

PEMODELAN DOWNSCALLING LUARAN GCM MENGGUNAKAN METODE PCR DAN PLS UNTUK PREDIKSI CURAH HUJAN DI JAWA TENGAH

Salwati¹⁾

¹⁾Peneliti Muda, Bidang Sumber Daya Lingkungan

Balai Pengkajian Teknologi Pertanian, Jambi

Abstrak

Statistical downscaling adalah metoda untuk memperoleh informasi pada skala lokal dari skala yang lebih besar dengan menarik kesimpulan dari hubungan-hubungan antar skala, sedangkan fungsi yang digunakan adalah fungsi acak atau deterministik. Metode ini dapat diaplikasikan diantaranya untuk prediksi curah hujan. Data GCM yang digunakan adalah peubah luaran curah hujan dari ECHAM3 dengan resolusi $2,8^{\circ} \times 2,8^{\circ}$. Sedangkan data lokal curah hujan bulanan digunakan dari 12 stasiun hujan di Kabupaten Subang, yaitu Stasiun Ciasem, Cibandung, Cinangling, Ciseuti, Curugagung, Dangdeur, Kasomalang, Pamanukan, Pawelutan, Ponggang, Subang dan Tanjungsari dengan periode data dari tahun 1986-2002. Model SD disusun dengan menggunakan data curah hujan dari tahun 1986-2000, demikian pula data GCMnya. Reduksi data global GCM dilakukan dua tahap. Tahap pertama adalah reduksi klimatologis dengan cropping domain (grid) prediktor dengan asumsi bahwa kondisi iklim di wilayah stasiun-stasiun di Kabupaten Subang dipengaruhi oleh data GCM pada $98^{\circ}44'BT - 118^{\circ}13'BT$ dan $103^{\circ}9'LU - 18^{\circ}14'LS$, dengan grid 8×8 sehingga terdapat 64 grid. Reduksi tahap dua dilakukan dengan metode statistik yaitu PCR (Principle Component Regression), PLS (Partial Least Squares) dan MARS (Multivariate Adaptive Regression Spline). Untuk analisis dua metode yang pertama digunakan Minitab ver. 14.1, sedangkan untuk metode yang terakhir digunakan MARS ver 2.0. Hasil uji prediksi curah hujan lokal dengan menggunakan prediktor curah hujan hasil ECHAM melalui pemodelan dengan PCR, PLS dan MARS, menunjukkan bahwa PLS memiliki nilai root mean square error prediction (RMSEP) dan mean absolute error prediction (MAEP) yang kecil dengan korelasi validasi yang lebih besar dibanding PCR ataupun MARS. Hal ini menunjukkan bahwa PLS merupakan model yang lebih akurat dibanding kedua metode lainnya. Sementara MARS masih lebih baik dibanding PCR meskipun pada nilai yang lebih bervariasi, tetapi dari nilai rata-rata, hasil RMSEP dan MAEP MARS lebih kecil dari PCR, dengan korelasi validasi yang sedikit lebih besar dibanding PCR. Sehingga untuk pembentukan model, PLS dapat dipilih, karena selain memiliki model yang lebih akurat, juga bersifat multi respon sehingga pelaksanaan pembentukan model pada banyak prediktan dapat dilakukan sekaligus, lebih cepat dibanding PCR ataupun MARS.

PENDAHULUAN

Statistical Downscaling (SD) merupakan model statistik yang menggambarkan hubungan antara data pada grid-grid berskala besar (GCM) dengan data pada grid berskala lebih kecil GCM merupakan alat prediksi utama iklim dan cuaca secara numerik dan sebagai sumber informasi primer untuk menilai pengaruh perubahan iklim

General Circulation Model (GCM) merupakan alat yang sangat penting untuk mempelajari perubahan iklim dan membuat peramalan iklim untuk masa yang akan datang (Benestad, 2004). GCM merupakan alat prediksi utama iklim dan cuaca secara

numerik dan sebagai sumber informasi primer untuk menilai pengaruh perubahan iklim (Wigena 2006). GCM mempunyai hubungan fungsional dengan data yang berasal dari stasiun cuaca (Uvo *et al.* 2001).

Model iklim GCM mempunyai bentuk luaran grid grid yang berukuran 100-500 km, menurut garis lintang dan bujurnya. Model ini dapat digunakan untuk menduga perubahan unsur-unsur cuaca (von Stroch *et al.* 1993 dalam Sutikno, 2008). GCM mensimulasikan tiap peubah iklim dengan tingkat akurasi yang berbeda-beda dan hanya menggambarkan keadaan iklim berskala besar (resolusi rendah), sehingga diperlukan suatu teknik untuk pengamatan peubah iklim secara lokal.

