

PENGELOMPOKAN WILAYAH CURAH HUJAN KALIMANTAN BARAT BERBASIS METODE *WARD* DAN *FUZZY CLUSTERING*

Ina Juaeni, Dewi Yullani, Rlsana Ayahbi, Noersomadi,
Teguh Hardjana, dan Nurzaman
Peneliti Pusat Pemanfaatan Sains Atmosfer dan Iklim, LAPAN
E-mail : inajuaeni@yahoo.com

ABSTRACT

In order to decrease the climate information gap, the rainfall clustering based on the TRMM data is presented. Then, the rainfall pattern could be determined for all region. Using the principal component analysis as the interface, clustering analysis, namely the Ward and the Fuzzy Clustering methods, the rainfall in West Kalimantan could be grouped in to four clusters according to its homogeneity. The first cluster consists of grids that gather in the Karimata strait, Java Sea, and some area in West Kalimantan Lands. The second group was built by grids in the land with homogeneous topography. The third group in the middle of West Kalimantan was a cluster in the land which has different elevation. The fourth group was a cluster in the northern Kalimantan and close to the South China Sea. The West Kalimantan rainfall derived from the TRMM data has the equatorial and monsoonal patterns. The monsoonal pattern was exhibited by the cluster in the ocean (first cluster), and the equatorial pattern was showed by clusters in the land (second, third, and fourth clusters). The equatorial rainfall pattern revealed that the first wet month occuring in April, whereas the second wet month occured in December or January. The month with lowest rainfall is August.

Key word: *Rainfall pattern, Rainfall cluster, Ward and Fuzzy Clustering methods*

ABSTRAK

Dalam rangka mengurangi kesenjangan informasi iklim, disajikan pengelompokan wilayah curah hujan berdasarkan data TRMM, sehingga pola curah hujan dapat ditentukan untuk seluruh wilayah. Dengan menggunakan analisis komponen utama sebagai analisis antara dan analisis klaster, yaitu metode *Ward* dan *Fuzzy Clustering*, curah hujan di Kalimantan Barat dapat dibagi menjadi empat kelompok sesuai homogenitasnya. Kelompok pertama terdiri dari grid-grid yang berkumpul di Selat Karimata, Laut Jawa dan sebagian daratan Kalimantan Barat. Kelompok kedua dibangun oleh grid-grid

yang berkumpul di wilayah darat dengan topografi homogen. Kelompok ketiga merupakan klaster yang berada di wilayah bagian tengah Kalimantan Barat dengan elevasi yang berbeda. Kelompok keempat merupakan klaster yang berada di daratan Kalimantan Utara dan berdekatan dengan Laut China Selatan. Curah hujan Kalimantan Barat berbasis data TRMM memiliki pola ekuatorial dan monsun. Pola curah hujan monsun ditunjukkan oleh klaster di lautan (klaster pertama), dan pola ekuatorial ditunjukkan oleh klaster di daratan (klaster 2, 3 dan 4). Untuk pola curah hujan ekuatorial, bulan terbasah pertama terjadi pada bulan April, sedangkan bulan basah kedua terjadi pada bulan Desember atau Januari. Bulan dengan curah hujan terendah adalah bulan Agustus.

Kata kunci: *Pola curah hujan, Klaster curah hujan, Metode Ward and Fuzzy Clustering*

1 PENDAHULUAN

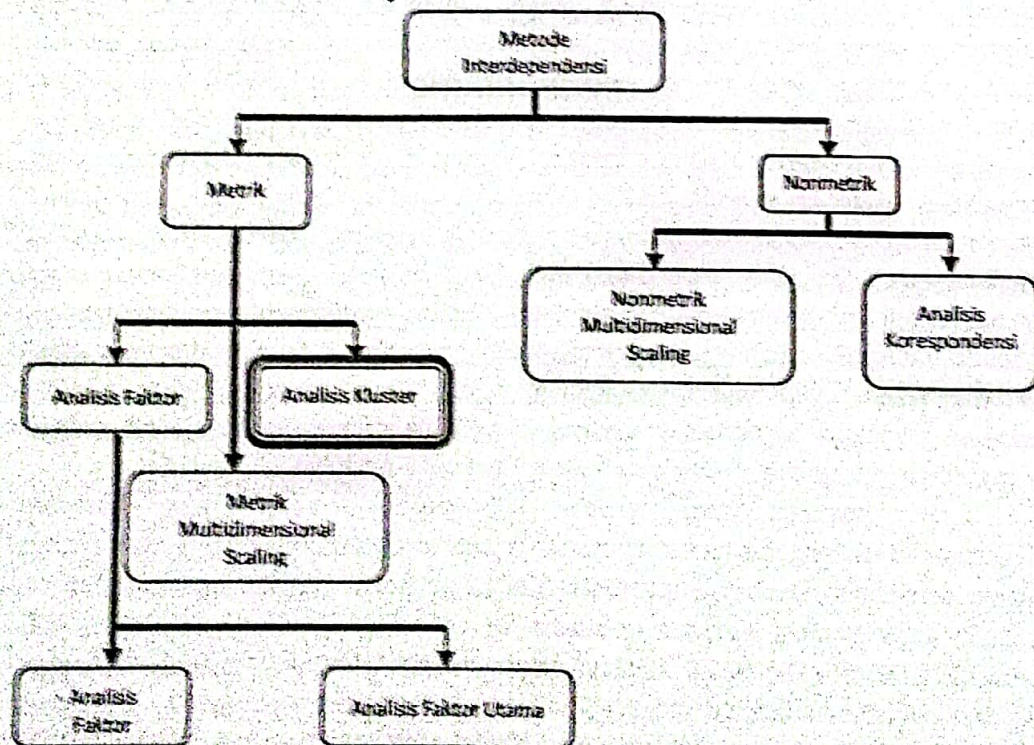
Wilayah Indonesia merupakan bagian wilayah tropis dengan intensitas curah hujan yang tinggi. Sumber energi panas radiasi Matahari yang selalu ada sepanjang tahun ditambah kelembaban dalam jumlah yang cukup tinggi, mendorong aktivitas konveksi dan proses pembentukan awan serta hujan menjadi sangat tinggi. Tidak hanya intensitasnya yang tinggi, curah hujan di wilayah Indonesia juga memiliki variasi spasial dan temporal yang tinggi, hal ini dapat dijelaskan sebagai akibat perbedaan kondisi permukaan, yaitu perbedaan relief dan perbedaan tata guna lahan. Variasi curah hujan yang tinggi ini belum ditunjang oleh sarana observasi yang memadai. Masih banyak lokasi terpencil yang miskin informasi cuaca dan iklimnya, padahal informasi ini penting. Data mentah iklim adalah sarana penunjang penelitian yang hasilnya dimanfaatkan oleh sektor yang terkait kegiatannya dengan kondisi cuaca/iklim.

