

Penentuan Zona Musim di Mojokerto Menurut Karakteristik Curah Hujan Dengan Metode *Time Series Based Clustering*

Iftitah Ayundari dan Sutikno

Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: sutikno@statistika.its.ac.id

Abstrak—Zona musim adalah daerah yang pos hujan rata-ratanya memiliki perbedaan yang jelas antara periode musim hujan dan musim kemarau. Metode analisis data yang diterapkan dalam pengelompokan ZOM ini adalah metode analisis cluster, yaitu mengelompokkan suatu objek yang memiliki kesamaan sifat tanpa menghilangkan struktur alami objek. Oleh karena kelompok yang dihasilkan memiliki makna, seperti pola atau klasifikasi. Data curah hujan merupakan data series sehingga dalam melakukan analisis cluster menggunakan jarak basis *time series*, yaitu pada penelitian ini menggunakan jarak *Dynamic Time Warping (DTW)* dan *autocorrelation function (ACF)*. Tujuannya adalah mengetahui karakteristik zona musim di seluruh wilayah Mojokerto dan memetakan pola zona musim di Mojokerto berdasarkan data curah hujan menggunakan *time series based clustering*. Pengelompokan dilakukan dengan membentuk 2 hingga 5 kelompok, dan selanjutnya akan dilakukan pemilihan kelompok optimum dengan *pseudo-f statistics* dan kriteria *silhouette*. Data curah hujan yang digunakan adalah curah hujan bulanan. Evaluasi kinerja zona musim dilakukan dengan rasio simpangan baku dimana metode yang dibandingkan adalah pengelompokan dengan jarak *ACF*, *DTW*, dan hasil pengelompokan oleh BMKG. Berdasarkan hasil evaluasi kinerja zona musim dengan rasio simpangan baku metode terbaik yang dipilih adalah pengelompokan dengan jarak *DTW*.
Kata Kunci : Curah Hujan, Clustering, Mojokerto, Zona Musim

analisis data yang sesuai adalah metode analisis *cluster*, yaitu mengelompokkan suatu objek yang memiliki kesamaan sifat tanpa menghilangkan struktur alami dari objek sehingga kelompok-kelompok yang dihasilkan memiliki makna, seperti pola atau klasifikasi. Badan Meteorologi Klimatologi dan Geofisika (BMKG) telah melakukan pengelompokan iklim (pola curah hujan) dengan membuat Zona Musim (ZOM). Pembagian zona tidak dibatasi oleh wilayah administrasi kabupaten atau kota, sehingga dalam satu zona tertentu bisa mencakup beberapa kabupaten/ kota. Untuk Propinsi Jawa Timur terbagi atas 60 ZOM [3].

Data curah hujan merupakan data *series* sehingga dalam melakukan analisis *cluster* menggunakan jarak basis *time series*, yaitu pada penelitian ini menggunakan jarak *Dynamic Time Warping (DTW)* dan *Autocorrelation function (ACF)*. Tujuan dari penelitian ini untuk mengetahui karakteristik curah hujan zona musim di seluruh wilayah Mojokerto dan mendapatkan zona musim yang optimal di wilayah Mojokerto berdasarkan karakteristik curah hujan menggunakan *time series based clustering*.

Pada analisis pengelompokan zona musim untuk wilayah Mojokerto dengan 25 pos hujan. Data yang digunakan memiliki periode 7 tahun terakhir. Pengelompokan dilakukan dengan membentuk 2 hingga 5 kelompok, dan selanjutnya akan dilakukan pemilihan kelompok optimum dengan *pseudo-f statistics* dan kriteria *silhouette*. Evaluasi kinerja zona musim dengan menggunakan rasio simpangan baku.

I. PENDAHULUAN

Dua unsur utama iklim adalah suhu dan curah hujan. Indonesia sebagai daerah tropis ekuatorial mempunyai variasi suhu yang kecil, sementara variasi curah hujannya cukup besar. Oleh karena itu curah hujan merupakan unsur iklim yang paling sering diamati dibandingkan dengan suhu. Dengan demikian iklim memiliki pengaruh terhadap pola curah hujan. Data pengamatan stasiun seringkali tidak homogen, karena adanya penggantian alat pengamatan (*instrumentasi*), perpindahan lokasi stasiun pengamatan, dan pengaruh urbanisasi [1]. Dalam upaya mempelajari iklim, data yang digunakan harus lengkap. Namun problematika yang ada pada saat ini adalah tidak semua data di BMKG tersedia dan lengkap. Misalkan dalam melakukan penentuan zona musim menggunakan pos hujan terdapat kendala tidak semua pos hujan memiliki data yang lengkap. Oleh karena itu dalam melakukan pendugaan dapat menggunakan lokasi lokasi yang memiliki karakteristik yang sama. Untuk itu dilakukan penelitian pengelompokan zona musim agar zona musim homogen. Data curah hujan bulanan yang akan digunakan dalam pengelompokan zona musim. Karena adanya ketidaklengkapan data pada semua pos hujan maka metode

II. TINJAUAN PUSTAKA

A. Analisis Cluster

Analisis cluster (analisis kelompok) adalah metode analisis yang digunakan untuk mengelompokkan obyek-obyek pengamatan menjadi beberapa kelompok sehingga akan diperoleh kelompok dimana obyek-obyek dalam satu kelompok tersebut mempunyai persamaan sedangkan dengan anggota kelompok lain mempunyai perbedaan [3].

B. Metode Hierarki

Analisis *cluster* metode hirarki diawali dengan mengelompokkan dua atau lebih obyek yang mempunyai kesamaan paling dekat. Selanjutnya dilanjutkan pada obyek lain yang mempunyai kedekatan kedua dan seterusnya. [4]. Berikut merupakan beberapa macam metode pada analisis *cluster* metode hirarki

— *Single Linkage*

Metode dengan prinsip jarak minimum. Langkah pertama menemukan jarak terdekat pada $D = \{d_{ik}\}, \{d_{jk}\}$ adalah nilai obyek ke-i pada variabel ke-k dengan $i=1,2,3,\dots,n$ dan menggabungkan obyek yang berkorespondensi, dikatakan U dan V dan sembarang kelompok lain W adalah

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (1)$$

d_{UW}, d_{VW} adalah jarak tetangga terdekat dari kelompok U dan W, dan jarak kelompok V dan W [4].

— *Complete Linkage*

Complete linkage menggunakan aturan jarak maksimum antar cluster. Dapat ditulis ke dalam rumus sebagai berikut [4].