Secara umum, ada dua pendekatan yang dapat dilakukan untuk memanfaatkan GCM menjadi informasi skala lokal yaitu *Dynamical Downscaling* dan *Statistical Downscaling*. (Benestad 2004). *Statistical Downscaling* (SD) merupakan model statistik yang menggambarkan hubungan antara data pada grid-grid berskala besar (GCM) dengan data pada grid berskala lebih kecil. Pada SD pola hubungan antara kedua peubah dapat diasumsikan linier dan nonlinier (Zorita dan Storch 1999).

Teknik SD membutuhkan suatu domain atau peubah iklim skala besar (peubah penjelas/prediktor : X) yang akan dihubungkan dengan data pada suatu stasiun cuaca atau peubah lokal (peubah respon/prediktan : Y). Domain berisi peubah-peubah curah hujan dari data GCM. Bentuk model SD secara umum adalah sebagai berikut : $y = f(x)$ dengan y merupakan vektor yang berisi peubah iklim lokal dalam jangka waktu t (misal: curah hujan) dan x merupakan matriks yang berisi gugus peubah luaran GCM (misal: curah hujan).

Salah satu tahapan dalam teknik SD adalah penentuan domain. Domain adalah lokasi dan luasan area permukaan atmosfer pada GCM yang akan dijadikan sebagai daerah yang berisi gugus peubah prediktor untuk menduga curah hujan di stasiun lokal. Pemilihan domain GCM akan menentukan hasil peramalan dan merupakan faktor kritis dalam pemodelan SD (Wilby & Wigley 2000 dalam Wigena 2006).

Adanya kompleksitas permasalahan dalam model *Statistical Downscaling*(SD) seringkali dilakukan penanganan satu persatu. Untuk mengatasi korelasi antar grid GCM dilakukan reduksi dimensi dengan analisis komponen utama (PCA), sehingga model regresinya $y=f(z) + \epsilon$, dimana z adalah skor komponen utama.

Metode *Partial Least Square* (PLS) merupakan metode statistik yang menggeneralisasi dan mengkombinasikan antara metode analisis faktor, *Principal Component Analysis* (PCA) dan *multiple regression* (Abdi, 2007). Tujuan dari metode PLS adalah membentuk komponen yang dapat menangkap informasi dari variabel independen yang digunakan untuk memprediksi variabel dependen. Model dari metode *Partial Least Square* (PLS) terdiri dari hubungan eksternal dan hubungan internal. Hubungan eksternal merupakan hubungan kelompok X dan Y secara individual. PLS merupakan prosedur regresi yang menghubungkan suatu set prediktor untuk dikalikan dengan variabel respon. PLS dikembangkan untuk digunakan pada kondisi data tertentu (prediktor mempunyai korelasi yang tinggi atau melebihi observasi).

Menurut Maitra dan Yan (2008), PCA sebagai metodologi dimensi reduksi, diaplikasikan tanpa menganggap adanya korelasi antara variabel bebas dan variabel tak bebas, sedangkan PLS diaplikasikan berdasarkan korelasi. Maka, PCA disebut sebagai *unsupervised dimension reduction methodology*, dan PLS disebut sebagai *supervised dimension reduction methodology*. Meskipun demikian, keduanya memiliki persamaan dalam analisis regresi, yaitu keduanya mempunyai teknik yang digunakan untuk merubah sebuah set variabel yang berkorelasi tinggi menjadi set dari variabel independen dengan menggunakan transformasi linier dan teknik yang digunakan keduanya adalah untuk mereduksi variabel.

Metode *statistical downscaling* merupakan metode yang cepat dan murah untuk dikerjakan dibanding metode *downscaling* lain. Tapi permasalahan yang sering muncul pada teknik SD yaitu menentukan domain (grid) dan pereduksian dimensi, mendapatkan variabel independen yang mampu menjelaskan keragaman variabel dependen, dan mendapatkan metode statistik yang sesuai karakteristik data, sehingga bisa menggambarkan hubungan antara variabel independen dan variabel dependen, serta dapat mengakomodasi kejadian ekstrim. Penggunaan metode statistik memungkinkan cepatnya diperoleh hasil yang dimaksud. Tulisan ini bertujuan melakukan pemodelan *downscaling* untuk memprediksi curah hujan lokal di wilayah Jawa Tengah dari data curah hujan luaran GCM menggunakan metode PCR dan PLS dan divalidasi dengan menggunakan data curah hujan 6 stasiun di Jawa Tengah.

