdasarkan hasil uji kohomogenan ragam pada lampiran 6 diperoleh bahwa nilai $2.2e - 16 \leq 0.05$ hal ini berarti bahwa data presipitasi $X_1 - X_{64}$ dari GCM secara nyata memiliki ragam yang heterogen. Oleh karena itu, AKU dibentuk berdasarkan matriks korelasi untuk mereduksi dimensi data GCM.

Jumlah KU dalam pemodelan SD berdasarkan pada proporsi kumulatif keragaman total yang dapat dijelaskan oleh KU. Selain itu, KU yang digunakan adalah KU dengan nilai akar ciri lebih dari 1 dan proporsi kumulatif lebih dari 80%. Tabel 2 menunjukkan

bahwa jumlah KU yang memiliki akar ciri lebih besar dari satu adalah sebanyak empat KU. Proporsi kumulatif keragaman KU1 sampai dengan KU4 bernilai 96%. Hal ini menunjukkan bahwa KU tersebut mampu menjelaskan dengan proporsi keragaman total sebesar 96% dari peubah asal. Dengan demikian, analisis selanjutnya sebagai peubah respon yaitu curah hujan di Kabupaten Pangkep akan dimodelkan dengan empat peubah prediktor yaitu KU1, KU2, KU3 dan KU4 dimana peubah prediktor merupakan peubah data GCM yang telah direduksi.

Tabel 2. Nilai Proporsi Kumulatif Keragaman Regresi Komponen Utama Data Kabupaten Pangkep

Komponen Utama	KU_1	KU_2	KU_3	KU_4	KU_5
Nilai akar ciri	46.62	8.65	4.40	1.60	0.84
proporsi keragaman (%)	0.73	0.14	0.07	0.03	0.01
Proporsi keragaman kumulative (%)	0.73	0.86	0.93	0.96	0.98

Berdasarkan kriteria pemilihan jumlah KU yaitu nilai eigen dan persentase keragaman kumulatif, diperoleh empat KU sebagai variabel baru. Model RKU dibangun dengan meregresikan KU terpilih dengan data curah hujan. Hasil yang diperoleh menunjukkan bahwa model RKU kurang baik dalam menggambarkan keragaman data curah hujan berdasarkan nilai koefisien determinasi (R^2) berkisar 20.70%–52.20% dan nilai *root mean square error* (RMSE) berkisar 271.03–211.61.

Selain itu, pola tebaran sisaan model RKU yang melibatkan 4 KU pertama (RKU_1) menggambarkan pola divergen. Hal ini mengindikasikan bahwa ragam sisaan dari model RKU_4 heterogen. Pola tebaran sisaan yang sama juga ditunjukkan pada model $RKU_1, RKU_2,$ dan RKU_3 yang dapat dilihat pada lampiran 8. Adapun hasil model KU dari metode RKU yaitu sebagai berikut:

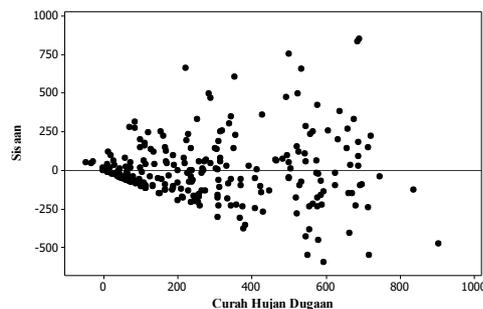
$$\hat{y} = -241 - 1.47 KU_1 - 1.03 KU_2 + 3.51 KU_3 + 9.20 KU_4$$

$$KU_1 = 0.135Z_1 + 0.142Z_2 + 0.144Z_3 + 0.144Z_4 + \dots + 0.124Z_{64}$$

$$KU_2 = 0.095Z_1 + 0.045Z_2 + 0.005Z_3 - 0.031Z_4 + \dots - 0.130Z_{64}$$

$$KU_3 = 0.097Z_1 + 0.079Z_2 + 0.067Z_3 + 0.021Z_4 + \dots - 0.155Z_{64}$$

$$KU_4 = 0.061Z_1 + 0.047Z_2 + 0.005Z_3 - 0.066Z_4 + \dots + 0.025Z_{64}$$



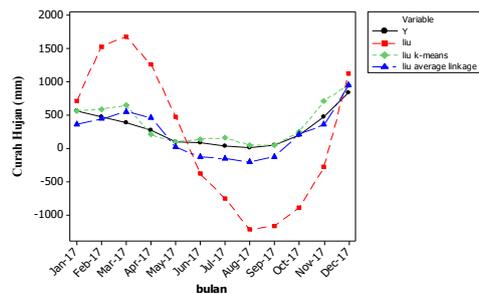
Gambar 5. Plot sisaan model RKU_4

Model SD dengan regresi komponen utama dengan penambahan peubah *dummy* mampu menghasilkan model yang bagus. Model rku yang ditentukan dengan teknik kluster *K-means* (RKUK-*Kmeans*) yaitu nilai koefisien determinasi (R^2) berkisar 95.40% – 95.80% dan nilai *root mean square error* (RMSE) berkisar 63.46 – 65.52. Model RKUK yang ditentukan dengan teknik kluster *average linkage* (RKU – *Average*) yaitu nilai koefisien determinasi (R^2) berkisar 93.90% – 94.20% dan nilai *root mean square error* (RMSE) berkisar 74.45 – 75.91.

