

# PEMODELAN STATISTICAL DOWNSCALING DENGAN LASSO DAN GROUP LASSO UNTUK PENDUGAAN CURAH HUJAN\*

M. Yunus<sup>1</sup>, Asep Saefuddin<sup>2</sup>, and Agus M Soleh<sup>3‡</sup>

<sup>1</sup>Sharia Economics Study Program, UIN Sulthan Thaha Saifuddin Jambi, Indonesia, myunus@uinjambi.ac.id

<sup>2</sup>Department of Statistics, IPB University, Indonesia, asaefuddin@gmail.com

<sup>3</sup>Department of Statistics, IPB University, Indonesia, agusms@apps.ipb.ac.id

‡corresponding author

**Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)  
Vol 4 No 4 (2020), 649 - 660**

Copyright © 2020 M. Yunus, Asep Saefuddin, and Agus M Soleh. 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

One of the rainfall prediction techniques is the Statistical Downscaling Modeling (SDS). SDS modeling is one of the applications of modeling with covariates conditions that are generally large and not independent. The problems that will be encountered is the problem of ill-conditional data i.e multicollinearity and the high correlation between variables. The case of highly correlated data causes a linear regression coefficient estimators obtained to have a large variance. This research was conducted to make the statistical downscaling modeling using the lasso and group lasso for the prediction of rainfall. Group of the covariate scenario is applied based on the adjacent area, the high correlation between covariates and correlation between covariates and responses, and also the addition of dummy variables. Scenario six (grouping which is done by considering the covariates that have a positive correlation to the response is divided into 3 groups, 1 individual and the covariates that are negatively correlated with the response are divided into 2 groups, 1 individual) is better than the other scenarios in linear modeling without a dummy. Then, linear modeling with a dummy is better than without a dummy for both techniques. In linear modeling with a dummy, the Group lasso technique can be considered more in SDS modeling, because the difference in the RMSEP statistical value and the correlation coefficient value is significant.

**Keywords:** group lasso, lasso, rainfall, statistical downscaling.

---

\* Received Nov 2020; Accepted Des 2020; Published online on Des 2020

## 1. Pendahuluan

Teknik pemodelan *Statistical Downscaling* untuk menduga curah hujan telah banyak digunakan seperti dalam penelitian-penelitian berikut (lihat: Permatasari et al., 2017a; Santri et al., 2016; Soleh et al., 2015a, 2015b; Wigena, 2006). Pemodelan *Statistical Downscaling* (SDS) merupakan pengaplikasian pemodelan dengan kovariat yang biasanya besar dan tidak independen satu sama lain. Masalah yang sering ditemukan yaitu masalah *ill-conditional data* seperti antar peubah memiliki korelasi yang tinggi dan multikolinieritas. Kondisi berkorelasi tinggi antar data dapat berakibat penduga koefisien regresi linier yang didapatkan sangat bervariasi (ragam yang tinggi). Penggunaan model linier terampat sebaran gamma menduga curah hujan bulanan yang telah dikembangkan sebagai model peramalan curah hujan menggunakan SDS (Soleh et al., 2015b), dan pendugaan curah hujan menggunakan regresi kuantil dengan regularisasi  $L_1$  (lasso) (Santri et al., 2016). Tibshirani (1996) telah memperkenalkan teknik *least absolute shrinkage and selection operator* (lasso), mengganti penalti pada regresi gulud menjadi norm  $L_1$  (regularisasi  $L_1$ ) dan metode algoritma LAR (*Least Angle Regression*) ditemukan dan mulai dikenalkan oleh Efron et al. (2004). Teknik ini memungkinkan mengeliminasi beberapa kovariat yang memiliki korelasi tinggi walaupun sebenarnya penting. Teknik yang mungkin dapat menanggulangi hal tersebut adalah teknik pengembangan dari lasso seperti teknik *group lasso*.