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (2)$$

Dengan

d_{UW} = jarak antara U dan W,

d_{VW} = jarak antara V dan W.

C. *Metode Non-Hierarki*

Metode non-hierarki digunakan apabila banyak kelompok yang akan dibentuk sudah diketahui terlebih dahulu. Salah satu contohnya adalah metode *K-means*. *K-Means* merupakan salah satu metode data clustering non-hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih klaster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu klaster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya.

D. *Analisis Faktor*

Analisis faktor mengasumsikan suatu struktur spesifik tertentu dalam sebuah model untuk variabel-variabel random.

Variabel random X yang diamati dengan p buah variabel komponen, yang memiliki rata-rata dan matriks kovarian, maka model faktor dari X yang merupakan kombinasi linier beberapa variabel saling bebas yang tidak teramati adalah disebut sebagai common factors dan ditambahkan dengan disebut specific factor, sehingga secara khusus dapat ditulis sebagai [4]:

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \varepsilon_p \end{aligned} \quad (3)$$

dengan : Common factor ke-j

= Loading factor ke-j dan variabel ke-i

= Specific factor ke-i

dimana : $i = 1, 2, \dots, p$ dan $j = 1, 2, \dots, m$

Dalam notasi matriks persamaan dapat ditulis sebagai :

$$X_{(p \times 1)} - \mu_{(p \times 1)} = L_{(p \times m)} F_{(m \times 1)} + \varepsilon_{(p \times 1)} \quad (4)$$

E. *Analisis Cluster Data Time Series*

Data *time series* pada dasarnya digolongkan sebagai data dinamis karena nilai fiturnya berubah sebagai fungsi waktu [6]

Definisi 1: Time Series Clustering, diberikan dataset n data time series $D = \{F_1, F_2, F_3, \dots, F_n\}$ dilakukan partisis Secara *unsupervised* D terhadap $C = \{C_1, C_2, C_3, \dots, C_3\}$ sedemikian rupa sehingga dalam kelompok *time series* homogen,

pengelompokkan dilakukan berdasarkan ukuran kesamaan tertentu. Selanjutnya, C_i disebut sebagai cluster, dimana $D = \bigcup_{i=1}^k C_i$ dan $C_i \cap C_j = \emptyset$ untuk setiap $i \neq j$ [7].

F. *Dynamic Time Warping (DTW) Distance*

Dynamic Time Warping (DTW) menggunakan teknik pemrograman dinamis untuk menemukan semua jalur yang mungkin, dan memilihnya untuk menghasilkan jarak minimum antara dua deret waktu menggunakan matriks jarak, dimana masing-masing elemen dalam matriks adalah jarak kumulatif nilai minimal dari tiga tetangga sekitarnya. Jika diberikan dua data time series $Q = q_1, q_2, q_4, \dots, q_m$, dimana dengan ukuran m dan $C = c_1, c_2, c_3, \dots, c_n$ berukuran n, sehingga terbentuk matriks $m \times n$. Setiap elemen (i, j) merupakan jarak kumulatif dari jarak (i, j) dan nilai minimum dari tiga elemen yang berdekatan dengan elemen (i, j) , dimana $0 < i \leq n$ dan $0 < j \leq m$, sehingga dapat didefinisikan elemen (ij) sebagai berikut

$$e_{ij} = d_{ij} + \min\{e_{(i-j)(j-1)}, e_{i(j-1)}\} \quad (5)$$

Dimana untuk e_{ij} merupakan elemen matriks (ij) . Nilai d_{ij} merupakan jarak antara q_i terhadap c_j . perhitungan nilai d_{ij} dapat dituliskan sebagai berikut

$$d_{ij} = (q_i - c_j)^2 \quad (6)$$

Jarak kumulatif terkecil atau jarak DTW pada (m, n) didefinisikan sebagai berikut:

$$d_{DTW}(Q, C) = \min_{P \in \mathcal{P}} \{\sqrt{\sum_{k=1}^K d_{wk}}\} \quad (7)$$

Dimana P adalah sekumpulan dari semua warping path yang mungkin, d_{wk} adalah elemen (ij) ke-k pada warpinhpath dan K adalah panjang dari warping path [8].

G. *Autocorrelation Function (ACF) Distance*

Ilustrasi untuk perhitungan jarak ACF adalah sebagai berikut, diberikan dua data time series $Q = q_1, q_2, q_3, \dots, q_m$, dimana dengan ukuran n dan $C = c_1, c_2, c_3, \dots, c_n$ dimana

$\hat{\rho}_Q = (\hat{\rho}_{1,Q}, \dots, \hat{\rho}_{L,Q})^t$ dan $\hat{\rho}_C = (\hat{\rho}_{1,C}, \dots, \hat{\rho}_{L,C})^t$ adalah vektor autokorelasi hasil pendugaan dari deret waktu Q dan C untuk beberapa L seperti dimana $\hat{\rho}_{i,Q} \approx 0$ dan $\hat{\rho}_{i,C} \approx 0$ untuk $i > L$. Jarak antara dua deret waktu kemudian dapat dibentuk pada persamaan sebagai berikut

$$d_{ACF}(Q, C) = \{(\hat{\rho}_Q - \hat{\rho}_C)^t \Omega (\hat{\rho}_Q - \hat{\rho}_C)\}^{1/2} \quad (8)$$

Dimana $d_{ACF}(Q, C)$ adalah jarak autokorelasi antara Q dan C sedangkan Ω adalah matriks pembobot, apabila jarak ACF tidak memiliki pembobot maka Ω adalah matriks identitas.

H. *Autocorelation Function*

Autocorrelation Function (ACF) merupakan suatu fungsi yang digunakan untuk menjelaskan korelasi antar Q_t dan $Q_{(t+k)}$ dari suatu proses yang sama dan hanya terpisah oleh lag waktu ke-k. Sehingga dengan menggunakan perhitungan ACF dapat diketahui apakah data pengamatan saling berhubungan disebut juga korelasi temporal. Perhitungan ACF untuk data time series Q_1, Q_2, \dots, Q_n dapat diperoleh melalui persamaan berikut [9]

$$\rho_k = \frac{\sum_{t=1}^{n-k} (Q_t - \bar{Q})(Q_{t+k} - \bar{Q})}{\sum_{t=1}^n (Q_t - \bar{Q})^2}, k = 0, 1, 2, \dots \quad (9)$$

Untuk mengetahui apakah koefisien autokorelasi signifikan atau tidak, perlu dilakukan uji dengan hipotesis sebagai berikut

$H_0: \rho_k = 0$ (koefisien autokorelasi tidak signifikan)

