

PENENTUAN DOMAIN SPASIAL NWP DALAM PEMBANGUNAN *MODEL OUTPUT STATISTICS*

DETERMINATION OF NWP SPATIAL DOMAIN ON THE MODEL OUTPUT STATISTICS DEVELOPMENT

Urip Haryoko¹, Hidayat Pawitan², Edvin Aldrian¹, Aji Hamim Wigena²

¹Badan Meteorologi Klimatologi dan Geofisika, Jl. Angkasa I No 2 Jakarta 10720

²Institut Pertanian Bogor, Kampus IPB Dramaga Bogor 16680

Naskah masuk: 15 Juli 2013; Naskah diperbaiki: 16 Oktober 2013; Naskah diterima: 24 Desember 2013

ABSTRAK

Model Output Statistics (MOS) adalah salah satu metode *statistical downscaling* pada tahap *post processing* luaran *Numerical Weather Prediction (NWP)* untuk mendapatkan nilai prakiraan parameter cuaca di sebuah titik stasiun pengamatan. Permasalahan yang timbul dalam MOS adalah penentuan domain spasial NWP yang akan digunakan sebagai prediktor. Pada makalah ini disajikan metode penentuan domain spasial untuk memprakirakan suhu maksimum di wilayah Jabodetabek menggunakan data luaran NWP *Global Forecast System (GFS)* dari *National Oceanic and Atmospheric Administration (NOAA)*. Data pengamatan suhu maksimum diambil dari delapan stasiun di Wilayah Jakarta, Jawa Barat dan Banten yang digunakan untuk kalibrasi. Pada tahap awal domain spasial NWP ditentukan berukuran 8x8 grid, selanjutnya dicobakan untuk beberapa domain, yaitu berukuran 2x2, 3x3, 3x4, 4x4 dan 5x5 grid. Tiga metode digunakan untuk menentukan domain spasial, yaitu metode analisis korelasi spasial, *singular value decomposition (SVD)* dan *partial least square regression (PLSR)*. Analisis ketiga metode secara umum menunjukkan hasil yang hampir sama, yaitu domain dengan ukuran 3x3 adalah yang paling baik. Analisis korelasi spasial menunjukkan luasan dengan korelasi lebih besar dari 0,4 hanya meliputi domain maksimal 3x3. Analisis SVD menunjukkan bahwa keeratan hubungan secara simultan antara data observasi dengan NWP hampir sama, yaitu pada ekspansi pertama. Sedangkan hasil verifikasi analisis PLSR menggunakan korelasi dan *root mean square error (RMSE)* menunjukkan bahwa grid berukuran 3x3 adalah domain terbaik.

Kata kunci : *statistical downscaling, numerical weather prediction, single value decomposition, partial least square.*

ABSTRACT

Model Output Statistics (MOS) is one of *statistical downscaling* method in *post-processing* of *Numerical Weather Prediction (NWP)* output to get weather forecasts at a point of observation stations. The problem in MOS is how to determine the spatial domain of NWP which will be used as predictor in development stage. This paper presented the methods for determining NWP spatial domain to predict the maximum temperature in the Greater Jakarta area using NWP output of *Global Forecast System (GFS)* produced by *National Oceanic and Atmospheric Administration (NOAA)*. Maximum temperature observation data was taken in eight stations, around West Java, Banten and Jakarta. In the first stage, spatial domain of NWP was defined as 8x8 grids, and then attempted for some domains, i.e. 2x2, 3x3, 3x4, 4x4 and 5x5 grids. Three methods for determining spatial domain were spatial correlation analysis, *singular value decomposition (SVD)* and *partial least square regression (PLSR)*. Those three analysis methods generally showed similar results, spatial domains with size 3x3 is the most excellent. Spatial correlation analysis shows that the size of the area which has correlation greater than 0.4 was only covers a maximum of 3x3 domain. SVD analysis suggests that the simultaneous relationship between the observation data with NWP is almost the same in the first expansion. While the results of the verification PLSR analysis using correlation and *root mean square error (RMSE)* indicates that 3x3 grid is the best domain.

Keywords : *statistical downscaling, numerical weather prediction, single value decomposition, partial least square.*

1. Pendahuluan

Peubah-peubah yang dihasilkan dari (NWP) mempunyai dimensi yang besar yaitu, dimensi spasial (S), dimensi waktu (T), dimensi vertikal (V) dan dimensi parameter itu sendiri (P). Pada setiap dimensi, peubah NWP tidak saling bebas dan mempunyai korelasi yang cukup tinggi. Pada pemodelan model statistik yang melibatkan banyak peubah bebas harus memenuhi persyaratan bahwa peubah bebasnya harus tidak saling berkorelasi.

Permasalahan yang muncul dalam pengembangan MOS adalah multikolinieritas antar peubah terutana bagi output NWP sebagai prediktor. Multikolinieritas yang terjadi meliputi autokorelasi masing-masing peubah, korelasi spasial untuk setiap peubah, korelasi setiap peubah pada satu ketinggian dengan ketinggian lainnya. Dapat disimpulkan bahwa peubah NWP saling berkorelasi menurut dimensi waktu dan ruang. Pada pemodelan regresi, adanya multikolinieritas pada peubah penjelas NWP akan memperbesar nilai *variance inflation factor* (VIP). Hal ini akan mempengaruhi performa model yang dihasilkan[1].

Untuk menangani permasalahan multikolinieritas dalam pemodelan *downscaling* atau MOS dapat dilakukan dengan mereduksi dimensi Ruang Vektor X (peubah bebas). Reduksi dimensi dimaksudkan untuk memperoleh Ruang Vektor X yang lebih kecil namun variasinya masih tetap dipertahankan. Dengan dimensi Vektor X yang lebih kecil akan memudahkan penanganan dalam pemodelan, sehingga hubungan antara Ruang Vektor Y dan Ruang Vektor X menjadi lebih sederhana. Misalkan $\dim X = n$ dan $\dim y = m$ ($n \gg m$). Permasalahan dalam pemodelan *statistical downscaling* atau MOS adalah mereduksi $\dim X = n$ menjadi n' dengan syarat variasi vektor X tetap dipertahankan dan $n' \ll n$ [1].