Penggunaan data satelit merupakan solusi yang banyak digunakan dalam rangka mengurangi kesenjangan informasi cuaca dan iklim tersebut. Berbagai metode dikembangkan untuk mengolah data satelit agar sesuai dengan rencana aplikasinya. Dalam penelitian ini, pengolahan data satelit dilakukan berbasis metode statistik dengan tujuan untuk memperoleh klaster-klaster curah hujan dengan karakteristik yang sama. Data yang digunakan adalah data curah hujan bulanan dari satelit TRMM (*Tropical Rainfall Measuring Mission*). Ada dua informasi penting yang akan diperoleh dari metode pengelompokan ini. Pertama, wilayah yang mempunyai karakter curah hujan yang sama, sehingga lokasi yang tidak memiliki sarana pengamatan dapat melakukan inisialisasi. Kedua, karena pengelompokan berdasarkan

karakter/pola curah hujan maka pola yang diperoleh dapat digunakan untuk membuat suatu kajian perilaku curah hujan atau bahkan untuk prediksi curah hujan di suatu lokasi/wilayah. Sebagai kasus lokasi penelitian dipilih wilayah Kalimantan Barat.

Metode pengklasifikasian yang tepat untuk pengelompokan curah hujan adalah metode interdependensi. Pembagian metode-metode statistik yang termasuk dalam metode interdependensi diperlihatkan pada Gambar 1-1. Dalam metode interdependensi, variabel-variabel yang digunakan tidak dapat diklasifikasikan baik ke dalam variabel bebas maupun tak bebas, semua variabel yang digunakan berstatus sama (Hair et al, 1998).

Data curah hujan bulanan TRMM termasuk data metrik. Dengan demikian ada tiga jenis analisis yang dapat digunakan, yaitu analisis faktor, analisis kluster dan analisis metrik *multidimensional scaling*. Tidak semua teknik statistik interdependensi bisa digunakan untuk data metrik. Masing-masing teknik memiliki tujuan yang berbeda. Analisis kluster merupakan teknik yang tepat untuk mengelompokkan karakteristik curah hujan bulanan TRMM.



Gambar 1-1: Bagan teknik analisis dalam metode Interdependensi

Analisis kluster adalah suatu teknik mengelompokkan variabel menjadi kelompok atau kluster-kluster berdasarkan kesamaan karakteristik variabel tersebut (Sharma, 1996). Hasil dari analisis kluster adalah ditemukannya kelompok-kelompok dengan kemiripan

(homogenitas) yang tinggi di dalam klasternya dan mempunyai ketidakmiripan (heterogenitas) yang tinggi antar klaster. Analisis klaster digunakan untuk mengelompokkan objek ke dalam beberapa kelompok yang memiliki karakteristik yang sama dalam lingkup klimatologi (Mimmack, 2000). Analisis klaster juga digunakan oleh Haryoko (2009) untuk mengelompokkan pos pengamatan hujan (stasiun) yang mempunyai kesamaan pola curah hujan dasarian (10 harian) ke dalam sub-sub kelompok. Analisis komponen utama dan analisis kluster juga digunakan oleh Degaetano (1996), untuk mengelompokkan grid yang memiliki iklim yang sama. Dalam penelitiannya, Degaetano menggunakan curah hujan dan temperatur bulanan beberapa tahun serta menerapkan analisis klaster *Average Linkage* dan metode *Ward*. Pola ukuran klaster yang dihasilkan untuk *Average Linkage* tidak memiliki karakteristik karena dari 47 klaster yang terbentuk, 40% klasternya (19 klaster) masing-masing hanya memiliki dua bahkan satu grid sebagai anggotanya, sedangkan metode *Ward* memberikan hasil yang seragam pada klaster yang terbentuk. Hasil tersebut menjadi latar belakang penelitian ini untuk menggunakan metode *Ward* sebagai teknik klasternya. Sebagai pembanding, pengelompokan curah hujan bulanan di Kalimantan Barat juga menggunakan metode klastering yang lain selain *Ward*. Metode *Ward* adalah salah satu metode klastering Hierarkhi, maka metode pembandingnya dipilih metode klastering non Hierarkhi. Metode non Hierarkhi yang dipilih adalah *Fuzzy Clustering*. Metode ini dipilih didasarkan pada hasil penelitian Klawonn dan Hoppner (2001) yang mengindikasikan bahwa metode *Fuzzy Clustering* merupakan metode yang bagus, karena pusat klaster dan hasil pengelompokan tidak berubah jika ada data baru yang ekstrim. Metode ini juga memberikan hasil yang *smooth* (halus) karena pembobotan yang digunakan berdasarkan himpunan *fuzzy* (Pravitasari, 2008). Kehalusan di sini berarti objek pengamatan tidak mutlak menjadi anggota satu kelompok saja, tapi mungkin menjadi anggota kelompok yang lain dengan ukuran tingkat keanggotaan yang berbeda-beda. Objek akan cenderung menjadi anggota kelompok tertentu dimana tingkat keanggotaan objek dalam kelompok itu paling besar dibandingkan dengan kelompok lainnya.

