PENGGEROMBOLAN DERET WAKTU DENGAN PENDEKATAN UKURAN KEMIRIPAN PICCOLO UNTUK PERAMALAN CURAH HUJAN PROVINSI BANTEN*

Sarah Fadhlia[1], I Made Sumertajaya[2‡], and Anik Djuraidah[3]

[1]Department of Statistics, IPB University, Indonesia, sarah.m.amin@gmail.com
[2]Department of Statistics, IPB University, Indonesia, imsjaya.stk@gmail.com
[3]Department of Statistics, IPB University, Indonesia, anikdjuraidah@apps.ipb.ac.id
‡corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)
Vol 4 No 2 (2020), 382 - 391

Copyright © 2020 Sarah Fadhlia, I Made Sumertajaya, and Anik Djuraidah. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Time series data modeling can be done by modeling each object one by one. Monthly rainfall data is an example of time series data. The purpose of time series analysis is to find patterns of past data and then forecast the future characteristics of data. The data used in this study is the Banten Province rainfall data which contained 19 rainfall stations. So it will require 19 models to forecast the rainfall data. The pattern of time series data in Banten Province monthly rainfall data in several locations has similarities. So that the similarity of this pattern can be considered in the clusters. In time series clustering, the idea is to investigate the similarity of time series in a cluster. The accuracy of distance similarity size measurements is performed on the generation data generated from 3 models, namely AR (1), AR (2), and AR (3). The piccolo method has an average accuracy of 0.62. While the maharaj method has an average accuracy of 0.41. This means that the Ward hierarchical clustering method using the Piccolo distance approach has a greater accuracy value than the Maharaj distance approach. Furthermore, the Piccolo method can be used as an alternative to the excellent distance method for grouping time series data in case data. The Banten Province rainfall station has 3 optimal clusters. Modeling individual level and cluster level has accuracy values that are not much different.

Keywords: clustering time series, maharaj distance, piccolo distance.

* Received Jan 2020; Accepted Jul 2020; Published online on Jul 2020

1. Pendahuluan

Pemodelan data deret waktu dengan observasi yang banyak dapat dilakukan dengan memodelkan setiap objek secara satu persatu. Akan tetapi, hal tersebut akan mengakibatkan proses pemodelan menjadi tidak efisien. Karena semakin banyak objek obsevasi maka semakin banyak model yang akan dihasilkan. Proses pemodelan dapat menjadi lebih efisien dengan melakukan penataan objek pengamatan yang cukup banyak kedalam bentuk gerombol-gerombol objek pengamatan. Menurut Liao (2005) terdapat tiga kategori analisis penggerombolan deret waktu, yaitu pendekatan berbasis data mentah, pendekatan dengan menyarikan data mentah dan pendekatan dengan berbasis model.

Pada pendekatan tidak berbasis model terdapat beberapa metode jarak atau ukuran ketidakmiripan yang dapat dilakukan, diantaranya yaitu jarak *Minkowski, Dynamic Time Warping (DTW), Frechet,* otokorelasi, dan *Periodogram*. Adinugroho (2018) membangun model ARIMA harga minyak goreng konsumen perdesaan di 32 provinsi Indonesia. Pada penggerombolan data deret waktu dengan pendekatan berbasis model, dapat menggunakan model ARIMA. Dari data deret waktu dibuat model ARIMA yang sesuai. Kriteria pemilihan model yang layak dapat menggunakan uji kebaikan model AIC. kemudian mengukur jarak antar pasangan model ARIMA, selanjutnya melakukan penggerombolan. Penggerombolan dapat dilakukan dengan metode hirarki. Untuk pendekatan berbasis model metode jarak yang dapat digunakan adalah jarak Piccolo, Maharaj, dan *linear predictive coding cepstrum (LPC-Cepstral)* (Montero & Vilar, 2014).

Piccolo (1990) melakukan penelitian pengukuran jarak antar data deret waktu dengan menggunakan model *Autoregresive intgrated Moving average* (ARIMA). Dari deret waktu dibuat sebuah model ARIMA berdasarkan kriteria penentuan model terbaik. Piccolo (1990) menjelaskan bahwa kesamaan pada deret waktu dapat diperoleh berdasarkan kesamaan struktur model sehingga jarak antara koefisien model dapat dilakukan untuk mengukur kesamaan dari deret waktu.

Curah hujan merupakan salah satu unsur cuaca yang sangat berpengaruh terhadap sektor pertanian maupun sektor lainnya. Keteraturan pola dan distribusi curah hujan di suatu wilayah merupakan jaminan berlangsungnya aktifitas warga. Data curah hujan bulanan merupakan salah satu contoh data deret waktu. Tujuan dari analisis deret waktu adalah untuk menemukan bentuk pola dari data di masa lalu dan menggunakan pengetahuan ini untuk melakukan peramalan terhadap sifat-sifat dari data di masa yang akan datang. Pemodelan dan peramalan deret waktu dapat dilakukan pada suatu peubah tunggal (*univariate variable*) dengan objek pengamatan cukup banyak.

