

**DAMPAK PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA)
DALAM CLUSTERING CURAH HUJAN DI PULAU JAWA, BALI,
DAN LOMBOK
[IMPACT OF PRINCIPAL COMPONENT ANALYSIS (PCA)
IMPLEMENTATION ON RAINFALL CLUSTERING OVER JAVA,
BALI AND LOMBOK ISLANDS]**

Ina Juaeni

Peneliti Pusat Sains dan Teknologi Atmosfer, Lapan

e-mail: inajuaeni@yahoo.com; ina_j@bdg.lapan.go.id

Diterima 10 April 2014; Disetujui 13 Juni 2014

ABSTRACT

PCA is a mathematical procedure that uses an orthogonal transformation to convert a set of data of possibly correlated variables into a set of values of linearly uncorrelated variables. This could have an impact on the rainfall clustering in Java, Bali, and Lombok islands. This study uses TRMM rainfall data every three hours for 13 years (1998 to 2010) that are grouped into annual average, monthly climatology (January to December), and average of the whole 13 years. Multicollinearity test is also conducted on those three groups before applying PCA on the clustering process. The results show that clustering with PCA reduces the number of clusters and changes the spatial distribution of clusters.

Key words: *Principal Component Analysis, Cluster, Rainfall*

ABSTRAK

Analisis komponen utama atau *Principal Component Analysis* (PCA) adalah prosedur matematik yang menggunakan teknik transformasi *orthogonal* untuk mengubah sekumpulan data dengan komponen yang mungkin saling berhubungan menjadi komponen yang tidak saling berkaitan. Hal ini bisa memberikan dampak pada *clustering* curah hujan di Jawa, Bali, dan Lombok. Penelitian ini menggunakan data curah hujan TRMM setiap 3 jam selama 13 tahun (dari tahun 1998 sampai 2010) yang diolah menjadi rata-rata tahunan, klimatologi bulanan (Januari – Desember), rata-rata selama 13 tahun. Tes multikolinieritas juga dilakukan pada ketiga jenis data ini sebelum melakukan *clustering* dengan PCA. Hasilnya menunjukkan bahwa *clustering* dengan PCA mengurangi jumlah *cluster* dan mengubah distribusi spasial *cluster* curah hujan.

Kata kunci: *Analisis Komponen Utama, Cluster, Curah hujan*

1 PENDAHULUAN

Data satelit *Tropical Rainfall Measuring Mission* (TRMM) sebagai data spasial dengan cakupan luas dan resolusi waktu yang cukup tinggi bisa menjadi solusi untuk kajian/penelitian tentang informasi curah hujan yang bermanfaat bagi sektor-sektor lainnya. Tersedianya data curah hujan TRMM dengan resolusi waktu 3 jam memberi peluang untuk mengidentifikasi proses-proses atmosfer dalam rentang waktu tersebut. Data TRMM ini memiliki resolusi $0,25^{\circ} \times 0,25^{\circ}$, maka setiap hari ada $8 \times 20 \times 16$ data curah hujan. Jika data ini akan dianalisis secara langsung, maka jumlah data yang sangat banyak bisa memberikan kesulitan dalam pengolahan data. Oleh karena itu

pengurangan jumlah data kadang perlu dilakukan. Salah satu cara untuk mengurangi volume data tanpa mengurangi informasi penting dari data tersebut adalah *Principal Component Analysis* (PCA). PCA adalah prosedur matematik yang menggunakan teknik transformasi *orthogonal* untuk mengubah sekumpulan data yang mungkin saling berhubungan menjadi komponen yang tidak saling berkaitan.

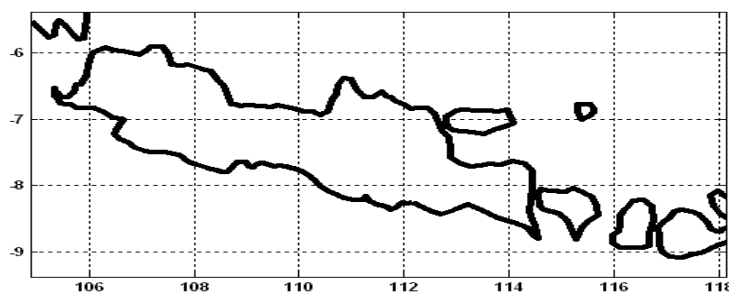
Dalam makalah ini, PCA digunakan sebagai salah satu tahap dalam proses pengelompokan (*clustering*) curah hujan. Selain PCA ada beberapa tahap lain yang harus dilakukan pada *clustering* curah hujan yaitu uji multi kolinieritas, penentuan jumlah *cluster* dan interpretasi. Masing-masing tahap dijelaskan lebih rinci pada seksi Data dan Metodologi. Pengelompokan curah hujan itu sendiri dilakukan dengan metode Ward (Ward, 1963), yaitu teknik untuk memperoleh *cluster* yang memiliki variansi internal sekecil mungkin. Juaeni dkk. (2010) telah menggunakan metode ini untuk pengelompokan curah hujan bulanan. Sebelumnya, Haryoko (2009), mengaplikasikan analisis *cluster* untuk mengelompokkan pos pengamatan hujan (stasiun) yang mempunyai kesamaan pola curah hujan dasarian (10 harian) ke dalam sub-sub *cluster*.

