

**PEMODELAN *STATISTICAL DOWNSCALLING* DENGAN PENDEKATAN
MARKOV CHAIN MONTE CARLO PCA
(STUDI KASUS : DATA GCM STASIUN AMBON)**

Ferry Kondo Lembang

Jurusan Matematika, FMIPA Universitas Pattimura

Jln. Ir. M. Putuhena, Kampus Poka Ambon

ferrykondolembang@yahoo.co.id

ABSTRAK

Masalah mendasar dari prediksi model curah hujan adalah keakuratan model berdasarkan proses stokhastik skala global maupun skala kecil. *Statistical Downscaling* (SD) merupakan salah satu alternatif untuk mengatasi masalah tersebut. SD adalah model yang menghubungkan skala global GCM dengan skala yang lebih kecil (lokal) dengan jalan pra-pemrosesan reduksi dimensi domain grid untuk mengatasi kasus multikolinearitas. Metode reduksi dimensi yang sering digunakan adalah *Principal Component Analysis* (PCA). Dari hasil reduksi dimensi domain grid tersebut selanjutnya diregresikan dengan variabel respon yaitu data curah hujan bulanan kota Ambon periode 1971-2000 dengan metode *Markov Chain Monte Carlo* (MCMC) untuk mendapatkan model SD, kemudian dilakukan validasi model SD dengan menggunakan kriteria kebaikan model RMSEP dan $R^2_{predict}$. Hasil penelitian diperoleh untuk domain grid 3x3 nilai RMSEP dan $R^2_{predict}$ berturut-turut sebesar 245 dan 45,59%, untuk domain grid 8x8 nilai RMSEP dan $R^2_{predict}$ berturut-turut sebesar 242 dan 67,90%. Sedangkan untuk domain grid 12x12 nilai RMSEP dan $R^2_{predict}$ berturut-turut sebesar 246 dan 69,21%.

Kata kunci: GCM, *statistical downscaling*, PCA, Regresi Bayes PCA

PENDAHULUAN

Analisis regresi merupakan analisis statistika yang bertujuan untuk memodelkan hubungan antara variabel bebas (X) dan variabel tidak bebas (Y). Metode *Ordinary Least Square* (OLS) merupakan salah satu metode estimasi parameter yang paling terkenal dalam model regresi karena relatif mudah. Kemudahan tersebut sebagai akibat adanya beberapa asumsi yang cukup ketat antara lain asumsi *error* identik independen dan berdistribusi normal yang harus dipenuhi sehingga akan diperoleh satu model taksiran untuk semua model data serta tidak terjadi kolinearitas ganda antara variabel bebas. Banyak metode estimasi parameter yang digunakan untuk mengatasi adanya multikolinearitas, antara lain: regresi komponen utama, regresi kuadrat terkecil parsial (PLS), regresi *ridge*, serta pendekatan regresi Bayes (Box and Tiao, 1973).

Jika pada metode *Ordinary Least squares* (OLS) parameter regresi (β) diasumsikan konstan, tetapi pada pendekatan Bayes parameter model diasumsikan memiliki sebaran tertentu. Informasi ini disebut informasi prior. *Update* informasi prior pada parameter θ menggunakan informasi sampel yang terdapat dalam data (melalui fungsi *likelihood*), sehingga diperoleh informasi posterior yang akan digunakan untuk pengambilan keputusan (Gelman, dkk., 1995 dalam Prastyo, 2008).

Pada beberapa kasus, korelasi diantara variabel independen terjadi dengan pola yang khusus (tertentu), misalnya pada model curah hujan dengan data luaran GCM. Namun informasi GCM sifatnya global dan tidak berlaku untuk informasi skala kecil, sehingga untuk menjembatani Skala GCM ke Skala Kecil dipakai Teknik *Downscaling* (Wigena, 2006) yang