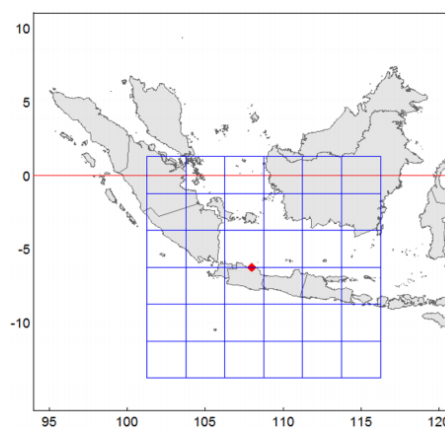
Bakin (1999) memberitahukan teknik *group lasso*, dan kemudian dilakukan generalisasi oleh (Yuan & Lin, 2006) (dalam (Hastie et al., 2009)). *Group lasso* digunakan pada data dengan peubah penjelas yang telah berkelompok sebelumnya. Jika dibandingkan metode kuadrat terkecil (MKT) dan lasso maka yang lebih baik adalah *Group lasso*, di saat ada kelompok  $\beta$  dan dalam satu kelompok,  $\beta_i \geq 1$  semua atau  $\beta_i < 1$  semua atau  $\beta_i = 0$  semua (Yunus et al., 2017). Berdasarkan konsep, jika suatu kelompok itu nyata maka akan berkontribusi terhadap respons untuk semua anggota kelompok. Analisis-analisis terkait teknik *group lasso* telah sering digunakan. Pada peubah yang sudah berkelompok dilakukan pemilihan model dan pendugaan (Yuan & Lin, 2006). Selanjutnya pendugaan *group lasso* digunakan pada model linier (Nardi & Rinaldo, 2008) dan hasil pendugaan *group lasso* diidentifikasi kekonsistenan (Wang & Leng, 2008), (Huang & Zhang, 2010) peubah dapat berkelompok dengan jumlah tidak merata dan dalam ukuran besar pada *Group lasso*. *Group lasso* memprediksi hasil lebih baik dibandingkan dengan lasso (Lounici et al., 2011). Mengecilkan jumlah kuadrat sisaan dengan kendala seperti dalam lasso tetapi di beberapa kovariat dikelompokkan untuk memperoleh penduga koefisien pada *group lasso* (Yuan & Lin, 2006). Oleh sebab itu, penelitian ini membuat pemodelan *Statistical Downscaling* dengan lasso dan *group lasso* untuk pendugaan curah hujan.

## 2. Metodologi

### 2.1 Data

Data yang digunakan pada tulisan ini adalah data curah hujan di Kabupaten Indramayu dan data presipitasi dari *Global Precipitation Climatology Project* (GPCP) versi 2.2. Data observasi presipitasi GPCP diberikan oleh "the NOAA/OAR/ ESRL PSD, Boulder,

Colorado, USA”, dari web sitenya di <http://www.esrl.noaa.gov/psd/>. Domain data yang digunakan disajikan pada Gambar 1 dengan mengambil 49 kovariat ( $7 \times 7$  grid) pada sistem koordinat  $101.25^{\circ}$ - $116.25^{\circ}$  BT dan  $13.75^{\circ}$  LS –  $1.25^{\circ}$  LU dengan lebar grid yaitu  $2.5^{\circ} \times 2.5^{\circ}$ . Kabupaten yang terletak di bawah grid tengah domain ditetapkan adalah Kabupaten Indramayu. Data curah hujan bulanan tahun 1981 sampai 2014 dari 11 pos hujan di Kabupaten Indramayu dan sekitarnya menjadi peubah respons. Pada lokasi ini terbagi empat daerah Zona Musim (ZOM), yaitu ZOM 77, ZOM 78, ZOM 79, dan ZOM 80. Penelitian dilakukan hanya menggunakan curah hujan bulanan dari pos hujan yang mewakili tiga ZOM yaitu ZOM 77 (Kr Anyar, Pusakanegara dan Tulang Kacang), ZOM 78 (Dempet, Indramayu, Juntinyuat dan Losarang), dan ZOM 79 (Gegesik, Karangkendal, Krangkeng dan Sukadana).



Gambar 1: Domain grid kovariat yang dipilih.

## 2.2 Prosedur Analisis

Prosedur analisis data curah hujan di Kabupaten Indramayu dan sekitarnya memakai pendekatan lasso dan *group* lasso adalah:

1. Mendapatkan informasi terkait pembuatan *group* (kelompok) yang diperlukan pada teknik *group* lasso melalui eksplorasi data.
2. Penentuan kelompok dengan cara memeriksa korelasi antara kovariat dengan respons dan antar peubah kovariat.
3. Data penelitian yang *missing value* dieliminasi untuk setiap peubah respons.
4. Berdasarkan daerah (grid) yang berdekatan dan hasil korelasi pada poin 2 dilakukan pengelompokan peubah kovariat dari data curah hujan.
5. Menganalisis data curah hujan melalui pemodelan SDS dengan regresi linier menggunakan lasso dan *group* lasso. Tahapan sebagai berikut:
  - a. Memilih besaran parameter pengontrol ( $\lambda$ ) lasso dan *group* lasso yang optimum. Parameter ( $\lambda \geq 0$ ) sebagai pengontrol besarnya penyusutan. Model akan menjadi bentuk standar ketika  $\lambda = 0$ , jika nilai  $\lambda$  semakin besar (menuju tak hingga) maka nilai dugaan koefisien akan semakin kecil menuju nol (Hastie *et al.*, 2009).

- b. Membuat pemodelan regresi linier mengambil  $\lambda$  optimum yang terpilih dari lasso memakai *software* komputasi statistik R dengan paket glmnet.
  - c. Membuat pemodelan regresi linier mengambil  $\lambda$  optimum yang terpilih dari *group* lasso memakai *software* komputasi statistik R dengan paket gglasso.
6. Memilih model terbaik dari beberapa model hasil pendekatan lasso dan *group* lasso dengan melihat  $\lambda$  yang menghasilkan nilai sisaan validasi silang minimum dan nilai korelasi. Model terbaik ditentukan berdasarkan nilai korelasi terbesar dan *root mean square error of prediction* (RMSEP) terkecil (Makridakis et al., 1999).

