JUDUL KARYA ILMIAH MAKSIMUM TIGA BARIS, LIMA BELAS KATA TIDAK TERMASUK KATA DEPAN DAN KATA SAMBUNG

AGUS MOHAMAD SOLEH

SEKOLAH PASCASARJANA
INSTITUT PERTANIAN BOGOR
BOGOR
2015

PERNYATAAN MENGENAI DISERTASI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa disertasi berjudul Judul Karya Ilmiah Maksimum Tiga Baris, Lima Belas Kata tidak Termasuk Kata Depan dan Kata Sambung adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir disertasi ini.

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

Bogor, September 2015

Agus Mohamad Soleh
NIM G161100021

RINGKASAN

AGUS MOHAMAD SOLEH. Judul Karya Ilmiah Maksimum Tiga Baris, Lima Belas Kata tidak Termasuk Kata Depan dan Kata Sambung . Dibimbing oleh PEMBIMBING 1, PEMBIMBING 2 dan PEMBIMBING 3.

Pemodelan *Statistical Downscaling* (SDS) merupakan suatu teknik dalam klimatologi yang menggunakan pemodelan statistika untuk menganalisis hubungan antara data iklim skala besar (global) dengan data iklim skala kecil (lokal). Pemodelan SDS umumnya melibatkan kovariat skala besar terkondisi buruk (*ill-conditioned*) (tidak bebas/korelasi tinggi). Teknik-teknik seperti pereduksian dimensi, seleksi peubah, dan penyusutan koefisien (*shrinkage*) dapat digunakan untuk mengatasinya. Teknik regularisasi L_1 merupakan salah satu teknik yang dikembangkan untuk menangani masalah kovariat terkondisi buruk oleh Tibshirani (1996) dengan cara seleksi peubah dan penyusutan koefisien. Penelitian yang dilakukan merupakan kajian tentang penggunaan dan pengembangan teknik regularisasi L_1 pada model linier untuk mendapatkan solusi bagi permasalahan kovariat terkondisi buruk dalam pemodelan SDS. Dalam hal ini peubah kovariat mengambil nilai dari luaran model GCM dari CMIP5 dan data observasi GPCP versi 2.2 pada grid domain 7×7 yang ditetapkan di atas wilayah Kabupaten Indramayu. Pemodelan yang digunakan merupakan pemodelan linier berbasis sebaran, yaitu respons diasumsikan berasal dari sebaran normal, sebaran Gamma dan sebaran pareto terampat.

Penelitian dibagi ke dalam dua kelompok, yaitu kelompok kajian pengembangan teknik regularisasi L_1 untuk pemodelan linier sebaran Gamma dan sebaran pareto terampat, dan kelompok kajian aplikasi pemodelan SDS untuk pendugaan curah hujan bulanan menggunakan pemodelan linier. Pengembangan teknik regularisasi L_1 dilakukan dengan menggunakan teknik optimisasi umum Nelder-Mead. Pada model linier terampat sebaran Gamma, nilai awal parameter diduga melalui teknik *iterative reweighted least square* (IRWLS), sedangkan pada model linier sebaran pareto terampat nilai awal diduga menggunakan metode IRWLS dan $\frac{\sqrt{6\ var(y)}}{\pi}$. Teknik optimisasi Nelder-Mead pada pemodelan linier terampat sebaran Gamma berhasil mendapatkan penduga parameter yang konvergen, tetapi pada pemodelan linier sebaran pareto terampat penduga parameter tidak konvergen ke parameter sebenarnya dengan menggunakan data simulasi.

Kata kunci: regularisasi L_1, *statistical downscaling*, model linier terampat sebaran Gamma, model linier sebaran pareto terampat, curah hujan ekstrim

SUMMARY

AGUS MOHAMAD SOLEH. Judul Karya Ilmiah Maksimum Tiga Baris, Lima Belas Kata tidak Termasuk Kata Depan dan Kata Sambung dalam Bahasa Inggris. Supervised by PEMBIMBING 1, PEMBIMBING 2 and PEMBIMBING 3.

Statistical Downscaling (SDS) modeling is a technique in climatology that uses statistical model to analyze the relationship between large-scale data (global) and small-scale (local) data. SDS models might involve large-scale ill-conditioned covariates (not independent/high correlation). Techniques such as dimensional reduction, selection, and shrinkage could be use to solve this problems. L_1 regularization is a technique for selection and shrinkage was proposed by Tibshirani (1996). This research is about the development and the use of L_1 regularization technique on linear model to obtain a solution for ill-conditioned covariates problem faced in SDS modeling. Covariates were taken from the output of CMIP5 and the GPCP version 2.2 in the 7×7 gridded domain above Indramayu. Linear modeling based on distribution was used in this research using normal, Gamma and generalized pareto distribution.

textbfKeywords: L_1 regularization, *statistical downscaling*, generalized linear model with Gamma distribution, generalized pareto distribution linear model, monthly extreme rainfall