Dalam penelitian ini *clustering* (jumlah dan distribusi spasial *cluster*) dilakukan pada data curah hujan dengan dan tanpa penerapan PCA untuk melihat apakah PCA mempunyai dampak terhadap hasil *clustering*. Secara lengkap, tujuan penelitian ini adalah membandingkan jumlah *cluster* dengan penerapan PCA dan jumlah *cluster* tanpa PCA dengan beberapa klasifikasi data (rata-rata bulan, rata-rata tahun dan rata-rata 13 tahun). Selain itu penelitian ini juga melakukan identifikasi faktor dominan yang mempengaruhi *cluster* curah hujan di Pulau Jawa, Bali, dan Lombok. Pengklasifikasian data dimaksudkan untuk mendapatkan informasi yang sebanyak-banyaknya tentang variasi jumlah *cluster*.

2 DATA DAN METODOLOGI

2.1 Data

Data yang digunakan dalam penelitian ini adalah data curah hujan per 3 jam (waktu sinoptik) dari satelit TRMM dengan resolusi $0,25^\circ \times 0,25^\circ$ yang diunduh dari http://disc2.nascom.nasa.gov/Giovanni/tovas/TRMM_V6.3B42.2.shtml. Rentang waktu datanya adalah dari Januari 1998 sampai dengan Desember 2010. Domain penelitian dibatasi pada wilayah yang meliputi Pulau Jawa, Bali dan Lombok ($5,5^\circ - 9,5^\circ$ LS, $105^\circ - 118^\circ$ BT) atau wilayah seluas $4^\circ \times 13^\circ$ ($444 \text{ km} \times 1443 \text{ km}$) seperti tampak pada Gambar 2-1.

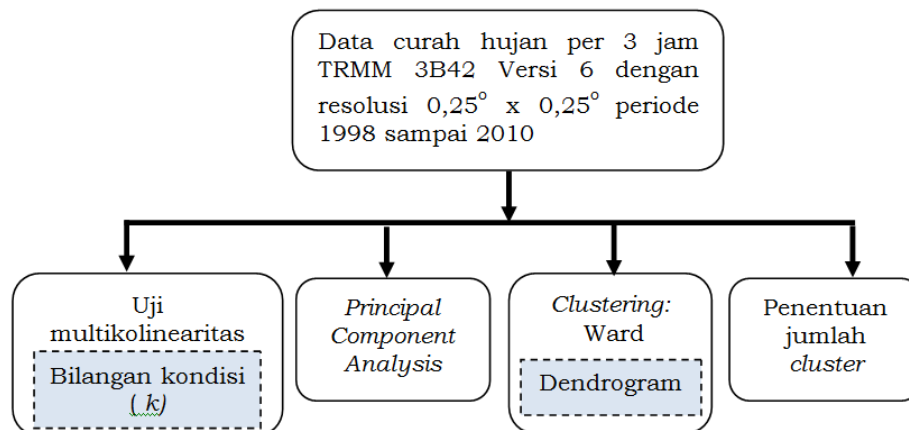


Gambar 2-1: Batasan wilayah penelitian

Data tersebut kemudian dibagi menjadi tiga klasifikasi, yaitu rata-rata tahunan (tahun 1998 sampai dengan tahun 2010), klimatologi bulanan (Januari sampai dengan Desember), dan rata-rata 13 tahun. Klasifikasi pertama dan kedua dibuat untuk menunjukkan variasi hasil *clustering* terhadap waktu (tahun dan bulan), sedangkan klasifikasi ketiga untuk melihat kondisi secara umum. Setelah pemrosesan awal ini, uji multikolinieritas dilakukan pada ketiga klasifikasi data ini untuk melihat keterkaitan

antar data. Tahap-tahap berikutnya adalah melakukan PCA, *clustering*, dan penentuan jumlah *cluster*. Tahap-tahap pengolahan data ini secara ringkas dapat dilihat pada Gambar 2-2.

Selain itu kaitan antara curah hujan dengan jumlah *cluster* dan antara jumlah *cluster* dengan *Southern Oscillation Index* (SOI) dikaji untuk menentukan faktor-faktor yang berpengaruh terhadap *clustering*. Dalam hal ini ada dua SOI yang digunakan, yaitu SOI Pasifik (SOI untuk lautan Pasifik) dan SOI Indonesia (SOI untuk laut di wilayah Indonesia). Data SOI (baik Pasifik maupun Indonesia) diperoleh dari <http://www.cpc.ncep.noaa/data/indices>.



Gambar 2-2: Bagan alur penentuan *cluster* curah hujan 3 jam-an dengan metode Ward

2.2 Principal Component Analysis (PCA)

PCA diterapkan pada data yang mempunyai multikolinieritas, karena multikolinieritas antar variabel adalah salah satu pelanggaran asumsi dalam analisis *cluster* (Hair *et al.*, 1998). Multikolinieritas adalah suatu keadaan dimana terdapat hubungan linier sempurna atau hampir sempurna antara beberapa atau semua variabel. Salah satu cara untuk mendeteksi ada tidaknya multikolinieritas adalah dengan menggunakan bilangan kondisi (k). Bilangan kondisi ditentukan dengan rumus:

$$k = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (2-1)$$

λ adalah nilai eigen dari matriks kovariansi variabel.

Batas-batas bilangan kondisi untuk mendiagnosa multikolinieritas (Hair *et al.*, 1998) adalah sebagai berikut:

- $k < 100$; terjadi multikolinieritas lemah
- $100 \leq k \leq 1000$; terjadi multikolinieritas sedang sampai kuat
- $k > 1000$; terjadi multikolinieritas sangat kuat

Jika data mempunyai multikolinieritas antar variabel, maka PCA diterapkan terlebih dahulu pada data curah hujan TRMM. Dengan demikian sejumlah komponen utama (*principal component*) yang saling *orthogonal* akan terbentuk. Komponen utama ini yang dijadikan sebagai variabel baru untuk input dalam analisis *cluster*.