Salah satu cara untuk meminimalisasi multikolinieritas adalah reduksi dimensi spasial atau mencari peubah baru yang merupakan kombinasi linier dari peubah aslinya dengan syarat peubah baru tidak saling berkorelasi. Rummukainen[2] mengemukakan bahwa metode reduksi dimensi yang sering digunakan diantaranya adalah analisis komponen utama (AKU), transformasi Regresi Fourier (RF), dekomposisi nilai singular (SVD), analisis Wavelets, *multidimensional scaling* (MS). Review metode *statistical downscaling* (SD) yang dilakukan Rummukainen[2] menyebutkan bahwa metode yang umumnya digunakan untuk mereduksi peubah bebasnya adalah AKU, sedangkan pemodelan untuk estimasi digunakan metode regresi dan analisis korelasi kanonik (AKK).

Haryoko[1] telah melakukan penelitian mengenai reduksi dimensi model global dengan membandingkan dua metode reduksi dimensi analisis komponen utama (AKU) dan metode regresi Fourier. Dari hasil penelitian tersebut disimpulkan bahwa metode reduksi dimensi regresi Fourier lebih baik dari metode AKU. Namun demikian metode regresi Fourier lebih rumit dan memerlukan penomoran urutan grid yang tepat.

Wigena[3] menentukan domain *Global Circulation Model* (GCM) dalam penyusunan model SD dengan mengambil domain spasial berbentuk segi bujur sangkar tepat di atas wilayah Kabupaten Indramayu. Segi bujur sangkar merupakan gabungan dari beberapa grid (satuan spasial GCM). Dalam penelitiannya, Wigena[3] menggunakan berbagai ukuran persegi yaitu 8x8 grid (segi 8), 10x10 grid (segi 10), 12x12 grid (segi 12), 14x14 grid (segi 14) dan 16 x 16 grid (segi 16) yang posisinya tepat pada daerah penelitian. Disamping itu digunakan pula domain grid pada wilayah di daerah 50°LU-40°LS dan 50°-185°BT terhadap beberapa persegi lokasi target pendugaan. Dari berbagai pilihan tersebut, Wigena[3] menyebutkan bahwa untuk menentukan grid masih dilakukan secara subyektif.

Maini *et al.*[4] membangun persamaan regresi menggunakan data TOGA dengan sembilan grid berukuran 2,5°x2,5° tahun 1985 – 1990 sebagai prediktor dan pengamatan suhu sebagai prediktan. Pada tahap awal ditentukan prediktor yang mungkin dapat digunakan. Untuk menentukan hasil kombinasi linier terbaik dari nilai-nilai prediktor di sembilan grid sekitar lokasi stasiun, dihitung korelasi kanonik antara prediktan (T_{maks} dan T_{min}) di stasiun dengan prediktor di sembilan grid. Kombinasi linier terbaik yang dihasilkan mempunyai korelasi maksimum dengan prediktan. Untuk mengeliminasi prediktor yang mempunyai informasi *redundance* digunakan regresi stepwise. Prediktor yang berpotensi menjelaskan seluruh varian dipilih sebagai prediktor.

Busuioc *et al* dalam Wigena[3] menyatakan bahwa salah satu syarat pemodelan SD adalah adanya hubungan yang erat antara prediktor dan prediktan. Dengan demikian maka penentuan domain spasial dalam pengembangan MOS perlu mempertimbangkan hubungan antara prediktor pada grid tertentu dengan prediktan.

Prakiraan cuaca jangka pendek mempunyai skala sinoptik yaitu skala waktu yang pendek (sampai dengan 7 hari) dan skala ruang yang cukup sempit (sampai dengan 100 km). Fenomena cuaca harian seperti *cloud cluster* dan *sea land breeze* mempunyai dimensi sampai dengan tujuh hari dan skala spasial sampai dengan 100 km[5].

Berdasarkan kondisi ini, maka pada makalah ini, sebagai domain awal dalam penentuan domain spasial NWP hanya mengambil sample grid 8 x 8 di sekitar wilayah penelitian, dan ukuran tiap gridnya 0,5° x 0,5°, sehingga luas wilayah penelitian meliputi 400 x 400 km. Tulisan ini bertujuan untuk mendapatkan domain spasial yang tepat dalam pemodelan MOS, yaitu menduga parameter cuaca di suatu titik berdasarkan prediktor NWP.

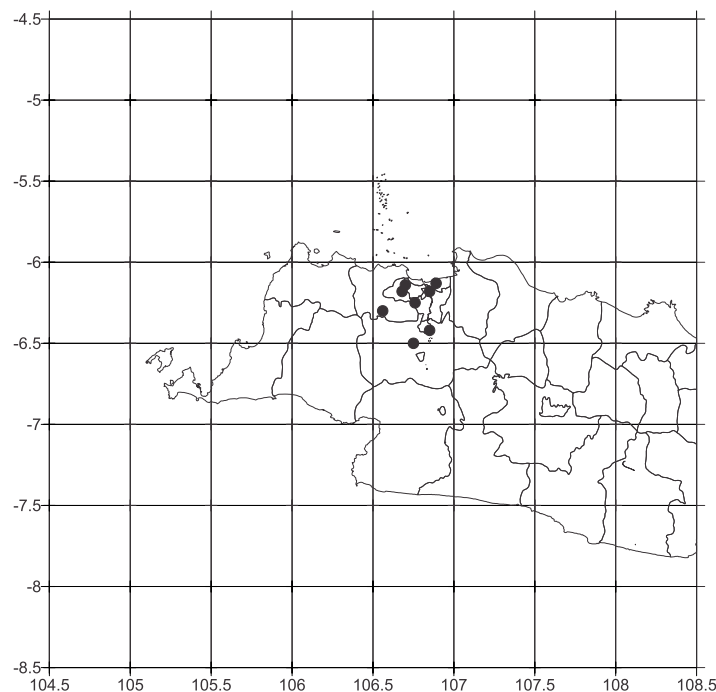
2. Metode Penelitian

Bahan. Data NWP yang digunakan dalam tulisan ini adalah Global Forecasting System (GFS) dengan resolusi grid $0,5^\circ$ selama dua tahun yaitu mulai September 2010 sampai dengan September 2012 (diambil dari situs <ftp://nomads.ncdc.noaa.gov/GFS/Grid4/>). Jumlah grid yang dianalisis hanya berukuran 8×8 yang berada di sekitar Jawa Barat, Banten dan Jakarta. Jumlah grid ini didasarkan skala berbagai fenomena cuaca harian sampai mingguan yang ditetapkan oleh WMO[5], yaitu variasi fenomena cuaca harian sampai mingguan mempunyai skala horizontal antara 10 km sampai dengan 100 km.

