

PENENTUAN KLASTER STASIUN HUJAN MENGGUNAKAN ANALISIS KOMPONEN UTAMA (Studi Kasus di Daerah Istimewa Yogyakarta)

*Determining The Cluster of Raingauge Stations Using Principal Component Analysis
(Case Study in Yogyakarta Special Region)*

Putu Sudira¹

ABSTRACT

The geographical condition of Yogyakarta Special Region is varied from the mountainous plateau of the southern part (Gunung Kidul) and the northern part (Sleman) as well as the boundary of the sea level along the southern part of the region. This condition causes the great variability of the amount of rainfall in this region. Related to this, the study aims at analyzing the rainfall data to determine the cluster of the raingauge stations based on the similar characteristics of the rainfall.

The data needed for the analysis is the 15 years daily rainfall data from 67 raingauge stations including the location of the stations which are unevenly distributed in 4 districts available in Yogyakarta, namely: Gunung Kidul, Sleman, Kulon Progo, and Bantul.

The Principal Component Analysis was used to analyze the data to determine the cluster of the raingauge stations. The results indicated that, by applying the similarity of 57 - 60%, the characteristics of rainfall in Yogyakarta could be categorized into 7 clusters.

Key words: *rainfall characteristics, principal component analysis, cluster*

PENDAHULUAN

Faktor iklim memegang peranan yang sangat penting dalam hampir semua kegiatan manusia. Curah hujan sebagai salah satu anasir iklim menjadi dasar perencanaan dalam kegiatan di bidang pertanian, bidang perternakan, bidang pariwisata, bidang industri, dan bidang yang lain.

Iklim di Indonesia ditentukan oleh kondisi geografis dan fisiografis. Formasi pegunungan yang ada hampir di seluruh pulau utama, berpengaruh terhadap sirkulasi udara yang terjadi. Hal ini mengakibatkan terdapat sejumlah tipe distribusi curah hujan yang bervariasi dari daerah yang curah hujannya tinggi sepanjang tahun sampai daerah yang mengalami musim kering berkepanjangan. Secara umum Indonesia dinyatakan sebagai daerah beriklim tropis lembab (*humid tropics*).

Pola sebaran hujan di Indonesia dicirikan oleh hujan secara musiman, yaitu musim basah atau musim hujan dan musim kering atau musim kemarau. Ada banyak cara menentukan klasifikasi iklim dan cara yang satu memiliki perbedaan dengan cara yang lain. Suatu klasifikasi iklim yang dibuat oleh para ahli sebenarnya berorientasi pada tujuan penggunaan tertentu, misalnya untuk keperluan tanaman perkebunan, tanaman pangan, dan lainnya (Oldeman dan Frere, 1982 dalam Sudira dan Tri Sudyastuti, 1999).

Curah hujan di Daerah Istimewa Yogyakarta sangat bervariasi baik variasi hujan dari utara ke selatan maupun dari timur ke barat. Hal ini disebabkan karena pengaruh

topografi maupun pengaruh laut yang membentang di selatan Yogyakarta. Untuk itu diperlukan untuk mengetahui karakteristik hujan di masing-masing stasiun atau wilayah dalam kaitannya dengan efisiensi pemanfaatan air khususnya untuk keperluan pertanian.

Dalam penelitian ini, karakteristik hujan ditentukan menggunakan metode Analisis Komponen Utama. Metode ini pernah digunakan untuk menentukan karakteristik hujan di Sardinia (Benzi, dkk.1997). Penelitian yang mengkaitkan faktor iklim atau hujan dengan berbagai parameter penunjang kehidupan masyarakat dan parameter gejala alami telah banyak dilakukan oleh para peneliti. Seperti misalnya Hill, dkk., (1979) memprediksi produksi tanaman kedelai menggunakan data iklim. Model yang menghubungkan antara hujan dan air limpasan sebuah DAS dibuat oleh O'Neals, dkk., (2000), dan Schwarzaier, dkk., (1992). Sedangkan model yang menentukan klasifikasi dan karakteristik hujan banyak dikerjakan oleh para peneliti, antara lain oleh Sevruk dan Zahlavova (1994), Jackson dan Weinand (1994), dan Benzi, dkk., (1997).