Johnson and Wichern (1992) mendefinisikan komponen utama sebagai salah satu bentuk transformasi variabel yang merupakan kombinasi linier dari variabel. Proses pembentukan komponen utama adalah dengan menentukan matriks yang merupakan data pengamatan curah hujan. Dari matriks X , kemudian menghitung matriks kovariansi, untuk menentukan nilai eigen (λ). Berdasarkan matriks eigen, maka komponen utama (PC) yang terbentuk adalah:

$$\begin{aligned}
 PC_1 &= z_j^v j_1 = z_1^v v_{11} + z_2^v v_{21} + \dots + z_p^v v_{j1} \\
 PC_2 &= z_j^v j_2 = z_1^v v_{12} + z_2^v v_{22} + \dots + z_p^v v_{j2} \\
 &\cdot \\
 &\cdot \\
 &\cdot \\
 PC_p &= z_j^v j_p = z_1^v v_{1p} + z_2^v v_{2p} + \dots + z_p^v v_{jp}
 \end{aligned}
 \tag{2-2}$$

di mana PC_1 = komponen utama pertama, PC_2 = komponen utama kedua dan seterusnya.

Kriteria yang digunakan untuk menentukan jumlah komponen yang dapat dibentuk adalah kriteria persentase variansi. Jumlah komponen utama yang akan digunakan dalam analisis *cluster* adalah yang memiliki persentase kumulatif variansi minimal 80% (Rencher, 2001). Tahap selanjutnya adalah menghitung komponen skor (PC_j) yang akan digunakan sebagai *input* untuk analisis *cluster*. Komponen skor yang diperoleh dari m komponen utama (dimana $m < p$) akan digunakan dalam analisis selanjutnya sebagai pengganti nilai data variabel awal. Komponen skor dari hasil analisis komponen utama dengan data asli (*raw data*) sebagai input analisis ditentukan dengan:

$$\begin{aligned}
 y_{i1} &= e_1 x_i \\
 y_{i2} &= e_2 x_i \\
 &\cdot \\
 &\cdot \\
 &\cdot \\
 y_{ik} &= e_k x_i
 \end{aligned}
 \tag{2-3}$$

(Johnson and Wichern, 1992)

2.3 Penentuan Jumlah Cluster

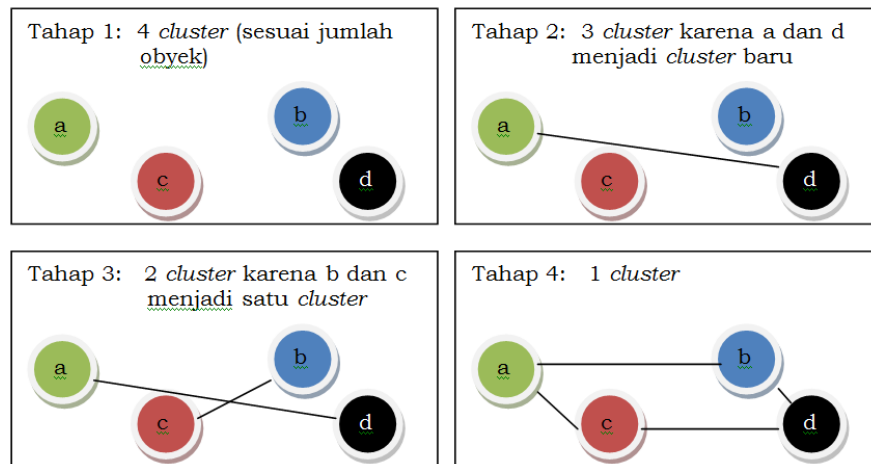
Metode Ward digunakan untuk menentukan *cluster*. Teknik ini memperoleh *cluster* dengan memilih wilayah yang memiliki variansi internal sekecil mungkin. Jumlah *cluster* awal dalam metode ini diperoleh dari dendrogram, yaitu berupa gambaran grafik (diagram pohon) di mana setiap obyek disusun pada satu sumbu dan sumbu lainnya menggambarkan langkah-langkah pada prosedur hierarkhi. Pada tahap awal, setiap obyek digambarkan sebagai *cluster* yang masih terpisah. Dendrogram menunjukkan secara grafik bagaimana *cluster-cluster* bergabung pada tiap tahap prosedur hingga semua obyek terkandung dalam satu *cluster*. Ukuran yang digunakan dalam metode Ward adalah *sum square error* (SSE) (Ward, 1963; Rencher, 2001) setiap observasi terhadap rata-rata *cluster* dimana observasi itu berada.

$$SSE = \sum_{j=1}^p \left(\sum_{i=1}^n X_{ij}^2 - \frac{1}{n} \left(\sum_{i=1}^n X_{ij} \right)^2 \right)$$

X_{ij} adalah nilai pengamatan ke i dalam variabel ke j
 p adalah banyaknya variabel yang diukur
 n adalah banyaknya objek dalam *cluster*

(2-4)

SSE akan bernilai nol untuk tahap pertama karena setiap obyek atau data akan membentuk *cluster*. Kemudian semua *cluster* dengan dua anggota akan digabungkan sehingga menghasilkan SSE yang paling kecil. Meminimumkan nilai SSE sama dengan meminimumkan jarak antar *cluster*. Gambar 2-3 memperlihatkan tahapan pembentukan *cluster* dengan metode Ward.



Gambar 2-3: Ilustrasi metode Ward

3 HASIL DAN PEMBAHASAN

Tingkat multikolinieritas data ditentukan oleh bilangan kondisi k yang diperoleh berdasarkan persamaan 2-1. Hasilnya ditunjukkan pada Tabel 3-1.