$H_1: \rho_k \neq 0$ (koefisien autokorelasi signifikan)

Statistik uji yang digunakan seperti pada persamaan sebagai berikut [9]

$$t = \frac{\widehat{\rho}_k}{SE(\rho_k)} \quad (10)$$

dengan nilai $SE(\rho_k)$ diperoleh pada persamaan berikut

$$SE(\rho_k) = \sqrt{\frac{1}{n}(1 + 2\widehat{\rho}_1^2 + 2\widehat{\rho}_2^2 + \dots + 2\widehat{\rho}_{k-1}^2)} \quad (11)$$

Kriteria keputusan tolak H_0 jika $|t_{hit}| > t_{\alpha/2, n-1}$, dimana nilai α merupakan tingkat kesalahan. Selain itu dapat menggunakan batas signifikansi dengan persamaan sebagai berikut

$$0 \pm t_{\alpha/2, n-1} SE(\rho_k) \quad (12)$$

Dikatakan signifikan ketika nilai $\widehat{\rho}_k$ berada di luar batas signifikansi dengan nilai α tertentu (Wei, 2006).

I. Pseudo-f Statistics

Penentuan jumlah kluster optimum pada pengelompokan data berskala numerik digunakan nilai *pseudo-f* [10]. Nilai *pseudo-f* tertinggi menunjukkan bahwa jumlah kelompok telah optimal. Kelompok dikatakan optimal apabila homogen antar anggota dalam kelompok yang sama dan heterogen antar anggota dalam kelompok lain.

$$f - Pseudo = \frac{(R^2 / (c-1))}{((1-R^2) / (n-c))} \quad (13)$$

$$R^2 = \left(\frac{SST - SSW}{SST} \right) \quad (14)$$

$$SST = \sum_{i=1}^c \sum_{j=1}^{n_c} \sum_{k=1}^p (x_{ijk} - \bar{x}_k)^2 \quad (15)$$

$$SSW = \sum_{i=1}^c \sum_{j=1}^{n_c} \sum_{k=1}^p (x_{ijk} - \bar{x}_{ik})^2 \quad (16)$$

J. Koefisien Silhouette

Salah satu ukuran ketepatan yang dapat digunakan dalam menentukan ketepatan pengelompokan deret waktu adalah koefisien silhouette yang digunakan untuk mengetahui kualitas pengelompokan [11]

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (17)$$

dimana $a(i)$ adalah rata-rata jarak antar anggota dalam kelompok, $b(i)$ adalah nilai minimum dari rata-rata jarak dari objek i dengan objek yang berada di cluster lainnya, sedangkan $S(i)$ merupakan nilai koefisien silhouette pada objek ke- i . Nilai koefisien silhouette secara keseluruhan merupakan rata-rata $S(i)$ setiap objek. Pada Tabel 1 ditampilkan kriteria koefisien silhouette yang ditetapkan oleh Kauffman dan Rousseeuw (1990) [12].

Tabel 1. Kriteria Koefisien Silhouette

Nilai Koefisien Silhouette	Keterangan
0,71 – 1,00	Strong
0,51 – 0,70	Good
0,26 – 0,50	Weak
0,00 – 0,25	Bad

K. Evaluasi Kinerja Terbaik

Untuk melihat kinerja terbaik digunakan kriteria dua nilai simpangan baku, yaitu dalam kelompok (S_w) dan antar kelompok (S_B) [1]. Rata rata simpangan baku dalam kelompok dinyatakan dengan:

$$S_w = \frac{1}{K} \sum_{k=1}^K S_k \quad (18)$$

Dimana, K adalah banyaknya kelompok yang terbentuk dan S_k merupakan simpangan baku kelompok ke- k . Sedangkan simpangan baku antar kelompok dinyatakan sebagai berikut:

$$S_B = \left[\frac{1}{c-1} \sum_{k=1}^c (\bar{X}_k - \bar{X})^2 \right]^{1/2} \quad (19)$$

Dimana, \bar{X}_k adalah rata rata kelompok ke- k dan \bar{X} adalah rata rata seluruh kelompok. Semakin kecil nilai S_w dan semakin besar nilai S_B maka metode tersebut memiliki kinerja yang baik, artinya mempunyai homogenitas yang tinggi.

L. Curah Hujan

Curah hujan merupakan jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi (mm) di atas permukaan horizontal bila tidak terjadi evaporasi, *run off*, dan infiltrasi. Curah hujan 1 (satu) millimeter, artinya dalam luasan satu meter persegi pada tempat yang datar tertampung air setinggi 1 (satu) millimeter atau tertampung air sebanyak 1 (satu) liter atau 1000 ml [2].

M. Zona Musim

Zona Musim (ZOM) adalah daerah yang pola hujan rata-ratanya memiliki perbedaan yang jelas antara periode musim kemarau dan periode musim hujan. Wilayah ZOM tidak selalu sama dengan luas daerah administrasi pemerintahan. Dengan demikian satu kabupaten/ kota dapat saja terdiri dari beberapa ZOM dan sebaliknya satu ZOM dapat terdiri dari beberapa kabupaten. Untuk wilayah Jawa Timur merupakan daerah yang memiliki pola monsoon, dan dibagi dalam 60 daerah zona musim [2].

N. Pengolahan Data Spasial

Overlay adalah proses tumpang – susun beberapa buah peta tematik dalam rangkaian kegiatan pengambilan kesimpulan secara spasial [13]. *Overlay* menampilkan suatu peta digital pada peta digital yang lain beserta atribut atributnya dan menghasilkan peta gabungan keduanya yang memiliki informasi atribut dari kedua peta tersebut.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Dinas Pekerjaan Umum dan Sumber Daya Air Provinsi Jawa Timur

B. Variabel Penelitian

Variabel penelitian yang digunakan adalah curah hujan bulan Januari 2011 hingga Desember 2017.

C. Langkah Analisis

- Melakukan *clustering* dengan menggunakan jarak *dynamic time warping* (DTW) pada data curah hujan Melakukan *clustering* dengan menggunakan jarak DTW dan ACF
- Memilih jumlah *cluster* terbentuk menggunakan kriteria *pseudo f statistics* dan koefisien *silhouette*.
- Melakukan analisis faktor dari data curah hujan
- Membuat visualisasi hasil kluster dalam bentuk peta dengan data lintang dan bujur
- Membuat kontur data curah hujan dari hasil analisis factor
- Peta yang sudah terbentuk kemudian di *overlay* atau ditumpangtindihkan dengan kontur curah hujan.