Penelitian ini bertujuan untuk menentukan karakteristik hujan di daerah Istimewa Yogyakarta, yang selanjutnya untuk mengelompokkan wilayah hujan (klaster) yang mempunyai karakteristik yang serupa

LANDASAN TEORI

Analisis Komponen Utama (Principal Component Analysis = PCA)

Principal Component Analysis (PCA) memungkinkan penyalinan fungsi diskrit dari ruang dan waktu, misalnya variabel iklim atau curah hujan dibuat menjadi beberapa bagian pola ruang bebas linier dan koefisien waktu tak bebas. Pola ruang adalah himpunan basis ortogonal lengkap dan disebut *Empirical Orthogonal Functions (EOF)*, (Benzi dkk., 1997).

Ditentukan $x_i(t)$ adalah fungsi diskrit atau sebuah variabel iklim atau hujan pada stasiun ke i dengan $i=1, \dots, L$, dan waktu t dengan $t=1 \dots N$, di mana L adalah jumlah total stasiun dan N adalah jumlah total hari-hari pengamatan. Dan C_{ij} adalah *cross-product matrix*, yang ekuivalen dengan matrik *co-varian* ketika pengamatan mempunyai rata-rata kosong (*null average*). Terlebih dahulu ditentukan nilai eigen yaitu nilai yang menggambarkan karakteristik masing-masing stasiun hujan yang diperoleh dari hasil perhitungan *cross-product matrix (Persamaan 1)*.

$$C_{ij} = \frac{1}{N} \sum_{t=1}^N x_i(t) \cdot x_j(t) \quad (1)$$

¹ Staf Pengajar Fakultas Teknologi Pertanian UGM, Yogyakarta

Cross-product matrix mempunyai nilai simetris dan positif, oleh karena itu semua nilai eigen (λ_k) harus positif dan berhubungan dengan "vektor eigen" yang ortogonal. Penyelesaian masalah nilai eigen menghasilkan Basis Ortogonal Lengkap, E^k :

$$\sum_{j=1}^L C_{ij} E_j^k = \lambda_k E_i^k \quad (2)$$

EOF_s yang digunakan dalam analisis ini adalah nilai "vektor eigen" yang dinormalkan terhadap batas ruang dan waktu. Dengan proyeksi akhir $x_i(t)$ terhadap EOF ke- k , menghasilkan hubungan komponen utama (PC) waktu tak-bebas.

$$a^k(t) = \sum_{i=1}^L x_j(t) E_j^k \quad (3)$$

Komponen utama ini menggambarkan bagian ke- k dari $a_i^k(t) = a^k(t) E_j^k$ yang menjelaskan satu bagian dari total varian yang sama dengan nilai eigen λ_k :

$$\frac{1}{N} \sum_{i=1}^L \sum_{t=1}^N [x_j^k(t)]^2 = \lambda_k = \frac{1}{N} \sum_{t=1}^L [a^k(t)]^2 \quad (4)$$

Komponen deterministik diperoleh dari penjumlahan bagian-bagian yang bersesuaian dengan nilai eigen tertinggi. Rekonstruksi dari bagian-bagian deterministik dapat ditulis sebagai berikut:

$$\tilde{x}_i(t) = \sum_{k \in A} a^k(t) E_i^k \quad (5)$$

Teknik Kluster

Teknik kluster juga dikenal dengan nama metode Complete Linkage. Prosedur ini dimulai pada keadaan kluster yang hanya mengandung satu nilai dan menggunakan penyatuan kelompok yang terdekat secara berturut-turut, sampai keadaan dimana hanya satu kelompok yang diakui mengandung semua nilai. Setiap penggabungan kluster dievaluasi derajat kemiripannya dengan mereduksi jumlah kluster.