METODOLOGI

Data

Data curah hujan yang digunakan sebagai peubah respon (prediktan) adalah data curah hujan lokal dari 6 stasiun klimatologi di Jawa Tengah yaitu : Stasiun Sukorejo, Kebumen, Banjar Negara, Lumbir, Ngadirejo, dan Brebes, merupakan data bulanan dari tahun 1990-2005. Sebagai peubah penjelas (prediktor) digunakan data curah hujan luaran GCM dari *National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP/NCAR) reanalysis* berukuran $2.5^{\circ} \times 2.5^{\circ}$ atau sekitar 300×300 km, dapat diakses melalui situs <http://www.esrl.noaa.gov/psd>.

Metode

Model SD disusun menggunakan data curah hujan lokal dan data curah hujan luaran GCM dari tahun 1990-2003. Reduksi data global GCM dilakukan dengan *cropping* domain dan metode statistik PCR dan PLS. Analisis metode PCR dan PLS dilakukan dengan menggunakan Minitab ver. 15.1.30.0

Cropping domain (grid) dilakukan untuk memilih domain GCM yang akan digunakan sebagai peubah prediktor. Pereduksian dilakukan terhadap dimensi *spasial* yaitu lintang dan bujur atau disebut grid pada semua variabel di setiap level serta pada setiap domain.

Analisis komponen utama dilakukan dengan menggunakan data luaran GCM tahun 1990-2003, lalu digunakan untuk menghitung skor komponen utama (z) dan diambil kumulatif eigenvalue hingga 90%. Selanjutnya skor komponen (z) yang didapat untuk tahun 1990-2003 digunakan sebagai prediktor pada pemodelan PCR. Dilakukan regresi untuk melihat keterkaitan antara z dengan curah hujan stasiun. Model yang diperoleh adalah regresi yang dihasilkan. Untuk memperoleh nilai dugaan model, dilakukan substitusi z data tahun 2004-2005 pada model yang dihasilkan. Nilai hasil dugaan model kemudian dibandingkan dengan data curah hujan observasi

PLS yang digunakan pada penelitian ini merupakan regresi satu respon (RKTP satu respon atau PLSR satu respon). Sebagai prediktan digunakan data curah hujan bulanan dari tahun 1990-2003 yang diaplikasikan sekaligus dengan prediktor data curah hujan hasil *NCEP-NCAR* (1990-2003), dengan memasukkan prediktor data curah hujan hasil *NCEP-NCAR* untuk prediksi (2004-2005). Hasil prediksi diperoleh dari nilai PFIT yang diperoleh dari keluaran. Nilai PFIT merupakan nilai dugaan dari model.

Untuk membandingkan hasil validasi antara PCR dan PLS digunakan tiga metode statistik, yaitu *Root Mean Square Error prediction* (RMSEP), *Mean Absolute Error Prediction*

(MAEP) dan korelasi (r) antara data observasi dan data hasil dugaan model, menggunakan data dua tahun terakhir 2004 dan 2005.

Root Mean Square Error of Prediction (RMSEP) menunjukkan besarnya perbedaan antara nilai dugaan dengan aktualnya. Semakin besar nilai RMSEP yang didapatkan akan menunjukkan bahwa nilai dugaan semakin kurang akurat. Nilai RMSEP didapatkan dengan melakukan perhitungan sebagai berikut :

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

dimana y_i = nilai observasi, \hat{y}_i = nilai dugaan, N_p = banyaknya data bebas yang digunakan untuk validasi model (Naes, 2002). Sedangkan MAEP dirumuskan

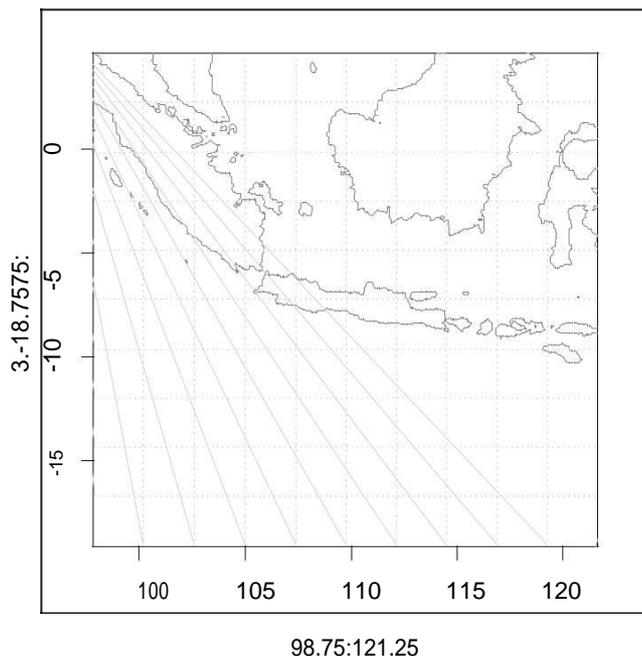
$$MAEP = \sum_{i=1}^{N_p} |y_i - \hat{y}_i| / N_p.$$

Nilai korelasi menunjukkan tingkat keeratan hubungan linier antara dua peubah atau lebih. Semakin besar nilai koefisien korelasi yang bernilai positif maka semakin kuat hubungan positif antara nilai dugaan dengan nilai aktual. Hal ini menunjukkan pola dari nilai dugaan semakin mendekati nilai aktual.