JUDUL KARYA ILMIAH MAKSIMUM TIGA BARIS, LIMA BELAS KATA TIDAK TERMASUK KATA DEPAN DAN KATA SAMBUNG

AGUS MOHAMAD SOLEH

Disertasi
sebagai salah satu syarat untuk memperoleh gelar
Doktor
pada
Program Studi Statistika

SEKOLAH PASCASARJANA
INSTITUT PERTANIAN BOGOR
BOGOR
2015

Penguji pada Ujian Tertutup:	Dr Ir Penguji 1, DEA Dr Penguji 2, MSi
Penguji pada Sidang Promosi:	Dr Ir Penguji 1, DEA Dr Penguji 2, MSi

Judul Disertasi	:	Judul Karya Ilmiah Maksimum Tiga Baris, Lima Belas Kata tidak Termasuk Kata Depan dan Kata Sambung
Nama	:	Agus Mohamad Soleh
NIM	:	G161100021

Disetujui oleh

Komisi Pembimbing

Dr Ir Pembimbing 1, MSc
Ketua

Dr Ir Pembimbing 2, MS
Anggota

Prof Dr Ir Pembimbing 3, MSc
Anggota

Diketahui oleh

Ketua Program Studi Statistika

Dr Ir I KaProdi, MSi

Dekan Sekolah Pascasarjana

Dr Ir Dekan SPs, MScAgr

Tanggal Ujian: 11 September 2015

Tanggal Lulus: 11 September 2015

PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala karunia-Nya sehingga disertasi ini dapat diselesaikan dengan baik. Tema yang dipilih dalam penelitian ini adalah pengembangan metode pemodelan linier, dengan judul ”Judul Karya Ilmiah Maksimum Tiga Baris, Lima Belas Kata tidak Termasuk Kata Depan dan Kata Sambung .

Terima kasih penulis ucapkan kepada Bapak Dr Ir Pembimbing 1, MSc, Ibu Dr Ir Pembimbing 2, MS dan Bapak Prof Dr Ir Pembimbing 3, MSc selaku pembimbing.

Semoga karya ilmiah ini bermanfaat.

Bogor,   September 2015

Agus Mohamad Soleh

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR ISTILAH

AKU	:	Analisis Komponen Utama (*Principal Component Analysis*)
GCM	:	*General Circulation Models*

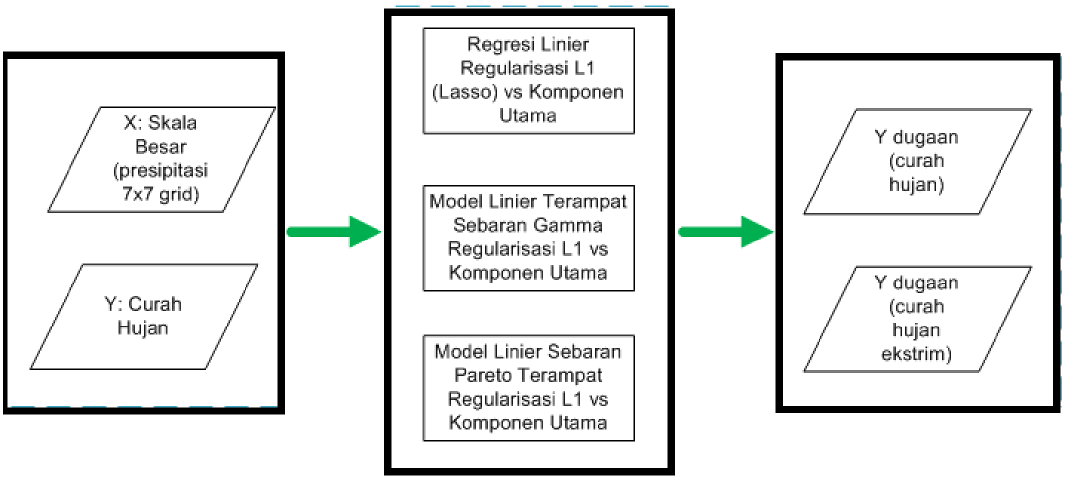
x

1 PENDAHULUAN

Latar Belakang

Pemodelan *Statistical Downscaling* (SDS) merupakan suatu teknik dalam klimatologi yang menggunakan pemodelan statistika untuk menganalisis hubungan antara data skala besar (global) dengan data skala kecil (lokal) (Benestad *et al.* 2008). Metode ini relatif baru walaupun perkembangan penerapan pendugaan menggunakan statistika dalam klimatologi sudah dimulai sejak Klein tahun 1948 (dalam Benestad *et al.* (2008)). Penggunaan istilah SDS merujuk pada keberadaan model iklim global yang direpresentasikan dengan luaran *General Circulation Model* (GCM) sebagai representasi data skala besar untuk pendugaan kasar iklim lokal seperti curah hujan pada suatu wilayah yang merepresentasi data skala kecil.