2 DATA DAN METODE

2.1 Data

Penelitian ini bertujuan untuk mengetahui karakteristik curah hujan di wilayah Kalimantan Barat berdasarkan data curah hujan bulanan TRMM 3B43 (<http://trmm.gsfc.nasa.gov>) dalam periode Januari 1998 sampai Desember 2007.

2.2 Metode

Analisis kluster adalah suatu teknik multivariat yang memiliki tujuan untuk mengelompokkan objek-objek yang mempunyai kesamaan karakteristik tertentu ke dalam kluster-kluster sehingga objek-objek memiliki homogenitas yang tinggi di dalam klasternya dan mempunyai heterogenitas yang tinggi antar kluster (Johnson dan Wichern, 1992).

Langkah-langkah dalam analisis kluster dimulai dengan pendeteksian *outlier*, uji multi kolinearitas, analisis komponen utama, penerapan analisis klasternya itu sendiri dan terakhir validasi dan interpretasi.

2.2.1 Pendeteksian *outlier*

Analisis kluster sensitif terhadap *outlier* (objek yang sangat berbeda dari objek-objek lainnya). Adanya *outlier* dapat menjadikan kluster yang diperoleh tidak merepresentasikan struktur populasi yang sebenarnya. Untuk alasan ini, pendeteksian terhadap *outlier* selalu diperlukan. Pendeteksian *outlier secara* multivariat dapat dilakukan dengan menggunakan jarak Mahalanobis (D^2) kemudian membaginya dengan derajat bebas (df) yang bernilai sama dengan jumlah variabel. Sehingga nilai (D^2/df) mengikuti nilai distribusi t . Kemudian dihitung nilai peluang (signifikansi) dari nilai (D^2/df) tersebut. Data yang signifikansinya lebih kecil dan sama dengan 0,001 dianggap sebagai *outlier* (Hair *et al.*, 1998). Data yang diidentifikasi sebagai *outlier* jika dinilai tidak merepresentasikan populasi, harus dibuang. Namun, perlu diperhatikan apakah penghapusan *outlier* dapat mengubah struktur data yang sebenarnya.

2.2.2 Uji Multikolinearitas

Multikolinearitas antar variabel adalah salah satu pelanggaran asumsi dalam analisis kluster (Hair, *et al.*, 1998). Multikolinearitas adalah suatu keadaan dimana terdapat hubungan linier sempurna atau hampir sempurna antara beberapa atau semua variabel. Salah satu cara untuk mendeteksi ada tidaknya multikolinearitas adalah dengan menggunakan bilangan kondisi. Bilangan kondisi ditentukan dengan rumus:

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

λ adalah nilai eigen dari matriks kovarians variabel.

Batas-batas bilangan kondisi untuk mendiagnosa multikolinearitas adalah sebagai berikut:

- bilangan kondisi < 100; terjadi multikolinearitas lemah
- 100 ≤ bilangan kondisi ≤ 1000; terjadi multikolinearitas sedang sampai kuat
- bilangan kondisi > 1000; terjadi multikolinearitas sangat kuat

Jika setelah dideteksi ternyata diketahui bahwa terdapat multikolinearitas antar variabel, maka diatasi dengan menerapkan analisis komponen utama terlebih dahulu pada data curah hujan bulanan TRMM, sehingga terbentuk sejumlah komponen utama yang saling orthogonal. Komponen utama ini yang dijadikan sebagai variabel baru untuk input dalam analisis kluster.

2.2.3 Analisis Komponen Utama

Johnson dan Wichern (1992) mendefinisikan komponen utama sebagai salah satu bentuk transformasi variabel yang merupakan kombinasi linier dari variabel. Proses pembentukan komponen utama adalah sebagai berikut:

- Matriks X merupakan data pengamatan curah hujan bulanan yang berukuran np dengan, n = objek (grid); n = 1, 2, ..., n

p = variabel (curah hujan bulanan) ; p = 1, 2, ..., p

$$X_{n \times p} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix} \quad (2-2)$$

- Dari matriks X, dicari matriks kovariansnya (S_{pp})

$$S = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{pmatrix} \quad (2-3)$$

- Tentukan nilai eigen dari matriks kovarians, misalkan $\lambda_1, \lambda_2, \dots, \lambda_p$, dengan $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$
- Tentukan vektor eigen ke-j untuk nilai eigen ke-j ($j=1, 2, \dots, p$), misalkan $u_j = u_{j1}, u_{j2}, \dots, u_{jp}$
- Berdasarkan matriks eigen, maka komponen utama yang terbentuk adalah:

$$PC_1 = z_j v_{j1} = z_1 v_{11} + z_2 v_{21} + \dots + z_p v_{p1}$$

$$PC_2 = z_j v_{j2} = z_1 v_{12} + z_2 v_{22} + \dots + z_p v_{p2}$$

.