HASIL DAN PEMBAHASAN

Cropping Domain

Cropping domain (grid) prediktor yang digunakan berukuran 10 x 10, sehingga terdapat 100 grid dengan asumsi bahwa kondisi iklim di wilayah stasiun-stasiun di Jawa Tengah dipengaruhi oleh data luaran GCM yang terletak pada 3.75oLU - 18.75oLS dan 25oBT - 98.75oBT-121. Penomoran grid dimulai dari 88.75LU dan 1.25 BT. Peta dari domain ini disajikan pada Gambar 1.



Gambar 1. Peta domain Jawa tengah

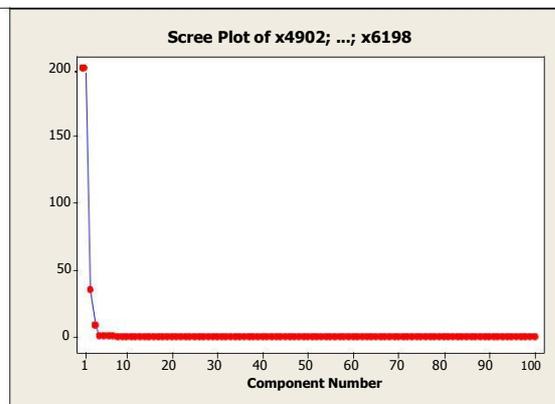
Pemilihan grid 10 x 10 sudah dapat mencangkup seluruh wilayah Jawa Tengah, tepat berada di atas lokasi target pendugaan, sehingga memberikan hasil yang lebih stabil atau konsisten dan tidak terlalu sensitif terhadap data pencilan. Ketepatan pemilihan domain, baik luasan maupun lokasinya, akan menghasilkan pendugaan curah hujan yang lebih akurat.

Principal Component Regression (PCR)

Analisis komponen utama/*Principal Component Analyses* (AKU/PCA) merupakan metode yang dilakukan sebelum pemodelan SD dengan metode PCR dan regresi. Hasil analisis PCA yang dapat menjelaskan keragaman lebih dari 90% sudah diperoleh pada 2 komponen utama. Nilai proporsi yang dapat dijelaskan dari kedua komponen utama ditunjukkan pada Tabel 1 dan *Screepplot* disajikan pada Gambar 2.

Tabel 1. Eigen analisis dari matriks ragam peragam pada PCA

Principal Component Analysis: x4902; x4903; x4904; x4905; x4906; x4907; x4908;									
<i>Eigenanalysis of the Covariance Matrix</i>									
<i>Eigenvalue</i>	196.97	35.03	8.27	0.94	0.81	0.68	0.35	0.18	0.17
<i>Proportion</i>	0.807	0.144	0.034	0.004	0.003	0.003	0.001	0.001	0.001
<i>Cumulative</i>	0.807	0.951	0.985	0.989	0.992	0.995	0.996	0.997	0.998



Gambar 2. *Screepplot* Komponen Utama

Pada *screepplot* terlihat bahwa titik belok terdapat pada Komponen Utama (KU) ke-3, namun di sini hanya digunakan 2 KU saja, karena proporsi keragaman kumulatif telah mencapai > 90% , dan pada *screepplot* dapat dilihat bahwa penurunan nilai akar ciri setelah KU-2 relatif kecil.

Tabel 1 menunjukkan bahwa keeratan hubungan antar grid diperoleh hingga 2 PCA. Sutikno (2008) menjelaskan bahwa banyaknya komponen utama menunjukkan hubungan keeratan antar peubah-peubah yang dilakukan reduksi dimensi. Semakin sedikit jumlah

komponen yang mempunyai persentase keragaman yang besar menunjukkan antar peubah yang dilakukan reduksi sangat erat dan sebaliknya.

Umumnya untuk suatu model yang dibangun dengan PCA membutuhkan paling sedikit 80% kontribusi PCA. Semakin tinggi kontribusi terhadap model, diharapkan *error* akan lebih rendah. Prosedur PCA pada dasarnya bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali atau yang biasa disebut dengan *principal component*. Setelah beberapa komponen hasil PCA yang bebas multikorelinearitas diperoleh, maka komponen-komponen tersebut menjadi variabel bebas baru yang akan diregresikan atau dianalisa pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi (Soemartini, 2008). Setelah metode PCA dilakukan metode PCR, keluaran dari model PCR disajikan pada Tabel 2.