GCM merupakan model numerik yang menghasilkan sejumlah data dari berbagai parameter iklim seperti presipitasi, temperatur, dan kelembaban untuk keperluan pendugaan iklim. Model GCM merepresentasikan cara kemungkinan terbaik mensimulasi kondisi iklim skala-besar dan memproyeksikan perubahan iklim ke depan akibat pengaruh kekuatan (*forcing*) yang diketahui seperti pengaruh gas rumah kaca. Kemampuan untuk menduga skala kecil menggunakan GCM sangat terbatas karena resolusi spasial dalam GCM umumnya kasar ($\pm$ 300 km $\times$ 300 km). Kerangka penelitian disajikan pada Gambar 1.1.



Gambar 1.1 Kerangka penelitian yang dilakukan

Kebaruan/*Novelty*

Penelitian ini dengan ”Judul Karya Ilmiah Maksimum Tiga Baris, Lima Belas Kata tidak Termasuk Kata Depan dan Kata Sambung ” memiliki kebaruan sebagai berikut:

1. Kajian 1, yaitu oleh Hammami *et al.* (2012) dan Gao *et al.* (2014).
2. Pengembangan metode

2 JUDUL BAB 2

Pendahuluan

Pendahuluan

Sub-Sub Bab

Tabel nilai RMSE disajikan pada Tabel 2.1

Tabel 2.1 Nilai RMSE pendugaan model linier untuk masing-masing ZOM

Model	GPD	RKU	RKU Dummy	Gamma-KU	Gamma-KU Dummy
ZOM 77	121.58	117.19	114.05	115.05	114.23
ZOM 78	114.22	110.12	105.88	105.20	105.66
ZOM 79	81.10	79.04	75.94	77.86	75.94

3 JUDUL BAB 3

Pendahuluan

Pendahuluan

Tinjauan Pustaka

Teknik lasso (*least absolute shrinkage and selection operator*) yang bertujuan mengatasi masalah dalam keakuratan pendugaan dan interpretasi dengan mempertahankan keuntungan-keuntungan metode regresi bertatar (*stepwise*) dan regresi gulud (*ridge*) dikembangkan oleh Tibshirani (1996). Pada regresi linier ganda, teknik lasso meminimumkan jumlah kuadrat sisaan dengan memberikan penalti L_1 pada koefisien parameternya. Misalkan terdapat vektor input $\mathbf{X}^T = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\mathrm{p})$ digunakan untuk memprediksi luaran nilai $\mathbf{Y}$ yang berupa bilangan riil. Model regresi linier memiliki bentuk:

$$f(\mathbf{X}) = \beta_0 + \sum_{j=1}^{p} \mathbf{x}_\mathrm{j}\beta_\mathrm{j} \tag{3.1}$$

Untuk menduga $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_\mathrm{p})^T$, metode kuadrat terkecil meminimumkan jumlah kuadrat sisaan (Hastie *et al.* 2008), yaitu dengan meminimumkan persamaan:

$$\mathrm{JKS}(\boldsymbol{\beta}) = \sum_{i=1}^{N} (y_i - f(\mathbf{x}_\mathrm{i}))^2 = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \tag{3.2}$$

yang dapat ditulis dalam catatan matriks, dengan $\mathbf{X}$ berukuran $N \times (p+1)$ dan $\boldsymbol{y}$ adalah vektor-N, sebagai :

$$\mathrm{JKS}(\boldsymbol{\beta}) = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})^T (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}). \tag{3.3}$$

$\mathrm{JKS}(\boldsymbol{\beta})$ minimum didapatkan dengan cara mendiferensialkan $\mathrm{JKS}(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ secara kalkulus, yang menghasilkan persamaan dalam bentuk:

$$\mathbf{X}^T\boldsymbol{y} = \mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \tag{3.4}$$

yang disebut sebagai persamaan normal.

Jika $\mathbf{X}^T\mathbf{X}$ adalah matriks berpangkat penuh, maka dugaan $\boldsymbol{\beta}$ akan menghasilkan solusi unik, yaitu:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T\boldsymbol{y}. \tag{3.5}$$

Apabila $\mathbf{X}^T\mathbf{X}$ tidak berpangkat penuh atau mendekati singular, maka $\hat{\boldsymbol{\beta}}$ yang diperoleh menjadi tidak stabil. Regresi gulud diperkenalkan oleh Hoerl dan

Kennard (1970) (dalam Draper dan Smith (1998)) diusulkan sebagai salah satu metode untuk menangani ketidakstabilan penduga kuadrat terkecil ini. Regresi gulud memberikan penalti koefisien regresi dalam norm L_2 atau secara spesifik menduga $\boldsymbol{\beta}$ dengan meminimumkan JKS$(\boldsymbol{\beta})$ dengan kendala:

$$\sum_{j=1}^{p} \beta_j^2 \leq t, \qquad t \geq 0. \tag{3.6}$$

Masalah regresi gulud ini dapat ditulis dengan cara lain dalam bentuk persamaan lagrange yaitu memininumkan jumlah kuadrat sisaan terkendala:

$$\text{JKS}(\boldsymbol{\beta},\lambda) = (\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_2^2 \qquad \lambda \geq 0. \tag{3.7}$$