.

.

$$PC_p = z_j v_{jp} = z_1 v_{1p} + z_2 v_{2p} + \dots + z_p v_{pp}$$

(2-4)

- Kriteria yang digunakan untuk menentukan berapa komponen yang dapat dibentuk adalah kriteria persen varian. Jumlah komponen utama yang digunakan memiliki persentasi kumulatif varian minimal 80% (Rencher, 2001).

- Menghitung komponen skor (PC_j) yang akan digunakan sebagai input untuk analisis kluster.

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 dapat dicari dengan:

$$y_{i1} = e_1 x_i$$

$$y_{i2} = e_2 x_i$$

.

.

.

$$y_{ik} = e_k x_i$$

(2-5)

2.2.4 Validasi kluster

Validasi dilakukan baik untuk metode *Fuzzy Clustering* maupun untuk metode *Ward*. Validasi bertujuan untuk meyakinkan bahwa solusi atau kelompok kluster yang diperoleh telah mewakili populasi penelitian, dan berlaku umum untuk objek lain serta stabil dari waktu ke waktu. Validasi pada metode *Ward* dilakukan dengan membagi data secara acak menjadi dua bagian. Kemudian lakukan analisis kluster dengan metode *Ward* pada setiap bagian. Hasil pengklasteran dikatakan valid apabila hasil pengklasteran pada 2 bagian tadi mirip dengan hasil pengklasteran pada data asli (Rencher, 2001). Validasi pada metode *Fuzzy Clustering* menggunakan persamaan 2-6. Semakin kecil S , kluster semakin valid.

$$S = \frac{\sum_{i=1}^n \sum_{j=1}^c (u_{ij}^m) d(X_i, \theta_j)}{n_{ij}^{\min} d(X_i, \theta_j)} \quad (2-6)$$

2.2.5 Analisis kluster

Metode *Ward* adalah teknik untuk memperoleh kluster yang memiliki varian internal sekecil mungkin. Untuk metode kluster *Ward*, jumlah kluster ditentukan terlebih dahulu berdasarkan dendrogram. Ukuran yang digunakan adalah *Sum Square Error* (SSE) variabel. Proses pengelompokan adalah melalui tahapan berikut ini:

- Langkah 1.* Dimulai dengan memperhatikan N kelompok subjek dengan satu subjek per kelompok. SSE akan bernilai nol untuk tahap pertama karena setiap objek atau individu akan membentuk kluster.
- Langkah 2.* Kelompok pertama dibentuk dengan memilih dua dari N kelompok ini yang bila digabungkan akan menghasilkan SSE dalam nilai fungsi tujuannya.
- Langkah 3.* N - 1 kumpulan kelompok kemudian diperhatikan kembali untuk menentukan dua dari kelompok ini yang bisa meminimumkan tujuan. Dengan demikian N kelompok secara sistematis dikurangi menjadi N - 1, lalu menjadi N - 2 dan seterusnya sampai menjadi satu kelompok. SSE dalam metode *Ward* ini dihitung berdasarkan persamaan berikut:

$$SSE = \sum_{j=1}^p \left(\sum_{i=1}^n X_{ij}^2 - \frac{1}{n} \left(\sum_{i=1}^n X_{ij} \right)^2 \right) \quad (2-7)$$

Dengan:

X_{ij} = adalah nilai variabel ke- j

p = adalah banyaknya variabel yang diukur

n = adalah banyaknya objek dalam kluster yang terbentuk

Analisis Kluster dengan menggunakan *Fuzzy Clustering* pada dasarnya adalah proses penghitungan secara iteratif dimana penghitungan klusternya direvisi secara iterasi. Adapun algoritma dari *Fuzzy Clustering* adalah sebagai berikut:

- Langkah 1.* Tentukan jumlah kluster, nilai *centroid* utama (*prototype*) $\theta_j^{(0)}$ secara random, nilai faktor koreksi $\epsilon > 0$.

Langkah 2. Hitung derajat keanggotaan $U_{ij}^{(k)}$ berdasarkan persamaan 2-8:

$$u_{ij} = \frac{\left(\frac{1}{|x_i - \theta_j|^2} \right)^{1/m-1}}{\left(\frac{1}{\sum_{i=1}^c |x_i - \theta_j|^2} \right)^{1/m-1}} \quad (2-8)$$

dengan

$$\theta_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (2-9)$$

Langkah 3. Hitung fungsi objektif $J^{(k)}$ berdasarkan persamaan

$$J = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m |x_i - \theta_j|^2$$

Langkah 4. Hitung centroid baru $\theta_j^{(k+1)}$ dengan persamaan 2-9

Langkah 5. Perbarui keanggotaan $U_{ij}^{(k)}$ menjadi $U_{ij}^{(k+1)}$ dan fungsi objektif $J^{(k)}$ menjadi $J^{(k+1)}$

Langkah 6. Jika $\max_{ij} \left\{ J^{(k+1)} - J^{(k)} \right\} < \varepsilon$ proses akan berhenti, jika tidak maka kembali ke langkah 4.

Dari algoritma di atas terlihat bahwa jumlah kluster dan nilai centroid-nya ditentukan terlebih dahulu dan ditentukan secara *a priori* (pendapat para ahli) atau ditentukan jumlah kluster awal beserta nilai centroid-nya dengan menggunakan teknik kluster hierarkikal. Nilai-nilai centroid inilah yang akan menjadi *prototype* (centroid awal) $\theta_j^{(0)}$.