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$$

7. Memberikan prediktor (kovariat) yaitu peubah *dummy*, dengan mengelompokan data curah hujan menjadi 6 kelompok (Permatasari et al., 2017), yaitu:
  - a. Kelompok 1 intensitas curah hujan pada kuantil ke-0.25.
  - b. Kelompok 2 intensitas curah hujan pada kuantil ke-0.50.
  - c. Kelompok 3 intensitas curah hujan pada kuantil ke-0.75.
  - d. Kelompok 4 intensitas curah hujan pada kuantil ke-0.90.
  - e. Kelompok 5 intensitas curah hujan pada kuantil ke-0.95.
  - f. Kelompok 6 intensitas curah hujan di atas kuantil ke-0.95.
8. Lakukan penambahan peubah *dummy* kelompok curah hujan untuk proses pemodelan seperti pada poin empat sampai poin enam.

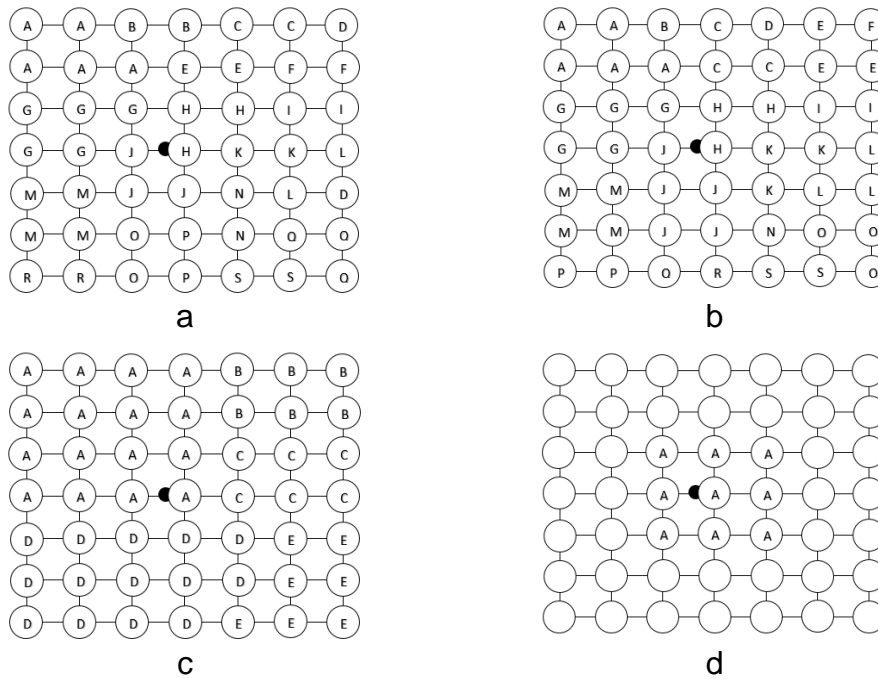
### 3. HASIL DAN PEMBAHASAN

#### 3.1 Eksplorasi Data

Teknik pemodelan linier yang digunakan adalah *group* lasso. Pada teknik *group* lasso terlebih dilakukan pengelompokan kovariat sebelum dilakukan pendugaan model linier. Pengelompokan kovariat menggunakan tujuh skenario *group* lasso dan sebagai pembandingan dilakukan juga pendugaan dengan teknik lasso. Pengelompokan dilakukan berdasarkan grid (daerah) yang berdekatan dan korelasi tinggi antar kovariat. Hal ini dilakukan karena kovariat yang berdekatan cenderung memiliki korelasi tinggi dan pengaruh yang relatif sama. Selanjutnya, pengelompokan juga dilakukan mempertimbangkan korelasi antara kovariat dengan peubah respons, yaitu dibedakan antara kovariat yang memiliki korelasi positif dan negatif. Peubah kovariat akan dikelompokan berdasarkan tiga hal, yaitu:

1. Pengelompokan berdasarkan daerah yang berdekatan dan korelasi tinggi antar kovariat  
 Pengelompokan ini dilakukan dengan melihat jauh dekatnya grid (daerah) dan kovariat yang bersebelahan cenderung memiliki korelasi tinggi. Peubah kovariat yang berdekatan dan memiliki korelasi yang tinggi ( $\rho > 0.7$ ) akan dikelompokan dalam satu grup yang sama. Penentuan batas korelasi yang membedakan antar grup dilakukan secara subjektif oleh peneliti. Terdapat 4 skenario yang dilakukan