7. Menentukan batas zona musim pada hasil (g) dengan cara di *overlay* dengan data ketinggian
8. Melakukan uji perbedaan rata rata dari pengelompokkan zona musim yang terbentuk.
9. Menghitung evaluasi kinerja pengelompokkan hasil penelitian dan pengelompokkan dari BMKG dengan rasio simpangan baku S_w dan S_B
10. Membandingkan ketepatan pengelompokkan hasil penelitian dengan hasil zona musim BMKG.

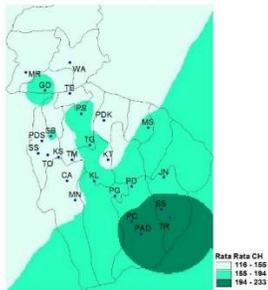
IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Curah Hujan di Mojokerto

Karakteristik curah hujan di Mojokerto meliputi rata rata curah hujan di Mojokerto pada bulan Januari 2011 hingga Desember 2017 dan pola curah hujan tiap bulan di Mojokerto. Terdapat 25 pos hujan yang terletak di Mojokerto.

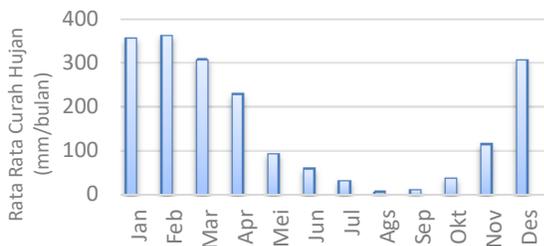
1) Rata Rata Curah Hujan di Mojokerto

Dari Gambar 2 dapat diperoleh informasi bahwa daerah yang memiliki curah hujan rendah di Mojokerto berwarna hijau muda dan curah hujan tertinggi berwarna hijau tua.



Gambar 2 Rata – rata Curah Hujan di Mojokerto

Gambar 3 memberikan informasi bahwa tipe curah hujan di Mojokerto berbentuk U atau dapat disebut tipe monsunial yang mengalami musim hujan di awal dan akhir tahun. Puncak curah hujan tertinggi terjadi pada bulan Desember, Januari, dan Februari. Hal ini dapat dimungkinkan karena pada bulan tersebut terjadi musim penghujan. Sedangkan pada bulan Juli hingga Oktober memiliki rata rata curah hujan terendah, hal ini dapat mengindikasikan bahwa pada pada saat itu terjadi musim kemarau.



Gambar 3 Rata- rata Curah Hujan di Mojokerto

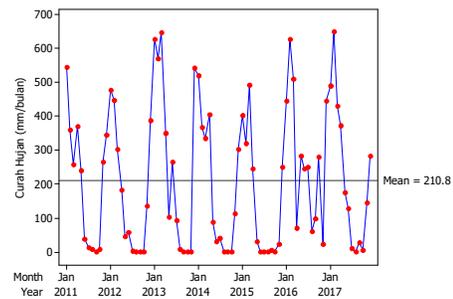
2) Time Series Plot

Salah satu karakteristik curah hujan bisa dilakukan dengan menggunakan *time series plot* untuk melihat bagaimana kondisi curah hujan di Mojokerto.

a. Time Series Plot Pos Hujan Pacet

Curah hujan tertinggi selalu berada di awal dan akhir tahun. Hal ini berarti bahwa musim penghujan terjadi pada setiap awal dan akhir tahun. Pada pertengahan tahun 2016 curah hujan di

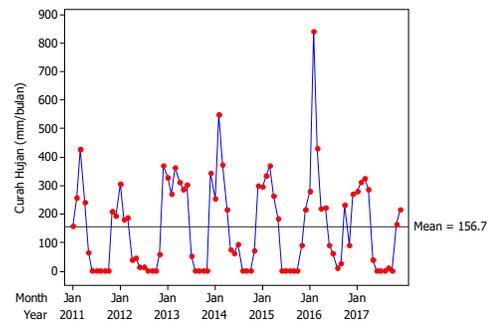
pos hujan Pacet mengalami peningkatan fluktuasi dibandingkan tahun tahun sebelumnya. Dimana terdapat rata rata curah hujan pada pertengahan tahun 2016 berada di atas rata rata



Gambar 4. Time Series Plot Curah Hujan di Pacet

b. Time Series Plot Pos Hujan Mojosari

Berikut merupakan *time series plot* curah hujan bulanan dari pos hujan di Mojosari dari tahun 2011 hingga 2017 atau selama 7 tahun.



Gambar 5 Time Series Plot Curah Hujan di Mojosari

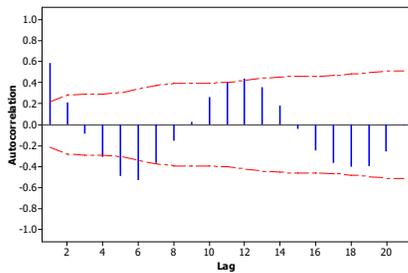
Dibandingkan dengan *time series plot* dari pos hujan Pacet, rata rata pada pos hujan Mojosari memiliki curah hujan lebih rendah dibandingkan dengan di Pacet. Selama periode 7 tahun yaitu dari tahun 2011 hingga 2017 curah hujan tertinggi terjadi pada tahun 2016 dimana mengalami kenaikan curah hujan yang cukup signifikan pada puncak curah hujan tertinggi dari tahun tahun sebelumnya .

B. Pengujian Autokorelasi Temporal

Pengujian autokorelasi temporal dilakukan untuk mengetahui data pengamatan curah hujan tiap bulan saling berhubungan. Pengujian ini menggunakan nilai ACF (*Autocorrelation Function*). Adanya autokorelasi saat terdapat lag (nilai ACF) keluar dari batas signifikansi. Berikut merupakan pengujian yang dilakukan pada beberapa pos hujan

a. ACF Pos Hujan Trowulan

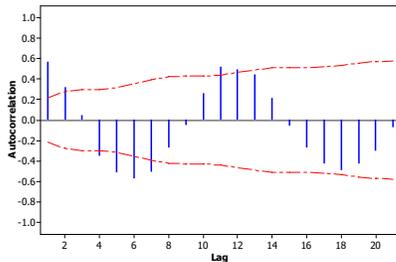
Plot ACF dari pos hujan Trowulan disajikan dalam Gambar 6. Pada Gambar 6 terdapat lag signifikan yang ditandai dengan keluarnya nilai ACF dari batas signifikan, maka data curah hujan di pos hujan Trowulan memiliki autokorelasi temporal.