merupakan teknik pereduksian dimensi. Metode reduksi dimensi dalam pra-pemrosesan yang digunakan antara lain: *Principal Component Analysis* (PCA), Transformasi *Wavelet Diskrit* (TWD) (Anggraeni, 2009), Kernel PCA (Manorang, 2009), dan ROBPCA (Khotimah, 2009). Hasil reduksi dimensi dalam pra-pemrosesan menggunakan *Principal Component Analysis* (PCA) untuk mendapatkan validasi model curah hujan bisa diselesaikan dengan pendekatan metode *Markov Chain Monte Carlo* (MCMC), alasan penggunaan metode ini disebabkan dapat model membentuk model yang sangat kompleks dan berdimensi tinggi walaupun antar variabel prediktor saling berkorelasi tinggi (multikolinearitas). Dalam perspektif statistika permasalahan ini merupakan pemodelan hubungan antara variabel iklim stasiun skala besar dengan komponen utama hasil reduksi pra-pemrosesan PCA. Komponen utama hasil reduksi dimensi PCA dibagi atas 2 komponen yaitu, data *in-sample* untuk mendapatkan model dan data *out-sample* untuk mendapatkan validasi model. Kriteria kebaikan model untuk perbandingan kinerja hasil reduksi dimensi PCA dengan model MCMC adalah RMSEP dan $R^2_{predict}$. Nilai RMSEP merupakan nilai dari *error* hasil taksiran sehingga model terbaik adalah model dengan RMSEP minimum yang menandakan nilai taksiran mendekati nilai sebenarnya sedangkan semakin besar nilai $R^2_{predict}$, maka semakin baik pula model yang didapatkan karena mampu menjelaskan lebih banyak data (Drapper dan Smith, 1996).

Makalah ini membahas pendekatan MCMC untuk pendugaan model regresi yang parameternya saling berkorelasi.

KAJIAN PUSTAKA

Principal Component Analysis (PCA)

Analisis komponen utama merupakan teknik statistik yang dapat digunakan untuk menjelaskan struktur variansi-kovariansi dari sekumpulan variabel melalui variabel baru dimana variabel baru ini saling bebas, dan merupakan kombinasi linier dari variabel asal (Johnson, 2002). Selanjutnya variabel baru ini dinamakan komponen utama (*principal component*). Secara umum tujuan dari analisis komponen utama adalah mereduksi dimensi data dan untuk kebutuhan interpretasi.

Misal Σ merupakan matriks kovariansi dari vektor acak $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ dengan pasangan nilai eigen dan vektor eigen yang saling ortonormal adalah $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$. Dimana $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, maka komponen utama ke- i didefinisikan sebagai berikut:

$$W_i = \mathbf{e}_i' \mathbf{X} = e_{i1} X_1 + e_{i2} X_2 + \dots + e_{ip} X_p, \quad i = 1, 2, \dots, p \quad (1)$$

dimana W_1 adalah komponen pertama yang memenuhi maksimum nilai $\mathbf{e}_1' \Sigma \mathbf{e}_1 = \lambda_1$. W_2 adalah komponen kedua yang memenuhi sisa keragaman selain komponen pertama dengan memaksimumkan nilai $\mathbf{e}_2' \Sigma \mathbf{e}_2 = \lambda_2$. W_p adalah komponen ke- p yang memenuhi sisa keragaman selain W_1, W_2, \dots, W_{p-1} dengan memaksimumkan nilai $\mathbf{e}_p' \Sigma \mathbf{e}_p = \lambda_p$. Urutan W_1, W_2, \dots, W_p harus memenuhi persyaratan $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Sementara itu, proporsi total variansi yang dijelaskan komponen utama ke- k adalah :

$$\left(\begin{array}{l} \text{proporsi total variansi populasi yang} \\ \text{dijelaskan oleh komponen utama ke - } k \end{array} \right) = \frac{\lambda_k}{\text{tr}(\Sigma)} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad (2)$$

Salah satu tujuan dari analisis komponen utama adalah mereduksi dimensi data asal yang semula terdapat p variabel bebas menjadi k komponen utama (dimana $k < p$). Adapun kriteria pemilihan k menurut Johnson (2002) yaitu proporsi kumulatif keragaman data asal yang dijelaskan oleh k komponen utama minimal 80 %, dan proporsi total variansi populasi bernilai cukup besar. Sedangkan menurut Rencher (1998) dengan menggunakan *scree plot* yaitu plot antara i dengan λ_i , pemilihan nilai k berdasarkan *scree plot* ditentukan dengan melihat

terjadinya belokan dengan menghapus komponen utama yang menghasilkan beberapa nilai eigen kecil membentuk pola garis lurus.