Tabel 2. Contoh keluaran PCR stasiun Sukorejo

Regression Analysis: sukorejo versus z1; z2					
<i>The regression equation is</i>					
sukorejo = 282 + 9.97 z1 - 13.6 z2					
<i>Predictor</i>		<i>Coef</i>	<i>SE Coef</i>	<i>T</i>	<i>P</i>
<i>Constant</i>	281.95	11.61	24.28	0.000	
z1	9.9659	0.8298	12.01	0.000	
z2	-13.590	1.968	-6.91	0.000	
S = 150.491 R-Sq = 53.8% R-Sq(adj) = 53.2%					
<i>Analysis of Variance</i>					
<i>Source</i>	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>
<i>Regression</i>	2	4347467	2173733	95.98	0.000

<i>Residual Error</i>	165 3736867	22648
Total	167 8084334	

Tabel 2 menyajikan contoh *output* PCR stasiun Sukorejo dengan persamaan regresi menggunakan z sebagai prediktor. Z merupakan skor komponen utama, yang merupakan hasil transformasi linier peubah bebas X menggunakan dua KU pertama yang diperoleh dari PCA. Setelah dilakukan transformasi, peubah bebas yang digunakan dalam regresi (dalam hal ini PCR) menjadi orthogonal, sehingga tidak lagi terdapat multikolinieritas antar peubah bebas x .

Berdasarkan hasil regresi diperoleh persamaan pada stasiun Sukorejo, $y = 282 + 9.97 z_1 - 13.6 z_2$, dengan y mewakili curah hujan lokal sebagai prediktan/peubah respon dan z_1 dan z_2 yang mewakili GCM sebagai prediktor/peubah penjelas. Tabel 3 menyajikan koefisien determinasi terkoreksi (R_2) dan RMSE dari model PCR.

Tabel 3. Model PCR dengan RMSE dan R_2 terkoreksi pada 6 stasiun di Jawa Tengah

Stasiun	Model PCR	RMSE	R_2 terkoreksi
Sukorejo	$y = 282 + 9.97 z_1 - 13.6 z_2$	22648	53.2%
Kebumen	$y = 282 + 9.60 z_1 - 13.8 z_2$	22142	52.6%
Banjar Negara	$y = 318 + 10.1 z_1 - 13.6 z_2$	40031	39.4%
Lumbir	$y = 253 + 7.56 z_1 - 12.3 z_2$	29524	35.4%
Ngadirejo	$y = 363 + 12.1 z_1 - 20.0 z_2$	47423	47.0%
Brebes	$y = 132 + 5.62 z_1 - 7.97 z_2$	11694	41.5%

Tabel 3 terlihat R_2 terkoreksi tertinggi terdapat pada stasiun Sukorejo (53.2%) dan terendah stasiun Brebes (41.5%). Dapat dijelaskan bahwa untuk Stasiun Brebes hanya sebanyak 41.5% yang diterangkan oleh data, sedangkan sisanya bisa disebabkan oleh faktor lain. Tabel 3 terlihat RMSE tertinggi terdapat pada stasiun Ngadirejo (47423) dan terendah berturut-turut terdapat pada stasiun Sukorejo (22648), Kebumen (22142), dan Brebes (11994). Hasil analisis dengan PCR dapat dikatakan bahwa stasiun Sukorejo mewakili stasiun yang R_2 terkoreksinya paling tinggi dengan RMSE yang juga cukup rendah.

Partial Least Square (PLS)

Hasil *analysis of variance* (Anova) pemodelan SD dengan metode PLS disajikan pada Tabel 4.

Tabel 4. Sidik ragam (Anova) keluaran PLSR

Stasiun	RMSE	F	nilai-p	R ₂
Sukorejo	17459	43.29	0.000	65.4%
Kebumen	16387	52.49	0.000	66.2%
Banjar Negara	25016	31.48	0.000	64.2%
Lumbir	26659	30.80	0.000	43.0%
Ngadirejo	34643	38.74	0.000	62.9%
Brebes	10190	41.20	0.000	50.3%

Hasil analisis dengan metode PLS juga dapat dikatakan bahwa stasiun Sukorejo mewakili stasiun yang R₂ paling tinggi sebesar 65.4% dengan RMSE yang juga cukup rendah. Dapat disimpulkan bahwa stasiun Sukorejo memiliki prediksi yang paling akurat dibanding stasiun lainnya. Hal ini kemungkinan disebabkan oleh kondisi stasiun atau kelengkapan data pada stasiun Sukorejo dibanding stasiun lainnya..

Pada contoh model seleksi dan validasi untuk stasiun Sukorejo diperoleh variasi X dari 0.55 sampai 0.99, sedangkan R₂ diperoleh dari 0.41 sampai 0.49 (Tabel 5).