Data cuaca permukaan yang digunakan dalam penelitian ini adalah data suhu maksimum di beberapa stasiun di wilayah Jawa Barat, Banten dan Jakarta. Data GFS yang digunakan untuk penentuan domain ini diambil beberapa peubah yang secara fisis mempunyai hubungan erat terhadap suhu udara maksimum. Lokasi penelitian dan grid seperti pada gambar 1.

Metode. Penentuan domain spasial grid NWP untuk menduga parameter cuaca di titik stasiun pengamatan dilakukan dengan beberapa metode yaitu analisis isokorelasi, *single value decomposition* (SVD) dan analisis regresi *partial least square* (PLSR). Dari ketiga metode tersebut akan ditentukan domain spasial yang paling tepat berdasarkan kuatnya hubungan melalui isokorelasi, kuatnya hubungan secara serentak dan nilai *root mean square error* (RMSE) terkecil dari pendugaan menggunakan PLSR.

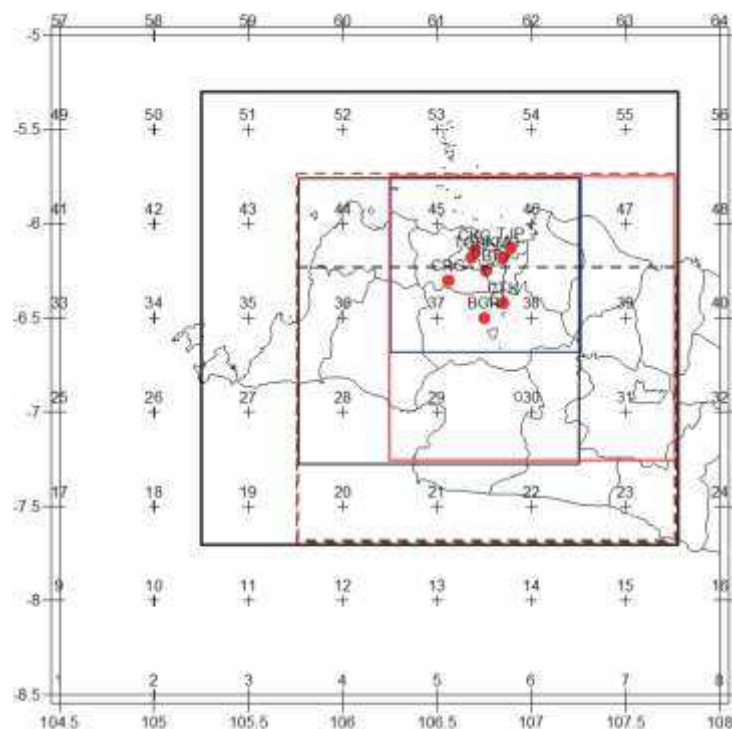
Sebagai langkah awal penentuan domain spasial adalah grid berukuran (8×8) dan selanjutnya dipilih beberapa ukuran grid yaitu jumlah grid 4 (2×2), 9 (3×3), 12 (3×4), 16 (4×4) dan 25 (5×5) di sekitar lokasi stasiun penelitian. Domain-domain spasial berbentuk persegi panjang dan bujur sangkar yang berada di sekitar wilayah penelitian dengan ukuran seperti pada tabel 1 dan gambar 2.



Gambar 1. Lokasi stasiun pengamatan cuaca dan domain spasial penelitian

Tabel 1. Domain spasial data NWP sekitar Jabodetabek

Domain	Ukuran (grid)	Anggota grid
domain_1	2 x 2	37 38 45 46
domain_2	3 x 3	28 29 30 36 37 38 44 45 46
domain_3	3 x 3	29 30 31 37 38 39 45 46 47
domain_4	3 x 4	20 21 22 23 28 29 30 31 36 37 38 39
domain_5	4 x 4	20 21 22 23 28 29 30 31 36 37 38 39 52 53 54 55
domain_6	5 x 5	19 20 21 22 23 27 28 29 30 31 35 36 37 38 39 43 44 45 46 47 51 52 53 54 55



Gambar 2. Domain spasial sekitar wilayah penelitian

Isokorelasi. Isokorelasi adalah garis yang menghubungkan tempat-tempat yang mempunyai nilai korelasi yang sama. Korelasi dalam analisis ini adalah korelasi antara suhu maksimum di titik stasiun terhadap hasil prakiraan GFS (NWP), yaitu nilai suhu maksimum dan suhu permukaan pada jam 06 UTC (13 waktu setempat) pada tanggal yang sama. Isokorelasi ini dimaksudkan untuk melihat seberapa jauh nilai suhu maksimum di suatu stasiun mempunyai korelasi yang signifikan terhadap NWP.

Parameter suhu maksimum di setiap stasiun dihitung korelasinya terhadap suhu maksimum NWP pada tanggal yang sama; dan suhu maksimum stasiun dengan suhu permukaan pukul 06 UTC. Data suhu permukaan NWP diambil pukul 06 UTC dengan pertimbangan bahwa suhu maksimum di wilayah Jawa Barat, Banten dan Jakarta terjadi sekitar pukul 13 – 14 WIB atau 06 – 07 UTC. Selanjutnya nilai korelasi untuk setiap grid diplot ke peta dan ditarik garis isokorelasinya. Nilai korelasi berkisar antara -1 s/d +1; sehingga dalam analisis ini hanya dilihat pada korelasi yang positif karena nilai prakiraan harus mempunyai korelasi yang positif terhadap nilai yang diprakirakan. Untuk membatasi tingkat signifikansi korelasi dipilih korelasi yang nilainya lebih dari 0,4.