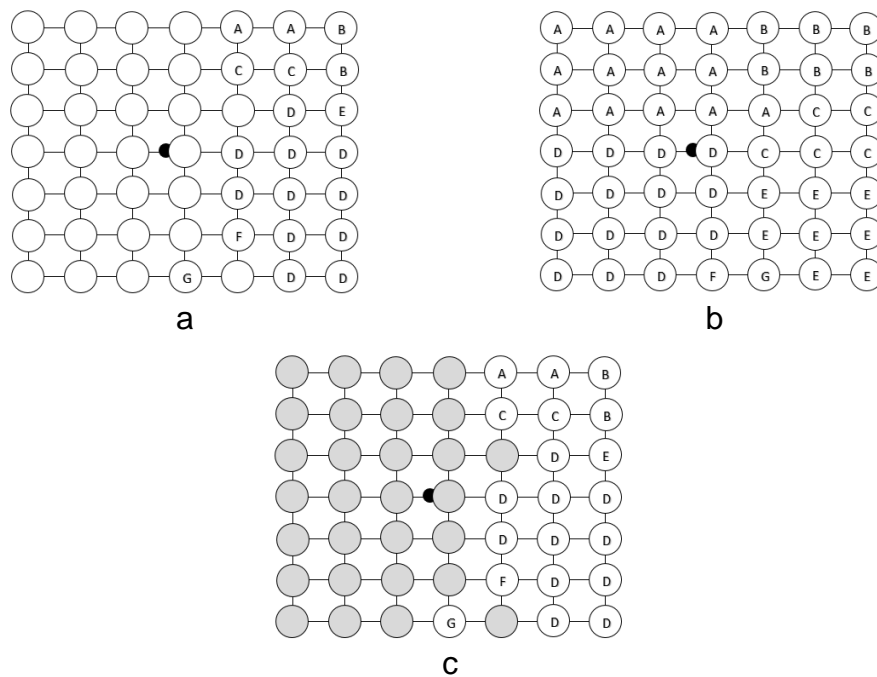
dalam pengelompokan berdasarkan daerah yang berdekatan dan korelasi, pembagian kelompok pada Gambar 2.



Gambar 2: Skenario berdasarkan daerah yang berdekatan dan korelasi tinggi antar kovariat (a) Skenario 1, (b) Skenario 2, (c) Skenario 3, dan (d) Skenario 4.

Pengelompokan yang dilakukan dengan mempertimbangkan setiap kelompok minimal memiliki 2 anggota kovariat, membentuk 19 kelompok (Gambar 2a). Pengelompokan yang dilakukan dengan mempertimbangkan semua kovariat dalam satu kelompok harus saling berkorelasi satu dengan yang lainnya dan terdapat kovariat yang tidak memiliki pasangan (individu), membentuk 13 kelompok dan 6 individu (Gambar 2b). Pengelompokan yang dilakukan dengan mempertimbangkan banyaknya anggota kovariat dalam kelompok menjadi perhatian utama, membentuk 5 kelompok (Gambar 2c). Pengelompokan yang dilakukan dengan mempertimbangkan bahwa curah hujan di Kabupaten Indramayu dipengaruhi oleh grid-grid yang berada di sekitarnya, membentuk 1 kelompok dan sisanya individu (Gambar 2d).

2. Pengelompokan berdasarkan daerah yang berdekatan, korelasi tinggi antar kovariat dan korelasi antara kovariat dan respons  
 Pada pengelompokan ini, dibentuk kelompok berdasarkan jauh atau dekatnya daerah, peubah kovariat yang memiliki korelasi yang tinggi ( $\rho > 0.7$ ) akan dikelompokkan dalam satu grup yang sama, dan kovariat yang memiliki pengaruh positif atau negatif akan dibagi dalam kelompok yang berbeda. Terdapat tiga skenario yang dilakukan dalam pengelompokan berdasarkan daerah yang berdekatan, korelasi antar kovariat, dan korelasi antara kovariat dan respons, pembagian kelompok pada Gambar 3.



Gambar 3: Skenario pengelompokan berdasarkan daerah yang berdekatan, korelasi tinggi antar kovariat dan korelasi antara kovariat dan respons (a) Skenario 5, (b) Skenario 6, dan (c) Skenario 7.

Pengelompokan yang dilakukan dengan mempertimbangkan kovariat yang memiliki korelasi positif terhadap respons dibagi ke dalam 4 kelompok, 3 individu dan sisanya yang berkorelasi negatif juga sebagai individu (Gambar 3a). Pengelompokan yang dilakukan dengan mempertimbangkan kovariat yang memiliki korelasi positif terhadap respons dibagi ke dalam 3 kelompok, 1 individu dan kovariat yang berkorelasi negatif terhadap respons dibagi ke dalam 2 kelompok, 1 individu (Gambar 3b). Pengelompokan yang dilakukan dengan mempertimbangkan kovariat yang memiliki korelasi positif terhadap respons dibagi ke dalam 4 kelompok, 3 individu dan sisanya yang berkorelasi negatif tidak diikutsertakan dalam pemodelan (Gambar 3c).