Tabel 5. Contoh model seleksi dan validasi untuk stasiun Sukorejo

<i>Model Selection and Validation for sukorejo</i>					
<i>Components</i>	<i>X Variance</i>	<i>Error SS</i>	<i>R-Sq</i>	<i>PRESS</i>	<i>R-Sq (pred)</i>
1	0.558238	4593121	0.431849	4783709	0.408274
2	0.942790	3547672	0.561167	3720179	0.539829
3	0.972033	3449681	0.573288	3694220	0.543040
4	0.984359	3309717	0.590601	3719020	0.539972
5	0.986309	3018719	0.626596	3904092	0.517079
6	0.988561	2846177	0.647939	3721057	0.539720
7	0.991886	2793495	0.654456	3677234	0.545141
8	2658284	0.671181	3812506	0.528408	
9	2517793	0.688559	3946580	0.511824	
10	2441289	0.698022	4100855	0.492741	

Validasi Model PCR dan PLS

Hasil validasi pemodelan SD dengan metode PCR dan PLS disajikan pada Tabel 6 dan nilai minimum, maksimum serta rata-rata RMSEP, MAEP dan korelasi validasi model berdasarkan PCR dan PLS disajikan pada Tabel 7.

Tabel 6. Hasil RMSEP, MAEP dan korelasi validasi (r) model 6 stasiun di Jawa Timur

Stasiun	RMSEP		MAEP		korelasi (r)	
	PCR	PLSR	PCR	PLSR	PCR	PLSR
Sukorejo	120.5322	125.1939	94.04336	94.77536	0.748019	0.76121
Kebumen	116.7194	120.405	90.89801	87.09273	0.739069	0.761906
Banjar Negara	159.4931	147.4078	128.8056	121.7362	0.718482	0.801105
Lumbir	168.3809	160.3837	132.5692	131.6034	0.526799	0.589326
Ngadirejo	191.8111	166.3313	155.6489	124.89	0.571334	0.728557
Brebes	101.4817	93.68691	61.60286	64.49059	0.801282	0.82086

Tabel 6 terlihat nilai RMSEP hasil pemodelan dengan metode PLS lebih rendah dibanding nilai PCR dengan nilai minimum sebesar 93.69, kecuali pada stasiun Sukorejo dan Kebumen. PLSR rata-rata MAEP juga lebih kecil dibanding PCR pada semua stasiun dengan nilai minimum sebesar 64.49. Korelasi r validasi juga menunjukkan metode PLS mempunyai nilai korelasi yang lebih tinggi dibanding PCR pada keseluruhan stasiun, dengan nilai tertinggi 0.82 dan terendah 0.59. Berdasarkan nilai RMSEP, MAEP, dan korelasi r yang diperoleh dari validasi kedua model, menunjukkan bahwa metode PLS lebih baik dari PCR

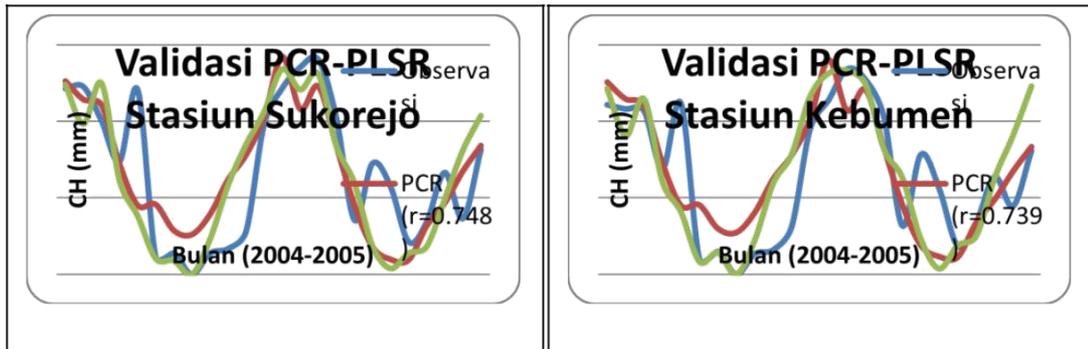
Tabel 7. Nilai minimum, maksimum dan rata-rata RMSEP, MAEP dan korelasi validasi model berdasarkan PCR dan PLS

Ukuran Statistik	PCR			PLSR		
	RMSEP	MAEP	r	RMSEP	MAEP	r
Minimum	101.4817	61.60286	0.526799	93.68691	64.49059	0.589326
Maximum	191.8111	155.6489	0.801282	166.3313	131.6034	0.82086
Rata-rata	143.0697	110.5947	0.684164	135.5681	104.0981	0.743827

