



BMKG

PROSIDING

SCIENTIFIC JURNAL CLUB

TAHUN 2011

Edisi ke : 6

BADAN METEOROLOGI KLIMATOLOGI DAN GEOFISIKA

Jl. Angkasa I No. 2 Kemayoran, Jakarta Pusat 10720.

Telp : (021) 4246321, Fax : (021) 4246703

P.O.Box : 3540 JKT, Website : <http://www.bmkg.go.id>

REGRESI KUADRAT TERKECIL PARSIAL UNTUK *STATISTICAL DOWNSCALING* (*Partial Least Square Regression for Statistical Downscaling*)

Aji Hamim Wigena

Departemen Statistika, FMIPA, Institut Pertanian Bogor

Abstrak

Dalam klimatologi statistical downscaling digunakan untuk memprediksi curah hujan lokal berdasarkan output dari model sirkulasi global (GCM). Regresi kuadrat terkecil parsial (PLSR) dapat digunakan sebagai alternatif teknik statistical downscaling. PLSR, serupa dengan regresi komponen utama (PCR), dalam mereduksi dimensi tetapi berbeda prosedurnya. Hasil analisis menunjukkan bahwa PLSR memberikan prediksi curah hujan bulanan yang lebih baik daripada PCR dan PLSR dapat memprediksi curah hujan di sejumlah lokasi secara simultan.

Kata kunci: *statistical downscaling, GCM, PLSR, PCR*

Abstract

In climatology statistical downscaling techniques have been used for predicting local rainfall based on GCM (Global Circulation Model) output. Partial Least Square Regression (PLSR) can be used as an alternative technique. PLSR is similar to the common statistical downscaling technique, Principal Component Regression (PCR), in reducing dimension but different procedures. The results of analysis show that PLSR performs better than PCR in monthly rainfall prediction and PLSR can predict rainfall in some locations simultaneously.

Keywords: *statistical downscaling, GCM, PCR, PLSR*

1. Pendahuluan

Dalam kajian klimatologi jangka panjang model sirkulasi global (*Global Circulation Model*, GCM) merupakan suatu model yang berorientasi spasial dan temporal, alat prediksi utama iklim secara numerik, serta sebagai sumber informasi primer untuk menilai pengaruh perubahan iklim. Namun informasi GCM masih berskala global dan tidak untuk fenomena skala lebih kecil (lokal), sehingga sulit untuk memperoleh langsung informasi berskala lokal dari GCM. Resolusi GCM terlalu rendah untuk memprediksi iklim lokal yang dipengaruhi oleh topografi dan tataguna lahan, tetapi GCM masih mungkin digunakan untuk memperoleh informasi skala lokal bila teknik *downscaling* digunakan.

Teknik *downscaling* adalah suatu proses transformasi data dari unit skala besar menjadi data pada unit skala yang lebih kecil. Wilby & Wigley (1997) menyatakan bahwa *downscaling* adalah suatu cara menginterpolasi peubah-peubah prediktor atmosfer berskala besar

terhadap peubah-peubah berskala lebih kecil. Salah satu jenis teknik ini adalah *statistical downscaling* (SD) yang menggunakan model statistik untuk menggambarkan hubungan antara data pada unit-unit berskala besar dengan data pada unit berskala lebih kecil dalam periode waktu tertentu.

Beberapa metode telah digunakan untuk SD (Wilby & Wigley 1997). Metode yang umum digunakan, antara lain regresi komponen utama (*Principal Component Regression*, PCR) berdasarkan analisis komponen utama untuk mereduksi dimensi atau mengatasi masalah multikolinieritas. Metode tersebut telah digunakan dalam pendugaan satu respon yaitu pendugaan curah hujan di suatu stasiun atau curah hujan rata-rata dari sejumlah stasiun di suatu wilayah.