Tabel 3-1: BILANGAN KONDISI (k) CURAH HUJAN TRMM PER 3 JAM PULAU JAWA, BALI, DAN LOMBOK

Data Rata-rata Tahun	Bilangan kondisi (k)	Data Rata-rata Bulan	Bilangan kondisi (k)
1998	123,45	Januari	117,99
1999	103,69	Februari	139,87
2000	58,04	Maret	162,78
2001	86,51	April	133,47
2002	53,15	Mei	68,06
2003	62,72	Juni	30,48
2004	62,93	Juli	27,82
2005	48,26	Agustus	36,27
2006	60,14	September	47,61
2007	73,79	Oktober	90,89
2008	63,19	November	147,28
2009	85,18	Desember	138,02
2010	100,00		
Seluruh data (rata-rata 13 tahun)	473,63		

Tabel 3-1 menunjukkan bahwa nilai k mempunyai rentang antara 48,26 sampai dengan 123,45 untuk data curah hujan rata-rata tahunan, antara 27,82 sampai dengan 162,78 untuk data curah hujan klimatologi bulanan, dan 473,63 untuk data curah hujan rata-rata 13 tahun (1998 sampai dengan 2010). Berdasarkan nilai k tersebut, multikolinieritas lemah terjadi pada klasifikasi rata-rata tahunan dan klimatologi bulanan, sedangkan data rata-rata 13 tahun mempunyai multikolinieritas sedang.

Analisis komponen utama kemudian diterapkan sebelum data curah hujan TRMM per 3 jam melalui tahap *clustering* untuk semua klasifikasi data dan tingkat multikolinieritas. Kriteria persentase variansi digunakan untuk menentukan berapa komponen yang dapat dibentuk. Kemudian jumlah *cluster* ditentukan dengan menggunakan jarak Euclidian dan dendrogram. Jumlah *cluster* untuk masing-masing klasifikasi data diperlihatkan pada Tabel 3-2.

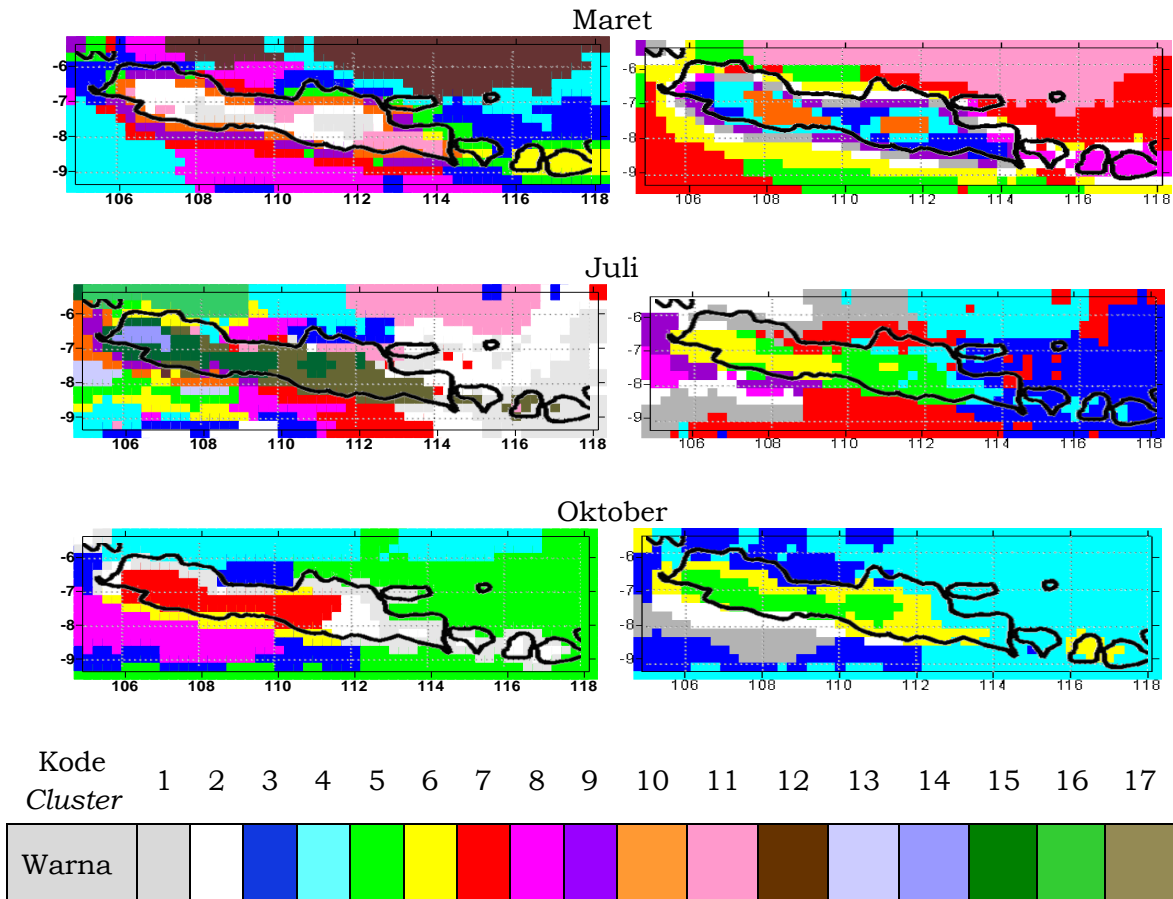
Jika *clustering* dilakukan tanpa PCA, hasilnya menunjukkan jumlah *cluster* yang lebih banyak (Tabel 3-2) dibandingkan jumlah *cluster* pada *clustering* dengan PCA. Jumlah *cluster* tahunan tanpa PCA adalah 6 sampai 14, sedangkan jumlah *cluster* dengan PCA untuk data yang sama antara 5 sampai 8. Untuk data klasifikasi klimatologi bulanan, jumlah *cluster* tahunan tanpa PCA adalah 7 sampai 14, sedangkan jumlah *cluster* dengan PCA untuk data yang sama antara 5 sampai 12. Demikian pula untuk seluruh data, jika diterapkan analisis komponen utama maka jumlah *cluster* berkurang menjadi 5. Dengan demikian, penerapan PCA mengurangi jumlah *cluster* sebesar 12 % sampai 64 % terhadap jumlah *cluster* tanpa PCA.