Markov Chain Monte Carlo (MCMC)

Untuk menyelesaikan suatu masalah statistik dengan mudah, dengan mengetahui dan memahami tentang distribusi prior “*conjugate*” maka digunakan statistik bayesian. Dalam memodelkan suatu masalah yang timbul dalam formulasi bayesian, tidak sedikit orang mengalami kesulitan. Langkah awal untuk mempermudah kerja adalah berikut ini :

$$p(\beta | y) \propto p(\beta)L(\beta | y) \quad (3)$$

Metode MCMC (*Markov Chain Monte Carlo*) merupakan suatu metode simulasi stokastik yang menirukan pola kerja *Markov Chain* dalam membantu penyelesaian pada estimasi model dengan membangkitkan data sampel berdasarkan skenario sampling tertentu atau pada data yang telah terstandarisasi. Rantai Markov merupakan proses stokastik bersifat khusus, yang berhubungan dengan karakteristik atau sifat dari sederet variabel random. Metode MCMC memiliki kegunaan salah satunya, yaitu membentuk model yang sangat kompleks, berdimensi tinggi, atau sifat data yang berkorelasi tinggi (multikolinearitas).

Jika dideteksi terdapat kasus multikolinearitas yang tinggi pada independent variabel randomnya, maka disarankan untuk menggunakan prior dengan struktur sebagai berikut (Zellner, 1971)

$$\begin{aligned} p(\theta) &= p(\beta_1, \beta_2, \dots, \beta_p, \sigma) \\ &= p(\beta_1 | \beta_2, \dots, \beta_p, \sigma) \times p(\beta_2 | \beta_3, \dots, \beta_p | \sigma) \times \dots \times p(\beta_p | \sigma) \times p(\sigma) \end{aligned} \quad (4)$$

atau dengan menggunakan metode independent sebagai berikut ((Box dan Tiao, 1973), (Charlin dan Chin, 1995), dan (Iriawan, 2003))

$$p(\theta) = p(\beta_1) \times p(\beta_2) \times \dots \times p(\beta_p) \times p(\sigma) \quad (5)$$

General Circulation Model (GCM)

GCM merupakan model matematis yang menggambarkan sirkulasi umum atmosfer di bumi. Model ini menduga perubahan unsur-unsur cuaca dalam bentuk luaran grid-grid yang berukuran 100-500 km menurut lintang dan bujur (von Stroch *et al.* 1993 dalam Sutikno, 2008). GCM merupakan suatu alat penting dalam studi keragaman iklim dan perubahan iklim (Zorita dan Storch, 1999). Namun informasi GCM masih berskala global, sehingga sulit untuk memperoleh langsung informasi berskala lokal dari GCM. Tetapi GCM masih mungkin digunakan untuk memperoleh informasi skala lokal atau regional bila teknik *downscaling* digunakan (Fernandez, 2005 dalam Wigena, 2006).

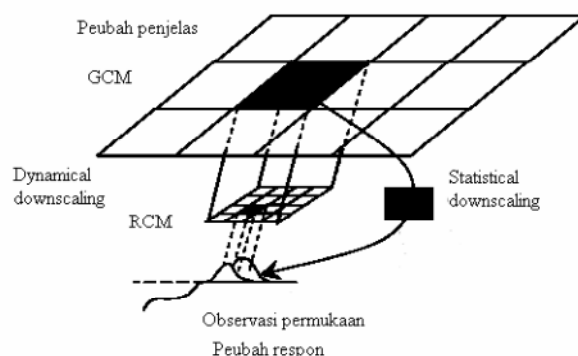
Downscaling adalah suatu metode untuk mendapatkan informasi tentang perubahan iklim atau iklim beresolusi tinggi dari model iklim global yang relatif kasar, dimana model iklim global memiliki resolusi 150-300 km. Akan tetapi, banyak dampak model memerlukan informasi pada skala 50 km atau kurang, sehingga beberapa metode dibutuhkan untuk memperkirakan informasi dengan skala kecil. Untuk menjembatani skala GCM yang besar dengan skala yang lebih kecil (kawasan yang menjadi studi) digunakan teknik *Statistical Downscaling* (SD).