3. Pengelompokan berdasarkan daerah yang berdekatan, korelasi tinggi antar kovariat dan korelasi antara kovariat dan respons serta penambahan peubah *dummy*

Pada pengelompokan ini, dibentuk kelompok berdasarkan jauh atau dekatnya daerah, peubah kovariat yang memiliki korelasi yang tinggi ( $\rho > 0.7$ ) akan dikelompokkan dalam satu grup yang sama, dan kovariat yang memiliki pengaruh positif atau negatif akan dibagi dalam kelompok yang berbeda. Terdapat 7 skenario yang dilakukan dalam pengelompokan berdasarkan daerah yang berdekatan, korelasi antar kovariat, dan korelasi antara kovariat dan respons, serta penambahan peubah *dummy*. Terdapat dua skenario yang diterapkan dalam penambahan peubah *dummy*. Skenario pertama peubah *dummy* dijadikan sebagai individu (D1) dan skenario kedua peubah *dummy* dibuat dalam satu kelompok (D2).

### 3.2 Pemodelan Lasso dan Group Lasso tanpa Dummy

Teknik pemodelan linier yang dibandingkan adalah lasso dan *group* lasso dianggap dapat menduga respons. Selanjutnya penentuan teknik terbaik menggunakan nilai statistik RMSEP dan nilai koefisien korelasi dinyatakan pada Tabel 1 dan Tabel 2.

Tabel 1: Nilai statistik RMSEP untuk setiap pendugaan model tanpa *dummy* kelompok curah hujan

Pos Hujan	LASSO	Skenario <i>Group</i> LASSO						
		1	2	3	4	5	6	7
Kr Anyar	<b>133.69</b>	134.67	136.20	136.01	135.87	135.90	138.17	147.79
Pusakanegara	100.08	99.05	100.26	100.28	99.48	98.98	<b>98.04</b>	109.98
Tulangcakang	<b>153.18</b>	154.83	154.55	153.89	156.57	154.06	155.53	165.68
Dempet	<b>112.20</b>	114.82	114.46	115.89	117.98	112.87	113.55	134.52
Indramayu	144.13	142.21	145.37	141.89	143.38	142.41	<b>140.68</b>	158.05
Juntinyuat	<b>113.23</b>	114.48	114.81	114.72	114.74	115.02	114.20	128.31
Losarang	114.08	113.51	113.84	114.17	114.94	115.03	<b>113.47</b>	122.75
Gegesik	95.68	95.17	95.00	94.61	94.73	96.96	<b>94.18</b>	99.05
Karangkendal	97.86	<b>97.71</b>	98.41	98.33	98.18	98.15	100.28	107.66
Krangkeng	91.43	89.65	89.75	88.27	91.38	90.35	<b>88.02</b>	100.57
Sukadana	94.96	93.82	<b>92.64</b>	94.56	93.66	94.42	93.22	104.09

Nilai statistik RMSEP dari pendugaan menggunakan *group* lasso yang diperoleh tidak selalu lebih kecil dibandingkan dengan nilai statistik RMSEP menggunakan teknik lasso pada model penduga curah hujan di setiap pos hujan, disajikan pada Tabel 1. Namun, ada kecenderungan *group* lasso lebih baik dibandingkan dengan teknik lasso. Hal ini terlihat bahwa sebagian besar nilai statistik RMSEP yaitu sebanyak tujuh pos hujan lebih kecil dibandingkan dengan teknik lasso. Namun secara umum, pendugaan model menggunakan *group* lasso relatif sama baik dibandingkan dengan menggunakan teknik lasso. Penggunaan *group* lasso pada pemodelan linier belum memberikan keunggulan mutlak sebagai teknik yang lebih baik dibandingkan dengan lasso.

Tabel 2: Nilai koefisien korelasi antara respons dengan dugaan, untuk setiap hasil pendugaan model tanpa *dummy* kelompok curah hujan

Pos Hujan	LASSO	Skenario <i>Group</i> LASSO						
		1	2	3	4	5	6	7
Kr Anyar	<b>0.546</b>	0.501	0.524	0.517	0.493	0.499	0.526	0.362
Pusakanegara	<b>0.653</b>	0.633	0.633	0.637	0.634	0.649	0.651	0.556
Tulangcakang	0.518	0.507	0.508	0.524	0.496	0.506	<b>0.530</b>	0.363
Dempet	<b>0.636</b>	0.595	0.605	0.602	0.581	0.616	0.616	0.385
Indramayu	0.545	0.526	0.547	0.557	0.534	0.539	<b>0.559</b>	0.351
Juntinyuat	0.571	0.556	0.565	0.572	0.550	0.556	<b>0.576</b>	0.405
Losarang	<b>0.592</b>	0.575	0.585	0.590	0.579	0.579	0.588	0.452
Gegesik	<b>0.670</b>	0.655	0.653	0.658	0.651	0.655	0.666	0.573
Karangkendal	<b>0.659</b>	0.642	0.634	0.636	0.631	0.635	0.631	0.539
Krangkeng	0.609	0.605	0.603	<b>0.623</b>	0.596	0.599	0.616	0.441
Sukadana	0.685	0.682	<b>0.689</b>	0.675	0.677	0.673	0.683	0.616

Berdasarkan nilai koefisien korelasi dari *group* lasso tidak memiliki pola yang dapat menentukan teknik ini lebih baik dibandingkan dengan lasso disajikan pada Tabel 2. Kedua teknik pendugaan memberikan koefisien korelasi yang bervariasi dan tidak ada teknik yang selalu memberikan koefisien korelasi lebih tinggi dibandingkan dengan teknik lain. Oleh karena itu, pemodelan linier tanpa *dummy* menggunakan teknik lasso dan *group* lasso sama-sama mampu menjadi alternatif pada pemodelan SDS. Perbedaan pada nilai statistik RMSEP dan nilai koefisien korelasi yang belum terlalu signifikan.

