

PEMODELAN STATISTICAL DOWNSCALING DENGAN PEUBAH DUMMY BERDASARKAN TEKNIK CLUSTER HIERARKI DAN NON-HIERARKI UNTUK PENDUGAAN CURAH HUJAN*

Sitti Sahriman¹, Anisa Kalondeng^{2‡}, Vieri Koerniawan³

¹Departemen Statistika, Universitas Hasanuddin (Unhas), Indonesia, sittisahriman@yahoo.com

²Universitas Hasanuddin (Unhas), Indonesia, nkalondeng@yahoo.com

³Universitas Hasanuddin (Unhas), Indonesia, vieri.koerniawan@gmail.com

‡corresponding author

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

Vol 3 No 3 (2019), 295 - 309

Copyright © 2019 Sitti Sahriman, Anisa Kalondeng, and Vieri Koerniawan. 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 (SD) is a statistical technique used to predict local scale rainfall based on global atmospheric circulation. The global scale climate variable used is precipitation from GCM (Global Circulation Model). However, the precipitation data of GCM outputs have a large dimension, giving rise to multicollinearity in the data. This problem is handled by the Principal Component Regression (PCR) method. In addition, the SD models have heterogeneous error variances. The dummy variable is added to the PCR models to solve the problem. Hierarchical (k-means) and non-hierarchical cluster techniques (average linkage, median linkage, and ward linkage) are used in modeling to determine rainfall data groups. Furthermore, the group formed is the basis of the formation of dummy variables. This study aims to estimate local rainfall data in Pangkep district as a salt-producing area in South Sulawesi. There are 4 dummy variables based on the 5 groups formed. Dummy variables are able to improve predictions from the PCR models. R² values of the PCR-dummy models (ranging from 89.89% to 95.58%) are relatively higher than the PCR models (ranging from 55.87% to 57.61%). This result is also consistent with the model validation stage. The PCR-dummy models based on non-hierarchical cluster techniques (k-means) are better than the PCR-dummy models based on cluster hierarchy techniques. In general, the best model is the PCR-dummy model of the non-hierarchical cluster technique (k-means) and involves 4 main components.

Keywords: cluster hierarchy techniques, dummy variables, non-hierarchical cluster techniques, principal component regression, statistical downscaling.

* Received Jul 2019; Accepted Aug 2019; Published online on Oct 2019

1. Pendahuluan

Sulawesi Selatan (Sulsel) adalah salah satu pusat produksi garam nasional. Terdapat empat kabupaten di Sulsel yang menjadi sentra produksi garam, yaitu Takalar, Jeneponto, Selayar, dan Pangkep. Kepala Bidang Pengembangan dan Penataan Ruang Laut, Pesisir dan Pulau-Pulau Kecil, Nasir Malawi, menjelaskan bahwa pada 2016 produksi garam Sulsel mengalami penurunan tajam ke angka 13 ribu ton per tahun dengan produksi sebelumnya di Tahun 2015 mencapai 115 ribu ton. Lebih lanjut Nasir menjelaskan bahwa penurunan produksi yang signifikan ini sangat dipengaruhi oleh kondisi cuaca di Tahun 2016 (Listyarini, 2017). Pada Tahun 2015 musim kemarau cukup panjang sehingga produksi garam bisa optimal, sementara di Tahun 2016 musim hujan mulai datang sejak Bulan September sehingga praktis hanya sekitar dua bulan waktu produksi. Tingginya curah hujan di Tahun 2016 menyebabkan penurunan produksi garam nasional di tahun tersebut.

Variabilitas curah hujan berdampak pada produksi garam nasional yang secara tidak langsung mempengaruhi kebutuhan garam baik untuk konsumsi maupun industri. Curah hujan adalah jumlah air hujan yang turun pada daerah tertentu dalam waktu tertentu. Curah hujan merupakan unsur iklim yang paling dominan dalam mencirikan kondisi iklim di Indonesia karena curah hujan di Indonesia memiliki keragaman dan fluktuasi yang tinggi dibandingkan suhu (Sipayung, 2005). Pendugaan model iklim di Indonesia masih terus dikembangkan dengan memanfaatkan informasi iklim global yang salah satunya dapat diperoleh dari *Global Climate Model*.

Global circulation model (GCM) merupakan representasi numerik dari gambaran perilaku sistem iklim dan interaksi antar komponen-komponennya, yakni atmosfer, lautan, kriosfer, biosfer, dan kemosfer. GCM adalah model iklim yang dapat dimanfaatkan sebagai alat untuk memahami iklim melalui simulasi dan membuat proyeksi perubahan iklim masa mendatang berdasarkan skenario perubahan emisi. GCM mensimulasi peubah-peubah iklim global pada setiap grid (berukuran $\pm 2.5^\circ$ atau $\pm 300 \text{ km}^2$) untuk setiap lapisan atmosfer yang selanjutnya digunakan untuk memprediksi pola-pola iklim dalam jangka waktu tahunan (Wigena, 2006). Model GCM telah menjadi pendekatan yang paling umum untuk pengembangan skenario perubahan iklim. Namun, informasi iklim luaran GCM masih berskala global. *Statistical downscaling* merupakan salah satu teknik yang dapat digunakan untuk mendapatkan informasi iklim dalam skala lokal dengan resolusi tinggi.

Statistical downscaling (SD) adalah model statistika yang menghubungkan peubah iklim luaran GCM yang berskala global dengan peubah iklim yang berskala lokal (Fernandez, 2005). Untuk dapat menjelaskan keragaman iklim dengan baik, model SD mensyaratkan korelasi yang kuat antara peubah iklim global dengan peubah iklim lokal (Busuioc et al., 2001). Oleh karena itu, fungsi korelasi silang (*cross correlation function*) dapat digunakan untuk meningkatkan korelasi antara data curah hujan berskala lokal dengan data presipitasi luaran GCM berskala global (Wigena et al., 2015).

Presipitasi luaran GCM memiliki dimensi besar dan korelasi yang tinggi (multikolinieritas) antar grid. Multikolinieritas menyebabkan standar error penduga parameter regresi menjadi besar dan selang kepercayaan menjadi lebar sehingga penggunaan metode kuadrat terkecil (MKT) menjadi tidak valid (Montgomery & Peck, 1992). Regresi komponen utama (RKU) merupakan salah satu metode yang paling sering digunakan dalam pemodelan SD. RKU diawali dengan analisis komponen utama (AKU) untuk mereduksi dimensi data dan mengatasi masalah multikolinieritas.