Single Value Decomposition (SVD). Menurut Baker[6], SVD dapat dilihat dari tiga sudut pandang. Pertama, SVD dapat dilihat sebagai alat untuk mentransformasi peubah yang saling berkorelasi menjadi sebuah set yang tidak saling berkorelasi sehingga akan dapat menjelaskan variasi data aslinya secara lebih baik. Kedua, pada saat yang sama, SVD adalah metode untuk mengidentifikasi dan mengurutkan dimensi variasi yang paling besar ke yang paling kecil. Ketiga, dengan mengetahui variasi

terbesar dari gugusan data tersebut selanjutnya dapat ditentukan keaslian gugus data melalui dimensi yang lebih kecil. Dengan demikian SVD dapat dipandang sebagai metode untuk mereduksi dimensi gugus data.

Dalam analisis klimatologi, SVD dapat digunakan sebagai alat untuk mengetahui hubungan antara dua gugus data, misalnya NWP dan data pengamatan cuaca lokal. Menurut Storch dan Navarra[7], untuk mencari telekoneksi antara dua gugus data diperlukan teknik seperti *empirical orthogonal function* (EOF) yang dapat diaplikasikan pada *crosscovariance* (peragam silang) antara gugus data yang berbeda. Pengembangan dari analisis EOF terhadap ragam silang didasarkan pada SVD dari ragam silangnya. Algoritma perhitungan nilai akar ciri dapat digunakan, namun perhitungannya cukup rumit. SVD adalah teknis aljabar yang cukup baik untuk dekomposisi sembarang matrik menjadi matrik-matrik ortogonal.

Sanjaya[8] menggunakan SVD untuk menyeleksi prediktor GCM dalam memprakirakan curah hujan bulanan di Wilayah Indramayu. Jumlah grid yang dipilih adalah 8x8 dengan ukuran tiap grid 2,5°x2,5°. Misalkan **X** dan **Y** adalah matrik data output NWP dan data pengamatan cuaca. Data NWP mencakup sebanyak waktu *t* dan sebanyak *p* grid, sehingga **X** adalah matrik berukuran *t* x *p*, sedangkan data pengamatan cuaca diambil sebanyak waktu *t* dan *q* lokasi, sehingga matrik **Y** berukuran *t* x *q*. Menurut Bjornson dan Venegas[9], dekomposisi nilai singular dapat diawali dengan membentuk matrik koragam silang sebagai berikut:

$$C_{xy} = Z'S \quad (1)$$

dengan

$$Z = \begin{bmatrix} X_{11} & \dots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{t1} & \dots & X_{tp} \end{bmatrix} - \begin{bmatrix} \bar{X}_1 & \dots & \bar{X}_p \\ \vdots & \ddots & \vdots \\ \bar{X}_1 & \dots & \bar{X}_p \end{bmatrix} \quad (2)$$

$$S = \begin{bmatrix} Y_{11} & \dots & Y_{1p} \\ \vdots & \ddots & \vdots \\ Y_{t1} & \dots & Y_{tp} \end{bmatrix} - \begin{bmatrix} \bar{Y}_1 & \dots & \bar{Y}_p \\ \vdots & \ddots & \vdots \\ \bar{Y}_1 & \dots & \bar{Y}_p \end{bmatrix} \quad (3)$$

Selanjutnya dapat dibentuk dekomposisi nilai singular terhadap matrix C dengan menemukan matrik U dan V serta matrik diagonal L sehingga diperoleh hubungan sebagai berikut :

$$C_{xy} = ULV^t \quad (4)$$

dengan :

U : matriks singular berukuran $p \times m$ dari matriks C

V : matriks singular berukuran $m \times q$ dari matriks C

L : matriks diagonal

m : $\min(p, q)$.

Vektor-vektor singular untuk X merupakan kolom dari matrik U , dan vektor-vektor singular untuk matrik Y merupakan kolom dari V . Kolom dari U biasanya disebut sebagai *left pattern* dan kolom V disebut *right pattern*.

Selanjutnya dapat diperoleh koefisien ekspansi, yaitu time series yang menjelaskan setiap mode osilasi. Untuk Z dihitung :

$$A = ZU \quad (5)$$

dan untuk S dihitung :

$$B = SV \quad (6)$$

Kolom-kolom dari matrik A dan B berisi koefisien setiap mode, dan dikarenakan U dan V saling ortogonal maka dapat direkonstruksi matrik data dengan menggunakan hubungan $Z = AU^t$ dan $S = BV^t$. Matriks L merupakan matriks diagonal tak negatif yang unsur-unsurnya adalah $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_m$. Total squared covariance dari C dihitung dengan menjumlahkan nilai kuadrat dari unsur matriks L . Jika λ_i adalah nilai singular ke- i , maka *fraction of squared covariance* (SCF) dapat dijelaskan dengan vektor singular yang berhubungan dengan vektor u dan v , sehingga

$$SCF_i = \frac{\lambda_i^2}{\sum_{i=1}^m \lambda_i^2} \quad (7)$$

SCF ke- i dapat diartikan sebagai proporsi ragam X dan Y yang diterangkan oleh pasangan *spatial pattern* ke- i .

Analisis regresi PLS. Termonia dan Deckmyn[10] menjelaskan bahwa MOS dapat mengoreksi output model NWP dengan cara menggunakan model

statistik. Biasanya model yang digunakan adalah model regresi linier. Namun demikian, hasil prakiraannya tidak akan optimal jika antar peubah bebas yang digunakan mempunyai korelasi yang kuat.