Gambar 6. ACF Data Curah Hujan Trowulan

b. ACF Pos Hujan Sukosari

Pada Gambar 8 terdapat lag signifikan yang ditandai dengan keluarnya nilai ACF dari batas signifikan. Taraf signifikan yang digunakan sebesar 5%. Salah satu lag yang keluar adalah pada lag 1, maka data curah hujan di pos hujan Sukosari memiliki autokorelasi temporal.



Gambar 7 ACF Data Curah Hujan Sukosari

C. Analisis Faktor

Sebelum melakukan pengelompokan stasiun curah hujan dilakukan pereduksian variabel dengan analisis faktor. Tujuannya untuk menggambarkan karakteristik curah hujan di Mojokerto.

Tabel 3. Nilai Component Rotated Matrix

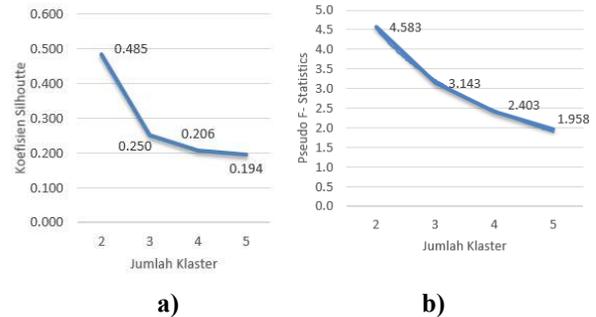
Bulan	F 1	F 2	F 3	F 4
Juni	0,818	0,272	0,176	-0,005
Mei	0,783	0,057	0,311	0,066
Juli	0,780	0,383	0,002	0,223
November	-0,597	0,094	0,393	0,487
Agustus	0,509	0,492	-0,198	0,257
Oktober	0,249	0,890	-0,068	0,088
September	0,144	0,860	-0,050	-0,013
Februari	0,121	0,663	0,516	-0,098
Maret	0,032	0,200	0,822	-0,001
Januari	0,190	-0,053	0,783	0,170
April	0,004	-0,230	0,685	0,169
Desember	0,168	0,094	0,213	0,869

Tabel 3 menunjukkan bahwa karakteristik curah hujan di Mojokerto apabila dilakukan analisis factor dihasilkan 4 faktor. Faktor 1 yang beranggotakan bulan Mei, Juni, Juli, Agustus, dan November menggambarkan musim transisi, Faktor 2 yang meliputi bulan Oktober, September, dan Februari. Faktor 3 yang menggambarkan musim penghujan beranggotakan bulan Januari, Maret, dan April. Dan faktor 4 yang merupakan bulan transisi dari musim kemarau ke musim penghujan beranggotakan bulan Desember.

Pengelompokan curah hujan di Mojokerto menggunakan analisis faktor dengan menghasilkan 4 faktor memiliki keragaman 73,37%.

D. Clustering dengan Jarak ACF

Pada tahap clustering digunakan jumlah cluster sebanyak 2 hingga 5 cluster. Untuk menentukan cluster optimal yang selanjutnya menjadi penentuan zona musim menggunakan kriteria koefisien silhouette dan pseudo-f statistics yang ditampilkan pada Gambar 9.

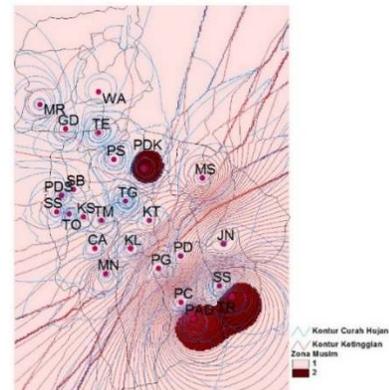


Gambar 8 (a) Koefisien Silhouette dan (b) Nilai Pseudo-f Statistics Hasil Pengelompokan dengan Jarak ACF

Gambar 8 memberikan informasi bahwa hasil kluster optimal berdasarkan kriteria silhouette dan pseudo-f statistics adalah 2 kluster atau menghasilkan 2 zona musim di Mojokerto. Dapat diketahui bahwa semakin banyak kluster yang terbentuk nilai koefisien silhouette dan pseudo-f statistics semakin kecil sehingga semakin tidak optimal dalam pengelompokan zona musim. Anggota kelompok masing masing zona musim adalah sebagai berikut

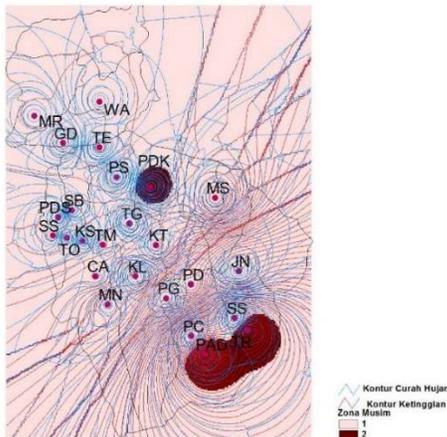
- Zona Musim 1 : Cakar Ayam, Gedeg, Janjing, Kasihan, Ketangi, Klegan, Mojosari, Pacet, Pandan, Pandansili, Pasinan, Pugeran, Sambiroto, Sumber Soko, Tampung, Tangunan, Trowulan, Manting, Sukosari, Mernung, dan Wates
- Zona Musim 2 : Pudaksari, Trawas, dan Padusan

Kontur yang berwarna biru menunjukkan kontur curah hujan dari faktor 1. Kontur ketinggian ditunjukkan dengan garis kontur berwarna merah. Pada kontur ketinggian menunjukkan bahwa semakin banyak garis kontur yang berdekatan atau berkumpul menunjukkan semakin tinggi suatu daerah. Pada Gambar 9 dapat dilihat bahwa pos hujan yang berada di dataran tinggi diantaranya adalah pos hujan Pacet, Trawas, Sukosari, dan Padusan.



Gambar 9 Pengelompokan Zona Musim Menggunakan Jarak ACF Pada Faktor 1

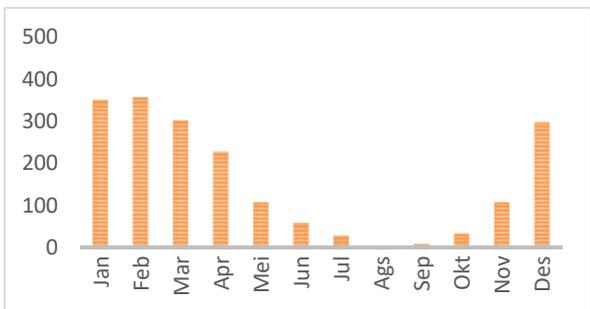
Berikut merupakan visualisasi pengelompokan zona musim dengan jarak ACF dimana dilakukan *overlay* pada ketinggian dan curah hujan dari faktor 3



Gambar 10 Pengelompokan Zona Musim Menggunakan Jarak ACF Pada Faktor 3

Dibandingkan dengan Gambar 9, kontur curah hujan pada faktor 3 lebih rapat dari faktor 1. Hal ini dikarenakan pada faktor 3 beranggotakan bulan-bulan musim penghujan.