Kajian ilmu tentang model curah hujan di Indonesia telah dilakukan oleh beberapa peneliti sebelumnya. Diantaranya adalah Estiningtyas & Wigena (2011) menggunakan regresi komponen utama (RKU) dan regresi kuadrat terkecil parsial (RKTP) untuk memprediksi curah hujan pada kondisi el-nino, la-nina, dan normal di Kabupaten Indramayu. Selanjutnya, Sahrinan (2014) memanfaatkan pergeseran waktu berdasarkan korelasi silang antara data presipitasi luaran GCM dengan data curah hujan kabupaten Indramayu dalam SD menggunakan RKU dan RKTP. Selain itu, Sahrinan & Kalondeng (2017) memperbaiki penelitian sebelumnya dengan menggunakan teknik cluster non-hierarki dalam menentukan kelompok atau peubah *dummy* data curah hujan. Hal ini dilakukan untuk mengatasi ragam sisaan model yang heterogen.

Analisis cluster termasuk teknik multivariat yang bertujuan untuk mengelompokkan objek-objek berdasarkan kesamaan karakteristik di antara objek-objek tersebut. Prinsip dari teknik *cluster* adalah homogenitas (kesamaan) yang tinggi antar kelompok dalam satu *cluster* dan heterogenitas (perbedaan) yang tinggi antar *cluster* yang satu dengan *cluster* yang lainnya. Secara umum terdapat dua teknik *cluster*, yakni analisis *cluster* hierarki dan non-hierarki (Mattjik & Sumertajaya, 2011).

Penelitian ini menggunakan model SD dengan menambahkan peubah *dummy* ke dalam model untuk mengatasi masalah keheterogenan ragam sisaan model. Peubah *dummy* ditentukan berdasarkan teknik *cluster* hierarki dan non-hierarki. Teknik *cluster* hierarki yang digunakan adalah *Average Linkage*, *Median Linkage*, dan *Ward Linkage*.

2. Metodologi

2.1 Data

Data observasi dalam penelitian ini adalah data curah hujan Kabupaten Pangkep periode Januari 1996 sampai dengan Desember 2017 sebagai peubah respon. Data peubah prediktor yang digunakan adalah presipitasi luaran GCM. Presipitasi luaran GCM merupakan data simulasi *climate model intercomparison project* (CMIP5) dalam satuan mm/bulan. 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°×2.5° untuk setiap grid) pada 119.57°BT s.d 129.37°BT dan -14.83°LS s.d 5.17°LU di atas sekitar wilayah Kabupaten Pangkep. Dengan demikian terdapat 64 peubah prediktor dengan jumlah observasi sebanyak 264 pengamatan yang digunakan dalam penelitian ini.

2.2 Metode Penelitian

Langkah awal yang dilakukan dalam penelitian ini adalah identifikasi multikolinearitas pada data presipitasi menggunakan *variance inflation factors* (VIF) sebagai berikut,

$$VIF_j = \frac{1}{1-R_j^2}$$

dengan R_j^2 adalah koefisien determinasi hasil regresi peubah prediktor ke- j dengan peubah prediktor lainnya.

Penentuan kelompok data curah hujan menggunakan teknik *cluster* hierarki dan non-hierarki. Hasil pengelompokan data curah hujan merupakan dasar penentuan peubah *dummy*. Metode *cluster* hierarki (Johnson & Wichern, 2007) yang digunakan dalam penelitian ini adalah sebagai berikut:

a. Average Linkage

Jarak antara dua *cluster* diukur dengan jarak rata-rata antara sebuah objek dalam *cluster* yang satu dengan sebuah objek dalam *cluster* yang lain.

$$d_{(UV)W} = \sum_i \sum_d d_{ik} / N_{(UV)} N_W$$

dengan d_{ik} jarak antara obyek i pada *cluster* (UV) dan obyek k pada *cluster* W , serta N_{uv} dan N_w berturut-turut adalah banyaknya item-item dalam klaster (UV) dan W (Johnson & Wichern, 2007).

b. Median Linkage

Jarak antar *cluster* didefinisikan sebagai jarak antar median, dan *cluster-cluster* dengan jarak terkecil akan digabungkan sebagai satu *cluster*.

c. Metode Ward

Jarak antar dua *cluster* adalah total jumlah kuadrat dua *cluster* pada masing-masing peubah. Metode ini berbeda dengan metode lainnya karena menggunakan pendekatan analisis variansi untuk menghitung jarak antar *cluster*. Fungsi objektif yang digunakan adalah *Error Sum of Square* (ESS) dengan rumus sebagai berikut:

$$ESS = \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2$$

dengan x_i adalah nilai pada objek ke- i , n adalah banyaknya objek dalam *cluster* terbentuk.

Selanjutnya, tahapan analisis teknik *cluster* non-hierarki dengan *K-Means* adalah sebagai berikut (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*.

Metode yang digunakan dalam model SD adalah RKU. Analisis RKU diawali dengan AKU untuk mereduksi dimensi data dengan mentransformasi peubah-peubah asal yang saling berkorelasi menjadi sekumpulan peubah baru (komponen utama, KU) yang tidak berkorelasi dengan tetap mempertahankan sebanyak mungkin keragaman yang mampu dijelaskan (Johnson & Wichern, 2007).

Jika $X' = [x_1, x_2, \dots, x_p]$ mempunyai matriks ragam-peragam Σ dengan nilai akar ciri $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, maka diperoleh w_j yang merupakan kombinasi linier peubah asal (Johnson & Wichern, 2007), yakni:

$$w_j = e_j' X = e_{j1}x_1 + e_{j2}x_2 + \dots + e_{jp}x_p \quad ; j = 1, 2, \dots, p$$

KU (w_1, w_2, \dots, w_p) merupakan kombinasi linear dari peubah asal X yang tidak berkorelasi dan mempunyai ragam maksimum. KU mempunyai ragam yang sama dengan nilai akar ciri dari matriks Σ , yakni

$$\begin{aligned} Var(w_j) &= e_j' \Sigma e_j = \lambda_j & j &= 1, 2, \dots, p \\ Cov(w_j, w_{j'}) &= e_j' \Sigma e_{j'} = 0 & j \neq j' &= 1, 2, \dots, p \end{aligned}$$