Statistical Downscaling

Statistical Downscaling (SD) adalah suatu proses *downscaling* yang bersifat statik dimana data pada grid-grid berskala besar dalam periode dan jangka waktu tertentu digunakan sebagai dasar untuk menentukan data pada grid berskala lebih kecil (Wigena, 2006). Pendekatan SD menggunakan data regional atau global untuk memperoleh hubungan fungsional antara skala lokal dengan skala global GCM, seperti model regresi. Pendekatan SD disusun berdasarkan adanya hubungan antara grid skala besar (prediktor) dan grid skala lokal (respon) yang dinyatakan dengan model statistik yang dapat digunakan untuk menterjemahkan anomali-anomali skala global menjadi anomali dari beberapa variabel iklim

lokal (Zorita & Storch, 1999). SD menjelaskan hubungan antara skala global dan lokal dengan lebih memperhatikan keakuratan model penduga untuk mempelajari dampak perubahan iklim (Yarnal *et al.*, 2001 dalam Wigena, 2006).

Pendekatan SD menggunakan data GCM untuk peramalan iklim lokal (Fuentes & Heimann, 2000 dalam Wigena, 2006). Dalam pendekatan ini perlu dilakukan pemilihan variabel-variabel yang akan dijadikan sebagai prediktor dan penentuan domain (lokasi dan jumlah grid), karena kedua hal ini merupakan faktor kritis yang akan mempengaruhi kestabilan peramalan (Wilby & Wigley, 2000). Dengan demikian dalam hal peramalan curah hujan, pemilihan variabel prediktor (data GCM) sebaiknya berdasarkan pada adanya korelasi yang kuat antara variabel tersebut dengan curah hujan. Metode SD didasarkan pada asumsi bahwa iklim regional dikendalikan oleh dua faktor yaitu: kondisi iklim skala besar (resolusi rendah) dan kondisi/karakteristik fisiografik regional/lokal (misal: topografi distribusi daratan-lautan dan tataguna lahan) (Von Stroch *et al.*, 1999 dalam Sutikno, 2008).



Gambar 1. Ilustrasi Pemodelan SD (Sumber:Sutikno, 2008)

Secara umum bentuk hubungan tersebut dinyatakan dengan:

$$Y = f(X) + \varepsilon$$

dimana:

$Y_{(t \times p)}$: variabel respon (curah hujan)

$X_{(t \times q \times g \times l)}$: variabel penjelas (luaran GCM)

p : banyaknya variabel y

q : banyaknya variabel x

g : banyaknya grid domain GCM

l : banyaknya level ketinggian

t : banyaknya waktu (misal: harian, dasarian atau bulanan)

ε : error

METODE PENELITIAN

Identifikasi variabel

➤ Variabel respon

Data curah hujan bulanan dengan satuan mili meter (mm) stasiun Ambon (4.66^0 LS, 129.38^0 BT).

➤ Variabel prediktor

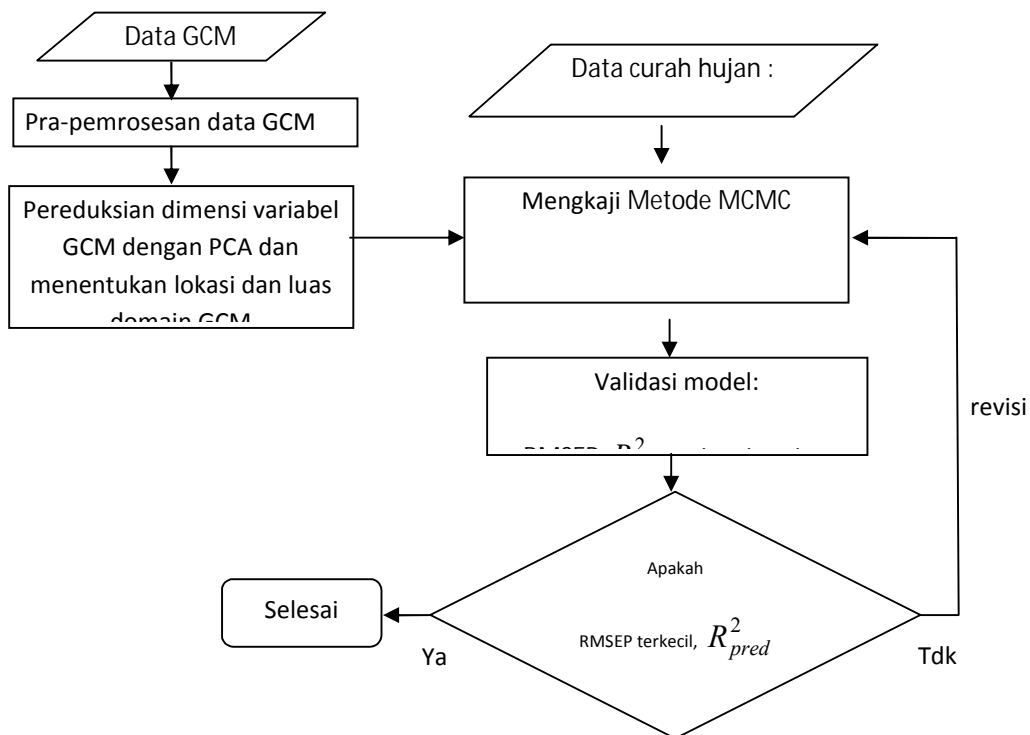
Variabel prediktor yang digunakan pada penelitian ini adalah variabel luaran CSIRO Mk3 yang digunakan meliputi: *precipitable water (prw)*, tekanan permukaan laut (*slp*), komponen angin meridional (*va*), komponen zonal (*ua*), ketinggian geopotensial (*zg*), dan kelembaban spesifik (*hus*). Ketinggian (level) yang digunakan dalam penelitian adalah 850 hPa, 500 hPa, dan 200 hPa (**Tabel 1.**). Data yang dipakai adalah data sekunder hasil reduksi dimensi menggunakan metode PCA.