Pemodelan menggunakan metode regresi komponen utama (PCR) masih menyisakan permasalahan jika komponen peubah bebas hanya memiliki pengaruh yang kecil dibandingkan dengan bagian yang tidak relevan. Variabilitas peubah bebas tidak muncul pada beberapa komponen pertama. Dengan demikian pada model PCR masih memunculkan permasalahan dalam memilih komponen. Untuk itu diperlukan metode untuk mengekstrak seluruh komponen agar dapat menjelaskan setiap penambahan komponen yang akan meningkatkan performance model. *Partial Least Square Regression* (PLSR) dapat menyelesaikan persoalan tersebut. Algoritma ini digunakan untuk menjelaskan kedua peubah X dan Y dan mengekstrak komponen (atau disebut sebagai faktor) yang relevan terhadap kedua peubah tersebut. Sehingga yang harus dilakukan adalah menentukan faktor-faktor model yang relevan mempengaruhi.

PLSR adalah metode untuk membangun model prediksi jika peubah bebasnya saling berkorelasi tinggi. Penekanan dari metode ini adalah memprediksi peubah respon, bukan pada mencari hubungan antar peubah. metode PLSR tidak cocok untuk mengeluarkan faktor-faktor yang tidak berpengaruh terhadap peubah respon.

Algoritmanya dapat dijelaskan sebagai berikut. Misalkan X adalah matriks pengamatan peubah bebas berukuran $n \times p$ dan Y adalah matriks data pengamatan peubah respon berukuran $n \times q$. Pada makalah ini n adalah jumlah pengamatan, p jumlah grid, dan q jumlah stasiun hujan. Prosedur yang harus dilalui adalah mengekstraksi faktor X dan Y sedemikian rupa sehingga faktor yang terekstraksi mempunyai kovarian yang maksimum.

Metode kuadrat terkecil parsial mendekomposisi secara linier matriks X dan Y .

$$X = TP^T + E \quad (8)$$

$$Y = UQ^T + F \quad (9)$$

dengan

$T_{n \times g}$: skor X
 $U_{n \times g}$: skor Y
 $P_{n \times g}$: loading X
 $Q_{n \times g}$: loading Y
 $E_{n \times p}$: error X
 $F_{n \times q}$: error Y

Kolom dari matriks T adalah vektor laten dan $U = TB$, ini merupakan regresi vektor t . Dengan demikian dapat diturunkan:

$$Y = TBQ^T + F \quad (10)$$

Dalam penerapan PLSR, akan dicari bobot w dan c untuk membentuk kombinasi linier pada kolom X dan

Y sehingga kombinasi linier ini akan mempunyai kovarian yang maksimal.

$$t = Xw \text{ dan } u = Yc \quad (11)$$

Pada penelitian ini analisis PLSR akan diterapkan pada setiap domain untuk menduga suhu maksimum di delapan stasiun pengamatan berdasarkan suhu maksimum keluaran NWP. Panjang data untuk membangun model adalah mulai tanggal 11 Nopember 2010 sampai dengan 31 Juli 2012, dan untuk verifikasi model digunakan data tanggal 1 Agustus 2012 sampai dengan 31 Oktober 2012. Pemilihan model terbaik didasarkan pada nilai korekasi antara data observasi dengan hasil prakiraan, dan nilai RMSE.

3. Hasil dan Pembahasan

Analisis peta korelasi. Untuk memudahkan analisis, dibuat peta korelasi (korelasi spasial) (lihat lampiran 1a, 1b, 2a dan 2b). Pada lampiran 1a dan 1b tampak bahwa korelasi antara suhu maksimum seluruh stasiun terhadap suhu maksimum NWP di atas 0,4 terjadi pada grid sekitar grid 30 dan 31. Untuk grid lain yang mempunyai korelasi lebih besar 0,4 terjadi pada tiga grid yang sebarannya tidak menyatu. Pada lampiran 2a dan 2b terlihat bahwa korelasi antara suhu maksimum seluruh stasiun dengan suhu permukaan pukul 06 UTC yang lebih besar 0,4 terjadi pada grid 29, 30, 31, 32 dan 36, 37, 38, 39, 40. Grid-grid NWP yang mempunyai korelasi lebih besar 0,4 tersebut terletak di sekitar stasiun.

Peta korelasi pada lampiran 1 dan 2 juga menunjukkan area grid yang mempunyai korelasi lebih besar 0,4 hampir sama untuk setiap stasiun, yaitu dengan luasan sekitar satu sampai tiga grid. Untuk korelasi dengan suhu permukaan, menunjukkan bahwa luas areanya sama untuk tiap stasiun yaitu sekitar enam sampai dengan 9 grid. Jika dibandingkan dengan korelasi dengan suhu maksimum NWP, luasannya yang lebih besar. Dengan melihat hasil korelasi ini, maka untuk tahap ini dapat disimpulkan sementara bahwa domain spasial yang mempunyai pengaruh terhadap suhu maksimum adalah grid sekitar lokasi stasiun dengan luas sekitar satu sampai dengan 9 grid.

Analisis SVD. Analisis SVD diaplikasikan pada data dengan 6 bentuk domain, masing-masing domain dihitung nilai SCF dan korelasi antara koefisien **A** dan **B** pada setiap ekspansi. Hasil lengkap perhitungan nilai SCF dan korelasi dapat dilihat pada tabel 2.

Keterkaitan spasial antara suhu maksimum observasi dengan suhu maksimum NWP dapat diidentifikasi dengan melihat proporsi ragamnya atau nilai SCF. SCF ekspansi ke-1 dapat diartikan sebagai proporsi ragam suhu maksimum observasi dengan suhu maksimum NWP yang diterangkan oleh pasangan *spatial pattern* ke-i. Pada tabel 2 dapat dilihat bahwa nilai SCF atau ragam yang dihasilkan pada ekspansi pertama semua domain telah mencapai nilai 99% dan nilai tertinggi terjadi pada domain 4 yaitu dengan ukuran 3x4 grid. Ini berarti bahwa ragam suhu

maksimum hasil pengamatan di delapan stasiun telah dapat dijelaskan oleh ekspansi pertama sebesar 99%.

Nilai SCF untuk semua domain tidak menunjukkan perbedaan yang signifikan, dan semuanya telah mencapai 99%. Hal ini menunjukkan bahwa dengan analisis SVD tidak tampak perbedaan yang signifikan antara masing-masing domain. Namun demikian jika dilihat dari nilai korelasi antara *expansion coefficient* pertama sampai dengan ke delapan untuk semua domain, maka pada ekspansi pertama terlihat bahwa besarnya korelasi meningkat sebanding dengan bertambahnya domain gridnya. Pada domain 1 (4 grid) hanya mencapai 0,41, domain 2 dan 3 (9 grid), mencapai rata-rata 0,5 dan grid 4 (12 grid) dan seterusnya mencapai di atas 0,66.