Selain menggunakan matriks ragam-peragam, KU juga dapat dibentuk menggunakan matriks korelasi. Penurunan KU menggunakan matriks korelasi terlebih dahulu dilakukan dengan mentransformasi peubah asal X menjadi bentuk baku Z dengan formula sebagai berikut:

$$Z_j = (X_j - \mu_j) / \sqrt{\sigma_j^2}$$

dengan notasi matriksnya dapat dituliskan sebagai

$$Z = (V^{1/2})^{-1} (X - \mu)$$

dengan $V^{1/2} = diag(\sqrt{\sigma_1^2}, \sqrt{\sigma_2^2}, \dots, \sqrt{\sigma_p^2})$, $E(Z) = \mathbf{0}$, Z adalah matriks peubah asal X yang telah dibakukan. Keragaman Z dapat dituliskan sebagai berikut:

$$Cov(Z) = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1} = R$$

dengan R adalah matriks korelasi peubah asal X . KU ke- j , yang dibentuk berdasarkan peubah-peubah yang dibakukan $Z' = [z_1, z_2, \dots, z_p]$, dapat ditentukan dari vektor ciri yang diperoleh melalui matriks korelasi peubah asal X dengan formula KU:

$$w_j = e_j' Z = e_{j1}z_1 + e_{j2}z_2 + \dots + e_{jp}z_p$$

Misalkan matriks P adalah matriks ortogonal berisi vektor ciri dari Σ peubah asal X yang memenuhi persamaan $P'P = PP' = I$. Proses pembentukan RKU dari regresi linier berganda dengan $W = XP$ dan $\alpha = P'\beta$ adalah (Jolliffe, 2002):

$$y = W\alpha + \varepsilon$$

Model RKU hasil reduksi menjadi r komponen adalah sebagai berikut:

$$y = \alpha_0 \mathbf{1} + W_r \alpha_r + \varepsilon$$

dengan $\varepsilon \sim N(0, \sigma^2 I)$ merupakan vektor galat berukuran $n \times 1$, X merupakan matriks peubah prediktor berukuran $n \times (p + 1)$, y merupakan vektor peubah respon berukuran $n \times 1$, α_0 adalah intersep, $\mathbf{1}$ adalah vektor bernilai satu berukuran $n \times 1$, W_r adalah matriks KU berukuran $n \times r$, dan α_r adalah vektor berisi koefisien KU berukuran $r \times 1$.

Pemilihan model terbaik menggunakan nilai RMSEP dan korelasi antara nilai aktual dan nilai dugaan tahun 2017. Performa model yang baik ditunjukkan dengan nilai RMSEP (*Root Mean Square Error of Prediction*) yang lebih kecil dan nilai korelasi yang lebih tinggi. Nilai RMSEP adalah sebagai berikut:

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

3. Hasil dan Pembahasan

3.1 Identifikasi Multikolinieritas

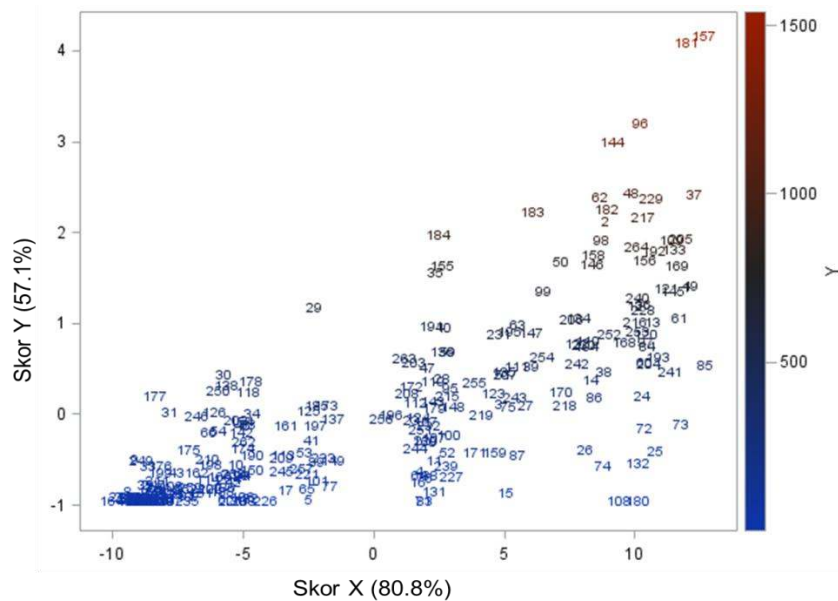
VIF merupakan salah satu statistik yang dapat digunakan untuk mendeteksi gejala multikolinieritas pada analisis regresi berganda. Jika nilai VIF lebih dari 10, menandakan adanya multikolinieritas yang signifikan. Hasil perhitungan nilai VIF setiap peubah prediktor menunjukkan data presipitasi GCM mengandung multikolinieritas yang perlu diperhatikan. Nilai VIF yang diperoleh berkisar antara 5.18–2949.55. Hal ini menunjukkan adanya hubungan yang kuat antar grid GCM data presipitasi. Oleh karena itu, metode yang digunakan dalam pemodelan SD untuk mengatasi masalah multikolinieritas pada data presipitasi GCM adalah metode RKU.

3.2 Penentuan Peubah *Dummy* dengan Non-Hierarki dan Hierarki

K-means adalah salah satu metode pengelompokan dengan pendekatan partisi yang mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam kelompok-kelompok yang ada. Gambar 1 adalah plot antara skor komponen data curah hujan (Y) dengan skor komponen data presipitasi (X) yang merupakan hasil analisis *partial least square*. Satu komponen terekstrak dapat menjelaskan sebesar 80.8% keragaman data presipitasi dan sebesar 57.1% keragaman data curah hujan. Gambar 1 menunjukkan 5 kelompok data curah hujan berdasarkan kelompok warna. Selanjutnya, data curah hujan dikelompokkan ke dalam 5 kelompok dengan teknik *cluster* hierarki dan non-hierarki.

Hasil pengelompokan dengan metode *k-means* diperoleh kelompok 1 umumnya terjadi pada bulan Mei hingga September dengan intensitas 0–176 mm/bulan terdiri atas 129 observasi. Kelompok 2 adalah kelompok curah hujan dengan intensitas 177–430 mm/bulan yang umumnya terjadi pada bulan Maret, April, dan Oktober (65 observasi). Kelompok 3 adalah kelompok curah hujan dengan intensitas 431–715 mm/bulan dan umumnya terjadi pada bulan Desember dan Januari (44 observasi). Kelompok 4 adalah kelompok curah hujan dengan intensitas 716–1019 mm/bulan dan umumnya juga terjadi pada bulan Januari dan Februari (22 observasi). Kelompok 5 adalah kelompok curah hujan dengan intensitas 1020–1541 mm/bulan dan umumnya juga terjadi pada bulan Januari dan Desember (4 observasi).