Tabel 3-2: JUMLAH *CLUSTER* CURAH HUJAN TRMM 3 JAM-AN PULAU JAWA, BALI DAN LOMBOK TANPA PCA DAN DENGAN PCA

Data Rata-rata Tahun	Jumlah <i>cluster</i>		Data Rata-rata Bulan	Jumlah <i>cluster</i>	
	tanpa PCA	dengan PCA		tanpa PCA	dengan PCA
1998	10	8	Januari	10	5
1999	7	6	Februari	8	7
2000	7	6	Maret	12	11
2001	8	8	April	8	6
2002	7	5	Mei	7	6
2003	7	6	Juni	13	12
2004	6	5	Juli	17	9
2005	9	8	Agustus	11	9
2006	8	5	September	14	7
2007	10	6	Oktober	8	6
2008	11	7	November	10	8
2009	9	5	Desember	14	10
2010	14	5			
Seluruh data (rata-rata 13 tahun)	8	5			

Faktor dominan yang berpengaruh pada jumlah dan distribusi spasial *cluster* diidentifikasi melalui distribusi spasial *cluster* yang terbentuk, pola temporal curah hujan dan SOI. Distribusi *cluster* secara spasial untuk tiga klasifikasi data yang dipakai dalam penelitian ini menunjukkan hasil serupa. *Cluster* yang terbentuk sesuai dengan bentuk dan lokasi permukaan yaitu *cluster* laut sebelah selatan, *cluster* pantai selatan, *cluster* daratan, *cluster* pantai utara dan *cluster* laut.

Gambar 3-1 memperlihatkan pola spasial *cluster* curah hujan rata-rata bulan untuk bulan Maret, Juli dan Oktober masing-masing untuk nilai *k* terbesar, terkecil dan sedang. Warna membedakan *cluster* satu dengan lainnya, *cluster* 1 diberi warna abu-abu, *cluster* 2 diberi warna putih dan seterusnya. Kode warna berlaku untuk semua gambar *cluster*. Dengan demikian jumlah warna menunjukkan jumlah *cluster*.



Gambar 3-1: Perbandingan distribusi dan jumlah cluster tanpa PCA (kiri) dengan distribusi dan jumlah cluster dengan PCA (kanan) untuk bulan Maret untuk bulan Juli dan Oktober

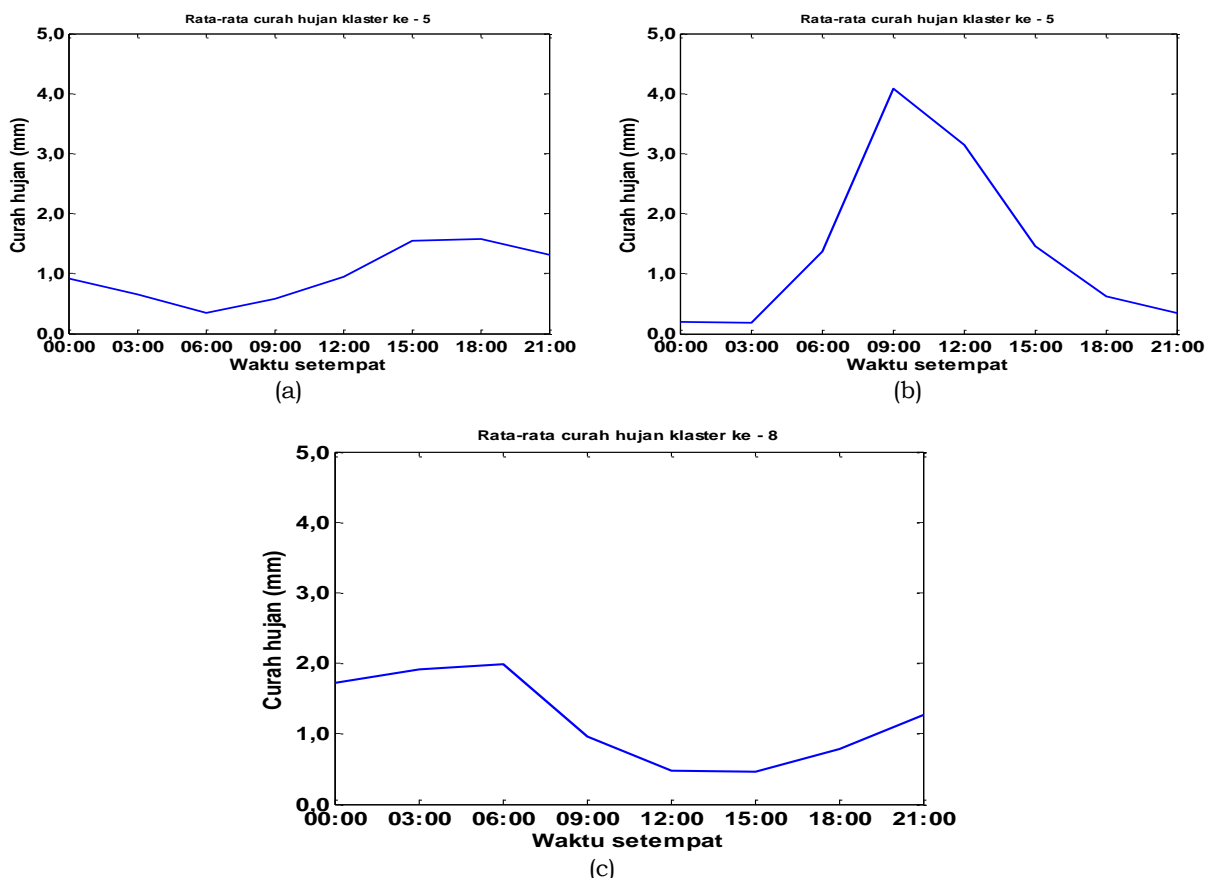
Gambar 3-1 juga menunjukkan adanya *cluster* yang mencakup dua wilayah berbeda yaitu daratan dan laut. Hal ini menunjukkan bahwa sifat hujan di dua wilayah tersebut sama, yang biasanya didominasi oleh hujan yang berasal dari awan-awan yang terbentuk di pinggiran pantai pada pagi hari. *Cluster* di daratan jumlahnya lebih banyak dibandingkan *cluster* di laut. Hal ini menunjukkan bahwa variasi bentuk permukaan (topografi dan jenis tutupan lahan) yang berkaitan dengan variasi curah hujan menekan meningkatnya jumlah *cluster* di daratan. Curah hujan di daratan terutama di Pulau Jawa berkelompok sesuai jaraknya terhadap laut atau dengan kata lain mengikuti garis pantai sedangkan *cluster* di laut di sekitar pulau-pulau tersebut hampir homogen (satu *cluster*). Ini menunjukkan bahwa komponen utama pola curah hujan spasial adalah perbedaan suhu di permukaan. Pola curah hujan yang mengikuti garis pantai juga ditemukan oleh Arakawa and Kitoh (2005) dengan data TRMM 3G68 yang mempunyai resolusi ruang $0,5^{\circ} \times 0,5^{\circ}$ dan resolusi waktu satu jam.