Dari tujuh skenario yang diterapkan pada *group* lasso, terlihat bahwa nilai statistik RMSEP dari pendugaan yang dihasilkan tidak ada skenario yang selalu lebih kecil dibandingkan dengan skenario yang lainnya. Namun, ada kecenderungan skenario enam lebih baik dibandingkan dengan skenario yang lainnya. Hal ini terlihat bahwa sebagian besar nilai statistik RMSEP yaitu sebanyak enam pos hujan lebih kecil dibandingkan dengan skenario lainnya. Demikian juga nilai koefisien korelasi dari tujuh skenario yang diterapkan tidak memiliki pola yang dapat menentukan skenario mana yang lebih baik dibandingkan dengan skenario lainnya. Nilai koefisien korelasi dengan menggunakan skenario enam cenderung lebih besar dibandingkan dengan skenario lainnya yaitu sebanyak enam pos hujan. Secara umum, ketujuh skenario pendugaan memberikan koefisien korelasi yang bervariasi dan tidak ada skenario yang selalu memberikan koefisien korelasi lebih tinggi dibandingkan dengan skenario lain. Namun secara lebih khusus, skenario enam lebih baik dibandingkan dengan skenario lainnya. Oleh karena itu, pemodelan linier tanpa *dummy* dengan *group* lasso menggunakan skenario enam bisa dijadikan sebagai alternatif pilihan yang lebih baik pada pemodelan SDS, walaupun perbedaan nilai statistik RMSEP dan nilai koefisien korelasi yang belum terlalu signifikan.

### 3.3 Pemodelan Lasso dan *Group* Lasso dengan *Dummy*

Penambahan peubah *dummy* pada kovariat diharapkan mampu memberikan hasil dugaan yang lebih baik dibandingkan dengan pemodelan tanpa *dummy*. Hasil analisis nilai statistik RMSEP dan nilai koefisien korelasi ditunjukkan dalam Tabel 3 dan Tabel 4. Hasil analisis nilai statistik RMSEP dari pendugaan menggunakan *group* lasso yang diperoleh tidak selalu lebih kecil dibandingkan dengan nilai statistik RMSEP menggunakan teknik lasso, disajikan pada Tabel 3. Secara umum, pendugaan model menggunakan *group* lasso relatif sama baik dibandingkan dengan menggunakan lasso. Penggunaan *group* lasso pada pemodelan linier belum memberikan keunggulan mutlak sebagai teknik yang lebih baik dibandingkan dengan lasso.



Tabel 3: Nilai statistik RMSEP untuk setiap hasil pendugaan model dengan *dummy* kelompok curah hujan

Pos Hujan	LASSO	Group LASSO													
		Skenario 1		Skenario 2		Skenario 3		Skenario 4		Skenario 5		Skenario 6		Skenario 7	
		D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2
Kr Anyar	84.56	84.30	86.15	85.31	84.37	87.01	85.77	87.00	83.55	86.39	86.88	84.49	88.81	81.04	82.24
Pusakanegara	<b>34.36</b>	36.57	36.73	36.34	37.25	37.45	36.66	37.99	38.94	37.57	37.26	37.21	37.99	35.74	35.24
Tulangkacang	<b>74.60</b>	79.29	77.26	82.11	77.02	82.55	76.65	78.89	80.76	80.10	78.75	78.70	78.42	76.66	78.36
Dempet	50.45	49.82	52.53	50.92	51.60	51.67	50.96	53.15	50.78	52.36	52.59	49.80	50.56	<b>49.25</b>	49.42
Indramayu	<b>75.83</b>	83.44	87.31	81.36	80.05	85.62	80.27	81.02	82.44	80.20	80.76	81.34	79.44	76.28	78.72
Juntinyuat	60.25	61.87	63.42	62.14	61.76	60.75	62.23	62.54	60.65	61.86	60.78	60.27	61.21	<b>59.93</b>	60.89
Losarang	<b>55.58</b>	58.72	58.10	58.61	56.60	58.28	57.55	59.68	57.63	57.94	58.88	57.48	58.60	58.55	57.75
Gegesik	<b>30.06</b>	33.11	32.38	32.64	32.09	31.50	31.95	33.97	33.30	35.16	32.89	32.69	31.42	30.27	30.15
Karangkendal	35.19	37.67	39.68	37.28	38.13	37.26	38.36	39.17	39.09	40.17	39.05	37.09	39.05	<b>34.73</b>	35.53
Krangkeng	45.09	45.32	45.59	44.37	<b>43.96</b>	45.42	45.61	46.16	45.41	45.33	44.94	45.48	44.33	43.98	45.41
Sukadana	33.73	36.63	36.97	36.14	35.46	36.02	35.28	37.18	36.88	37.33	36.40	35.32	35.49	<b>33.50</b>	33.61