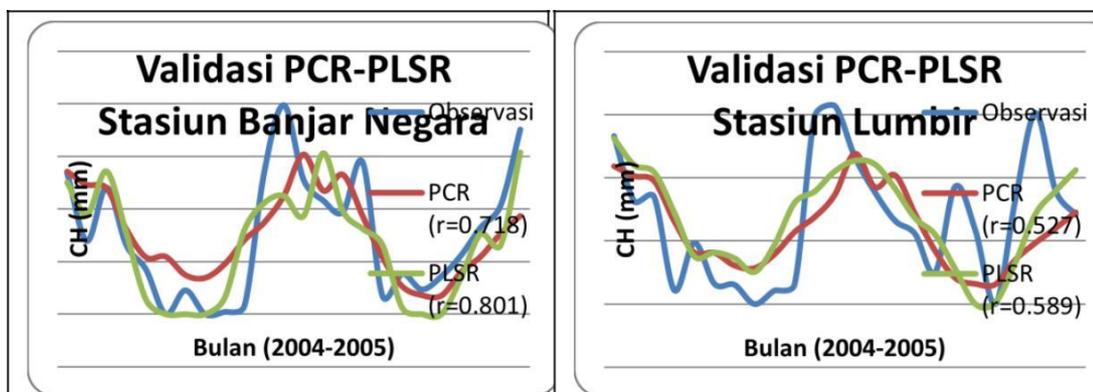
Dari Tabel 7 juga terlihat nilai minimum metode PLS yang lebih rendah dari metode PCR, serta nilai korelasi r maksimum yang lebih tinggi dari metode PCR.

Berdasarkan hasil pengujian dengan menggunakan RMSEP, MAEP dan korelasi validasi dapat dikatakan bahwa pemodelan curah hujan menggunakan data prediktor curah hujan luaran GCM dengan prediktor curah hujan lokal menunjukkan bahwa PLS merupakan metode yang lebih baik dari metode PCR untuk prediksi. Model dikatakan lebih baik jika RMSEP dan MAEP minimum dan korelasi maksimum.

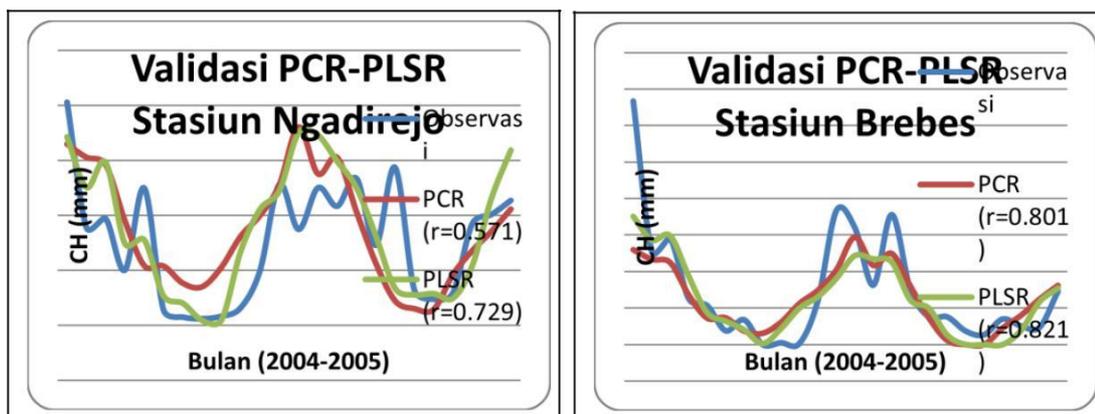
Gambar 2a sampai 2c memperlihatkan pola curah hujan yang terbentuk berdasarkan model PCR dan PLS yang dibandingkan dengan data hasil observasi. Tabel 8 memperlihatkan besarnya nilai sisaan yang merupakan hasil dari pengurangan data observasi dengan data dugaan yang diperoleh melalui PCR dan PLS untuk stasiun Sukorejo.



Gambar 2a. Hasil validasi dan korelasi PCR dan PLS Stasiun Sukorejo dan Kebumen



Gambar 2b. Hasil validasi dan korelasi PCR dan PLS stasiun Banjar Negara dan Lumbir



Gambar 2c. Hasil validasi dan korelasi PCR dan PLS stasiun Sukorejo dan Brebes

Tabel 8. Nilai observasi, dugaan dan sisaan stasiun Sukorejo dengan menggunakan PCR dan PLS