Solusi regresi gulud didapat dengan cara yang sama seperti pada metode kuadrat terkecil, yaitu dengan meminimumkan jumlah kuadrat sisaan JKS$(\boldsymbol{\beta},\lambda)$ sehingga memperoleh persamaan dalam bentuk:

$$\mathbf{X}^T\boldsymbol{y} = (\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I})\boldsymbol{\beta}. \tag{3.8}$$

Dengan cara ini dapat dijamin $(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I})$ selalu berpangkat penuh walaupun $\mathbf{X}^T\mathbf{X}$ tidak berpangkat penuh dengan cara mengambil $\lambda \geq 0$. Untuk $\lambda = 0$ persamaan ini adalah persamaan normal seperti yang diperoleh menggunakan metode kuadrat terkecil. Solusi yang unik dapat diperoleh dalam bentuk tertutup:

$$\hat{\boldsymbol{\beta}}^{gulud} = \left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\boldsymbol{y}. \tag{3.9}$$

Penduga koefisien yang diperoleh menggunakan metode regresi gulud tidak *equivariant* (Hastie *et al.* 2008), artinya penduga koefisien tersebut tidak dapat diperbandingkan hasilnya jika peubah asal tidak dibakukan. Oleh karena itu untuk pendugaan $\hat{\boldsymbol{\beta}}^{gulud}$ ini sebelumnya disarankan untuk membakukan skala dari peubah asal sehingga memiliki nilai harapan nol dan ragam satu (Hastie *et al.* 2008). Penduga koefisien regresi hasil dari regresi gulud akan disusutkan ke arah nol seiring dengan peningkatan nilai λ. Tetapi, penyusutan ini tidak dapat dilakukan untuk seleksi peubah secara otomatis dikarenakan secara simultan koefisien yang diduga mungkin tidak bernilai nol.

Tibshirani (1996) mengembangkan metode lasso yang mengubah kendala dalam regresi gulud menjadi dalam bentuk norm L_1, yaitu: $\sum_{j=1}^{p}|\beta_i| \leq t$ atau disebut juga dengan istilah regularisasi L_1. Solusi dari lasso yang dituliskan dalam bentuk persamaan lagrange adalah meminimumkan:

$$\text{JKS}(\boldsymbol{\beta},\lambda) = (\boldsymbol{y}-\beta_0-\mathbf{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\beta_0-\mathbf{X}\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1. \tag{3.10}$$

Untuk mendapatkan solusi penduga koefisien tidak dapat diperoleh dalam bentuk tertutup, tetapi harus menggunakan pemrograman kuadratik (Tibshirani 1996) yang merupakan bagian dari optimisasi *convex* Boyd dan Vandenberghe (2004). Dampak yang terjadi dari pengubahan kendala ini sangat besar, yaitu menyebabkan koefisien menyusut ke arah nol seperti dalam regresi gulud dan beberapa koefisien menghasilkan nilai nol secara tepat.

Ide dasar metode lasso berasal dari *Non-negative Garrotte* (Breiman 1995) yang meminimumkan fungsi berikut terhadap $c = c_j$:

$$\sum_{i=1}^{N}(y_i - \sum_{j=1}^{p} c_j x_{ij}\hat{\beta}_j)^2 \qquad \text{dengan kendala} \qquad c_j \geq 0, \sum_{j=1}^{p} c_j \leq t, \qquad (3.11)$$

dalam hal ini $\hat{\beta}_j$ adalah penduga kuadrat terkecil biasa. Metode NN-Garrotte ini tidak terdefinisikan ketika $p > N$ (yang bukan merupakan topik panas pada tahun 1995) (Tibshirani 2011). Pada sekitar tahun tersebut, beberapa metode yang mirip dengan lasso telah dikembangkan berdasarkan penalti L_1, seperti *bridge regression* oleh Frank dan Friedman tahun 1993 dan *basis pursuit* oleh Chen *et al.* (1998) (dalam Tibshirani (2011)). Setelah publikasi pertama tahun 1996 sampai tahun 2002, makalah metode lasso dengan pendekatan pemrograman kuadratik ini tidak mendapatkan perhatian. Tetapi setelah tahun 2002, metode lasso mulai menjadi perhatian setelah dikembangkan algoritma lar (*Least Angle Regresion*) oleh Efron, Hastie, Johnstone dan Tibshirani yang dipublikasikan tahun 2004 (Tibshirani 2011).

Efron *et al.* (2004) mengembangkan algoritma lar yang digunakan untuk menduga model regresi linier dalam bentuk model umum:

$$\boldsymbol{E}(Y|\mathbf{X} = \mathbf{x}) = f(\mathbf{x}) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_M\phi_1(\mathbf{x}) + \boldsymbol{\beta}_M\phi_2(\mathbf{x}) + \ldots + \boldsymbol{\beta}_M\phi_M(\mathbf{x}), \qquad (3.12)$$

dalam hal ini ϕ_M adalah fungsi nonlinier dari prediktor $\mathbf{X}$ asli. Modifikasi dari lar untuk lasso menghasilkan efisiensi algoritma dalam menduga solusi penduga koefisien lasso dengan komputasi yang lebih cepat dibandingkan pemrograman kuadratik. Selain untuk menduga koefisien lar dan lasso, algoritma lar ini juga dimodifikasi untuk digunakan dalam menduga koefisien regresi *forward stagewise* dan regresi bertatar, sehingga kemudian namanya dikenal sebagai lars (untuk lar, lasso, *forward stagewise* dan regresi bertatar).