Regresi Kuadrat Terkecil Parsial (*Partial Least Square Regression*, PLSR) dapat juga digunakan dalam *statistical downscaling* (Bergant & Kajfez-Bogataj 2005). PLSR serupa dengan PCR dalam hal mereduksi dimensi dan mengatasi masalah

multikolinearitas. Metode ini dapat digunakan untuk pendugaan baik satu respon maupun multi respon. Pendugaan multi respon mempertimbangkan adanya hubungan antar respon. Metode PLSR mulai dikembangkan pada tahun 1960an oleh Herman Wold dalam bidang ekonometrika dan telah digunakan secara luas dalam bidang kemometrika (Wold et.al 2001). Metode ini juga digunakan dalam bioinformatika, pangan, kedokteran, farmakologi, ilmu-ilmu sosial, fisiologi (Rosipal dan Kramer 2006), *neuroimaging* (McIntosh dan Lobaugh 2004), dan pengenalan pola (Wang et.al 2002).

Makalah ini membahas penggunaan PLSR dalam SD dan membandingkannya dengan PCR dalam hal pendugaan satu respon. PLSR multi respon akan dibandingkan juga dengan PLSR satu respon. Dalam makalah ini pendugaan multi respon yaitu pendugaan curah hujan di sejumlah stasiun di suatu wilayah secara simultan yang mempertimbangkan adanya hubungan spasial antar stasiun wilayah itu yang akan berpengaruh terhadap prediksi curah hujan.

2. Metode Penelitian

Data yang digunakan dalam kajian ini adalah presipitasi skala global dari tahun 1966 sampai dengan 2001 yang dibangkitkan dengan GCM dari ECHAM dan data curah hujan di enam stasiun (Sukadana, Bondan, Jatibarang, Kedokan Bunder, Tugu, Ujung Garis) di kabupaten Indramayu pada periode tahun yang sama. Lokasi Indramayu terletak pada 6.3° - 6.5° LS dan 107.8° - 108.4° BT. Ukuran domain untuk *downscaling* adalah sejumlah grid GCM berukuran 8x8 (Bergant *et al.* 2002) pada 1.4° LU - 18.1° LS dan 98.4° - 118.1° BT, yaitu di atas sekitar wilayah Indramayu.

Pendekatan SD secara umum mencari informasi skala lokal berdasarkan skala global melalui hubungan fungsional antara kedua skala tersebut, yang mempunyai bentuk umum:

$$Y = f(X) \text{-----} (1)$$

dimana:

$Y_{(b \times s)}$ = peubah skala lokal atau respon (seperti curah hujan bulanan),

$X_{(b \times g)}$ = peubah skala global atau prediktor (output GCM),

b = banyaknya waktu (seperti bulanan),

g = banyaknya grid skala global GCM,

s = banyaknya stasiun curah hujan.

Model seperti pada persamaan (1) dapat diduga dengan PCR dan PLSR. Tujuannya adalah untuk memprediksi peubah respon (Y) berdasarkan sejumlah peubah prediktor (X).

Analisis data dilakukan dengan PCR dan PLSR. Keduanya berbeda dalam proses pereduksian dimensi. Dalam PCR pereduksian dimensi dilakukan dulu dengan analisis komponen utama (*Principal Component Analysis*, PCA) terhadap data presipitasi tahun 1976-2000. Sejumlah komponen yang dihasilkan PCA dijadikan sebagai prediktor dan data curah hujan stasiun sebagai respon sehingga diperoleh penduga model SD. Dalam PLSR pereduksian dimensi dilakukan secara iteratif dengan melibatkan data presipitasi dan data curah hujan secara simultan sehingga diperoleh penduga model SD.

Dekomposisi dengan pendekatan PLSR serupa dengan gabungan analisis komponen utama dan regresi ganda. Metode PLSR akan mendapatkan sejumlah komponen dari X yang juga relevan dengan Y. Hal ini dilakukan dengan cara dekomposisi X dan Y secara bersamaan dengan batasan bahwa komponen-komponen tersebut dapat menjelaskan sebesar mungkin peragam (*covariance*) antara X dan Y. Proses dekomposisi ini diikuti dengan tahapan pendugaan model sehingga diperoleh hubungan antara X dan Y dan dapat digunakan untuk memprediksi Y.