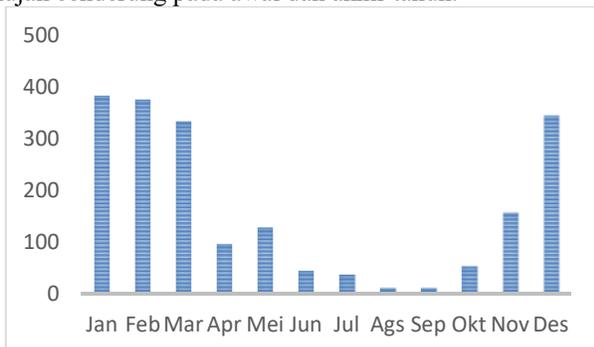
Sebagian besar sebaran pos hujan di Mojokerto termasuk ke dalam zona 1. Dari 3 pos hujan yang berada di zona musim, 2 pos hujan terletak di dataran tinggi dan sisanya pos hujan yang terletak di dataran rendah. Dengan demikian secara visual pengelompokan dengan jarak ACF tidak begitu terlihat pola pengelompokan zona musim yang terbentuk. Pola curah hujan pada setiap zona dilihat berdasarkan rata-rata curah hujan bulanan setahun pada Gambar 11 dan Gambar 12.



Gambar 11

11 Pola Curah Hujan Zona Musim 1 dengan Jarak ACF

Pola curah hujan di zona musim 1 berbentuk U atau memiliki tipe monsunal. Tipe monsunal mengalami musim penghujan cenderung pada awal dan akhir tahun.

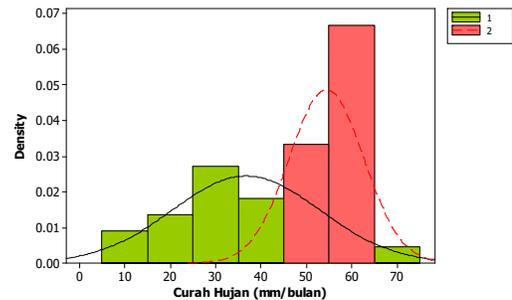


Gambar 12 Pola Curah Hujan Zona 2 Pada Klaster dengan Jarak ACF

Pada zona 2 musim kemarau dimulai lebih awal dari zona 1 yaitu pada bulan April hingga Oktober dengan puncak musim

kemarau pada bulan Agustus dan september dimana memiliki rata-rata curah hujan paling rendah. Sedangkan puncak dari musim penghujan terjadi pada bulan Januari.

Perbandingan rata-rata curah hujan zona 1 dan zona 2 menggunakan jarak ACF ditampilkan pada Gambar 13. Pada Gambar 13 diketahui bahwa pola curah hujan di zona 1 memiliki perbedaan dengan zona 2. Pada zona 1 curah hujan cenderung lebih stabil dibandingkan dengan zona 2.



Gambar 13 Perbandingan Rata Rata Curah Hujan Zona 1 dan Zona 2 dengan Jarak ACF

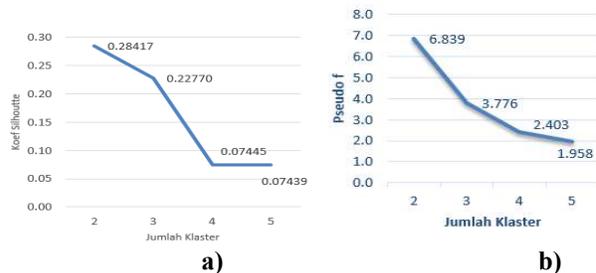
Untuk melihat apakah ada perbedaan antara kedua zona musim maka dilakukan uji t. Dengan taraf signifikansi 5%, perhitungan uji t menghasilkan nilai *p-value* 0,04. Karena 0,04 kurang dari 0,05 maka dapat disimpulkan adanya perbedaan rata-rata curah hujan pada pengelompokan zona musim menggunakan jarak ACF.

1) ANOVA

Uji ANOVA dilakukan untuk menganalisis variabilitas data menjadi dua sumber variasi yaitu variasi di dalam kelompok (*within*) dan variasi antar kelompok (*between*). Variabel yang digunakan merupakan rata-rata curah hujan tiap bulan. Dengan menggunakan taraf signifikansi 5% variabel rata-rata curah hujan pada bulan November memiliki keputusan tolak H_0 . Artinya rata-rata curah hujan bulan November untuk zona musim 1 dan zona musim 2 dengan menggunakan jarak ACF memiliki perbedaan. Sedangkan untuk rata-rata curah hujan di bulan yang lain selain bulan November tidak memiliki perbedaan.

E. Clustering dengan Jarak DTW

Pada tahap *clustering* dengan jarak DTW digunakan jumlah *cluster* sebanyak 2 hingga 5 *cluster*. Untuk menentukan *cluster* optimal yang selanjutnya menjadi penentuan zona musim menggunakan kriteria koefisien *silhouette* dan *pseudo-f-statistics* yang ditampilkan pada Gambar 14.



Gambar 14 (a) Koefisien *Silhouette* dan (b) Nilai *Pseudo-f-Statistics* Hasil Pengelompokan dengan Jarak DTW

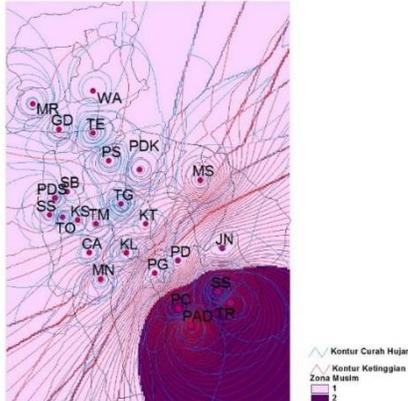
Dari Gambar 14 dapat diketahui bahwa semakin banyak

kelompok yang terbentuk maka nilai Koefisien *Silhouette* dan *pseudo-f statistics* semakin kecil. Sehingga yang dipilih adalah 2 klaster karena memiliki nilai koefisien *silhouette* dan *pseudo-f statistics* yang lebih tinggi dibandingkan 3,4,dan 5 klaster.