Pendugaan Parameter dengan Metode Iterasi

Perhatikan kembali permasalahan lasso sebagai berikut:

$$\arg\min_{\beta_k} \left\{ (\mathbf{y} - \beta_0 - \sum_{k=1}^{p} \beta_k \mathbf{x}_k)^T (\mathbf{y} - \beta_0 - \sum_{k=1}^{p} \beta_k \mathbf{x}_k) + \lambda \sum_{k=1}^{p} |\beta_k| \right\}.$$

Misalkan $f(\beta_k, \lambda) = (\mathbf{y} - \sum_{k=1}^{p} \beta_k \mathbf{x}_k)^T (\mathbf{y} - \sum_{k=1}^{p} \beta_k \mathbf{x}_k) + \lambda \sum_{k=1}^{p} |\beta_k|$, solusi dari lasso untuk setiap β_j diperoleh dengan mendiferensialkan $f(\beta_k, \lambda)$ terhadap β_j sama dengan nol yang akan memberikan teorema sebagai berikut.

Teorema 1: Terdapat λ_j yang membuat $\hat{\beta}_j$ bernilai nol, yaitu: $\lambda_j \geq |2x_j^T r_{-j}|$, dalam hal ini $r_{-j} = y - \sum_{k \neq j} \hat{\beta}_k x_k$.

Bukti.

$$\frac{\partial}{\partial \beta_j} f(\beta_k, \lambda) = 0 = \frac{\partial}{\partial \beta_j}\left\{ \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\sum_{k=1}^{p}\beta_k\mathbf{x}_k + (\sum_{k=1}^{p}\beta_k\mathbf{x}_k)^T(\sum_{k=1}^{p}\beta_k\mathbf{x}_k) + \lambda\sum_{k=1}^{p}|\beta_k| \right\}$$

$$= -2\mathbf{y}^T\mathbf{x}_j + 2\mathbf{x}_j^T\sum_{k=1}^{p}\beta_k\mathbf{x}_k + \lambda\ \mathrm{sign}(\beta_j)$$

$$= -\mathbf{x}_j^T\mathbf{y} + \mathbf{x}_j^T\sum_{k=1}^{p}\beta_k\mathbf{x}_k + \frac{\lambda}{2}\ \mathrm{sign}(\beta_j)$$

$$= \mathbf{x}_j^T\left(\sum_{k=1}^{p}\beta_k\mathbf{x}_k - \mathbf{y}\right) + \frac{\lambda}{2}\ \mathrm{sign}(\beta_j)$$

$$= \beta_j\mathbf{x}_j^T\mathbf{x}_j + \mathbf{x}_j^T\left(\sum_{k\neq j}\beta_k\mathbf{x}_k - \mathbf{y}\right) + \frac{\lambda}{2}\ \mathrm{sign}(\beta_j)$$

$$= \beta_j\mathbf{x}_j^T\mathbf{x}_j - \mathbf{x}_j^T\left(\mathbf{y} - \sum_{k\neq j}\beta_k\mathbf{x}_k\right) + \frac{\lambda}{2}\ \mathrm{sign}(\beta_j)$$

Notasi: $\mathbf{x}_j^T\mathbf{x}_j = \|\mathbf{x}_j\|^2$, sehingga:

$$= \beta_j\|\mathbf{x}_j\|^2 - \mathbf{x}_j^T\left(\mathbf{y} - \sum_{k\neq j}\beta_k\mathbf{x}_k\right) + \frac{\lambda}{2}\ \mathrm{sign}(\beta_j)$$

$$= \beta_j - \frac{\mathbf{x}_j^T\left(\mathbf{y} - \sum_{k\neq j}\beta_k\mathbf{x}_k\right)}{\|\mathbf{x}_j\|^2} + \frac{\lambda\ \mathrm{sign}(\beta_j)}{2\|\mathbf{x}_j\|^2}$$

Misalkan $\mathbf{r}_{-j} = \mathbf{y} - \sum_{k\neq j}\beta_k\mathbf{x}_k$, maka penduga dari β_j adalah:

$$\hat{\beta_j} = \frac{\mathbf{x}_j^T\mathbf{r}_{-j}}{\|\mathbf{x}_j\|^2} - \frac{\lambda}{2\|\mathbf{x}_j\|^2}\ \mathrm{sign}(\beta_j)$$

Perhatikan λ dan $\|\mathbf{x}_j\|^2$ selalu positif, sedangkan $\mathbf{x}_j^T\mathbf{r}_{-j}$ searah tandanya dengan koefisien β_j.