Kemudian, dihitung derajat keanggotaan U_{ij} , lalu penghitungan kembali nilai-nilai centroid baru. Nilai centroid dari masing-masing kluster akan bergerak menemukan posisi yang sesungguhnya seiring dengan penambahan langkah iterasi.

3 HASIL DAN PEMBAHASAN

3.1 Hasil Pendeteksian Outlier

Hasil pendeteksian outlier secara multivariat untuk data 900 grid di Kalimantan Barat, menunjukkan tidak terdapat objek (grid) yang memiliki nilai signifikansi lebih kecil dari 0,001, sehingga dapat disimpulkan tidak terdapat outlier pada data curah hujan 900 grid di Kalimantan Barat. Dengan demikian, seluruh objek dapat digunakan untuk tahap analisis selanjutnya.

3.2 Hasil Pendeteksian Multikolinearitas

Dari matriks X yang merupakan matriks pengamatan curah hujan TRMM bulanan di Kalimantan Barat, didapat matriks kovariansnya (S_{pp}) dan nilai eigen (λ) dari S_{pp} .

Untuk mendeteksi ada tidaknya multikolinearitas digunakan bilangan kondisi (k) seperti pada persamaan 2-1. Diperoleh:

$$k = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{346805}{135} = 2568,926$$

Karena $k (=2568,926) > 1000$ maka dapat disimpulkan terdapat gejala multikolinearitas yang sangat kuat, sehingga digunakan analisis komponen utama untuk mendapatkan komponen-komponen utama dari variabel (curah hujan bulanan selama 10 tahun) yang ortogonal dan tidak berkorelasi, yang selanjutnya dapat dijadikan variabel baru untuk dasar pengelompokan pada analisis kluster dengan metode Ward dan metode Fuzzy Clustering.

3.3 Hasil Analisis Komponen Utama

Jumlah komponen utama yang harus dibentuk ditentukan melalui kriteria persen varian, hasilnya adalah sebagai berikut:

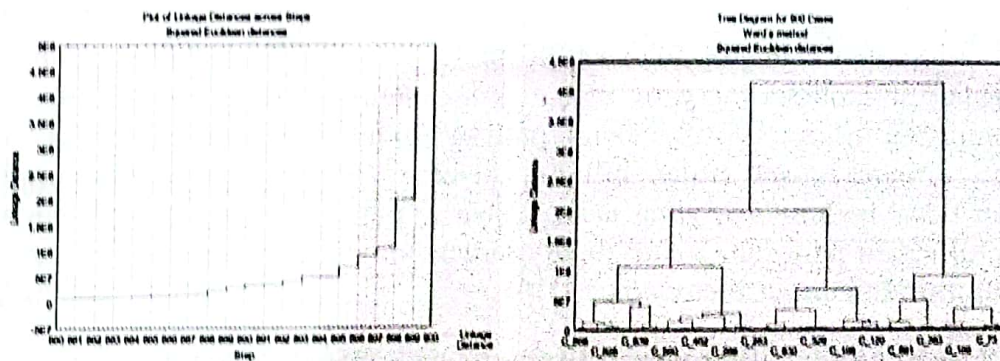
Tabel 3-1: OUTPUT ANALISIS KOMPONEN UTAMA

	PC ₁	PC ₂	PC ₃	PC ₄	PC ₅	PC ₆
Eigenvalue	346805	152895	103447	62295	30123	29352
Proportion	0,337	0,148	0,100	0,060	0,029	0,029
Cumulative	0,337	0,485	0,586	0,646	0,675	0,704
	PC ₇	PC ₈	PC ₉	PC ₁₀	PC ₁₁	PC ₁₂
Eigenvalue	22920	19622	18365	15638	14046	11411
Proportion	0,022	0,019	0,018	0,015	0,014	0,011
Cumulative	0,726	0,745	0,763	0,778	0,792	0,803

Pada Tabel 3-1 terlihat bahwa nilai kumulatif persen varian untuk komponen utama pertama (PC_1), kedua (PC_2), ketiga (PC_3) sampai dengan ke dua belas (PC_{12}) bernilai 0,803. Persentase kumulatif varian untuk 12 komponen ini sudah memenuhi batas minimal persentase kumulatif varian yaitu 80% (Rencher, 2001). Hal ini berarti apabila seluruh variabel ($X_1, X_2, X_3, \dots, X_{120}$) direduksi menjadi 12 variabel, maka variabel yang baru dapat menjelaskan 80,3 % karakteristik seluruh variabel.