Anggota dari masing masing zona musim adalah sebagai berikut

- Zona Musim 1 : Cakar Ayam, Gedeg, Janjing, Kasihan, Ketangi, Klegen, Mojosari, Pudaksari, Pandan, Pandansili, Pasinan, Pugeran, Sambiroto, Sumber Soko, Tampung, Tangunan, Trowulan, Manting, Mernung, dan Wates
- Zona Musim 2 : Pacet, Trawas, Sukosari, dan Padusan

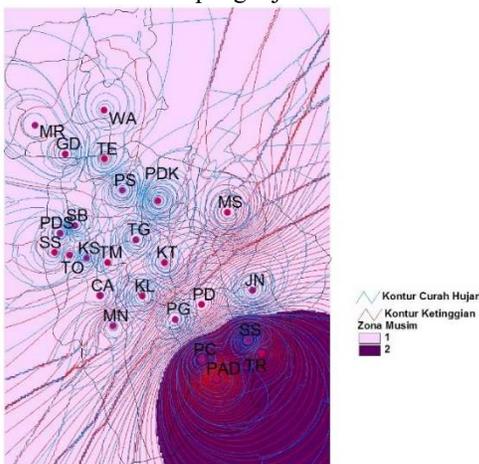
Berikut merupakan visualisasi dari pembentukan zona musim dengan jarak DTW yang ditampilkan pada Gambar 15.



Gambar 15 Pengelompokan Zona Musim Menggunakan Jarak DTW dengan Faktor 1

Sebagian dari pos hujan di Mojokerto berada di zona 1, sedangkan hanya ada 4 pos hujan saja yang berada di zona 2 yaitu Sukosari, Pacet,Padusan, dan Trawas. Ke empat pos hujan tersebut memiliki ketinggian yang lebih tinggi dibandingkan dengan pos hujan yang berada di zona 1. Secara visual pengelompokkan dengan jarak DTW dapat menangkap pola pengelompokkan dari zona musim yang terbentuk. Zona musim yang terbentuk dibedakan dari ketinggian tempat pos hujan..

Gambar 15 dan Gambar 16 menunjukkan perbedaan kontur curah hujan dari faktor 1 dan faktor 3. Factor 1 memiliki kategori musim penghujan sedangkan faktor 3 merupakan kategori musim transisi dari penghujan ke musim kemarau.

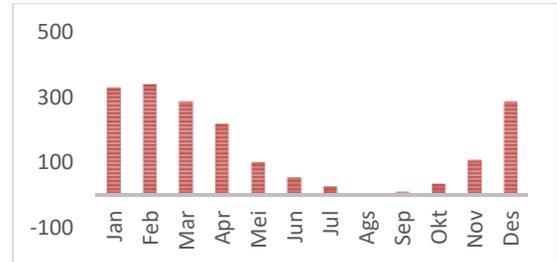


Gambar 16. Pengelompokan Zona Musim Menggunakan Jarak DTW dengan Faktor 3

Pada Gambar 17 . Zona musim 1 memiliki pola curah hujan monsunal, dimana musim hujan terjadi pada awal dan akhir

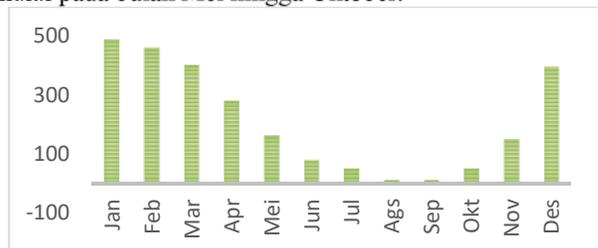
tahun. Puncak musim penghujan terjadi di bulan Februari karena memiliki curah hujan tertinggi. Musim kemarau di zona 1 dimulai pada bulan Mei hingga November.

Gambar 17 menunjukkan pola curah hujan pada zona 1, dimana pada zona 1 memiliki tipe U atau tipe monsunal. Tipe zona tersebut akan mengalami musim hujan di awal dan di akhir tahun.



Gambar 17. Pola Curah Hujan Zona Musim 1 dengan Jarak DTW

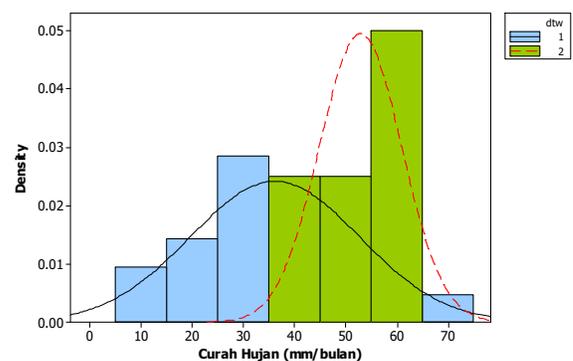
Pada zona 1 memiliki puncak curah hujan tertinggi atau puncak musim hujan pada bulan Januari dan Februari. Sedangkan musim kemarau diprediksi pada pertengahan bulan dimulai pada bulan Mei hingga Oktober.



Gambar 18. Pola Curah Hujan Zona Musim 2 dengan Jarak DTW

Di zona musim 2 yang meliputi Pacet, Trawas, Padusan, dan Sukosari memiliki pola curah hujan monsunal. Puncak dari musim penghujan terjadi pada bulan Januari dan puncak dari musim kemarau pada bulan Agustus dan September.

Perbandingan rata rata curah hujan zona 1 dan zona 2 menggunakan jarak DTW ditampilkan pada Gambar 20. dari Gambar 20 diketahui bahwa rata rata curah hujan pada zona 2 lebih tinggi dibandingkan dengan zona 1.



Gambar 19 Perbandingan Rata Rata Curah Hujan Zona 1 dan Zona 2 dengan Jarak DTW

Untuk melihat apakah ada perbedaan antara kedua zona musim maka dilakukan uji t. Dengan taraf signifikansi 5%, perhitungan uji t menghasilkan nilai *p-value* 0,015. Karena nilai *p-value* kurang dari 0,05 maka dapat disimpulkan adanya perbedaan rata rata curah hujan pada pengelompokkan zona musim menggunakan jarak ACF.

1) ANOVA

Berikut merupakan hasil dari uji ANOVA dari variabel curah

hujan pada pengelompokkan menggunakan jarak DTW dengan taraf signifikansi 5%. Kesimpulannya rata rata curah hujan pada bulan September dan Oktober di zona 1 maupun zona 2 tidak memiliki perbedaan. Hal ini dapat terjadi karena pada bulan tersebut terjadi musim kemarau sehingga tidak ada perbedaan rata rata curah hujan. Pada bulan lainnya zona 1 dan zona 2 memiliki perbedaan rata rata curah hujan.