Berdasarkan analisis di atas, maka dengan menggunakan analisis SVD dapat ditentukan domain yang sebaiknya digunakan, yaitu domain ke-4 dengan ukuran grid 3x4.

Analisis PLSR. Gambar 3 (kiri) menunjukkan verifikasi nilai korelasi data pengamatan dan hasil prakiraan model PLSR. Pada grafik tersebut tampak bahwa terdapat perbedaan korelasi yang cukup signifikan antar beberapa domain. Korelasi tertinggi secara umum terjadi pada domain 3, yaitu domain dengan ukuran grid 3 x 3 (grid nomor : 29 30 31 37 38 39 45 46 47). Demikian pula dengan grafik RMSE (kanan) tampak bahwa nilai RMSE terkecil secara umum dicatat pada domain 3.

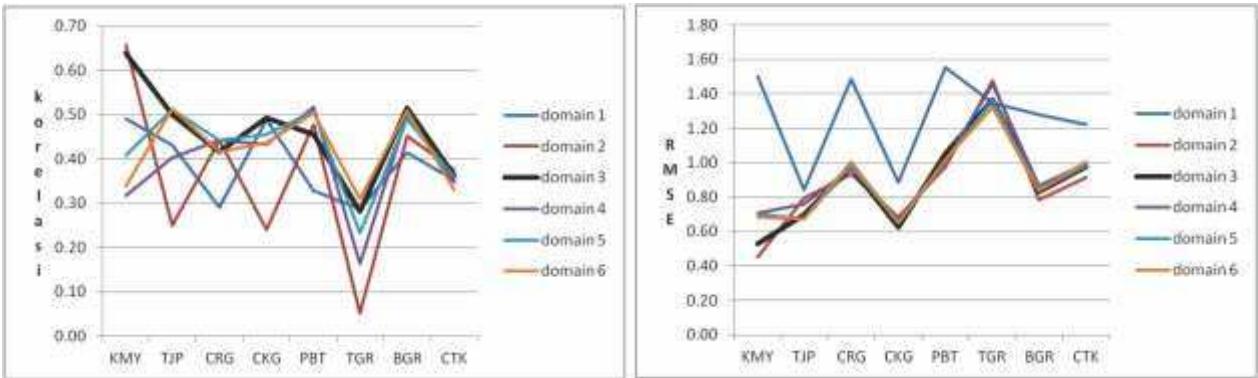
Gambar 4 menunjukkan plot antara nilai korelasi dan RMSE dalam bentuk kuadran dengan ketentuan sebagai berikut :

1. Kuadran 1 menunjukkan nilai korelasi kecil dan RMSE besar (model tidak baik)
2. Kuadran 2 menunjukkan nilai korelasi besar dan RMSE besar (model kurang baik)
3. Kuadran 3 menunjukkan nilai korelasi besar dan RMSE kecil (model paling baik)
4. Kuadran 4 menunjukkan nilai korelasi kecil dan RMSE kecil (model kurang baik).
5. Angka pada titik-titik di dalam diagram menunjukkan nomor domain.

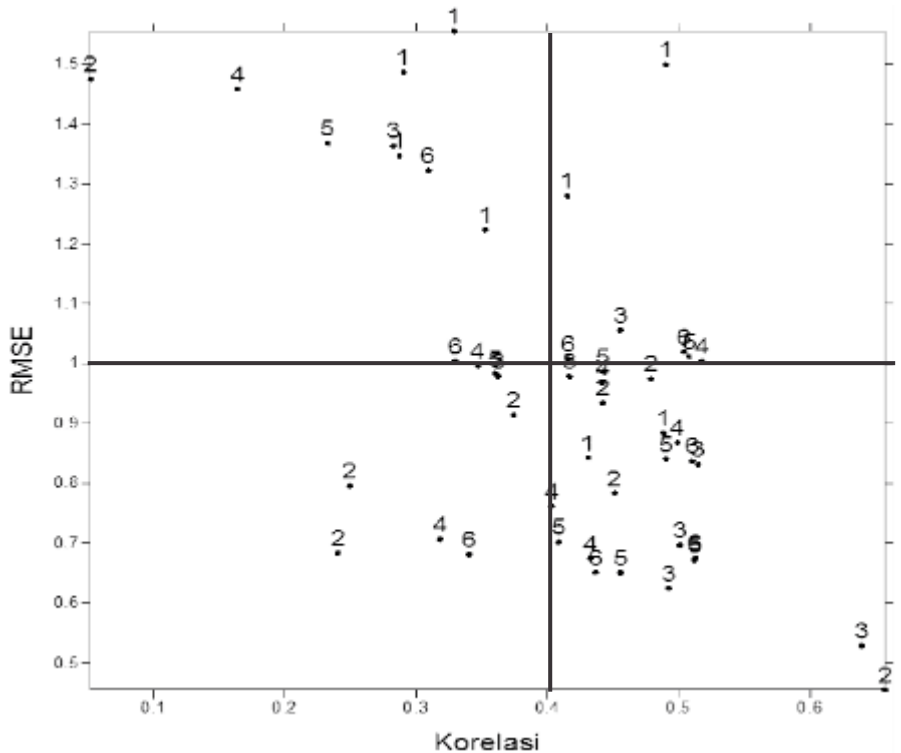
Pada gambar 4 terlihat bahwa nomor domain yang terletak pada kuadran 3 adalah domain 3. Dengan demikian dapat dikatakan bahwa dari enam domain yang dianalisis, domain 3 merupakan domain terbaik untuk memprakirakan suhu maksimum.