Perhatikan daerah sebagai berikut:

- $\frac{\mathbf{x}_j^T\mathbf{r}_{-j}}{\|\mathbf{x}_j\|^2} > \frac{\lambda}{2\|\mathbf{x}_j\|^2} \Rightarrow \mathrm{sign}(\beta_j)$ bernilai +. Hal ini berimplikasi

$$\hat{\beta_j} = \frac{\mathbf{x}_j^T\mathbf{r}_{-j}}{\|\mathbf{x}_j\|^2} - \frac{\lambda}{2\|\mathbf{x}_j\|^2}$$

- $\frac{\mathbf{x}_j^T\mathbf{r}_{-j}}{\|\mathbf{x}_j\|^2} < -\frac{\lambda}{2\|\mathbf{x}_j\|^2} \Rightarrow \mathrm{sign}(\beta_j)$ bernilai -. Hal ini berimplikasi

$$\hat{\beta_j} = \frac{\mathbf{x}_j^T\mathbf{r}_{-j}}{\|\mathbf{x}_j\|^2} + \frac{\lambda}{2\|\mathbf{x}_j\|^2}$$

- $0 < \frac{\mathbf{x}_j^T \mathbf{r}_{-j}}{\|\mathbf{x}_j\|^2} < \frac{\lambda}{2\|\mathbf{x}_j\|^2} \Rightarrow$ sign(β_j) bernilai +. Hal ini berimplikasi sign($\hat{\beta_j}$) memiliki tanda - yang berkebalikan dengan sign(β_j). Oleh karena itu, maka $\hat{\beta_j}$ secara asimtotik sama dengan 0

- $-\frac{\lambda}{2\|\mathbf{x}_j\|^2} < \frac{\mathbf{x}_j^T \mathbf{r}_{-j}}{\|\mathbf{x}_j\|^2} < 0 \Rightarrow$ sign(β_j) bernilai -. Hal ini berimplikasi sign($\hat{\beta_j}$) memiliki tanda + yang berkebalikan dengan sign(β_j). Oleh karena itu, maka $\hat{\beta_j}$ secara asimtotik sama dengan 0.

Sehingga $\lambda_j \geq |2x_j^T r_{-j}|$ akan membuat $\hat{\beta_j}$ bernilai nol.

Teorema 2: Terdapat λ minimum yang membuat semua $\hat{\beta}_j$ bernilai nol, yaitu $\lambda = 2max(|x_j^T y|)$.

Bukti. Berdasarkan fakta bahwa $\lambda_j \geq |2x_j^T r_{-j}|$ menghasilkan $\hat{\beta}_j$ bernilai nol dan jika semua $\hat{\beta}_j$ bernilai nol maka $r_{-j} = y$. Maka λ minimum adalah sebesar $\max \lambda_j = 2max(|x_j^T y|)$.

Dari pembuktian Teorema 1 diperoleh solusi dari dari lasso sebagai berikut:

$$\hat{\beta}_j = \begin{cases} \frac{\mathbf{x}_j^T \mathbf{r}_{-j}}{\|\mathbf{x}_j\|^2} - \frac{\lambda}{2\|\mathbf{x}_j\|^2} & ,\lambda < 2x_j^T r_{-j} \\ \frac{\mathbf{x}_j^T \mathbf{r}_{-j}}{\|\mathbf{x}_j\|^2} + \frac{\lambda}{2\|\mathbf{x}_j\|^2} & ,-\lambda > 2x_j^T r_{-j} \\ 0 & ,\lambda \geq |2x_j^T r_{-j}| \end{cases} \tag{3.13}$$

Perhatikan solusi dari lasso tidak dapat dilakukan dengan cara langsung menggunakan formula tersebut, tetapi harus dilakukan secara iterasi karena masing-masing $\hat{\beta}_j$ tergantung pada $\hat{\beta}_k$ lain. Algoritma untuk solusi iterasi kemudian diusulkan sebagai berikut:

1. Bakukan kovariat (X)
2. Tetapkan i=0, $\hat{\beta}^0 = 0$
3. Untuk $\lambda = 0$ sampai $\lambda = 2max(|x_j^T y|)$
   (a) i = i+1
   (b) Untuk j=1 sampai p
      i. Hitung $r_{-j} = y - \sum_{k \neq j} \beta_k^{i-1} x_k$
      ii. Hitung $\hat{\beta}^j$ menggunakan formula pada Persamaan 2.13.
   (c) ulangi (a) dan (b) sampai $(\hat{\beta}^i - \hat{\beta}^{i-1}) < ie^{-6}$
4. Penduga akhir $\hat{\beta}$ ditentukan dengan pendekatan validasi silang.

4 JUDUL BAB 4

Pendahuluan

5 JUDUL BAB 5

Pendahuluan

6 JUDUL BAB 6

7 SIMPULAN DAN SARAN

Simpulan

DAFTAR PUSTAKA

Acero FJ, Garcia JA, Gallego MC. 2010. Peaks-over-Threshold Study of Trends in Extreme Rainfall over the Iberian Peninsula. *J Climate* 24:10891105.

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P. 2003. The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *J Hydrometeor*, 4:1147-1167.