Perbandingan gambar spasial *cluster* setiap bulan pada Gambar 3-1 juga menunjukkan bahwa penerapan PCA mengubah jumlah *cluster* dan berakibat pada berubahnya distribusi spasial *cluster*. Distribusi *cluster* tanpa PCA (Gambar 3-1 sebelah kiri) memberikan gambaran yang lebih rinci tentang kelompok atau *cluster* curah hujan di Pulau Jawa, Bali, dan Lombok. Dengan demikian penggunaan analisis komponen utama atau PCA selain ditentukan oleh kriteria bilangan kondisi (persamaan 2-1) dapat ditentukan berdasarkan kebutuhan analisis yang diinginkan.

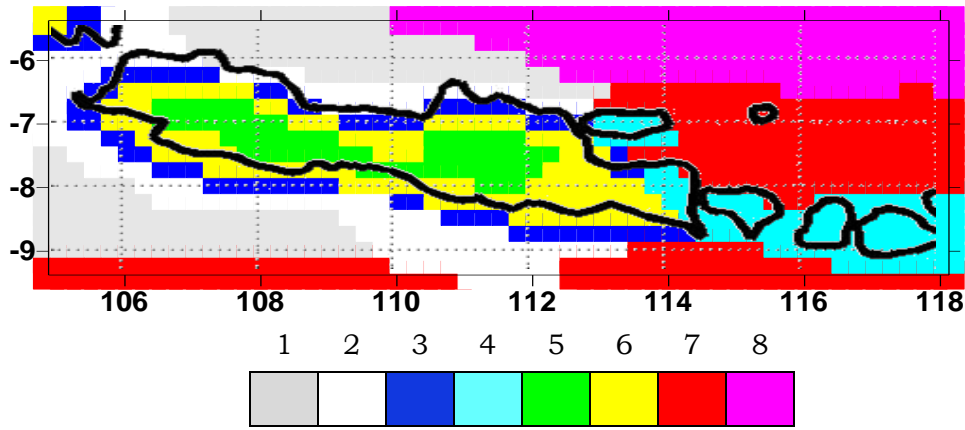
Pola temporal curah hujan untuk *cluster-cluster* darat mengikuti pola sore hari (*late afternoon*). Pola sore hari adalah pola curah hujan yang nilai maksimumnya terjadi pada sore hari, sekitar pukul 15.00 WIB. Contohnya adalah *cluster* 5 atau *cluster* warna hijau pada *cluster* tanpa PCA bulan Maret, seperti diperlihatkan pada Gambar 3-2a. Pola

sore hari menjadi indikator bahwa proses konveksi terjadi akibat perbedaan panas radiasi matahari di permukaan yang mencapai maksimum pada sore hari. Pola curah hujan di pantai adalah pola jelang siang hari (*late morning*), sedangkan curah hujan di laut mempunyai pola larut malam atau *late night* (Lim and Kwon, 1998, Lim and Suh, 2000). Pola larut malam adalah pola curah hujan yang nilai maksimumnya terjadi larut malam (sekitar pukul 01.00 sampai pukul 06.00 WIB) seperti terjadi pada *cluster* 5 (*cluster* warna hijau) (Gambar 3-2b) untuk *cluster* tanpa PCA rata-rata bulan Januari, sedangkan pola *late morning* adalah pola curah hujan yang nilai maksimumnya terjadi pagi hari (setelah pukul 06.00 sampai 11.00 WIB), seperti pada *cluster* 8 atau *cluster* merah muda dalam *cluster* tanpa PCA rata-rata bulan Februari (Gambar 3-2c).