Tabel 2. Nilai fraction of square covariance (SCF) dan korelasi (rho) pada ekspansi 1 sampai 8 hasil analisis SVD

Statistik	Ekspansi ke	Domain					
		1	2	3	4	5	6
SCF	1	98.56%	98.97%	99.04%	99.37%	99.09%	98.69%
	2	1.24%	0.83%	0.83%	0.46%	0.69%	0.89%
	3	0.15%	0.15%	0.09%	0.12%	0.15%	0.32%
	4	0.05%	0.03%	0.02%	0.03%	0.03%	0.05%
	5		0.01%	0.01%	0.01%	0.02%	0.03%
	6		0.01%	0.00%	0.01%	0.01%	0.01%
	7		0.00%	0.00%	0.00%	0.00%	0.00%
	8		0.00%	0.00%	0.00%	0.00%	0.00%
Rho (korelasi antara A1 dan B1)	1	0.41	0.57	0.65	0.66	0.67	0.67
	2	0.35	0.33	0.32	0.31	0.33	0.33
	3	0.30	0.31	0.34	0.18	0.24	0.26
	4	0.16	0.12	0.15	0.10	0.09	0.09
	5		0.09	0.08	0.11	0.16	0.17
	6		0.08	0.09	0.10	0.12	0.13
	7		0.07	0.09	0.10	0.09	0.08
	8		0.07	0.01	0.03	0.08	0.10



Gambar 3. Korelasi suhu maksimum observasi dengan prakiraan (kiri) dan RMSE suhu maksimum observasi dengan prakiraan (kanan).



Gambar 4. Plot nilai korelasi dan RMSE Model PLSR

4. Kesimpulan

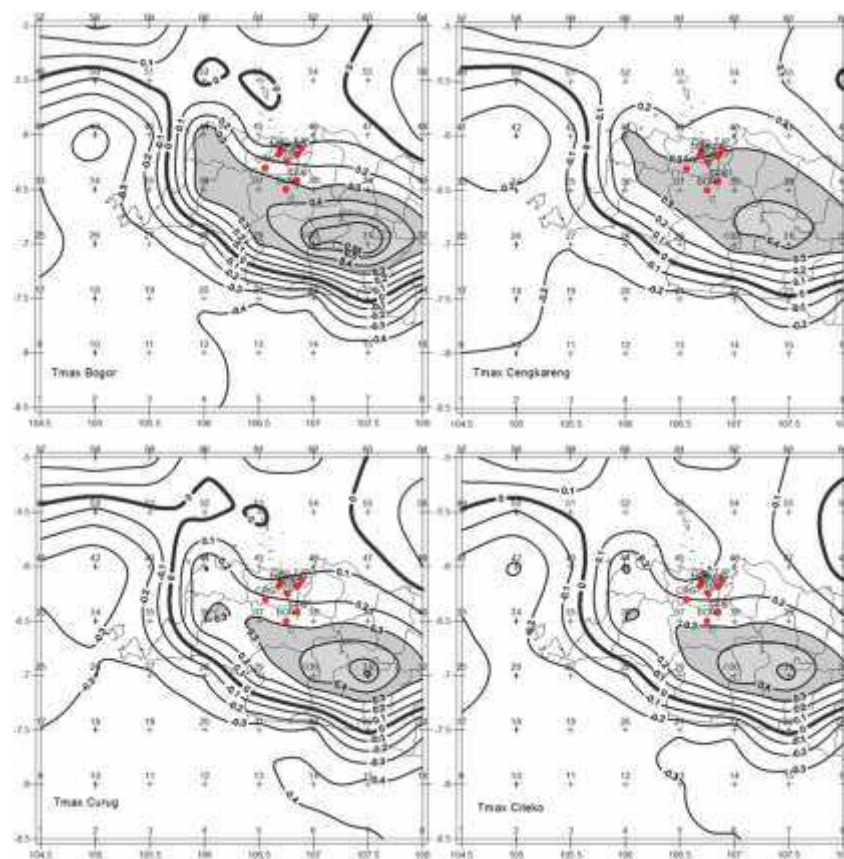
Domain grid berukuran 3x3 pada posisi grid 29, 30, 31, 37, 38, 39, 45, 46, dan 47 berpotensi untuk dijadikan sebagai domain spasial dalam pengembangan model MOS sekitar Jawa Barat, Banten dan Jakarta menggunakan data NWP GFS.

Model analisis korelasi spasial, SVD dan PLSR untuk penentuan domain spasial menunjukkan hasil yang hampir sama. Namun demikian, model yang paling mudah untuk penentuan domainnya adalah metode PLSR karena dapat dengan mudah untuk memilihnya yaitu dengan memverifikasi berdasarkan nilai korelasi dan RMSE.