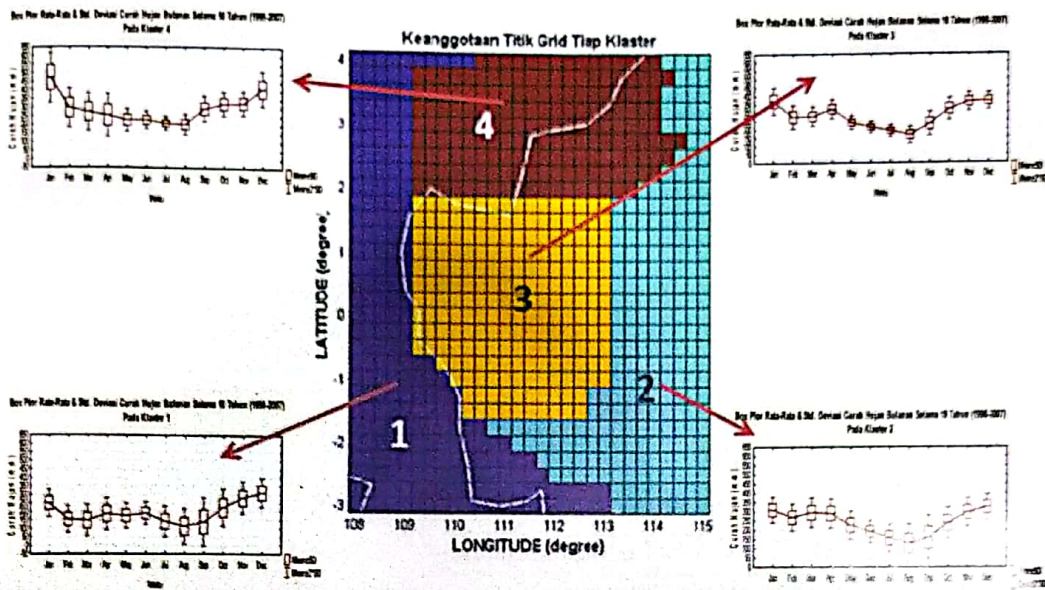
3.4 Hasil Analisis Kluster dengan Metode Ward

Skor komponen dari 12 komponen utama yang terbentuk dijadikan sebagai input data dalam pengelompokan karakteristik curah hujan di wilayah Kalimantan Barat dengan menggunakan metode Ward. Berdasarkan hasil pengelompokan dengan metode Ward menggunakan software Statistica 8, maka diperoleh output plot *Squared Euclidean Distance* pada setiap langkah pengelompokannya. Output plot langkah 880 hingga 900 diperlihatkan pada Gambar 3-1 (kiri), sedangkan output dendrogram diperlihatkan pada Gambar 3-1 (kanan).



Gambar 3-1: Plot jarak *Squared Euclidean* pada tiap langkah pengelompokan (langkah 880-900) (kiri), Dendrogram (kanan)

Terlihat pada Gambar 3-1 (kiri) bahwa pada jarak 1×10^8 atau langkah ke 898 terdapat garis yang cukup terjal. Pada posisi garis tajam ini dendrogram dipotong untuk menentukan berapa kluster yang harus dibentuk, maka didapat 4 kluster optimum pada jarak 1×10^8 dan step 898. Kluster yang terbentuk diperlihatkan pada Gambar 3-2. Validasi dalam metode *Ward* dilakukan dengan cara membagi data menjadi dua bagian kemudian dilakukan analisis kluster *Ward* pada masing-masing bagian dan hasilnya menunjukkan kesesuaian lebih dari 80 %, maka proses klustering ini valid.



Gambar 3-2: Kluster yang terbentuk dengan metode Ward dan pola curah hujan masing-masing Kluster

3.5 Hasil Analisis Kluster dengan Metode Fuzzy Clustering

Untuk menentukan jumlah kluster optimum dalam metode Fuzzy, hasil pengelompokan divalidasi dengan rumus Xie dan Benni (1991) dan hasilnya diperlihatkan pada Tabel 3-2. Tampak bahwa 4 kluster memberikan nilai validasi terkecil, maka jumlah kluster yang optimum adalah empat. Kluster yang terbentuk diperlihatkan pada Gambar 3-3.

Tabel 3-2: JUMLAH KLAS TER DAN NILAI VALIDASI

Jumlah Kluster Yang dibentuk	Fungsi Objektif	Nilai Validasi
2 Kluster	357.533.759.561	42.826
3 Kluster	234.566.463.704	35.859
4 Kluster	175.287.750.841	22.559
5 Kluster	139.514.948.165	35.313
6 Kluster	115.663.133.613	48.574
7 Kluster	99.015.618.448	56.300

intensitas minimum sebesar 150 mm terjadi pada bulan Agustus. Klaster 4 pada *Fuzzy* atau klaster 2 pada *Ward* berada paling selatan di Kalimantan Barat. Kisaran intensitas curah hujan hampir sama dengan klaster 3. Di klaster ini terdapat pegunungan yang lebih banyak dibandingkan klaster lainnya.

Perbedaan pola dan intensitas curah hujan antar klaster disebabkan perbedaan geografi dan relief. Meskipun seluruh wilayah Kalimantan Barat terletak di wilayah ekuatorial dengan jumlah panas radiasi yang relatif sama, namun perbedaan relief permukaan (daratan, lautan, dataran tinggi, dataran rendah pedalaman atau tepi pantai) menghasilkan proses atmosferis yang berbeda yang selanjutnya menghasilkan intensitas curah hujan yang berbeda. Intensitas curah hujan yang relatif tinggi di klaster 2 atau 4 (*Fuzzy/Ward*) dipengaruhi oleh suplai kelembaban dari laut China Selatan karena klaster 2 atau 4 berada dekat dengan laut tersebut.

Pada Gambar 3-2 dan 3-3, nampak bahwa curah hujan di Kalimantan Barat berdasarkan TRMM mengikuti pola ekuatorial untuk klaster 1, 2 dan 3 dengan metode *Ward* atau klaster 1, 3 dan 4 dengan metode *Fuzzy* serta pola musonal untuk klaster 2 (metode *Fuzzy*) atau klaster 4 (metode *Ward*) dengan puncak bulan basah terjadi pada bulan Januari/Desember dan bulan April sedangkan bulan dengan curah hujan terkecil adalah bulan Agustus. Pola ini sama untuk semua klaster, kecuali intensitasnya. Untuk klaster di daratan, pola curah hujan mempunyai kesesuaian dengan observasi. Untuk klaster di lautan (klaster 1 atau 1) pola curah hujan tidak sesuai dengan observasi. Berdasarkan observasi pola curah hujan di lautan seharusnya musonal. Ini menunjukkan bahwa di klaster 1/1 sebagai wilayah perbatasan darat dan laut, pengaruh monsun lebih kuat dibandingkan pengaruh posisi matahari di atas ekuator/ekinoks.