Y-observasi	PCR		PLSR	
	Y-dugaan	Sisaan	Y-dugaan	Sisaan
485	505.0774	-20.0774	495.6649	-10.6649
490	456.7755	33.22446	393.0697	96.9303
401	442.4315	-41.4315	498.8237	-97.8237
295	288.1909	6.809138	244.8668	50.1332
482	180.4903	301.5097	152.9514	329.0486
45	182.7434	-137.743	41.34544	3.654558
56	114.9783	-58.9783	36.00762	19.99238
0	104.1904	-104.19	0	0
55	156.5183	-101.518	91.41332	-36.4133
67	247.1749	-180.175	232.6735	-165.673
112	312.0272	-200.027	339.0892	-227.089
392	411.8655	-19.8655	424.511	-32.511
486	568.3732	-82.3732	536.0791	-50.0791
541	431.6876	109.3124	482.4273	58.57273
561	491.0601	69.9399	523.2826	37.71736
408	335.4393	72.5607	336.1742	71.82579
142	180.8713	-38.8713	238.5139	-96.5139
290	71.90312	218.0969	73.48335	216.5167
229	39.22821	189.7718	15.6115	213.3885
85	37.7631	47.2369	55.55311	29.44689
137	127.1593	9.840701	73.09896	63.90104
266	186.7575	79.24247	196.9146	69.08545
146	269.7453	-123.745	323.9917	-177.992
326	336.499	-10.499	414.7696	-88.7696

Gambar 2a, 2b, dan 2c di atas memperlihatkan pola curah hujan validasi yang hampir berhimpit antara nilai observasi dengan pola curah hujan yang dihasilkan dari model PCR

dan PLS untuk stasiun Sukerejo dan Brebes dibanding stasiun lainnya. Hal ini menunjukkan bahwa stasiun Sukorejo dan Brebes memberikan hasil validasi yang baik dibanding stasiun lainnya.

Dari Gambar 2a, 2b, dan 2c juga terlihat PLS memberikan nilai korelasi r yang tinggi dibanding PCR untuk semua stasiun yang digunakan. Analisis dengan metode PLS lebih cepat memberikan hasil, karena keseluruhan prediktan maupun prediktor dapat langsung dimasukkan dalam proses analisis dan hasil keseluruhannya dapat langsung diperoleh.

Data merupakan gabungan dari komponen sistematis dengan komponen sisaan (acak). Tabel 8 terlihat bahwa baik pada PLS maupun PCR nilai sisaan yang diperoleh setiap bulan bervariasi. Hal ini dapat dikatakan bahwa PLS dan PCR sama-sama baik digunakan untuk prediksi.

KESIMPULAN

Hasil uji prediksi curah hujan lokal menggunakan prediktor curah hujan luaran GCM melalui pemodelan DS dengan metode PCR dan PLS menunjukkan bahwa $RMSEP_{PLSR} < RMSEP_{PCR}$, dan $MAEP_{PLSR} < MAEP_{PCR}$ serta dan $korelasi_{PLSR} > korelasi_{PCR}$. Sehingga dapat disimpulkan jika dilihat dari kriteria **RMSEP, MAEP dan korelasi (r)** model yang dibangun dengan metode PLSR menunjukkan hasil yang lebih unggul dibandingkan dengan PCR.

Pemodelan DS dengan metode PLS merupakan model yang akurat digunakan untuk prediksi curah hujan. Metode PLS selain memiliki model yang lebih akurat, juga bersifat multi respon sehingga pelaksanaan pembentukan model pada banyak prediktan dapat dilakukan sekaligus, lebih cepat dibanding PCR. Metode PCR dengan pra-pemrosesan PCA dapat digunakan untuk mengatasi masalah multikolinearitas pada pemodelan SD untuk peramalan curah hujan bulanan.

DAFTAR PUSTAKA

- Haryoko U. 2004. Pendekatan reduksi dimensi luaran GCM untuk penyusunan model *Statistical Downscaling*. Tesis. Sekolah Pasca Sarjana IPB. Bogor.
- Maitra S, Yan J. 2008. *Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression*. Casualty Actuarial Society, 2008 Discussion Paper Program
- Sutikno. 2002. Penggunaan Regresi Splines Adaptif Berganda untuk Peramalan Indeks ENSO dan Hujan Bulanan. Tesis. Sekolah Pasca Sarjana IPB. Bogor.
- Sutikno. 2008. *Statistical downscaling* luaran GCM dan pemanfaatannya untuk peramalan produksi padi. Disertasi. Sekolah Pasca Sarjana IPB. Bogor.
- Wigena AH. 2006. Pemodelan *Statistical Downscaling* dengan *Regresi Projection Pursuit* untuk Peramalan Curah Hujan Bulanan, Kasus Curah Hujan Bulanan di Indramayu. Disertasi. Sekolah Pasca Sarjana IPB. Bogor.

Soemartini, 2008. *Principal Component Analysis (PCA) Sebagai Salah Satu Metode Untuk Mengatasi Masalah Multikolaritas*. Jurusan Statistika. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Padjajaran.