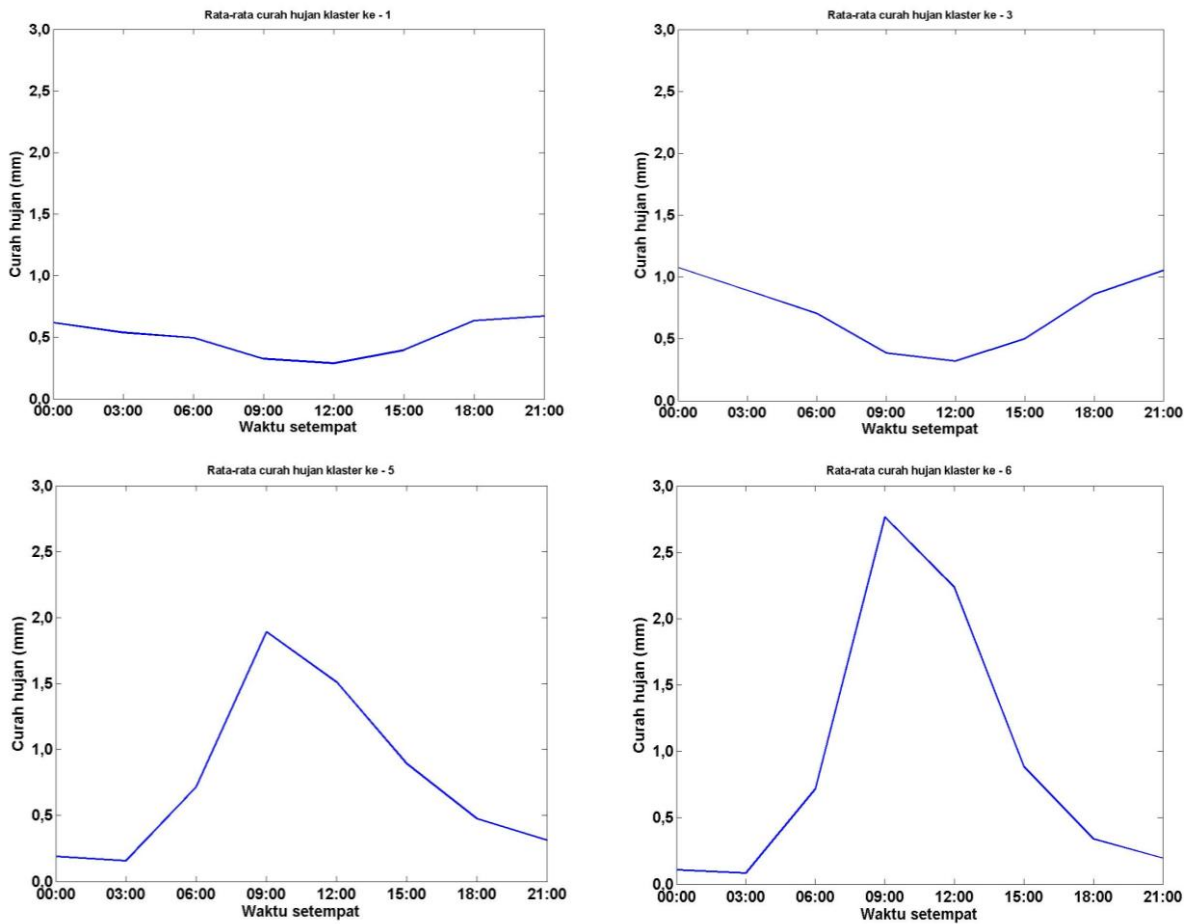
Di wilayah pantai, terjadi perbedaan panas antara daratan dan laut pada pagi hari. Perbedaan ini menjadi pemicu terbentuknya awan dan hujan di wilayah tersebut. Sedangkan di wilayah lepas pantai, proses konveksi aktif terjadi pada malam hari dan mencapai puncaknya pada dini hari, dengan ditandai oleh curah hujan maksimum pada dini hari sampai pagi hari. Curah hujan maksimum di daratan lebih tinggi dibandingkan di laut, sebagai contoh Gambar 3-3 dan Gambar 3-4 memperlihatkan pola curah hujan di seluruh *cluster* untuk klasifikasi seluruh data tanpa PCA (8 *cluster*, lihat Tabel 3-2). *Cluster* 1, 2, 3, 7 dan 8 atau *cluster* yang berwarna abu-abu, putih, biru tua, merah dan merah muda yang berada di laut mempunyai curah hujan maksimum yang lebih rendah dibandingkan *cluster-cluster* 4, 5, dan 6 atau *cluster-cluster* berwarna hijau, kuning dan biru muda yang berada di daratan. Hal yang sama juga ditunjukkan oleh Arakawa and Kitoh (2005).



Gambar 3-2: (a) Pola *late afternoon* di *cluster* 5 untuk data rata-rata bulan Maret, (*cluster* warna hijau pada Gambar 3-1, Maret), (b) Pola *late morning* di *cluster* 5 untuk data rata-rata bulan Januari (*cluster* warna hijau pada Gambar 3-1, Januari), (c) Pola *late night* di *cluster* 8 untuk data rata-rata bulan Februari (*cluster* warna merah muda pada Gambar 3-1, Februari)



Gambar 3-3: Distribusi spasial *cluster* untuk klasifikasi seluruh data tanpa PCA



Gambar 3-4: Pola curah hujan (mm) per 3 jam di masing-masing *cluster* (*cluster* 1 dan 3 yang berada di laut serta *cluster* 5 dan 6 yang berada di daratan)

Dengan demikian, dapat diidentifikasi tiga faktor yang mempengaruhi curah hujan per 3 jam yang nampak dari hasil *cluster*, spasial maupun temporal, yaitu suhu permukaan, bentuk permukaan dan kontras suhu di permukaan.