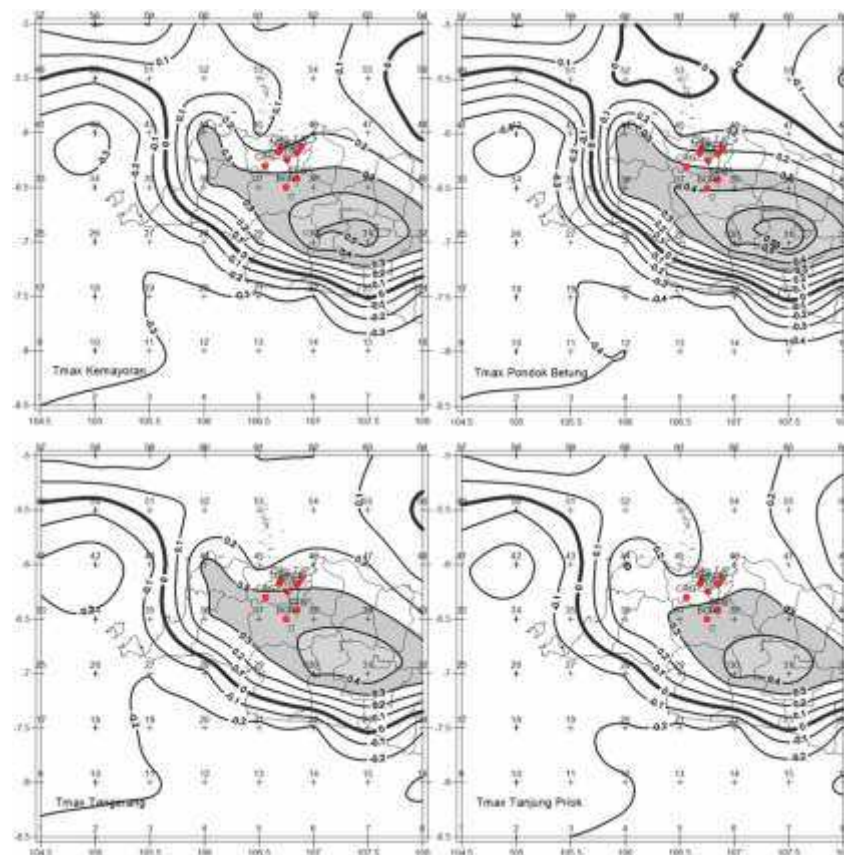
Daftar Pustaka

- [1] Haryoko U. (2004). *Pendekatan reduksi dimensi luaran GCM untuk penyusunan model statistical downscaling*. Thesis, Sekolah Pasca Sarjana: Institut Pertanian Bogor.
- [2] Rummukainen M. (1997). *Methods for statistical downscaling of GCM simulations*. SMHI Reports Meteorology and Climatology. Rossby Centre. SMHI.
- [3] Wigena AH. (2006). *Pemodelan statistical downscaling dengan regresi projection pursuit untuk peramalan curah hujan bulanan*. Disertasi, Sekolah Pasca Sarjana: Institut Pertanian Bogor.
- [4] Maini P., Ashok Kumar, L.S. Rathore, dan S.V.Singh. (2003). Forecasting maximum and minimum temperatures by statistical interpretation of numerical weather prediction model output. *Journal of Weather and Forecasting* 18:938-952.
- [5] World Meteorological Organization. (1999). *Technical Note No. 195, Method of Interpreting Numerical Weather Prediction Output for Aeronautical Meteorology*. WMO-No 770, Secretariat WMO, Geneva, Switzerland.
- [6] Baker, Kirk. (2005). Singular Value Decomposition Tutorial, diakses 4 September 2012.
- [7] Storch H von and Navarra. (1993). *Analysis of statistical variability, Application of Statistical Techniques*. Proceeding of an Autumn School 30 to November 6.
- [8] Sanjaya I. (2010), *Penentuan prediktor pada statistical downscaling dengan singular value decomposition (studi kasus di Stasiun Meteorologi Indramayu)*. Skripsi, Fakultas Matematika dan Ilmu Pengetahuan Alam: Institut Pertanian Bogor.
- [9] Bjornsson H. and Venegas SA. (1997). *A Manual for EOF and SVD Analyses of Climatic Data*. Department of Atmospheric and Oceanic Sciences and Center for Climate and Global Change Research, McGill University.
- [10] Termonia P. and Deckmyn A. (2007). Model-inspired predictors for Model Output Statistics (MOS). *Monthly Weather Review Vol.* 135:3496-3505, American Meteorological Society.

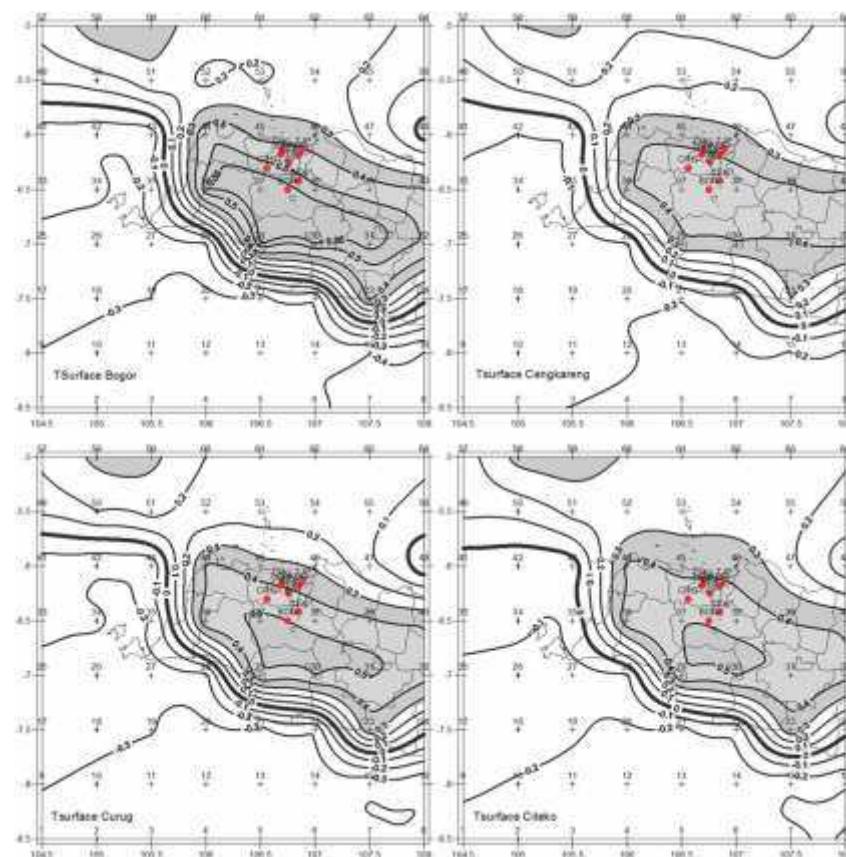
Lampiran 1a. Isokorelasi suhu maksimum Stasiun Bogor, Cengkareng, Curug dan Citeko dengan suhu maksimum NWP GFS pukul 06 UTC



Lampiran 1b. Isokorelasi suhu maksimum Kemayoran, Pondok Betung, Tangerang dan Tanjung Priok dengan suhu maksimum NWP GFS pukul 06 UTC



Lampiran 2a. Isokorelasi suhu maksimum Stasiun Bogor, Cengkareng, Curug dan Citeko dengan suhu permukaan NWP GFS pukul 06 UTC



Lampiran 2b. Isokorelasi suhu maksimum Kemayoran, Pondok Betung, Tangerang dan Tanjung Priok dengan suhu permukaan NWP GFS pukul 06 UTC