Tabel 1. Variabel prediktor yang diambil dari model luaran CSIRO-Mk3 dan pengkodingan

No.	Variabel	Satuan	Ketinggian/level			
			Permukaan	850 hPa	500 hPa	200 hPa
1.	<i>Precipitable water</i>	Kg m ⁻²	prw	-	-	-
2.	Tekanan permukaan laut	Pa	slp	-	-	-
3.	Komponen angin meridional	m s ⁻¹	vas	va850	va500	va200
4.	Komponen angin zonal	m s ⁻¹	uas	ua850	ua500	ua200
5.	Ketinggian geopotensial	m	-	zg850	zg500	zg200
6.	Kelembaban spesifik	ltr	huss	hus850	hus500	hus200

Metode Analisis

Adapun tahapan analisis diperlihatkan pada *flowchart* berikut dibawah ini :



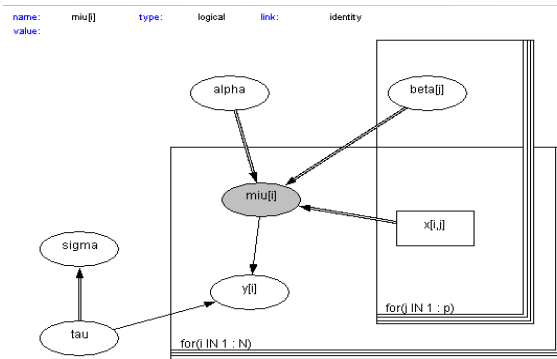
Gambar 2. Tahapan Umum Metode Analisis

HASIL DAN PEMBAHASAN

Estimasi Model SD menggunakan metode MCMC dalam program Winbugs.

Dalam penelitian ini prior untuk setiap parameter dalam program winbugs berdistribusi normal dengan mean dan varians dari masing-masing parameter dari model diduga menggunakan metode OLS. Simulasi stochastik MCMC dari model SD dalam winbugs ditentukan kekonvergenan pada iterasi 1000 kali (Cordas dan Casella, 1996 dalam Sutanto, 2007) dengan burn-in dimulai dari iterasi 2001. Langkah pertama yang dilakukan

adalah menyusun doodle dari data GCM yang sudah di PCA. Gambar 3 berikut menunjukkan struktur doodle general untuk data yang ada.



Gambar 3. Struktur Doodle General

Sedangkan listing program dari doodle diatas untuk masing-masing domain tiap stasiun curah hujan yang menjadi lokasi penelitian dilengkapi dengan data dan inisialnya pada lampiran D-lampiran L.

Dengan menggunakan data *pre-processing* PCA diperoleh estimasi model SD dengan metode *Markov Chain Monte Carlo* yang ditunjukkan pada **Tabel 2** berikut dibawah ini :

Tabel 2. Model SD menggunakan Metode MCMC Stasiun Ambon untuk Domain 3x3, Domain 8x8, dan Domain 12x12