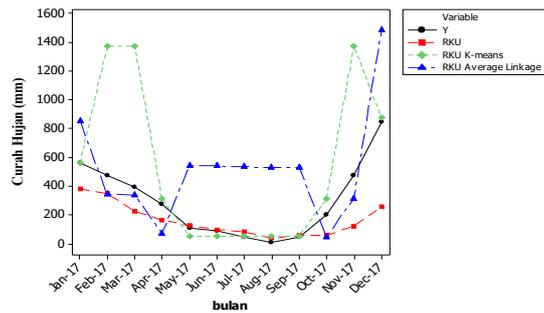
Data curah hujan periode Januari hingga Desember Tahun 2017 digunakan sebagai data validasi. Hasilnya menunjukkan bahwa metode estimasi parameter model regresi liu-*dummy* (*k-means*) mampu menjelaskan keragaman curah hujan yang akurat. metode estimasi parameter model regresi liu *k-means* memperlihatkan nilai rmsep sekitar 122.09 dan nilai korelasi 0.955. secara grafik estimasi parameter model regresi liu *k-means* menunjukkan performa yang lebih baik dari pada RKU tanpa *dummy*, RKU dengan *dummy*, estimasi parameter model regresi liu, dan estimasi parameter model regresi liu *average linkage*.

Gambar 6 memperlihatkan bahwa plot curah hujan dugaan model estimasi parameter model regresi liu dan plot curah hujan dugaan model estimasi parameter model regresi liu dengan peubah *dummy* yang ditentukan dari teknik cluster *k-means* dan *average linkage*. Hasil model estimasi parameter model regresi liu memperlihatkan bahwa model estimasi parameter regresi liu dapat memprediksi intensitas curah hujan lebih baik. Pada musim hujan pada bulan febuari, maret, april, dan desember, nilai-nilai prediksi berada lebih tinggi dari nilai-nilai aktual. Sedangkan pada musim kemarau pada bulan juni sampai November, nilai-nilai prediksi berada lebih rendah dari curah hujan actual.

Gambar 7 memperlihatkan bahwa plot curah hujan dugaan RKU plot curah hujan dugaan model RKU dengan peubah *dummy* yang ditentukan dari teknik cluster *k-means* dan *average linkage*. Hasil model RKU memperlihatkan bahwa model RKU-*dummy* (RKU *K-means*) dapat memprediksi intensitas curah hujan lebih baik. Pada musim hujan pada bulan febuari, maret, april, oktober dan November, nilai-nilai prediksi berada lebih tinggi dari nilai-nilai aktual. Sedangkan pada musim kemarau pada bulan mei, juni, juli, agustus, dan september, nilai-nilai prediksi berada lebih rendah dari curah hujan aktual.



Gambar 6. Plot Curah Hujan Aktual Semua model dengan Estimasi Parameter Model Regresi Liu Kota Pangkep



Gambar 7. Plot Curah Hujan Aktual Semua model dengan RKU Kota Pangkep

4. Kesimpulan

Penambahan variabel *dummy* pada estimasi parameter model regresi liu dan regresi komponen utama untuk model statistical downscaling digunakan untuk menangani masalah ragam sisaan yang heterogen. Adanya 4 peubah *dummy* yang disertakan ke dalam model merupakan hasil pengelompokan teknik *cluster* berdasarkan *k-means* yang digunakan jarak *Euclidian*, untuk menetapkan nilai kedekatan antara objek dan teknik kluster berdasarkan *average linkage* yang menggabungkan kluster-kluster menurut jarak rata-rata pasangan anggota masing-masing pada himpunan antara dua kluster. Model *statistical downscaling* dengan Estimasi Parameter Model Regesi Liu *k-means* mampu memperbaiki hasil dugaan data curah hujan. Model Estimasi Parameter Model Regesi Liu merupakan model terbaik berdasarkan nilai korelasi (0.955) yang tinggi dan RMSEP (122.09) yang dihasilkan.

Daftar Pustaka

- [1] Haidawati. *Evaluasi Program Pemberdayaan Usaha Garam Rakyat*. Makassar: Universitas Hasanuddin. 2014.
- [2] Effendi B. *Hybrid MARS-Time series pada Data Pemodelan Statistical Downscaling*. Surabaya: Institut Teknologi Sepuluh November. 2009.
- [3] Purnomandi H. *Pemodelan Statistical Downscaling dengan Pendekatan Regresi Kontinum-PCA (Studi Kasus : Stasiun Pontianak dan Ambon)*. Seminar Statistik Nasional IX. 2010.
- [4] Liu K. A new class of biased estimate in linear regression. *Communications in Statistics* 22(2): 393-402. 1993.
- [5] Imamdullah M. A., Altaf, M., & Liureg, S. A Comprehensive R Package for the Liu Estimation of Linear Regression Model with Collinear Regressors. *The R Journal*. Vol. 9/2. 2017
- [6] Johnson, R. A. & Wichern, D. W. *Applied Multivariate Statistical Analysis, Sixth Edition*. New Jersey : United States of America. 2007.