Metode hirarki yang biasa digunakan adalah metode pemusatan. Metode pemusatan yang digunakan dalam penelitian ini adalah *average linkage*, *median linkage*, dan metode Ward. Hasil pengelompokan data curah hujan dengan metode *average linkage* adalah kelompok 1 terjadi pada bulan Juli hingga September dengan intensitas 0–176 mm/bulan (129 observasi). Kelompok 2 adalah kelompok curah hujan dengan intensitas 177–652 mm/bulan yang umumnya terjadi pada bulan Februari, Maret, April, November, dan Desember (103 observasi). Kelompok 3 adalah kelompok curah hujan dengan intensitas 653–1019 mm/bulan dan umumnya terjadi pada bulan Januari (28 observasi). Kelompok 4 adalah kelompok curah hujan dengan intensitas 1020–1253 mm/bulan dan umumnya terjadi pada bulan Desember (2 observasi). Kelompok 5 adalah kelompok curah hujan dengan intensitas 1254–1541 mm/bulan dan umumnya juga terjadi pada bulan Januari (2 observasi).



Gambar 1: Kelompok Data Curah Hujan.

Metode *median linkage* berdasarkan pada jarak median antar *cluster*. Kelompok 1 umumnya terjadi pada bulan April hingga Oktober dengan intensitas 0–364 mm/bulan (180 observasi). Kelompok 2 adalah kelompok curah hujan dengan intensitas 365–607 mm/bulan yang umumnya terjadi pada bulan Januari, Februari, Maret, November, dan Desember (48 observasi). Kelompok 3 adalah kelompok curah hujan dengan intensitas 608–1019 mm/bulan dan umumnya terjadi pada bulan Januari dan Desember (32 observasi). Kelompok 4 adalah kelompok curah hujan dengan intensitas 1020–1253 mm/bulan dan umumnya terjadi pada bulan Desember (2 observasi). Kelompok 5 adalah kelompok curah hujan dengan intensitas 1254–1541 mm/bulan dan umumnya juga terjadi pada bulan Januari (2 observasi).

Metode Ward merupakan metode analisis *cluster* yang didasarkan adanya informasi yang hilang akibat penggabungan objek menjadi *cluster*. Kelompok 1 umumnya terjadi pada bulan Mei hingga September dengan intensitas 0–176 mm/bulan (129 observasi). Kelompok 2 adalah kelompok curah hujan dengan intensitas 177–497 mm/bulan yang umumnya terjadi pada bulan Maret, April, Oktober, dan November (78 observasi). Kelompok 3 adalah kelompok curah hujan dengan intensitas 498–652 mm/bulan dan umumnya terjadi pada bulan Desember (25 observasi). Kelompok 4 adalah kelompok curah hujan dengan intensitas 653–1253 mm/bulan dan umumnya terjadi pada bulan Januari, Februari, dan Desember (30 observasi). Kelompok 5 adalah kelompok curah hujan dengan intensitas 1254–1541 mm/bulan dan umumnya juga terjadi pada bulan Januari (2 observasi). Hal ini menunjukkan bahwa terdapat perbedaan kelompok data curah hujan pada setiap metode *cluster* hierarki maupun non-hierarki. Dengan demikian ada 4 peubah *dummy* yang digunakan dalam pemodelan SD. Setiap data curah hujan dalam kelompok sama diberikan nilai 1 pada peubah *dummy* dan 0 untuk kelompok yang lain. Sebagai contoh, penentuan peubah *dummy* berdasarkan teknik *cluster* non-hierarki *K-Means* disajikan dalam Tabel 1.

Tabel 1: Kelompok dan Peubah *Dummy* dengan *K-Means*.

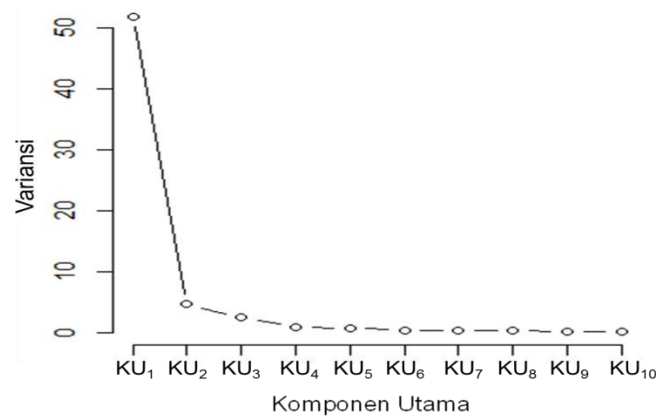
Waktu	Kelompok	D ₁	D ₂	D ₃	D ₄
Jan-96	1	1	0	0	0
Feb-96	2	0	1	0	0
Mar-96	4	0	0	0	1
Apr-96	5	0	0	0	0
May-96	5	0	0	0	0
Jun-96	5	0	0	0	0
Jul-96	5	0	0	0	0
Aug-96	5	0	0	0	0
Sep-96	5	0	0	0	0
Oct-96	5	0	0	0	0
Nov-96	5	0	0	0	0
Dec-96	1	1	0	0	0
⋮	⋮	⋮	⋮	⋮	
Sep-17	5	0	0	0	0
Oct-17	4	0	0	0	1
Nov-17	1	0	0	0	0
Dec-17	2	1	1	0	0

3.3 Pemodelan *Statistical Downscaling* dengan Regresi Komponen Utama

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. Tabel 2 menunjukkan bahwa KU₁ memiliki keragaman terbesar, yakni 81.1%. Dari Gambar 2 terlihat bahwa KU₁ memiliki varians (nilai akar ciri) paling besar, yaitu 51.925. KU₁, KU₂, KU₃, dan KU₄ memiliki nilai akar ciri lebih dari 1. Proporsi keragaman total dari KU₁, KU₂, KU₃, dan KU₄ adalah sebesar 94% (Tabel 2). Hal ini berarti bahwa 4 KU pertama (KU₁-KU₄) dapat menjelaskan keragaman data presipitasi sebesar 94%. Simulasi model R_{KU} berdasarkan pada jumlah KU yang digunakan dalam model.