METODE PENELITIAN

Penyiapan data

- Data hujan harian selama 15 tahun dikumpulkan dari 67 stasiun hujan yang tersebar di empat kabupaten di Daerah Istimewa Yogyakarta. Data 15 tahun sebagai acuan karena rangkaian data terlengkap di seluruh stasiun hujan di Yogyakarta hanya 15 tahun. Lebih banyak ketersediaan data akan memberikan hasil yang lebih baik.
- Data curah hujan yang tersedia tidak secara langsung digunakan dalam analisis, tetapi harus diseleksi dengan kriteria sebagai berikut: (i) data hujan yang digunakan untuk pengelompokan adalah data hujan harian yang

mempunyai ketebalan lebih besar daripada 1 mm, (ii) data yang nol (kosong) tidak dapat dikenal oleh PCA, oleh karena itu hanya menggunakan data observasi hari-hari dalam katagori hujan. Hari disebut hari hujan jika ketebalan hujan lebih besar dari 1 mm (Benzi, dkk. 1997). (iii) untuk mengurangi variabilitas yang besar, digunakan Transformasi Logaritmik sebagai berikut :

$$x_i(t) = \begin{cases} \ln\left(\frac{p_i(t)}{p_s}\right) & \text{apabila } p_i(t) \geq p_s \\ 0 & \text{apabila } p_i(t) < p_s \end{cases} \quad (6)$$

$$i=1, \dots, L; \quad t=1, \dots, N$$

dimana $p_i(t)$ adalah ukuran curah hujan (mm), p_s adalah 1 mm (curah hujan mula-mula), N adalah jumlah hari hujan dan L adalah jumlah stasiun.

Analisis data

Setelah dilakukan transformasi logaritmik terhadap data yang akan dianalisis, tentukan nilai cross product matrix menggunakan Persamaan 1, (ii) nilai eigen (karakteristik hujan) untuk masing-masing komponen utama dari cross product matrix dihitung dengan Persamaan 2, (iii) menentukan standar deviasi dari masing-masing komponen utama, dan (iv) mereduksi data menjadi beberapa komponen utama dan diurutkan menurun berdasarkan nilai eigen. Untuk mendapatkan susunan kluster yang stabil dilakukan analisis iteratif (berulang) dengan langkah-langkah sbb: (i) tetapkan jari-jari atau jarak tertentu (r). Dalam masing-masing nilai tahapan ruang, sebuah cakupan $I(x(t))$ didefinisikan sebagai kesepakatan oleh semua nilai yang ada pada jarak r tersebut, (ii) tiap-tiap nilai sebuah fungsi kerapatan $x(t)$ didefinisikan sebagai jumlah antar nilai untuk cakupan $I(x(t))$, dan (iii) sebuah nilai dijadikan sebagai kerapatan lokal maksimum jika nilai itu merupakan kerapatan terbaik dari semua nilai yang lain pada cakupan itu. Tiap-tiap kerapatan lokal maksimum adalah nilai pertama kluster baru. Semua nilai diurutkan dalam kerapatan yang semakin menurun.

Teknik ini diulang terus dengan memasukan jari-jari yang berbeda untuk mencari susunan kluster yang dianggap stabil. Karakteristik sebuah kluster c diilustrasikan dalam Persamaan 7.

$$\langle x_i \rangle_c = \sum_{k \in I} \langle a^k(t) \rangle_c E_i^k \quad (7)$$

dimana $\langle a^k(t) \rangle_c$ adalah koordinat ke- k pada bidang kluster c, dan pembentukan rata-rata komponen utama ke- k berada di antara nilai-nilai pada kluster c.

HASIL DAN PEMBAHASAN

Iklim

Hasil analisis data sekunder pada seluruh stasiun curah hujan yang ada di DIY selama 15 tahun pengamatan (1985-1999) yang mencakup 67 stasiun curah hujan, menunjukkan rata-rata hujan tahunan sebesar 2.179,7 mm. Curah hujan tahunan terbesar terjadi di stasiun hujan Ngipiksari, Kecamatan Ngaglik, Sleman yaitu sebesar 3.878,3 mm, sedangkan curah hujan tahunan terkecil terdapat di stasiun hujan Kepek, Kecamatan Wonosari, Gunung Kidul yaitu sebesar 1.547,5 mm. Penentuan tipe iklim pada daerah pengamatan dilakukan dengan tiga metode yaitu metode Mohr, Schmidt-Ferguson dan Oldeman. Menurut metode Mohr wilayah DIY mempunyai bulan kering rata-rata 3-4 bulan termasuk Iklim Golongan III (agak kering). Sedangkan menurut metode Schmidt-Ferguson wilayah DIY termasuk tipe iklim D (sedang) dengan nilai Q sebesar 0,71. Oldeman mengklasifikasikan iklim Yogyakarta dengan tipe D3 yang mempunyai bulan basah berturut-turut 3 – 4 bulan dan bulan kering berturut-turut 4 – 6 bulan.

