

REGRESI KUADRAT TERKECIL PARSIAL MULTI RESPON UNTUK *STATISTICAL DOWNSCALING*

(*Multi Response Partial Least Square for Statistical Downscaling*)

Aji Hamim Wigena
Departemen Statistika IPB
e-mail: ajiwigena@gmail.com

Abstract

In climatology partial least square regression (PLSR) can be used as an alternative technique in statistical downscaling based on global circulation model (GCM) output. PLSR is the technique to forecast not only one response but also multi responses to accommodate the correlation among responses. PLSR is compared to PCR (Principal Component Regression). The results show that PLSR is better than PCR and can be used to forecast rainfall simultaneously in more than one rainfall stations relatively as well as in one station.

Keywords: statistical downscaling, PLSR, PCR, multi responses

PENDAHULUAN

Dalam kajian klimatologi jangka panjang GCM merupakan suatu model yang berorientasi spasial dan temporal. Output GCM dijadikan sebagai sumber informasi primer untuk menilai pengaruh perubahan iklim. Namun informasi dari GCM ini masih berskala global sehingga sulit digunakan 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 (Fernandez 2005).

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

Beberapa metode telah digunakan untuk SD (Wilby & Wigley 1997). Regresi komponen utama (PCR) sering digunakan untuk pemodelan SD yang berdasarkan analisis komponen utama untuk mereduksi dimensi atau mengatasi masalah multikolinieritas. Serupa dengan metode PCR, regresi kuadrat terkecil parsial (PLSR) juga dapat digunakan untuk mengatasi multikolinieritas. Metode ini telah digunakan oleh Bergant dan Kajfez-Bogataj (2005). Makalah ini membahas penggunaan PLSR dan membandingkannya dengan PCR dalam hal keakuratan peramalan curah hujan.

STATISTICAL DOWNSCALING

Pendekatan SD mencari informasi skala lokal dari data skala global melalui hubungan fungsional antara kedua skala tersebut (Storch *et al.* 2001), atau berupa suatu fungsi transfer (Sailor *et al.* 2000). Namun untuk keadaan skala global yang sama, keadaan skala lokalnya bisa bervariasi.

Secara umum bentuk modelnya adalah:

$$\underline{Y} = f(\underline{X})$$

dengan:

$\underline{Y}_{(t)}$ = peubah skala lokal (curah hujan bulanan).

$\underline{X}_{(t \times g)}$ = peubah skala global (output GCM).

t = banyaknya waktu (seperti bulanan).

g = banyaknya grid skala global GCM.

Hasil dari model SD terkait langsung dengan karakteristik iklim pada waktu sebelumnya dan data deret waktu yang homogen dalam berbagai kondisi iklim. Model SD juga akan memberikan hasil yang baik jika ketiga syarat berikut terpenuhi, yaitu (1) Hubungan erat antara respon dengan prediktor yang menjelaskan keragaman iklim lokal dengan baik; (2) Peubah prediktor disimulasi baik oleh GCM, dan (3) Hubungan antara respon dengan prediktor tidak berubah dengan perubahan waktu dan tetap sama meskipun ada perubahan iklim (Busuioc *et al.* 2001).

REGRESI KUADRAT TERKECIL PARSIAL

Regresi kuadrat terkecil parsial (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 bidang bioinformatika, penelitian makanan, kedokteran, farmakologi, ilmu-ilmu sosial, fisiologi

(Rosipal & Kramer 2006), *neuroimaging* (McIntosh & Lobaugh 2004), dan pengenalan pola (Wang *et.al.* 2002).

Metode PLSR mengkombinasikan analisis komponen utama dan regresi ganda. Tujuannya adalah untuk memprediksi suatu gugus peubah respon (**Y**) berdasarkan gugus peubah prediktor (**X**). Prediksi ini diperoleh dengan mengekstraksi sejumlah komponen, yang disebut peubah *latent*, dari peubah prediktor.

Bila suatu vektor **Y** dan suatu matriks **X** berpangkat penuh, prediksi dapat dilakukan dengan regresi kuadrat terkecil. Tetapi bila jumlah prediktor lebih banyak daripada jumlah pengamatan dan **X** berpangkat tidak penuh atau bersifat *singular* maka pendekatan regresi kuadrat terkecil tidak lagi tepat karena ada masalah multikolinearitas dalam **X**.