Domain	Model SD menggunakan metode MCMC
3x3 grid	$\hat{Y} = 288,6 + 22,33HUS200 - 17,57HUS500 - 150,0HUS850 - 20,23HUS1 + 140,7PRW - 17,84PSL + 21,28UA200 - 6,112UA500 + 20,76UA850 + 9,297VA200 + 16,71VA500 - 8,781VA850 - 17,56VAS + 6,155ZG200 + - 10,64ZG500 + 14,48ZG850$
8x8 grid	$\hat{Y} = 272,7 - 0,5792HUS200 + 27,95HUS500_1 - 29,72HUS500_2 + - 28,46HUS850_1 + 20,62HUS850_2 + 2,987HUSS_1 - 7,358HUSS_2 + - 34,52HUSS_3 - 8,953PRW_1 + 7,842PRW_2 - 41,63PSL + 9,447UA200 + - 4,282UA500_1 + 12,02UA500_2 + 19,26UA850_1 + 11,06UA850_2 + - 1,912VA200 + 4,19VA500_1 - 2,542VA500_2 - 9,268VA500_3 + 2,18VA850_1 - 2,244VA850_2 + 0,5062VA850_3 - 1,196VAS1 + 8,779VAS2 + 1,151ZG200 - 5,365ZG500 + 38,37ZG850$
12x12 grid	$\hat{Y} = 279,3 - 4,531HUS200_1 - 2,173HUS200_2 + 19,12HUS500_1 + - 16,89HUS500_2 + 3,474HUS500_3 - 11,37HUS850_1 - 8,232HUS850_2 + 0,6502HUSS_1 - 2,891HUSS_2 + 9,142HUSS_3 - 15,41HUSS_4 - 5,40PRW_1 + 15,28PRW_2 - 27,74PSL + 5,17UA200_1 + 4,2UA200_2 - 3,99UA500_1 + - 8,175UA500_2 - 0,9139UA500_3 - 7,716UA850_1 - 4,722UA850_2 + - 2,741VA200_1 + 8,448VA200_2 - 2,525VA500_1 - 0,7073VA500_2 + - 0,1523VA500_3 - 5,476VA500_4 - 3,691VA500_5 + 4,531VA850_1 + 0,6397VA850_2 + 1,341VA850_3 + 0,8417VA850_4 - 2,16VA850_5 + 1,97VAS_1 + 10,76VAS_2 + 11,49VAS_3 + 0,96ZG200 - 0,9171ZG500 + 22,13ZG850$

Model SD yang dihasilkan ini selanjutnya akan divalidasi dengan menggunakan kriteria kebaikan model RMSEP dan $R^2_{predict}$. Pada validasi model ini juga akan ditunjukkan prediksi curah hujan tiap bulan di kota Ambon yang akan dibandingkan dengan curah hujan aktual kota Ambon lewat visualisasi.

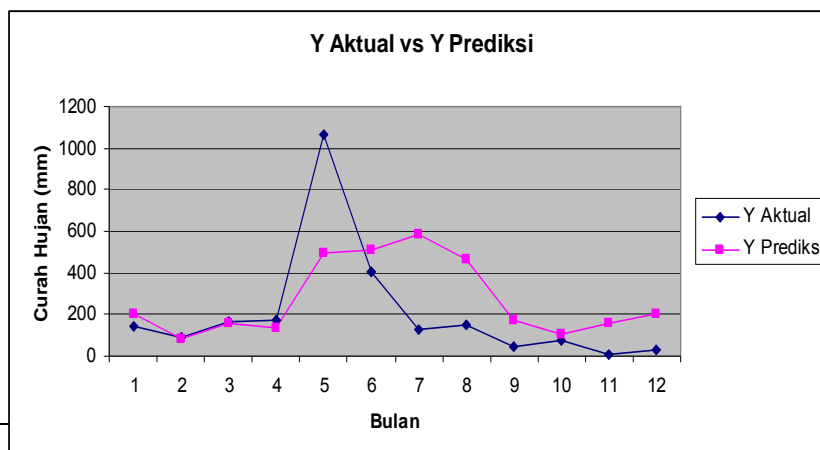
Validasi Model SD menggunakan metode MCMC

Model SD metode MCMC menggunakan fungsi prior berdistribusi normal. Validasi model SD metode MCMC menggunakan kriteria kebaikan model RMSEP, R^2_{pred} , dan nilai dugaan curah hujan berdasarkan domain grid yang telah ditentukan. Hasil validasi model SD ditunjukkan pada tabel 3 dibawah ini :

Tabel 3. Curah Hujan Aktual dan dugaan berdasarkan domain (3x3, 8x8, dan 12x12) dengan metode MCMC stasiun Ambon