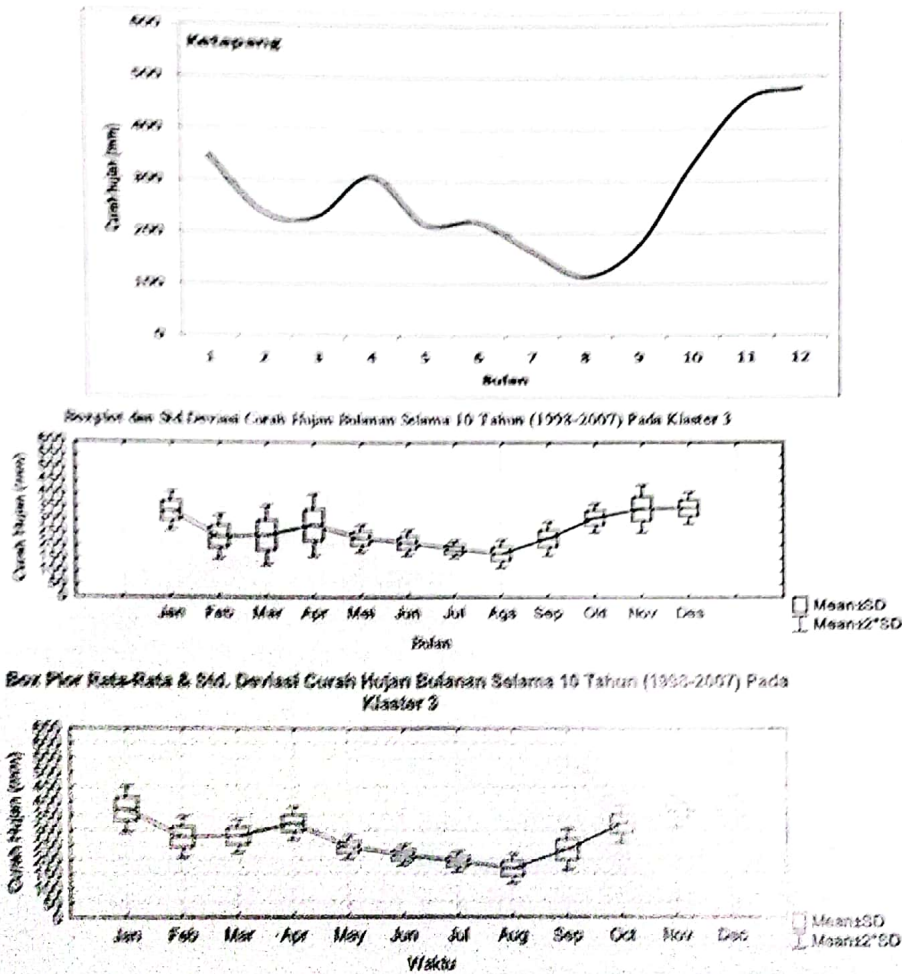
Perbedaan lain antara metode *Ward* dan metode *Fuzzy* diperlihatkan pada luas daerah/jumlah grid untuk setiap klaster. Jumlah grid pada masing-masing klaster 1, 2, 3 dan 4 berturut-turut adalah 288, 272, 150 dan 190 untuk metode *Fuzzy Clustering*. Klaster 1 metode *Ward* terdiri dari 260 grid, klaster 2 berisi 256 grid, klaster 3 mengandung 207 grid dan klaster 4 terdiri dari 177 grid. Perbedaan ini bersumber dari *overlapping* pada metode *Fuzzy*, sebagai contoh grid-grid yang berada di sebelah utara pada klaster 2 metode *Ward* ternyata dapat juga dikelompokkan kedalam klaster 3 pada metode *Fuzzy*. Inilah keunikan dari metode *Fuzzy* yang menganut sistem *overlapping*. Hal ini dibenarkan karena pola dan intensitas klaster 2 dan 3 tidak terlalu

berbeda, jika kemudian grid-grid itu dimasukkan dalam klaster 3 semata-mata karena derajat keanggotaan grid-grid tersebut di klaster 3 lebih tinggi dibandingkan di klaster 2.

Tabel 3-3: KOEFISIEN KORELASI POLA CURAH HUJAN ANTARA DATA TRMM DAN CURAH HUJAN OBSERVASI

Lokasi	Koefisien korelasi antara pola curah hujan metode Fuzzy clustering dengan curah hujan observasi (TRMM)	Koefisien korelasi antara pola curah hujan metode Ward dengan curah hujan observasi (TRMM)
Ketapang	0,99	0,90
Sambas	0,91	0,77
Sintang	0,85	0,86
Pangsuma (Kapuas Hulu)	0,83	0,73