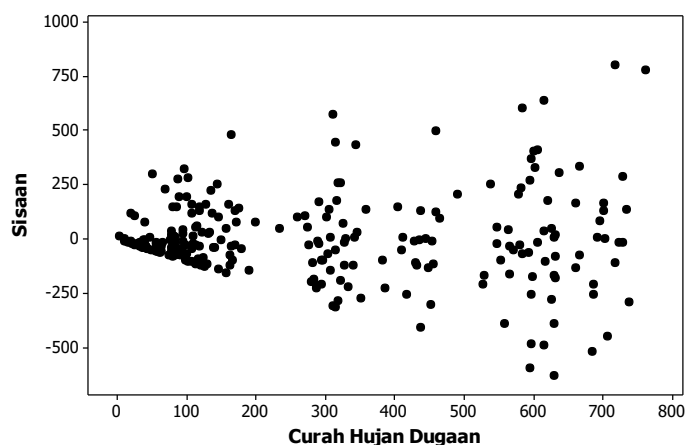
Tabel 2: Nilai Akar Ciri dan Proporsi Keragaman 5 KU.

	KU ₁	KU ₂	KU ₃	KU ₄	KU ₅
Nilai akar ciri	51.925	4.685	2.563	1.009	0.770
Proporsi keragaman	0.811	0.073	0.004	0.016	0.012
Proporsi kumulatif	0.811	0.885	0.925	0.940	0.952



Gambar 2: Plot KU dari Data Presipitasi.

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 55.87%–57.61% dan nilai *root mean square error* (RMSE) berkisar 229.156–233.817 (Tabel 3). Selain itu, pola tebaran sisaan model RKU yang melibatkan 4 KU pertama (RKU4) menggambarkan pola divergen. Hal ini mengindikasikan bahwa ragam sisaan dari model RKU4 heterogen (Gambar 3). Pola tebaran sisaan yang sama juga ditunjukkan pada model RKU1, RKU2, dan RKU3 .



Gambar 3: Plot sisaan model RKU4.

3.4 Pemodelan *Statistical Downscaling* dengan RKU-Dummy

Keheterogenan ragam sisaan model RKU dapat ditangani salah satunya dengan penambahan peubah *dummy* sebagai peubah prediktor. Penambahan peubah *dummy* mampu memperbaiki model RKU dengan nilai R^2 yang lebih tinggi dan RMSE yang lebih rendah (Tabel 3). Hal ini berarti bahwa model RKU dengan peubah *dummy* (RKUK) lebih baik dalam menjelaskan keragaman data. Model RKUK yang ditentukan dari teknik *cluster* hierarki *average linkage* (RKUK-Average) memiliki nilai R^2 berkisar 92.31%-92.42% dengan nilai RMSE berkisar 97.959-98.675. Model RKUK yang

ditentukan dari teknik *cluster* hierarki *median linkage* (RKUK-Median) memiliki nilai R^2 berkisar 89.86%-90.32% dengan nilai RMSE berkisar 110.721-113.283. Lebih lanjut, dalam teknik *cluster* hierarki model RKUK berdasarkan pada metode *ward linkage* (RKUK-Ward) merupakan model terbaik dengan nilai R^2 berkisar 94.23%-94.27% dan nilai RMSE berkisar 85.178-85.483. Namun secara keseluruhan, model RKUK dengan *dummy* berdasarkan pada teknik *cluster* non-hierarki *k-means* (RKUK-Kmeans) lebih baik dari model RKUK berdasarkan teknik *cluster* berhierarki. Model RKUK-Kmeans lebih baik dalam menjelaskan keragaman data curah hujan dengan R^2 berkisar 95.49%-95.58% dan nilai RMSE berkisar 74.786-75.574.

Diagnostik dilakukan terhadap sisaan model RKUK. Gambar 4 memperlihatkan 5 kelompok sisaan pada model RKU-Kmeans. Pola tebaran sisaan pada model RKUK-Kmeans relatif lebih homogen daripada pola tebaran sisaan model lainnya (RKUK-Average, RKUK-Median, dan RKUK-Ward). Penambahan peubah *dummy* berdasarkan teknik *cluster* K-Means ke dalam model RKU mampu menghasilkan model dengan pola tebaran sisaan yang lebih homogen.

Tabel 3: Nilai R^2 dan RMSE model RKU dan RKUK.

Metode	Teknik Cluster	Komponen	R^2	RMSE
RKU tanpa Dummy	-	KU1	55.87%	233.817
	-	KU1, KU2	57.59%	229.202
	-	KU1, KU2, KU3	57.60%	229.189
	-	KU1, KU2, KU3, KU4	57.61%	229.156
RKU dengan Dummy	K-Means (non-hierarki)	KU1, D1, D2, D3, D4	95.49%	75.574
		KU1, KU2, D1, D2, D3, D4	95.49%	75.536
		KU1, KU2, KU3, D1, D2, D3, D4	95.51%	75.396
		KU1, KU2, KU3, KU4, D1, D2, D3, D4	95.58%	74.786
	Average Linkage (Berhierarki)	KU1, D1, D2, D3, D4	92.31%	98.675
		KU1, KU2, D1, D2, D3, D4	92.40%	98.101
		KU1, KU2, KU3, D1, D2, D3, D4	92.42%	97.965
		KU1, KU2, KU3, KU4, D1, D2, D3, D4	92.42%	97.959
	Median Linkage (Berhierarki)	KU1, D1, D2, D3, D4	89.86%	113.283
		KU1, KU2, D1, D2, D3, D4	90.29%	110.879
		KU1, KU2, KU3, D1, D2, D3, D4	90.31%	110.725
		KU1, KU2, KU3, KU4, D1, D2, D3, D4	90.32%	110.721
	Ward Linkage (Berhierarki)	KU1, D1, D2, D3, D4	94.23%	85.483
		KU1, KU2, D1, D2, D3, D4	94.25%	85.315
		KU1, KU2, KU3, D1, D2, D3, D4	94.27%	85.199
		KU1, KU2, KU3, KU4, D1, D2, D3, D4	94.27%	85.178

Hasil model dugaan terbaik dari setiap metode dalam teknik *cluster* hierarki dan non-hierarki dapat dituliskan sebagai berikut:

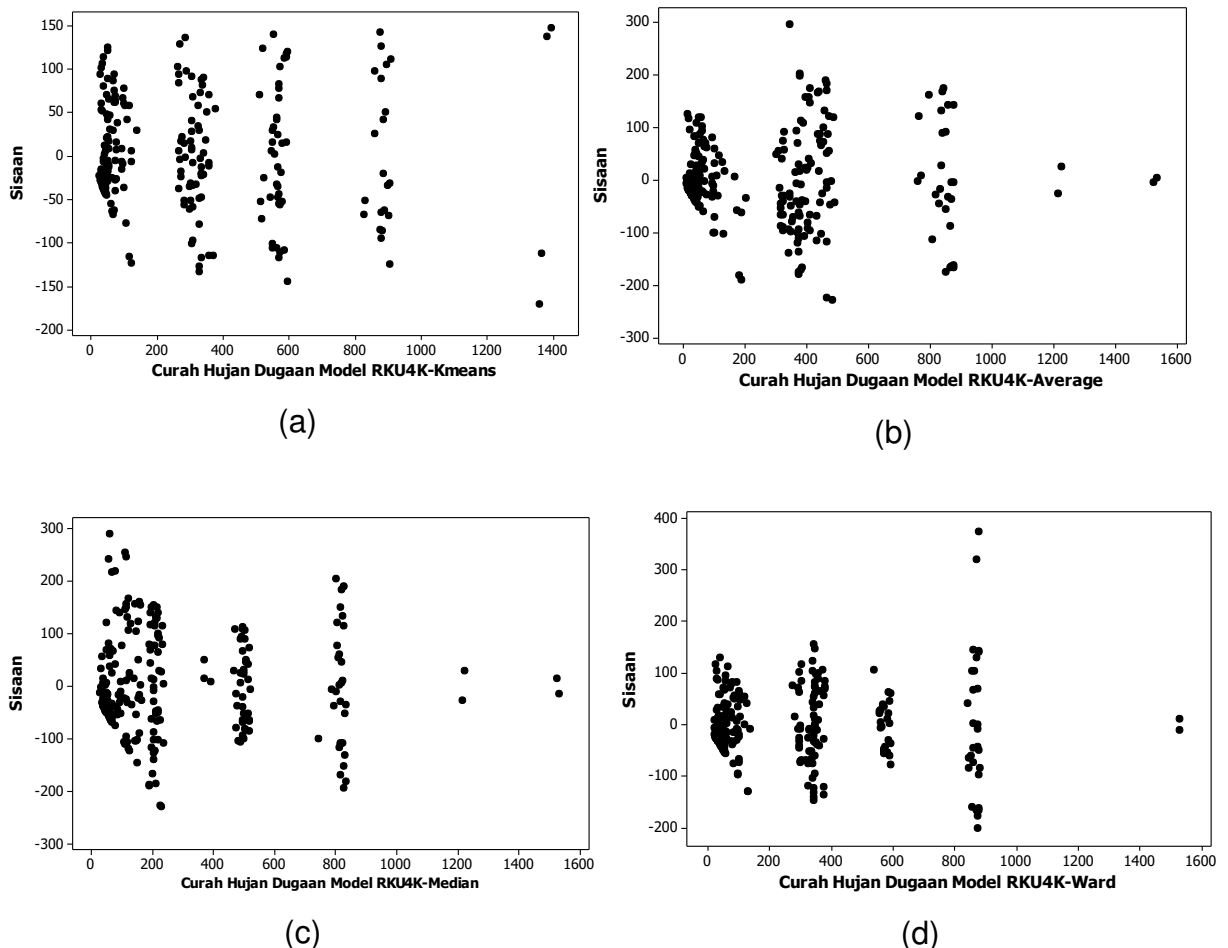
- Teknik *cluster* non-hierarki *K-Means*: Model RKU4K-*Kmeans*

$$\hat{y} = 77.26 + 4.91 \text{ KU1} + 1.09 \text{ KU2} + 2.48 \text{ KU3} + 8.12 \text{ KU4} + 446.76 \text{ D}_1 + 759.32 \text{ D}_2 + 1241.12 \text{ D}_3 + 232.47 \text{ D}_4$$
- Teknik *cluster* berhierarki *Average Linkage*: Model RKU4K-*Average*

$$\hat{y} = 1408.25 + 8.56 \text{ KU1} + 4.56 \text{ KU2} + 2.75 \text{ KU3} + 0.90 \text{ KU4} - 651.87 \text{ D}_1 - 1038.23 \text{ D}_2 - 1313.82 \text{ D}_3 - 273.12 \text{ D}_4$$
- Teknik *cluster* berhierarki *Median Linkage*: Model RKU4K-*Median*

$$\hat{y} = 1449.81 + 9.62 \text{ KU1} - 9.78 \text{ KU2} - 3.16 \text{ KU3} - 0.79 \text{ KU4} - 705.48 \text{ D}_1 - 1299.72 \text{ D}_2 - 1020.33 \text{ D}_3 - 313.71 \text{ D}_4$$
- Teknik *cluster* berhierarki *Ward Linkage*: Model RKU4K-*Ward*

$$\hat{y} = 1477.20 + 5.08 \text{ KU1} - 2.31 \text{ KU2} + 2.35 \text{ KU3} - 1.66 \text{ KU4} - 653.19 \text{ D}_1 - 1154.09 \text{ D}_2 - 1397.93 \text{ D}_3 - 942.07 \text{ D}_4$$



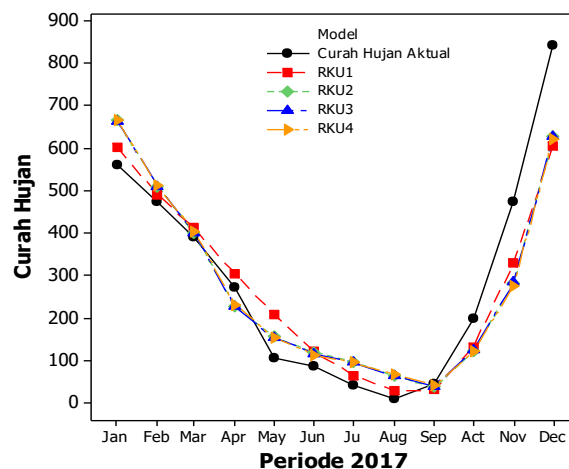
Gambar 4: Plot sisaan model (a) RKU4K-*Kmeans*, (b) RKU4K-*Average*, (c) RKU4K-*Median*, (d) RKU4K-*Ward*.