PLSR akan mendapatkan komponen-komponen dari **X** yang juga relevan dengan **Y**. Hal ini dilakukan dengan cara dekomposisi **X** dan **Y** secara simultan dengan batasan bahwa komponen-komponen tersebut dapat menjelaskan sebesar-besarnya peragam (*covariance*) antara **X** dan **Y**. Proses dekomposisi ini diikuti dengan tahapan regresi di mana hasil dekomposisi **X** digunakan untuk memprediksi **Y**.

Bila **X** berukuran $N \times K$ (N adalah jumlah pengamatan dan K adalah jumlah peubah prediktor), yang terdiri dari vektor x_k , $k = 1, \dots, K$, dan **Y** berukuran $N \times M$ (M adalah jumlah peubah respon), yang terdiri dari vektor y_m , $m = 1, \dots, M$. Metode PLSR menghasilkan sejumlah komponen baru yang akan memodelkan **X** dan **Y**, sehingga diperoleh hubungan antara **X** dan **Y**. Komponen-komponen baru ini disebut skor **X**, yang dicatat dengan t_a , $a=1, 2, \dots, A$.

Skor **X** merupakan kombinasi linier peubah-peubah asal x_k dengan koefisien yang disebut pembobot, dicatat dengan vektor w_{ka} ($a = 1, 2, \dots, A$). Proses tersebut dapat diformulasikan sebagai berikut: (Wang *et.al.* 2002)

$$\begin{cases} t_{ia} = \sum_k x_{ik} w_{ka}, i = 1, \dots, N \\ T = XW \end{cases} \quad (1)$$

Skor **X** (yaitu t_a) digunakan sebagai prediktor untuk **Y** dan model dari **X**, skor tersebut mempunyai sifat-sifat berikut:

(a) Skor **X** dikalikan dengan p_{ak} , sehingga residunya, e_{ik} , kecil:

$$\begin{cases} x_{ik} = \sum_a t_{ia} p_{ak} + e_{ik} \\ X = TP' + E \end{cases}$$

(b) Skor **X** adalah prediktor bagi **Y**:

$$\begin{cases} y_{im} = \sum_a t_{ia} r_{am} + f_{im} \\ Y = TR' + F \end{cases} \quad (2)$$

Residu **Y**, yaitu f_{im} , menyatakan simpangan antara respon pengamatan dengan respon dugaan. Berdasarkan persamaan (1), persamaan (2) dapat dituliskan sebagai model regresi ganda berikut:

$$\begin{cases} y_{im} = \sum_a r_{ma} \sum_k w_{ka} x_{ik} + f_{im} = \sum_k b_{mk} x_{ik} + f_{im} \\ Y = XWR' + F = XB + F \end{cases}$$

Koefisien model PLSR, b_{mk} , adalah:

$$\begin{cases} b_{mk} = \sum_a r_{ma} w_{ka} \\ B = WR' \end{cases}$$

Prediksi bagi data pengamatan yang baru dapat diperoleh berdasarkan data **X** dan matriks koefisien **B**.

HASIL DAN PEMBAHASAN

Data yang digunakan dalam kajian ini adalah presipitasi skala global (GCM) dari tahun 1966 sampai dengan 2002 dari ECHAM dan data curah hujan beberapa stasiun di kabupaten Indramayu pada periode tahun yang sama. Data diperoleh dari Departemen Geofisika dan Meteorologi, FMIPA – IPB. Lokasi Indramayu terletak pada $6.3^\circ - 6.5^\circ$ LS dan $107.8^\circ - 108.4^\circ$ BT. Ukuran *domain* untuk *downscaling* adalah sejumlah grid GCM berukuran 8×8 pada 1.4° LU - 18.1° LS dan $98.4^\circ - 118.1^\circ$ BT, yaitu di atas sekitar wilayah Indramayu (Wigena 2006).