PCR dan PLSR dibandingkan berdasarkan nilai RMSEP (*Root Mean Square Error of Prediction*) dan korelasi antara nilai dugaan dengan data aktual tahun 2001. Performa model yang baik ditunjukkan dengan nilai RMSEP yang lebih kecil dan nilai korelasi yang lebih tinggi.

3. Hasil dan Pembahasan

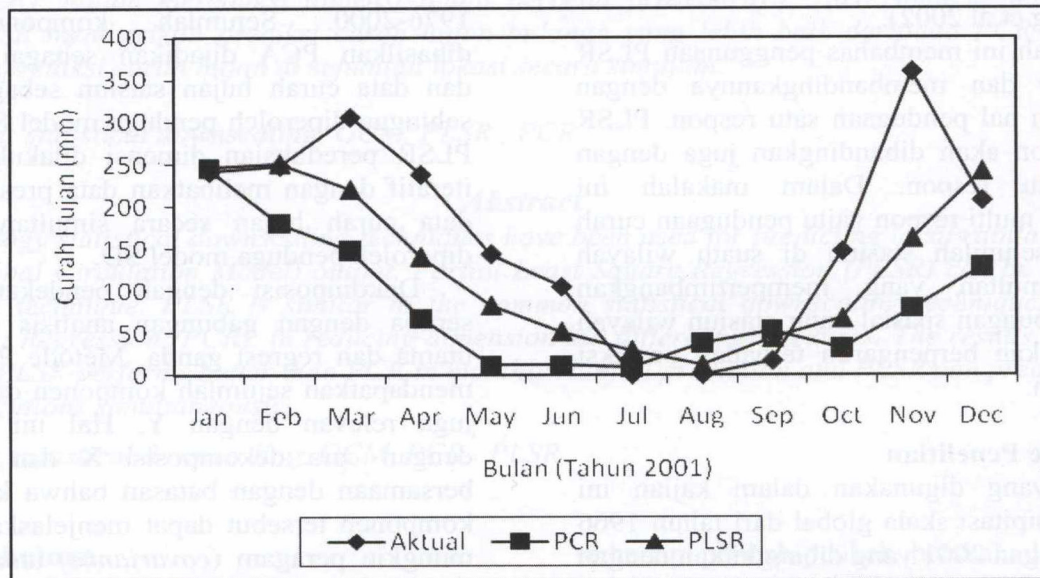
PCR dan PLSR digunakan untuk menduga model berdasarkan data 35 tahun (1966-2000) untuk memprediksi curah hujan pada tahun 2001. Hasil dugaan kedua model tercantum pada Tabel 1 dan Gambar 1. Pada Tabel 1 terdapat curah hujan bulanan aktual dan hasil prediksi kedua model berdasarkan data 35 tahun. Nilai RMSEP model PCR (125) lebih besar dari nilai RMSEP PLSR (77). Nilai

korelasi yang dihasilkan model PCR (0.60) lebih kecil dari nilai korelasi yang dihasilkan model PLSR (0.81). Gambar 1 menunjukkan perbandingan curah hujan aktual dan prediksi

dengan PCR dan PLSR. Model PLSR memberikan prediksi curah hujan yang lebih mendekati curah hujan aktual

Table 1. Curah hujan aktual dan prediksi (dengan PCR dan PLSR) tahun 2001

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	RMSEP	R
Actual	241	248	306	238	144	105	0	0	17	147	360	207	-	-
PCR	245.2	180.3	147.9	67.2	11.3	11.5	9.4	37.8	53.0	32.0	79.2	128.0	125	0.60
PLSR	245.7	251.1	220.3	140.7	82.6	51.5	29.3	11.4	43.9	67.3	162.1	241.6	77	0.81



Gambar 1. Curah hujan aktual dan prediksi (dengan PCR dan PLSR) tahun 2001

Analisis data juga dilakukan dengan menggunakan PLSR untuk satu dan multi respon. Nilai korelasi antara curah hujan aktual

dengan curah hujan dugaan dan RMSEP dari pendugaan dengan PLSR untuk satu respon dan multi respon tercantum pada Tabel 2.