F. Perbandingan Hasil Cluster Optimal dengan Zona Musim BMKG

Terdapat 3 zona musim di Mojokerto berdasarkan pengelompokkan yang dilakukan oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Untuk melihat evaluasi kinerja zona musim hasil *time series based clustering* dengan zona musim milik BMKG digunakan kriteria dua nilai simpangan baku, yaitu dalam kelompok (S_w) dan antar kelompok (S_B).

Tabel 5 Perbandingan Hasil Cluster Optimal dengan Zona Musim BMKG

Metode	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Ags	Sep	Okt	Nov	Des	
ACF	SW	106.51	99.83	73.16	53.11	46.47	17.98	17.19	6.81	4.33	12.29	29.94	67.11
	SB	400.17	384.32	349.93	229.78	139.16	37.39	40.47	13.98	12.85	63.90	183.12	369.38
	Rasio	0.27	0.26	0.21	0.23	0.33	0.48	0.42	0.49	0.34	0.19	0.16	0.18
DTW	SW	43.16	49.74	36.75	28.75	31.31	25.08	14.85	4.22	5.32	12.28	31.71	42.48
	SB	574.06	526.74	464.60	315.12	200.36	93.69	67.68	18.84	14.59	61.83	174.06	453.43
	Rasio	0.08	0.09	0.08	0.09	0.16	0.27	0.22	0.22	0.36	0.20	0.18	0.09
BMKG	SW	41.93	47.20	39.79	27.72	32.59	20.83	13.20	4.68	5.53	13.79	31.06	39.78
	SB	312.05	335.35	256.17	158.18	100.23	57.06	34.12	8.21	8.05	35.50	93.78	261.26
	Rasio	0.13	0.14	0.16	0.18	0.33	0.37	0.39	0.57	0.69	0.39	0.33	0.15

Telah diketahui bahwa semakin kecil nilai S_w dan semakin besar nilai S_B maka metode tersebut memiliki kinerja yang baik. Dari Tabel 5 diketahui bahwa metode pengelompokkan dengan jarak DTW memiliki nilai Rasio S_w/S_B paling kecil diantara pengelompokkan dengan jarak ACF dan milik BMKG. Hal ini berarti metode dengan jarak DTW merupakan metode terbaik untuk mengelompokkan zona musim di Mojokerto. Karena simpangan baku dalam kelompok untuk metode pengelompokkan dengan jarak DTW lebih kecil berarti keragaman antar objek semakin homogen. Sedangkan simpangan baku antar kelompok lebih besar yang berarti keragaman antar kelompok semakin heterogen.

V. KESIMPULAN DAN SARAN

Dari hasil analisis yang telah dilakukan, diperoleh hasil diantaranya, dari rata rata curah hujan per bulan selama 7 tahun dapat diketahui bahwa curah hujan di Mojokerto memiliki pola bahwa daerah dataran tinggi memiliki intensitas curah hujan lebih tinggi dibandingkan dengan daerah dataran rendah. Di Mojokerto puncak curah hujan tertinggi berada di bulan Januari dan Februari. Dan puncak curah hujan terendah pada bulan Agustus dan September. Pola curah hujan di Mojokerto memiliki tipe monsun. Pada jarak ACF klaster optimum yang dihasilkan sebanyak 2 zona musim dan pada jarak DTW klaster optimum yang dihasilkan juga 2 zona musim. Evaluasi kinerja zona musim dilakukan dengan rasio simpangan baku dimana metode yang dibandingkan adalah pengelompokkan dengan jarak ACF, DTW, dan hasil pengelompokkan oleh BMKG. Berdasarkan hasil evaluasi kinerja zona musim dengan rasio simpangan baku metode terbaik yang dipilih adalah pengelompokkan dengan jarak DTW.

• Zona Musim 1: Cakar Ayam, Gedeg, Janjing, Kasihan, Ketangi, Klegen, Mojosari, Pandan, Pandansili, Pasinan,

Pudaksari, Pugeran, Sambiroto, Tampung, Tangunan, Terusan, Trowulan, Manting, Mernung, dan Wates

• Zona Musim 2 : Pacet, Trawas, Padusan, dan Sukosari.

Berdasarkan kesimpulan yang diperoleh dapat diberikan saran dalam penelitian ini penggunaan ACF secara visualisasi kurang menangkap pola pengelompokkan dari zona musim, oleh karena itu pada penelitian selanjutnya bisa menggunakan jarak yang lain dalam menganalisis pengelompokkan zona musim dengan metode time series based clustering.

DAFTAR PUSTAKA

- [1] W. Bunkers, J. Miller and A. DeGaetano, "Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique," *J. Climate*, vol. 9, pp. 130-146, 1996.
- [2] BMKG, 2012. [Online]. Available: <http://balai3.denpasar.bmkg.go.id/daftar-istilah-musim>. [Accessed 19 October 2018].
- [3] Rencher and C. Alvin, *Methods of Multivariate Analysis Second Edition*, United States of America: Wiley-Interscience, 2002.
- [4] R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis*, New Jersey: Pearson Education, 2007.
- [5] R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis*, New Jersey: Pearson Education, 2007.
- [6] S. Rani and G. Sikka, "Recent Techniques of Clustering of Time Series Data: A survey," *International Journal of Computer Applications*, vol. 52, 2012.
- [7] S. Aghabozorgi, A. Shirkhorshidi and T. Wah, "Time Series Clustering - A decade review," *Information System*, pp. 16-38, 2015.
- [8] V. Niennattrakul and C. Ratanamahatana, "On Clustering Multimedia Time and Dynamic Time Warping," *Computer Society*, pp. 733-738, 2007.
- [9] W. W. S. Wei, *Time Series Analysis*, New York: Addison Wesley, 2006.
- [10] A. Orpin and V. Kostylev, *Towards a Statistically Valid Method of Textural Sea Floor Characterization of Benthic Habitats*, 2006.
- [11] J. Hair, W. C. Black, W. J. Babin and R. Anderson, *Multivariate Data Analysis.*, New Jersey: Pearson Education International, 2010.
- [12] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data-An Introduction to Cluster Analysis*, New York: Wiley, 1990.
- [13] E. Budiyo, *Sistem Informasi Geografis dengan ArcView GIS*, Yogyakarta: Penerbit ANDI, 2010.