Tabel 4: Koefisien korelasi antara nilai respons dengan dugaan, untuk setiap hasil pendugaan model dengan *dummy* kelompok curah hujan

Pos Hujan	Group LASSO															
	LASSO		Skenario 1		Skenario 2		Skenario 3		Skenario 4		Skenario 5		Skenario 6		Skenario 7	
	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2
Kr Anyar	0.8731	0.8747	0.8786	0.8659	0.8757	0.8697	0.8758	0.8703	0.8818	0.8661	0.8815	0.8663	0.8619	0.8764	<b>0.8889</b>	<b>0.8889</b>
Pusakanegara	0.9686	0.9681	0.9675	0.9686	0.9680	0.9685	0.9680	0.9670	0.9683	0.9671	0.9675	0.9684	0.9678	0.9692	<b>0.9699</b>	<b>0.9699</b>
Tulangkacang	0.9234	0.9191	0.9270	0.9210	0.9255	0.9172	0.9284	0.9226	<b>0.9294</b>	0.9224	0.9266	0.9241	0.9279	0.9292	0.9241	0.9241
Dempet	0.9484	0.9498	0.9483	0.9492	0.9497	0.9511	0.9515	0.9465	0.9510	0.9463	0.9471	<b>0.9522</b>	0.9511	0.9493	0.9520	0.9520
Indramayu	0.9089	0.8931	0.8968	0.9019	0.9044	0.8963	<b>0.9123</b>	0.9022	0.9001	0.9016	0.9039	0.8955	0.9036	0.9121	0.9061	0.9061
Juntinyuat	0.9075	0.9065	0.9150	0.9049	0.9128	0.9126	0.9135	0.9080	<b>0.9177</b>	0.9097	0.9171	0.9157	0.9118	0.9092	0.9157	0.9157
Losarang	0.9274	0.9232	0.9261	0.9207	<b>0.9310</b>	0.9217	0.9286	0.9242	0.9303	0.9240	0.9268	0.9278	0.9267	0.9235	0.9284	0.9284
Gegesik	0.9743	0.9742	0.9752	0.9751	0.9753	0.9757	0.9754	0.9720	0.9740	0.9712	0.9739	0.9754	0.9750	0.9750	<b>0.9757</b>	<b>0.9757</b>
Karangkandal	0.9655	0.9663	0.9672	0.9677	0.9675	<b>0.9677</b>	0.9665	0.9658	0.9668	0.9657	0.9656	0.9673	0.9672	0.9671	0.9666	0.9666
Krangkeng	0.9270	0.9313	<b>0.9381</b>	0.9324	0.9353	0.9305	0.9339	0.9293	0.9336	0.9317	0.9354	0.9297	0.9344	0.9298	0.9345	0.9345
Sukadana	0.9684	0.9675	0.9690	0.9698	0.9692	0.9682	0.9704	0.8703	0.8818	0.8661	0.8815	0.8663	0.8619	0.8764	<b>0.8889</b>	<b>0.8889</b>

Berdasarkan nilai koefisien korelasi, ada kecenderungan *group* lasso lebih baik dibandingkan dengan lasso, disajikan pada Tabel 4. Teknik *group* lasso selalu memberikan koefisien korelasi lebih tinggi dibandingkan dengan teknik lasso. Teknik *Group* lasso lebih bisa dipertimbangkan pada pemodelan SDS, karena perbedaan nilai koefisien korelasi sudah signifikan. Secara umum, pemodelan linier dengan *dummy*, lasso dan *group* lasso sama-sama bisa dipertimbangkan pada pemodelan SDS, karena perbedaan nilai statistik RMSEP yang belum terlalu signifikan.

Dari tujuh skenario yang diterapkan pada *group* lasso, terlihat bahwa nilai statistik RMSEP dari pendugaan yang dihasilkan tidak ada skenario yang selalu lebih kecil dibandingkan dengan skenario yang lainnya. Namun secara khusus, skenario tujuh D1 lebih baik dibandingkan dengan skenario lainnya. Hal ini terlihat bahwa sebagian besar nilai statistik RMSEP yaitu sebanyak enam pos hujan lebih kecil dibandingkan dengan skenario lainnya. Demikian juga nilai koefisien korelasi dari tujuh skenario yang diterapkan tidak memiliki pola yang dapat menentukan skenario mana yang lebih baik dibandingkan dengan skenario lainnya. Nilai koefisien korelasi dengan menggunakan skenario D2 yaitu apabila *dummy* dikelompokkan cenderung lebih besar dibandingkan dengan D1 yaitu *dummy* tidak dikelompokkan. Secara umum, ketujuh skenario pendugaan memberikan koefisien korelasi yang bervariasi dan tidak ada skenario yang selalu memberikan koefisien korelasi lebih tinggi dibandingkan dengan skenario lain.