Aldrian E, Susanto RD. 2003. Identification of Three Dominant Rainfall Regions within Indonesia and Their Relationship to Sea Surface Temperature. *Int J Climatol* 23:1435-1452.

As-syakur AR, Prasetia R. 2010. Pola Spasial Anomali Curah Hujan Selama Maret Sampai Juni 2010 di Indonesia; Komparasi Data TRMM Multisatellite Precipitation Analysis (TMPA) 3B43 dengan Stasiun Pengamat Hujan. Di dalam: Penelitian Masalah Lingkungan di Indonesia. Prosiding Seminar Ilmiah Tahunan; Denpasar, 29 Juli 2010. Denpasar: Program Magister Ilmu Lingkungan UNUD. hlm 505-515.

Begueria S, Vicente-Serrano SM. 2006. Mapping the Hazard of Extreme Rainfall by Peaks over Threshold Extreme Value Analysis and Spatial Regression Techniques. *J Appl Meteor Climatol* 45:108124.

Benestad RE, Chen D, Hanssen-Bauer I. 2008. *Empirical-Statistical Downscaling*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Breiman L. 1995. Better Subset Regression Using the Nonnegative Garrote. *J Technometrics* 37: 373-384.

Boyd S, Vandenberghe L. 2004. *Convex Optimization*. New York (USA): Cambridge University Press.

Buono A, Faqih A, Boer R, Santikayasa IP, Ramadhan A, Muttqien MR, Agmalaro MA. 2010. A Neural Network Architecture for Statistical Downscaling Technique: A Case Strudy in Indramayu District. Di Dalam: *The Quality Information for Competitive Agricultural Based Production System and Commerce. Proceeding AFITA 2010 International Conference*; Bogor, 4-7 Oktober 2010. Bogor: Indonesian Society for Agricultural Information. hlm 99-104.

Busuioc A, Tomozeiu R, Cacciamani C. 2008. Statistical downscaling model based on canonical correlation analysis for winter extreme precipitation events in the Emilia-Romagna region. *Int J Climatol* 28:449-464.

Coles S. 2001. *An Introduction to Statistical Modeling of Extremes Values*. London (UK): Springer-Verlag.

Djuraidah A, Wigena AH. 2011. Regresi Kuantil untuk Eksplorasi Pola Curah Hujan di Kabupaten Indramayu. *J Ilmu Dasar* 12:50-56.

Dobson AJ. 2002. *An Introduction to Generalized Linear Models*. Ed ke-2. Washington DC (USA): Chapman & Hall/CRC.

Draper NR, Smith H. 1998. *Applied Regression Analysis*. Ed. ke-3. New York (USA): John Wiley & Sons Inc.

Efron B, Hastie T, JohnstoneI, Tibshirani R. 2004. Least Angle Regression. *Ann Statist* 32:407-840.

Faraway JJ. 2006. *Extending the Linear Model with R*. London: Chapman & Hall/CRC Taylor & Francis Group.

Friederichs P. 2010. Statistical downscaling of extreme precipitation events using extreme value theory. *Extremes* 13:109-132.

Friederichs P, Hense A. 2007. Statistical Downscaling of Extreme Precipitation Events Using Censored Quantile Regression. *Mon Wea Rev* 135:23652378.

Friedman J, Hastie T, Tibshirani R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Soft* [internet]. [diunduh 2014 Mar 14]; 33:1-22. Tersedia pada: http://www.jstatsoft.org/v33.

Gao L, Schulz K, and Bernhardt M. 2014. Statistical Downscaling of ERA-Interim Forecast Precipitation Data in Complex Terrain Using LASSO Algorithm. *Advances in Meteorology* 2014: http://dx.doi.org/10.1155/2014/472741.

Hammami D, Lee TS, Ouarda TBMJ, Lee J. 2012. Predictor selection for downscaling GCM data with LASSO. *J Geophys Res* 117. doi:10.1029/2012JD017864.

Haryoko U. 2015. Pewilayahan Hujan untuk Menentukan Pola Hujan. Contoh Kasus Kabupaten Indramayu. Badan Meteorologi Dan Geofisika. http://www.staklimpondokbetung.net/publikasi/PengelompokanPolaHujan.pdf. Diakses tanggal 11 Mei 2015.

Hastie T, Tibshirani R, Friedman J. 2008. *The Elements of Statistical Learning. Data mining, Inference, and Prediction*. Ed. ke-2 [internet]. [diunduh 2014 Mar 14]; Springer. Tersedia pada: http://www.stanford.edu/~hastie/pub.htm.

Krishnamoorthy K. 2006. *Handbook of Statistical Distributions with Applications*. New York (USA): Chapman & Hall/CRC.

Koenker R. 2005. *Quantile Regression*. New York (USA): Cambridge University Press.

Mannshardt-Shamseldin EC, Smith RL, Stephan SR, Mearns LO, Colley D. 2010. Downscaling Extremes: A Comparison Of Extreme Value Distributions In Point-Source And Gridded Precipitation Data. *Ann Appl Stat* 4: 484-502.

McCullagh P, Nelder JA. 1989. *Generalized Linear Models*. Ed ke-2. London (UK): Chapman & Hall/CRC.