Konfirmasi pola dan intensitas curah hujan TRMM dengan curah hujan observasi di beberapa lokasi sampel menunjukkan bahwa metode *Fuzzy Clustering* dan metode *Ward* dapat menggambarkan pola dan intensitas curah hujan dengan baik. Ini diperlihatkan dengan korelasi pola curah hujan yang berada dalam kisaran di atas 0,73 sampai 0,99 (Tabel 3-3). Perbandingan korelasi tersebut juga menunjukkan metode *Fuzzy* sedikit lebih baik dibandingkan dengan metode *Ward*. Konfirmasi visual curah hujan TRMM dengan curah hujan observasi salah satu lokasi (Ketapang) diperlihatkan pada Gambar 3-4. Perbandingan spasial klaster yang terbentuk menunjukkan bahwa metode *Ward* lebih mendekati observasi untuk klaster 2/4, 3/3 dan 4/2 sementara untuk klaster 1 baik untuk metode *Fuzzy* maupun *Ward* berbeda dengan observasi. Berdasarkan observasi (Bayong, 2004) pola curah hujan di daratan Kalimantan Barat adalah ekuatorial dan di lautan sekitarnya adalah monsun. Dengan demikian, data TRMM menunjukkan bahwa di wilayah lautan sekitar Kalimantan Barat pengaruh posisi Matahari di atas ekuator/ekinoks lebih kuat dibandingkan pengaruh monsun.



Gambar 3-4: Perbandingan pola dan intensitas di Ketapang antara curah hujan observasi (atas) dengan curah hujan TRMM metode *Fuzzy Clustering* (tengah) dan curah hujan metode *Ward* (bawah)

4 KESIMPULAN

Dengan data curah hujan bulanan TRMM dari tahun 1998 sampai 2007, penggunaan analisis komponen utama sebagai analisis awal dan analisis kluster dengan metode *Ward/Fuzzy Clustering* untuk mengelompokkan karakteristik curah hujan di Kalimantan Barat menghasilkan empat kelompok grid yang memiliki karakteristik curah hujan yang homogen. Jumlah kluster bisa berubah jika ditambahkan data baru yang membentuk kluster dengan karakteristik yang sangat berbeda dengan data yang sudah ada, jika derajat keanggotaannya lebih tinggi untuk kluster yang sudah ada maka jumlah kluster tetap.

Perbedaan jumlah grid antar kluster antara metode *Ward* dan metode *Fuzzy Clustering* disebabkan oleh adanya penerapan *overlapping*

pada metode *Fuzzy*, sehingga grid tertentu bisa menjadi anggota di dua klaster atau lebih dengan derajat keanggotaan yang berbeda.

Perbandingan pola dan intensitas curah hujan antara data TRMM dan data observasi menunjukkan hasil yang baik dengan koefisien korelasi $> 0,7$. Perbandingan intensitas curah hujan berdasarkan dua metode menunjukkan bahwa metode *Fuzzy Clustering* lebih baik dibandingkan metode *Ward*, namun perbandingan spasial klaster di daratan menunjukkan bahwa metode *Ward* lebih mendekati observasi.

Pola curah hujan musonal ditunjukkan oleh klaster 2/4 (gabungan daratan dan lautan) dan pola ekuatorial ditunjukkan oleh klaster 1/1, 3/3 dan 4/2 (daratan dan gabungan daratan dan lautan). Pola ekuatorial di daratan Kalimantan Barat sesuai dengan observasi, namun pola curah hujan di klaster 1/1 yang seharusnya musonal ternyata ekuatorial. Dengan demikian, data TRMM menunjukkan bahwa di wilayah lautan sekitar Kalimantan Barat pengaruh posisi matahari di atas ekuator/ekinoks lebih kuat dibandingkan pengaruh monsun.

DAFTAR RUJUKAN

- <http://trmm.gsfc.nasa.gov/> didownload Maret 2009.
- Bayong, T. H. K., 2004. *Klimatologi*, Edisi kedua, Penerbit ITB.
- Degaetano, A. T., 1996. *Delineation of Mesoscale Climate Zones in The Northeastern United States using a Novel Approach to Cluster Analysis*, *Journal of Climate*, 9.
- Hair, J. F.; Anderson, R. E.; Tatham, R. L.; and Black, W. C., 1998. *Multivariate Data Analysis*, Fifth edition, Prentice Hall International, Inc., Upper Saddle River, New Jersey.
- Haryoko, U., 2009. *Pewilayahan Hujan untuk Menentukan Pola Hujan (contoh kasus Kabupaten Indramayu)*, <http://www.staklimpondokbetung.net/publikasi/didownload> Juli 2009.
- Johnson, R. A., and Wichern, D. W., 1992. *Applied Multivariate Statistical Analysis*, Fifth edition, Prentice Hall, Englewood Cliffs, New Jersey.
- Klawonn, F. and Hoppner, F., 2001. A New Approach to Fuzzy Partitioning, *Proc. of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference, Vancouver, Canada*.
- Mimmack, G. M.; Mason, S. J. and Galphin, J. S., 2000. Choice of Distance Matrices in Cluster Analysis: Defining Regions, *Journal of Climate*, 14.
- Pravitasari, A. A., 2008. Analisis Pengelompokan dengan Fuzzy Z-Means Cluster (Kasus Pengelompokan Kecamatan di Kabupaten

Tuban berdasarkan Tingkat Partisipasi Pendidikan), *Tesis Master*, ITS.

Rencher, A. C., 2001. *Methods of Multivariate Analysis*, Second Edition, A Wiley-Interscience Publication, United States.

Sharma, S., 1996. *Applied Multivariate Techniques*, A Wiley-Interscience Publication, United States.

Xie, X. L. and Benni, G., 1991. A Validity Measure for Fuzzy Clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13.