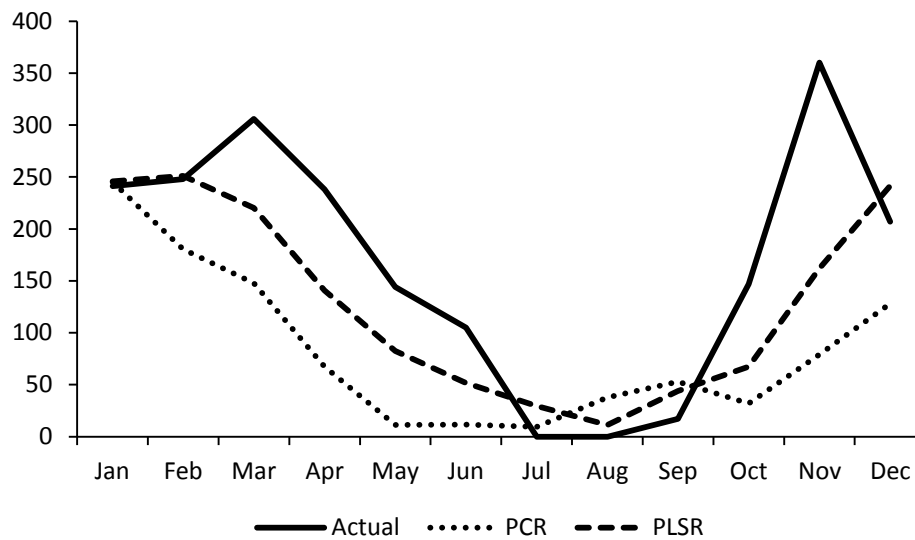
PCR dan PLSR digunakan untuk menduga model berdasarkan data 35 tahun (1966-2000) untuk memprediksi curah hujan pada tahun 2001. Model PLSR digunakan juga untuk berbagai periode data 30 tahunan. Model-model tersebut dibandingkan berdasarkan nilai RMSEP (*Root Mean Square Error of Prediction*) dan korelasi (r) antara data prediksi dengan data aktual.

Hasil dugaan kedua model tercantum pada Tabel 1 dan Gambar 1. Pada Tabel 1 terdapat curah hujan bulanan aktual dan hasil prediksi dengan kedua model berdasarkan data 35 tahun. Nilai RMSEP model PCR (125) lebih besar dari nilai RMSEP model 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 lebih baik daripada model PCR.

Selain itu metode PLSR digunakan juga untuk pendugaan multi respon. Pada Tabel 2 tercantum nilai korelasi (r) antara curah hujan aktual dengan curah hujan dugaan dan RMSEP dari pendugaan dengan PLSR untuk satu respon dan multi respon.

Tabel 1. Curah hujan aktual dan prediksi dengan PCR dan PLSR

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	RMSEP	r
Aktual	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

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

Stasiun	satu respon		multi respon	
	r	RMSEP	r	RMSEP
Sukadana	0.78	84.0	0.70	102.8
Bondan	0.61	77.7	0.64	76.5
Jatibarang	0.74	71.6	0.70	81.0
Kedokan Bunder	0.66	90.5	0.61	95.0
Tugu	0.52	125.1	0.40	138.5
Ujung Garis	0.77	76.9	0.76	74.9

Di stasiun-stasiun tersebut nilai korelasi sebagai hasil dari 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 nilai korelasi relatif sama sedangkan nilai RMSEP dengan PLSR satu respon lebih besar dari RMSEP dengan PLSR multi respon. Secara umum PLSR satu respon memberikan hasil dugaan yang lebih baik daripada PLSR multi respon. Hal ini menunjukkan bahwa adanya korelasi spasial antar stasiun PLSR dapat digunakan untuk menganalisis sejumlah stasiun secara simultan.

KESIMPULAN

Metode PLSR dapat digunakan untuk pemodelan multi respon meskipun memberikan hasil dugaan yang relatif kurang baik dibandingkan dengan untuk satu respon tetapi proses pendugaannya dapat dilakukan secara simultan. Namun suatu kajian simulasi tentang penggunaan PLSR untuk multi respon perlu dilakukan dengan memperhatikan berbagai nilai korelasi spasial antar stasiun untuk mengetahui berapa besaran nilai korelasi sehingga metode ini efektif.

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.
- Busuioac A, Chen D, Hellstrom C. 2001. Performance of statistical downscaling models in GCM validation and regional climate change estimates: Application for Swedish precipitation. *Int. J. of Climatol*, 21: 557-578.
- Fernandez E. 2005. On the influence of predictors area in statistical downscaling of daily parameters. *Report no.09/2005*. Norwegian Meteorological Institute, Oslo.
- 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 Square. 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].
- Sailor DJ, Hu T, Li X, Rosen JN. 2000. A neural network approach to local downscaling of GCM output for assessing wind power implications of climate change. *Renewable Energy*, 19: 359-378.
- Storch H, Zwiers FW. 1999. *Statistical Analysis in Climate Research*. London: Cambridge University Press.
- 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].
- Wigena AH. 2006. *Pemodelan Statistical Downscaling dengan Regresi Projection Pursuit untuk Peramalan Curah Hujan Bulanan. Studi Kasus: Curah Hujan Bulanan di Indramayu*. Disertasi. Sekolah Pascasarjana, Institut Pertanian Bogor.
- 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, Sjöström M, Eriksson L. 2001. PLS regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58: 109-130.