3.5 Validasi Model *Statistical Downscaling*

Berdasarkan hasil pemodelan, model *RKUK-Kmeans* adalah model terbaik dibandingkan dengan model *RKU* dengan *dummy* dari teknik *cluster* hierarki. Hal ini juga konsisten dalam hasil proses validasi model menggunakan data periode 2017. Tabel 4 menjelaskan bahwa penambahan peubah *dummy* dalam model *RKU* (berdasarkan teknik *cluster* berhierarki dan non-hierarki) dapat menurunkan nilai *RMSEP* sekitar 31.058. Model *RKU* menghasilkan dugaan curah hujan dengan nilai korelasi berkisar 0.923-0.941 dengan nilai *RMSEP* yang relatif tinggi yakni berkisar 90.339-98.499.

Tabel 4 juga menyajikan hasil validasi model *SD* dengan peubah *dummy* berdasarkan teknik *cluster* non-hierarki dan berhierarki. Nilai korelasi dan *RMSEP* yang dihasilkan relatif sama pada setiap model tersebut. Model *RKUK-Average* memiliki nilai korelasi berkisar 0.964-0.969 dan nilai *RMSEP* berkisar 62.171-66.726. Model *RKUK-Ward* juga memiliki nilai korelasi (berkisar 0.964-0.966) dan *RMSEP* (berkisar 66.260-67.441). Selain itu, model *RKUK-Median* memberikan hasil dugaan curah hujan akurat daripada model *RKUK-Average* dan *RKUK-Ward* berdasarkan nilai nilai korelasi dan *RMSEP*. Namun secara umum, model *RKUK-Kmeans* merupakan model terbaik berdasarkan nilai korelasi berkisar 0.987-0.989 dan *RMSEP* yang lebih rendah yakni berkisar 44.200-48.036.

Gambar 5 menjelaskan bahwa model *RKU* tidak dapat melakukan pendugaan curah hujan yang dekat dengan curah hujan aktual. Jarak antara curah hujan hasil dugaan model *RKU* dengan curah hujan aktual relatif jauh. Model *RKU* dapat menduga curah hujan aktual dengan tepat pada bulan Februari, Maret, dan September. Namun, model *RKU* menghasilkan dugaan curah hujan yang melebihi nilai sebenarnya pada bulan Januari, Mei hingga Agustus dan menghasilkan dugaan curah hujan yang lebih rendah dari curah hujan aktual pada bulan April, Oktober hingga Desember.

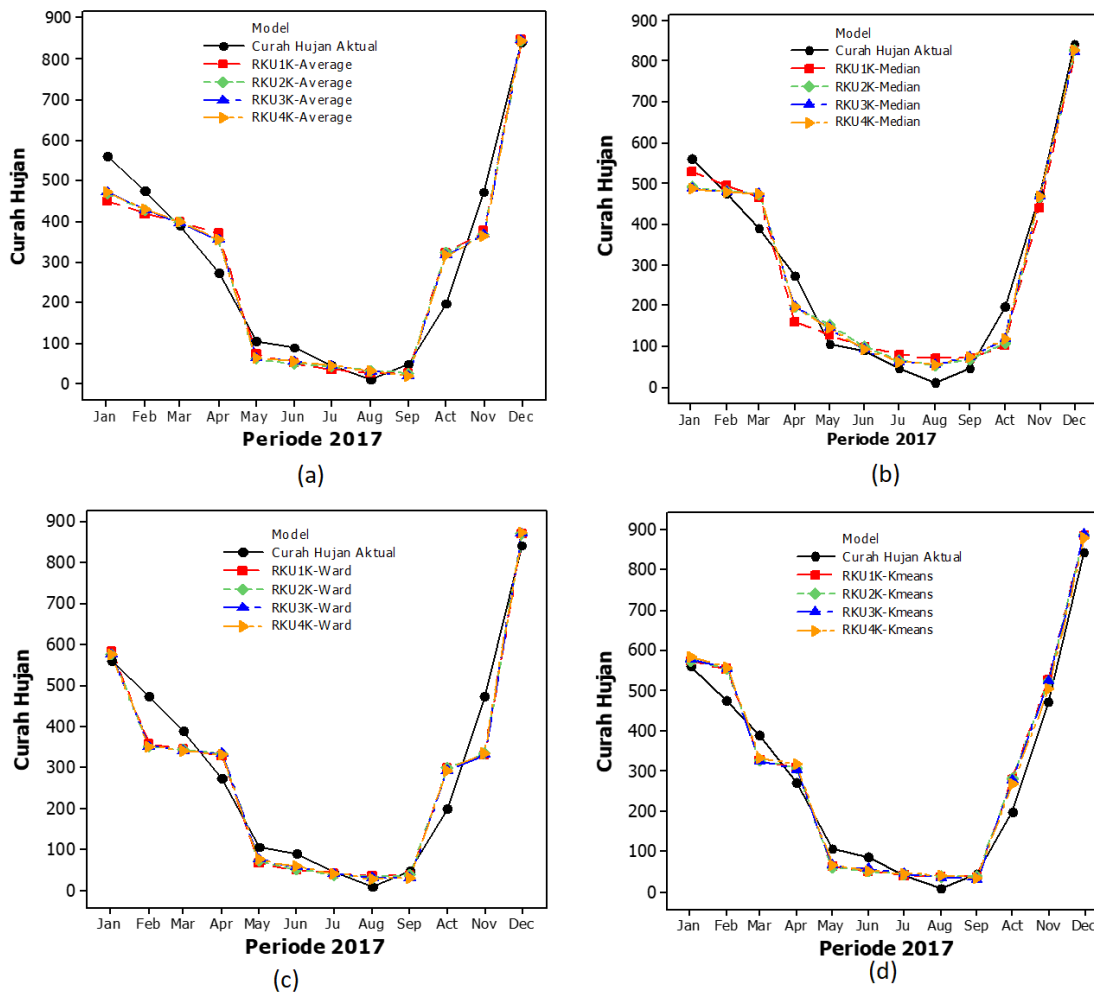


Gambar 5: Plot Curah Hujan Aktual dan Curah Hujan Dugaan Model *RKU*.

Tabel 4: Nilai korelasi dan RMSEP Model RKU.