Analisis Komponen Utama

Data yang dianalisis yaitu data yang berasal dari 67 buah stasiun curah hujan yang ada di seluruh Daerah Istimewa Yogyakarta. Tahap pertama, metode PCA mentransformasi data dengan transformasi logaritmik sehingga dihasilkan data hujan transformasi. Penyeleksian data curah hujan harian dengan metode transformasi logaritmik berfungsi untuk menjamin data tidak nol. Selain itu proses ini juga untuk memperkecil variabilitas data yang terlalu tinggi.

Nilai eigen atau nilai karakteristik masing-masing stasiun hujan yang diperoleh dari hasil perhitungan *Cross Product Matrix* (Persamaan 2) dapat dilihat pada Tabel 1. Pada landasan teori di atas, PCA mereduksi semua variabel data menjadi beberapa komponen utama (PC). Persentase total varian yang dihasilkan oleh PC pertama diurutkan menurun berdasarkan nilai eigen. Dua PC pertama pada

Tabel 1 tersebut memberikan varian kumulatif sebesar 95,41% variabel hujan. Oleh karena itu curah hujan sudah dapat dikarakteristik oleh dua PC tersebut tanpa mengurangi keberadaan data lain karena dua PC pertama sudah mewakili lebih dari 95,41%. Hal ini sesuai dengan kaidah Kaiser (Kaiser, 1960, dalam Benzi, dkk. 1997) yang menyatakan bahwa hanya Prinsip Komponen dengan nilai eigen lebih besar dari pada 1 (satu) yang bisa digunakan untuk menafsirkan hasil dari PCA.

Table 1. Ten (10) of the first eigen values at each of component principal

PC	Eigen Values	Percentage	Cumulative Percentage
1	61.6766	92.0500	92.0500
2	2.2488	3.3600	95.4100
3	0.6600	0.9900	96.4000
4	0.4251	0.6300	97.0300
5	0.3256	0.4900	97.5200
6	0.1935	0.2900	97.8100
7	0.1560	0.2300	98.0400
8	0.1259	0.1900	98.2300
9	0.1037	0.1500	98.3800
10	0.0986	0.1500	98.5300

Setelah itu diperoleh dua fungsi eigen untuk mempelajari masa curah hujan. Dua fungsi eigen ini dikelompokkan dengan teknik kluster menggunakan metode *Complete Linkage*. Data yang dikluster adalah koefisien PC yang merupakan koefisien dari fungsi eigen. Pengelompokkan dimulai dengan derajat similaritas 71%, dari 67 parameter sehingga terbentuk 11 kluster (Tabel 2).

Table 2. Number of clusters consist of several rainfall stations

No	Cluster	No. of stations
1	A	1, 5
2.	B	2, 25, 18, 31, 3, 39, 12, 13, 15, 16, 23
3	C	17, 21, 19, 29, 22, 4, 38, 14, 20, 44, 47, 48, 51, 61
4	D	6, 27, 28
5	E	7, 9
6	F	10, 35, 30, 8, 24, 37, 33
7	G	11
8	H	26, 36, 53, 62, 56, 57, 65
9	I	32
10	J	34, 46, 49, 50, 40, 4, 41, 52, 59, 54, 60
11	K	42, 55, 43, 64, 58, 66, 67

Dalam pengelompokkan data di atas masih ada dua kluster yang hanya mengandung satu variabel yaitu kluster G dan I. Untuk menghindari terjadinya satu variabel dalam satu kluster, maka analisis dilakukan terus menerus secara

berulang (iterasi) sampai diperoleh satu kluster mengandung paling sedikit lebih daripada satu variabel. Iterasi terakhir menghasilkan 7 (tujuh) kluster dengan derajat kemiripan antara 57% sampai dengan 60% (Tabel 3)