Mondiana YQ. 2012. Pemodelan *Statistical Downscaling* dengan Regresi Kuantil untuk Pendugaan Curah Hujan Ektrim; Studi Kasus Stasiun Bangkir Kabupaten Indramayu) [tesis]. Bogor (ID): Institut Pertanian Bogor.

Olsson J, Uvo CB, Jinno K. 2001. Statistical Atmospheric Downscaling of Short-Term Extreme Rainfall By Neural Networks. *Phys Chem Earth (B)* 26: 695-700.

Park MY, Hastie T. 2007. L1-regularization Path Algoritm for Generalized Linear Models. *J R Statisc Soc (B)* 69:659-677.

Park T, Casella G. 2008. The Bayesian Lasso. *J Amer Statist Assoc* 103:681-686.

Soleh AM, Aunuddin. 2013. LASSO : Solusi Alternatif Seleksi Peubah dan Penyusutan Koefisien Model Regresi Linier. *FSK Indones J Statist* 18:21-27.

Stephenson DB, Kumar KR, Doblas-Reyes FJ, Royer JF, Chauvin E, Pezzulli S. 1999. Extreme Daily Rainfall Events and Their Impact on Ensemble Forecasts of the Indian Monsoon. *Monthy Weather Review* 127:1954-1966.

Sutikno. 2008. *Statistical Downscaling* Luaran GCM dan Pemanfaatannya untuk Peramalan Produksi Padi [disertasi]. Bogor (ID): Institut Pertanian Bogor.

Sutikno, Setiawan, Purnomoadi H. 2010. Statistical Downscaling Output GCM Modeling with Continuum Regression and Pre-Processing PCA Approach. *IPTEK J Tech Sci 21(3)*: 109-117.

Taylor KE, Stouffer RJ, Meehl GA. 2012. An Overview of CMIP5 and the Experiment Design. *Bull Amer Meteor Soc* 93: 485-498.

Tibshirani R. 1996. Regression Shrinkage and Selection via The LASSO. *J R Statist Soc (B)* 58: 267-288.

Tibshirani R. 2011. Regression Shrinkage and Selection via The LASSO: a retrospective. *J R Statist Soc (B)* 73:273-282.

Tryhorn L, DeGaetanoa A. 2011. A comparison of techniques for downscaling extreme precipitation over the Northeastern United State. *Int J Climatol* 31: 19751989.

Vimont DJ, Battisti DS, Naylor RL. 2010. Downscaling Indonesian precipitation using large-scale meteorogical fields. *Int J Climatol* 30:1706-1722.

Wigena AH. 2006. Pemodelan *Statistical Downscaling* dengan Regresi Projection Pursuit untuk Peramalan Curah Hujan [disertasi]. Bogor (ID): Institut Pertanian Bogor.

Wigena AH. 2011. Regresi Kuadrat Terkecil Parsial untuk *Statistical Downscaling*. Di dalam: *Prosiding Scientific Jurnal Club BMKG Ed. Ke-6*. Jakarta: Badan Meteorologi Klimatologi dan Geofisika. hlm 10-13.

Yee TW, Stephenson AG. 2007. Vector generalized linear and additive extreme value models. *Extremes* 10:1-19. DOI 10.1007/s10687-007-0032-4

Yee TW, Wild CJ. 1996. Vector Generalized Additive Models. *J R Statist Soc (B)* 73:481-493.

LAMPIRAN

Lampiran 1. Implementasi algoritma metode iterasi regresi linier dengan regularisasi L_1 dalam perangkat lunak komputasi statistik R

```
softthresh <- function(x,y) {
  if (is.na(x) || is.na(y)) stop("argumen x dan y tidak boleh NA")
  if (y < 0) stop("argumen y harus non negatif")
  if (x > y) return(x-y) else {
    if (x < -y) return(x+y) else return(0)
  }
}

sol.beta <- function(x,y,v.lambda,err.max=1e-4, max.it=1e3) {
  xscale <- as.matrix(scale(x,T,T))
  sb <- attr(xscale,"scaled:scale")
  p <- ncol(xscale)
  #inisialisasi
  jkxj <- apply(xscale^2,2,sum)
  v.out <- NULL; beta <- rep(0,p)
  for (lambda in v.lambda) {
    konverg <- F; j <- 0
    while (!konverg && j<=max.it){
      j <- j+1
      betaawal <- beta
      for (i in 1:p) {
        rmj <- y - (xscale[,-i]%*%beta[-i])
        tmp <- (t(xscale[,i]) %*% rmj)
        beta[i] <- softthresh(tmp/jkxj[i],lambda/(2*jkxj[i]))
      }
      e <- (beta-betaawal)
      if (max(e) < err.max) konverg <- T
    }
    if (!konverg & i>= 1e4) print(paste("tidak konvergen pada lambda=",lambda))
    beta <- beta / sb
    v.out <- cbind(v.out,c(lambda,beta))))
  }
  return(v.out)
}
```

RIWAYAT HIDUP

Penulis dilahirkan di Tenpat Lahir.