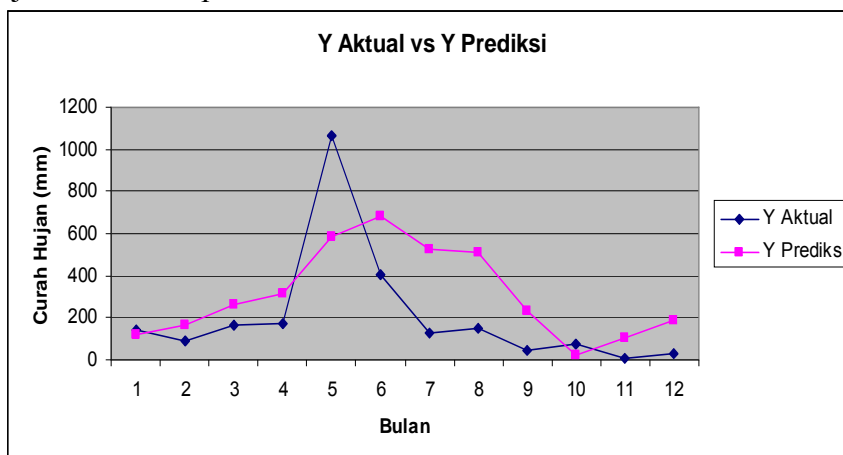
Bulan	Curah hujan Aktual (mm) tahun 2000	Prediksi Curah hujan yang (mm)		
		Domain grid 3x3	Domain Grid 8x8	Domain grid 12x12
Januari	140	206	118	108
Februari	91	79	163	180
Maret	168	160	264	182
April	172	134	313	287
Mei	1068	496	585	588
Juni	404	508	681	710
Juli	125	586	523	495
Agustus	152	464	510	526
September	47	174	235	294
Oktober	72	105	25	76
November	11	158	108	201
Desember	30	205	190	91
RMSEP		245	242	246
R^2_{pred}		45,59%	67,90%	69,21%

Dari **Tabel 3.** diatas terlihat bahwa nilai RMSEP terkecil berada pada domain 8x8 yaitu bernilai sebesar 242, sedangkan R^2_{pred} terbesar pada domain 12x12. Secara umum dapat disimpulkan bahwa semakin luas domain grid tingkat akurasi prediksi curah hujan semakin besar apabila dilihat dari kriteria R^2_{pred} . Sedangkan untuk kriteria RMSEP tidak dapat disimpulkan demikian sebab nilai RMSEP terkecil berada pada domain grid 8x8. berikut dibawah ini akan ditunjukkan visualisasi nilai curah hujan aktual dan nilai prediksi curah hujan.



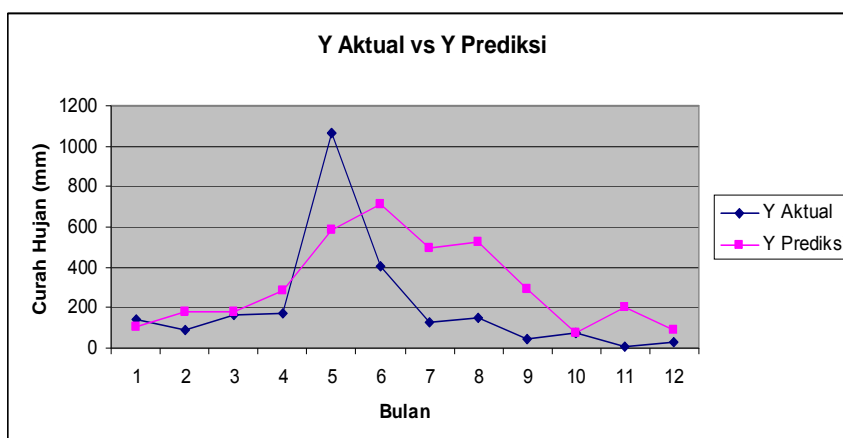
Gambar 4. Curah Hujan Aktual vs Curah hujan Prediksi metode MCMC domain grid 3x3

Pada **Gambar 4.** terlihat bahwa hasil prediksi yang hampir tepat terletak pada bulan Januari-April dan pada bulan Oktober. Untuk bulan lainnya terlihat error yang sangat besar antara curah hujan aktual dan prediksi.