Table 3. Clusters of rainfall stations for the last iteration

No	Cluster	Numbers and Name of Stations (see the map enclosed)
1	A	1 (Barongan), 5 (Dogongan), 11 (Gedangan)
2	B	2 (Kalijoho), 25 (Semanu), 18 (Giriwungu), 31 (Beji Wates), 3 (Karang Ploso), 39 (Wijilan), 12 (Karangmojo), 13 (Beji), 15 (Kedung Keris), 16 (Nglipar), 23 (Wonodoyo).
3	C	17 (Paliyan), 21 (Wonogomo), 19 (Giriharjo), 29 (Kepek), 22 (Playen), 4 (Tirtonirmolo), 38 (Samigaluh), 14 (Ngawen), 47 (Pathuk), 44 (Kolombo), 47 (Patukan), 48 (Godean), 51 (Cebongan), 61 (Sayegan).
4	D	6 (Palbapang), 27 (Sidoharjo), 28 (Tepus), 10 (Terong), 35 (Kokap), 30 (Wonosari), 8 (Pleret), 24 (Rongkop), 37 (Tambak), 33 (Sentolo).
5	E	7 (Piring), 9 (Pandak), 32 (Galur).
6	F	26 (Semin), 36 (Kalibawang), 53 (Prumpung), 62 (Beran), 56 (Kemput), 57 (Ngipiksari), 65 (Angin-angin).
7	G	34 (Girimulyo), 46 (Jambon), 49 (Juwangen), 50 (Sendang Pitu), 40 (Berbah), 45 (Maguwoharjo), 41 (Tanjungtirto), 52 (Gondangan), 59 (Prambanan), 54 (Dolo), 60 (Trukan), 42 (Banjarharjo), 55 (Jangkang), 43 (Bronggang), 63 (Medari), 64 (Pakem), 66 (Dadapan), 67 (Ledoknongko).

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan uraian di atas diperoleh beberapa kesimpulan :

1. Masih lemahnya pencatatan data hujan dan pengelolaan data hujan di Daerah Istimewa Yogyakarta, sehingga banyak kekurangan data dalam stasiun pengamat yang memerlukan pengisian data dengan analisis data hilang.
2. *Principal Component Analysis* (PCA) tidak dapat dipergunakan untuk menentukan klasifikasi iklim tetapi hanya untuk mengelompokkan jaringan stasiun hujan yang mempunyai kemiripan karakteristik hujannya.

Saran

Perlu adanya keseragaman spesifikasi alat penakar hujan agar data yang terkumpul memberikan indikator yang sama sehingga analisis bisa lebih akurat dan dapat dijadikan acuan bagi pengembangan wilayah serta sangat diperlukan *network* penangan data curah hujan.

Pengklasteran stasiun hujan menggunakan *Principal Component Analysis* masih terlalu sedikit dilakukan, oleh karena itu diperlukan penelitian dengan analisis PCA dengan mempertimbangkan faktor-faktor geografik lahan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Direktorat Jenderal Pendidikan Tinggi lewat Lembaga Penelitian Universitas Gadjah Mada yang telah memberikan kepercayaan untuk melaksanakan penelitian ini. Ucapan terima kasih juga disampaikan kepada sdr. Siti Mutmainah yang telah membantu dalam penyelesaian analisis.

DAFTAR PUSTAKA

- Benzi, R, Deidda, R., Marrocu, M. 1997. Characterization of temperature and precipitation field over Sardinia with Principal Component Analysis and Singular Spectrum Analysis. *Int. J. Climatology*, (17) :1231-1262
- Hill, R. W. D. R. Johnson and K.H. Ryan. 1979. A model for predicting soybean yields from climate data. *Agronomy J.* (71): 251 - 256
- Jackson, I.J. dan H.Weinand.1994. Towards a classification of tropical rainfall stasions. *Int. J. Climatology*, (14) : 263-286
- O' Neals, M. I. , J. R. Frankerberger, D. R. Ess, and R. H. Grant. 2000. Precipitation variability at the farm scale during crop phenological phases. *Trans. ASAE.* 43 (6) : 1449 - 1458
- Scharzmaier, G., H. Mayer dan U. Beyer 1992. Climatological criteria for estimation of runoff by linear regression *Int.J. Climatology*, (12) :103-110
- Sevruk, B dan L Zahlanova 1994. Classification system of precipitation gauge site exposure: evaluation and application *In.J. Climatology*, (14):681-689
- Sudira, P dan Tri Sudyastuti, 1999. Dktat Kuliah Klimatologi. Fakultas Teknologi Pertanian UGM, Yogyakarta.