Koefisien korelasi antara curah hujan dengan jumlah *cluster* dan antara jumlah *cluster* dengan SOI (SOI Pasifik dan SOI Indonesia) digunakan untuk mengidentifikasi faktor lain yang berpengaruh terhadap *clustering* curah hujan di Pulau Jawa, Bali, dan Lombok. Hasilnya diperlihatkan pada Tabel 3-3. Jumlah curah hujan tahunan mempunyai korelasi yang cukup tinggi (0,74) dengan jumlah *cluster* tahunan tanpa PCA. Sementara jumlah *cluster* tahunan tanpa PCA relatif lebih berkaitan dengan SOI lautan Pasifik yang ditunjukkan dengan koefisien korelasi 0,50.

Tabel 3-3: KOEFISIEN KORELASI

Koefisien korelasi	Jumlah curah hujan tahunan	Jumlah curah hujan bulanan	SOI Indonesia	SOI Pasifik
Jumlah <i>cluster</i> tahunan dengan PCA	0,17	-	0,12	0,03
Jumlah <i>cluster</i> tahunan tanpa PCA	0,74	-	-0,25	0,50
Jumlah <i>cluster</i> bulanan dengan PCA	-	-0,15	-0,23	-0,11
Jumlah <i>cluster</i> bulanan tanpa PCA	-	-0,30	-0,14	0,24

4 KESIMPULAN

Jumlah *cluster* tanpa PCA lebih banyak dibandingkan jumlah *cluster* hasil *clustering* dengan penerapan PCA baik dengan data rata-rata bulan, data rata-rata tahun maupun untuk seluruh data (rata-rata 13 tahun). Penerapan PCA mengurangi jumlah *cluster* dari 12 % sampai 64 % terhadap jumlah *cluster* tanpa PCA. Sementara itu distribusi spasial dari *cluster* juga berubah akibat penerapan PCA. Dengan demikian maka dampak penerapan PCA pada *clustering* curah hujan TRMM adalah berkurangnya jumlah *cluster* curah hujan dan berubahnya distribusi spasial.

Kajian terhadap faktor dominan yang mempengaruhi *cluster* curah hujan di Pulau Jawa, Bali dan Lombok yang ditentukan secara kualitatif dan kuantitatif. Secara kualitatif, distribusi spasial *cluster* menunjukkan bahwa suhu permukaan, keragaman atau bentuk permukaan dan kontras suhu permukaan adalah faktor penting dalam pembentukan *cluster* curah hujan. Sementara untuk kajian kuantitatif, faktor suhu muka laut baik lautan Pasifik maupun laut di sekitar wilayah Indonesia tidak memberikan pengaruh yang berarti pada *clustering*. Kajian faktor dominan yang berpengaruh terhadap *clustering* dengan koefisien korelasi menunjukkan bahwa korelasi tertinggi ditunjukkan oleh korelasi antara jumlah *cluster* tahunan tanpa PCA dengan jumlah curah hujan tahunan. Jadi fenomena atmosfer yang terjadi di laut (baik lautan Pasifik maupun laut wilayah Indonesia) tidak akan berdampak pada *clustering* selama fenomena tersebut tidak mempengaruhi jumlah curah hujan dalam skala tahunan.

DAFTAR RUJUKAN

- Arakawa, O. and Kitoh, A.; 2005. *Rainfall Diurnal variation over the Indonesian Maritime Continent Simulated by 20 km-mesh GCM, SOLA, 1*, 109-112.
- Hair, J. F.; Rolph, E. A.; Ronald, L. T. and William, C. B., 1998. *Multivariate Data Analysis*, Fifth Edition, Prentice Hall, New Jersey, 768 pp.
- Haryoko, U., 2009. *Pewilayahan Hujan untuk Menentukan Pola Hujan (contoh kasus Kabupaten Indramayu)*; <http://www.staklimpondokbetung.net/publikasi> diunduh pada Juli 2009.
- Johnson, R. A. and Wichern, D. W., 1992. *Applied Multivariate Statistical Analysis*, Prentice

- Hall, New Jersey, 642 pp.
- Juaeni, I.; Yuliani, D.; Ayahbi, R.; Noersomadi; Harjana, T. dan Nurzaman, 2010. *Pengelompokan Wilayah Curah Hujan Kalimantan Barat Berbasis Metoda Ward dan Fuzzy Clustering*, Jurnal Sains Dirgantara, 7, 82-99.
- Lim, G. H. and Kwon, H. J., 1998. *Diurnal Variation of Precipitation Over South Korea and its Implication*, Journal Korean Meteorological Society, 34, 222 – 237.
- Lim, G. H. and Suh, A. E., 2000. *Diurnal and Semidiurnal Variations in the Time Series of 3-Hourly Assimilated Precipitation by NASA GEOS-1*, American Meteorological Society, Journal Climate, 13, 2923–2940.
- Rencher, A. C., 2001. *Method of Multivariate Analysis*. Second Edition. A Wiley-Interscience Publication, United States, 727 pp.
- Ward Jr., J. H., 1963. *Hierarchical Grouping to Optimize an Objective Function*, Journal of the American Statistical Association, 58, 236–244.