Metode	Teknik Cluster	Komponen	Korelasi	RMSEP
RKU tanpa Dummy	-	KU1	0.941	90.339
	-	KU1, KU2	0.926	96.555
	-	KU1, KU2, KU3	0.927	95.935
	-	KU1, KU2, KU3, KU4	0.923	98.499
RKU dengan Dummy	K-Means (non-hierarki)	KU1, D1, D2, D3, D4	0.987	48.036
		KU1, KU2, D1, D2, D3, D4	0.987	47.965
		KU1, KU2, KU3, D1, D2, D3, D4	0.988	46.583
		KU1, KU2, KU3, KU4, D1, D2, D3, D4	0.989	44.200
	Average Linkage (Berhierarki)	KU1, D1, D2, D3, D4	0.964	66.726
		KU1, KU2, D1, D2, D3, D4	0.967	63.760
		KU1, KU2, KU3, D1, D2, D3, D4	0.969	62.171
		KU1, KU2, KU3, KU4, D1, D2, D3, D4	0.969	62.394
	Median Linkage (Berhierarki)	KU1, D1, D2, D3, D4	0.975	55.212
		KU1, KU2, D1, D2, D3, D4	0.979	51.132
		KU1, KU2, KU3, D1, D2, D3, D4	0.980	50.109
		KU1, KU2, KU3, KU4, D1, D2, D3, D4	0.980	49.968
	Ward Linkage (Berhierarki)	KU1, D1, D2, D3, D4	0.965	67.201
		KU1, KU2, D1, D2, D3, D4	0.964	67.441
		KU1, KU2, KU3, D1, D2, D3, D4	0.965	66.698
		KU1, KU2, KU3, KU4, D1, D2, D3, D4	0.966	66.260

Gambar 6 adalah plot curah hujan dugaan model RKUK dengan *dummy* ditentukan dari teknik *cluster* berhierarki. Gambar 6a menunjukkan bahwa model RKUK-Average secara tepat menduga data curah hujan aktual pada bulan Maret, Juli, Agustus, dan November. Model RKUK-Median dengan tepat menduga data curah hujan pada bulan Maret, Juni, Agustus, dan Desember (Gambar 6b). Sedangkan, model RKUK-Ward menghasilkan curah hujan dugaan yang tepat bulan Januari dan Desember yang merupakan curah hujan dengan intensitas tinggi (Gambar 6c). Namun, secara umum model RKUK-Kmeans lebih baik dalam menduga data curah hujan daripada model lain (Gambar 6d). Model RKUK-Kmeans dapat melakukan pendugaan curah hujan yang mengikuti pola curah hujan aktual. Dengan demikian, model RKUK-Kmeans adalah model terbaik berdasarkan hasil pemodelan SD maupun validasi model dalam menggambarkan dan menduga data curah hujan di Kabupaten Pangkep.



Gambar 6: Plot Curah Hujan Aktual dan Curah Hujan Dugaan Model: (a) RKUK-Average (b) RKUK-Median (c) RKUK-Ward (d) RKUK-k-means.

4. Simpulan

Peubah *dummy* ditentukan berdasarkan pada teknik *cluster* hierarki dan non-hierarki. Terdapat 5 peubah *dummy* yang digunakan dalam model RKU berdasarkan kelompok yang terbentuk pada data curah hujan. Penambahan peubah *dummy* mampu memperbaiki dugaan curah hujan model RKU. Secara umum, model RKU dengan *dummy* dari teknik *cluster* non-hierarki K-means adalah model terbaik berdasarkan hasil pemodelan dan validasi model. Dalam tahapan pemodelan, model RKUK-Kmeans yang melibatkan KU1 adalah model terbaik berdasarkan nilai R^2 (95.58%) tertinggi dan nilai RMSE (74.786) terendah. Model ini juga konsisten pada tahapan validasi model dengan nilai korelasi (0.989) tertinggi dan RMSEP (44.200) terendah.

Ucapan Terima Kasih. Kami berterima kasih kepada pimpinan Universitas Hasanuddin melalui LP2M (Lembaga Penelitian dan Pengabdian pada Masyarakat) Universitas Hasanuddin sebagai pemberi dana penelitian internal Unhas. Penelitian ini mendapatkan pembiayaan melalui Penelitian Dosen Penasehat Akademik (PDPA) tahun anggaran 2019. Kami juga berterima kasih kepada seluruh staf yang telah menyediakan fasilitas demi terselesainya penelitian ini.

Daftar Pustaka

- Busuioc, A., Chen, D., & Hellström, C. (2001). Performance of statistical downscaling models in GCM validation and regional climate change estimates: application for Swedish precipitation. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **21**(5): 557–578.
- Estiningtyas, W., & Wigena, A. H. (2011). Teknik statistical downscaling dengan regresi komponen utama dan regresi kuadrat terkecil parsial untuk prediksi curah hujan pada kondisi el nino, la nina, dan normal. *Jurnal Meteorologi dan Geofisika*, **12**(1): 65–72.
- Fernandez, E. (2005). On the influence of predictors area in statistical downscaling of daily parameters. *Norwegia Meteorological Institute*, **9**: 1–21.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (Vol. 5). New Jersey (US): Prentice hall Upper Saddle River.
- Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York (US): Springer-Verlag.
- Listyarini, T. (2017). Pemerintah jadikan sulsel pusat produksi garam. Retrieved March 4, 2019, from beritasatu.com website: <https://www.beritasatu.com/agribisnis/446948-pemerintah-jadikan-sulsel-pusat-produksi-garam.html>
- Mattjik, A. A., & Sumertajaya, I. M. (2011). *Sidik peubah ganda dengan menggunakan SAS*. Bogor (ID): IPB Press.
- Montgomery, D. C., & Peck, E. A. (1992). *Introduction to linear regression analysis* (2nd ed.). New York (US): John Wiley & Sons.
- Sahrman, S. (2014). *Model statistical downscaling dengan time lag data global circulation model untuk peramalan curah hujan* (Tesis). IPB University, Bogor (ID).
- Sahrman, S., & Kalondeng, A. (2017). *Pemodelan statistical downscaling dengan peubah boneka berdasarkan teknik cluster untuk peramalan curah hujan* (Laporan penelitian). Universitas Hasanuddin, Makassar (ID).
- Sipayung, S. B. (2005). Dampak variabilitas iklim terhadap produksi pangan di Sumatera. *Jurnal Sains Dirgantara*, **2**(2): 111–126.
- Wigena, A. H. (2006). *Pemodelan statistical downscaling dengan regresi projection pursuit untuk peramalan curah hujan bulanan* (Disertasi). IPB University, Bogor (ID).
- Wigena, A. H., Djuraidah, A., & Sahrman, S. (2015). Statistical downscaling dengan pergeseran waktu berdasarkan korelasi silang. *Jurnal Meteorologi Dan Geofisika*, **16**(1): 19–24.