Pemodelan linier dengan *dummy* lebih baik dibandingkan dengan tanpa *dummy* di kedua teknik tersebut. Pemodelan dengan *dummy* menghasilkan nilai RMSEP lebih kecil dibandingkan dengan tanpa *dummy* dan pemodelan dengan *dummy* juga menghasilkan nilai koefisien korelasi lebih besar dibandingkan dengan tanpa *dummy*. Oleh karena itu, pemodelan dengan *dummy*, bisa dijadikan sebagai alternatif pilihan yang lebih baik pada pemodelan SDS, walaupun perbedaan nilai statistik RMSEP dan nilai koefisien korelasi pada setiap skenario yang diterapkan belum terlalu signifikan.

#### 4. Simpulan

Terdapat kecenderungan teknik *group* lasso lebih baik dibandingkan dengan teknik lasso, di saat diterapkan skenario pengelompokan kovariat berdasarkan daerah yang berdekatan, korelasi tinggi antar kovariat dan korelasi antara kovariat dan respons, serta penambahan peubah *dummy*. Skenario enam (pengelompokan yang dilakukan dengan mempertimbangkan kovariat yang memiliki korelasi positif terhadap respons dibagi ke dalam 3 kelompok, 1 individu dan kovariat yang berkorelasi negatif terhadap respons dibagi ke dalam 2 kelompok, dan 1 individu lebih baik dibandingkan dengan skenario lainnya pada pemodelan linier tanpa *dummy*. Kemudian pemodelan linier dengan *dummy* lebih baik dibandingkan dengan tanpa *dummy* di kedua teknik tersebut. Pada pemodelan linier dengan *dummy*, teknik *Group* lasso lebih dapat dipertimbangkan pada pemodelan SDS, karena perbedaan nilai statistik RMSEP dan nilai koefisien korelasi sudah signifikan.

#### Daftar Pustaka

Bakin, S. (1999). *Adaptive regression and model selection in data mining problems [thesis]*. Canberra (AU): The Australian National University.

- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least Angle Regression. *The Annals of Statistics*, 32(2): 407–499.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York (US): Springer Science & Business Media.
- Huang, J., & Zhang, T. (2010). The Benefit of Group Sparsity. *The Annals of Statistics*, 38(4): 1978–2004.
- Lounici, K., Pontil, M., Van De Geer, S., Tsybakov, A. B., & others. (2011). Oracle Inequalities and Optimal Inference Under Group Sparsity. *The Annals of Statistics*, 39(4): 2164–2204.
- Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1999). *Metode dan Aplikasi Peramalan*. Jakarta (ID): Erlangga.
- Nardi, Y., & Rinaldo, A. (2008). On the Asymptotic Properties of the Group Lasso Estimator for Linear Models. *Electronic Journal of Statistics*, 2: 605–633.
- Permatasari, S. M., Djuraidah, A., & Soleh, A. M. (2017). Statistical Downscaling with Gamma Distribution and Elastic Net Regularization. *The 2nd International Conference On Applied Statistics (ICAS 2016)*, 121–129. Bandung (ID): Department of Statistics, UNPAD.
- Santri, D., Wigena, A. H., & Djuraidah, A. (2016). Statistical Downscaling Modeling with Quantile Regression Using Lasso to Estimate Extreme Rainfall. *AIP Conference Proceedings*, 1707(1), 080005. AIP Publishing LLC.
- Soleh, A. M., Wigena, A. H., Djuraidah, A., & Saefuddin, A. (2015a). Pemodelan Statistical Downscaling untuk Menduga Curah Hujan Bulanan Menggunakan Model Linier Terampat Sebaran Gamma. *Jurnal Informatika Pertanian*, 24(2): 215–222.
- Soleh, A. M., Wigena, A. H., Djuraidah, A., & Saefuddin, A. (2015b). Statistical downscaling to predict monthly rainfall using linear regression with L1 regularization (LASSO). *Applied Mathematical Sciences*, 9(108): 5361–5369.
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1): 267–288.
- Wang, H., & Leng, C. (2008). A Note on Adaptive Group Lasso. *Computational Statistics & Data Analysis*, 52(12): 5277–5286.
- Wigena, A. (2006). *Pemodelan statistical downscaling dengan regresi projection pursuit untuk peramalan curah hujan bulanan: kasus curah hujan bulanan di indramayu [disertasi]*. Bogor (ID): Institut Pertanian Bogor.
- Yuan, M., & Lin, Y. (2006). Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1): 49–67.
- Yunus, M., Saefuddin, A., & Soleh, A. M. (2017). Characteristics of Group Lasso in Handling High Correlated Data. *Applied Mathematical Sciences*, 11(20): 953–961.