Tabel 2. Nilai korelasi (r) dan RMSEP dari model PLSR

Stasiun	PLSR (satu respon)		PLSR (multi respon)	
	r	RMSEP	R	RMSEP
Sukadana	0.78	84.04	0.70	102.80
Bondan	0.61	77.69	0.64	76.50
Jatibarang	0.74	71.55	0.70	81.01
Kedokan Bunder	0.66	90.46	0.61	94.98
Tugu	0.52	125.13	0.40	138.52
Ujung Garis	0.77	76.92	0.76	74.93

Di keenam stasiun (Sukadana, Bondan, Jatibarang, Kedokan Bunder, Tugu, dan Ujung Garis) nilai korelasi (r) dengan PLSR satu respon relatif lebih besar dari nilai korelasi dengan PLSR multi respon, dan nilai RMSEP dengan PLSR satu respon relatif lebih kecil dari nilai RMSEP dengan PLSR multi respon, kecuali untuk stasiun Bondan dengan nilai r relatif sama sedangkan nilai RMSEP dengan PLSR satu respon lebih besar dari RMSEP dengan PLSR multi respon. Dalam hal ini secara umum PLSR satu respon menunjukkan hasil dugaan yang lebih baik dari PLSR multi respon. Namun demikian PLSR multi respon dapat digunakan karena dalam proses analisisnya memperhatikan hubungan spasial antar stasiun.

Secara umum nilai RMSEP masih cukup besar. Hal ini kemungkinan terkait dengan tipe data yang umumnya data iklim bersifat *uncertainty* dan nonlinier. Selain tipe data, kualitas prediksi dari suatu metode tergantung pada ukuran dan lokasi *domain* GCM. Salah satu perbaikan yang diperlukan adalah mencari metode yang dapat mengantisipasi tipe data tersebut.

4. Kesimpulan dan Prospek

1. Model PLSR lebih baik daripada model PCR sehingga dapat dijadikan sebagai alternatif teknik *statistical downscaling*.
2. Metode PLSR untuk satu respon dan multi respon memberikan hasil dugaan yang relatif sama, namun PLSR multi respon lebih baik digunakan untuk pendugaan secara simultan.
3. Metode *statistical downscaling* masih terus berkembang untuk memperoleh hasil pendugaan yang lebih baik. Kajian lebih lanjut diperlukan untuk mengkaji berbagai teknik *statistical downscaling* sehingga diperoleh metode terbaik.

Daftar Pustaka

- Bergant K, Kajfez-Bogataj L, Crepinsek Z. 2002. Downscaling of general-circulation-model-simulated average monthly air temperature to the beginning of flowering of dandelion (*Taraxacum officinale*) in Slovenia. *Int. J. Biometeorol*, 46: 22-32.
- Bergant K, Kajfez-Bogataj L. 2005. N-PLS regression as empirical downscaling tool in climate change studies. *Theor. App. Climatol*, 81: 11-23.
- McIntosh AR, Lobaugh NJ. 2004. Partial Least Square analysis of neuroimaging data: applications and advances. *NeuroImage*, 23: S260-S263.
- Rosipal R, Kramer N. 2006. Overview and Recent Advances in Partial Least Squares. C. Saunders et.al. (eds.): SLSFS 2005, LNCS 3940, 34-51, Springer-Verlag, Berlin.
- www.ofai.at/~roman.rosipal/Papers/pls_book06.pdf. [27 July 2007].
- Wang J, Zhenzhen K, Liang J. 2002. Facial Feature Point Extraction by partial least square regression. <http://citeseer.ist.psu.edu/wang02facial.html>. [25 July 2007].
- Wilby RL, Wigley TML. 1997. Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography*, 21,4:530-548.
- Wold S, Sj"ostr"om M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58: 109-130.