Gambar 5. Curah Hujan Aktual vs Curah hujan Prediksi metode MCMC domain grid 8x8

Pada **Gambar 5.** terlihat bahwa hasil prediksi yang hampir tepat terletak pada bulan Januari dan Oktober. Untuk bulan lainnya terlihat error yang sangat besar antara curah hujan aktual dan prediksi.



Gambar 6. Curah Hujan Aktual vs Curah hujan Prediksi metode MCMC domain grid 12x12

Pada **Gambar 6.** terlihat bahwa hasil prediksi yang hampir tepat terletak pada bulan Januari-April, Oktober, dan Desember. Untuk bulan lainnya terlihat *error* yang sangat besar antara curah hujan aktual dan prediksi.

PENUTUP

Kesimpulan

Berdasarkan hasil dan pembahasan diatas dapat disimpulkan bahwa :

1. Semakin besar domain grid maka semakin baik pula ketepatan validasi model SD untuk kriteria kebaikan model R^2_{pred} . Hal ini ditunjukkan dengan nilai R^2_{pred} berturut-turut 45,59%, 67,90%, dan 69,21%
2. Keakuratan prediksi curah hujan untuk model SD menggunakan metode MCMC dominan pada periode bulan Januari-April dan Bulan Oktober. Sedangkan untuk bulan yang lain tingkat akurasi prediksi belum memuaskan.

Saran

Berdasarkan hasil simpulan diatas maka dapat disarankan beberapa hal sebagai berikut :

1. Penentuan metode dalam memodelkan SD sebaiknya dipilih metode non linear yang sesuai dengan karakteristik data curah hujan.
2. Penentuan kriteria kebaikan model sebaiknya dipilih salah satu saja agar tidak terjadi kesimpangsiuran pengambilan keputusan model mana yang baik.

DAFTAR PUSTAKA

- Anggraeni, S., 2009. *Pra-Pemrosesan Data Luaran GCM (General Circulation Models) CSIRO-Mk3 dengan Metode Transformasi Wavelet Diskrit*. Tugas Akhir. Surabaya: Program Sarjana. Institut Teknologi Sepuluh Nopember.
- Box, G.E.P. and Tiao G.C., 1973. *Bayesian Inference in Statistical Analysis*. Reading Mass. Addison- Wesley Publishing Co. London.
- Draper, N.R. and Smith, H., 1992. *Analisis Regresi Terapan, Edisi kedua*. Jakarta: PT. Gramedia Pustaka Utama.
- Johnson, R.A. and Wichern, D.W., 2002. *Applied Multivariate Statistical Analysis*, 5th End. Prentice Hall. New Jersey.
- Khotimah, K., 2009. *Reduksi Dimensi Robust Dengan Estimator MCD Untuk Pra-Pemrosesan Data Pemodelan Statistical Downscaling*. Tugas Akhir. Program Sarjana, Institut Teknologi Sepuluh Nopember. Surabaya.
- Manorang, J., 2009. *Analisis Komponen Utama Kernel Untuk Pra Pemrosesan Pemodelan Statistical Downscaling*. Tugas Akhir. Program Sarjana. Institut Teknologi Sepuluh Nopember. Surabaya.
- Prastyo, D.D., 2008. *Pemodelan Farmakokinetika Populasi dan Individu Menggunakan Algoritma EM-Nonparametrik dan Analisis Bayesian*. Tesis. Sekolah Pascasarjana, Institut Tekonologi Sepuluh Nopember. Surabaya.
- Rencher, A.C., 1998. *Multivariate Statistical Inference and Application*. Willey-Interscience Publication. Brigham.
- Setiawan, 2003. *Pendekatan Bayes Dengan Prior Normal Dalam Kalibrasi*. Makalah Seminar Nasional Matematika dan Statistika VI, 11 Oktober 2003. FMIPA-ITS Surabaya.
- Sutikno, 2008. *Statistical Downscaling Luaran GCM dan Pemanfaatannya untuk Peramalan Produksi Padi*. Disertasi. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Zorita, E. and von Storch, H., 1999. The Analog Method as a Simple Statistical Downscaling technique: comparison with more complicated method. *Journal of Climate*, 12, 2474-2489.
- Wigena, A.H., 2006. *Pemodelan Statistical Downscaling dengan Regresi Projection Pursuit untuk Peramalan Curah Hujan Bulanan*. Disertasi. Bogor: Program Pascasarjana, Institut Pertanian Bogor.