Berdasarkan uraian diatas, penelitian ini akan melakukan pemodelan ARIMA terhadap data deret waktu dengan membandingkan ukuran kemiripan Piccolo dan Maharaj. Pada data aplikasi akan dilakukan pada data curah hujan di Provinsi Banten di 19 stasiun, dan untuk memberikan efisiensi dalam pemodelan maka akan dilakukan penggerombolan untuk menentukan peramalan curah hujan. Tujuan dari penelitian ini adalah Mengevaluasi keakuratan analisis gerombol pada data deret waktu dengan menggunakan jarak Piccolo dan Maharaj dan Membangun model peramalan pada level gerombol dengan menggunakan jarak terbaik.

2. Metodologi

2.1 Data

Data yang digunakan dalam penelitian ini adalah data simulasi dan data riil. Data simulasi adalah data yang dibangkitkan dari model deret waktu AR. Kajian simulasi tersebut dibedakan atas nilai parameter dari setiap model yang dibangkitkan dengan batasan yang berbeda-beda. Data simulasi dibangkitkan sebanyak 30 peubah data deret waktu yang akan dikelompokkan ke dalam 3 gerombol. Masing-masing gerombol terdiri dari 10 data deret waktu dengan 100 periode waktu.

Data riil yang digunakan adalah data deret waktu intensitas curah hujan bulanan dari 19 stasiun yang terdapat di provinsi Banten pada tahun 1998 sampai dengan 2018 yang terdiri dari 252 periode waktu. Data ini diperoleh dari Badan Meteorologi, klimatologi, dan geofisika provinsi Banten.

2.2 Metode Penelitian

a. Kajian Simulasi

1. Membangkitkan data deret waktu sebanyak 3 gerombol yaitu gerombol AR (1), AR (2), dan AR (3). Setiap gerombol terdapat 10 objek pengamatan dan panjang waktu 100 titik. Prosedurnya adalah sebagai berikut:
   a. Gerombol 1 merupakan model AR (1) dengan parameternya adalah dibangkitkan dari sebaran seragam (0,0.5).
   b. Model AR (2) mempunyai 2 parameter (ϕ_1, ϕ_2). ϕ_1 adalah data bangkitan dari sebaran seragam (-0.8,0). Sedangkan ϕ_2 adalah data bangkitan dari sebaran (0,0.5).
   c. Parameter ϕ_1 yang digunakan pada model AR (3) adalah data bangkitan dari sebaran seragam (-0.9,0). Sedangkan ϕ_2 adalah data bangkitan dari sebaran (0.5,1), dan ϕ_3 adalah sebaran seragam (0,0.3) (Tabel 1).

Tabel 1: Parameter model AR

Objek	AR (1)	AR (2)		AR (3)		
	ϕ_1	ϕ_1	ϕ_2	ϕ_1	ϕ_2	ϕ_3
1	0,29	-0,36	0,43	-0,39	0,75	0,11
2	0,19	-0,57	0,04	-0,83	0,86	0,17
3	0,34	-0,05	0,39	-0,86	0,82	0,06
4	0,40	-0,12	0,26	-0,12	0,58	0,15
...			...			
9	0,41	-0,19	0,41	-0,61	0,64	0,21
10	0,18	-0,27	0,38	-0,29	0,61	0,24

2. Membangkitkan data deret waktu model AR (1), AR (2), dan AR (3) dengan parameter sesuai pada Tabel 1.
3. Menghitung ukuran ketidakmiripan dengan menggunakan jarak Piccolo.

Jika $\widehat{\prod X_T} = \left(\hat{\pi}_{1,X_T}, \hat{\pi}_{2,X_T,} \ldots., \hat{\pi}_{k_1,X_T}\right)'$ adalah vektor dari AR(k_1) dan $\widehat{\prod Y_T} = \left(\hat{\pi}_{1,Y_T}, \hat{\pi}_{2,Y_T,} \ldots., \hat{\pi}_{k_2,Y_T}\right)'$ adalah vektor dari AR(k_2) penduga parameter untuk X_T dan Y_T, maka jarak Piccolo menjadi:

$$d_{PIC}(X_T, Y_T) = \sqrt{\sum_{j=1}^{k} \left(\hat{\pi}'_{j,X_T} - \hat{\pi}'_{j,Y_T}\right)^2} \quad (1)$$

dengan:
$\mathbf{X}_T$ = vektor objek x.
$\mathbf{Y}_T$ = vektor objek y
k_1 = ordo *AR* untuk $\mathbf{X}_T$.
k_2 = ordo *AR* untuk $\mathbf{Y}_T$.
k = *max* (k_1,k_2)
$\hat{\pi}'_{j,X_T} = \hat{\pi}_{j,X_T}$, jika $j \leq k_1$ dan $\hat{\pi}'_{j,X_T} = 0$ untuk lainnya
$\hat{\pi}'_{j,Y_T} = \hat{\pi}_{j,Y_T}$, jika $j \leq k_2$ dan $\hat{\pi}'_{j,Y_T} = 0$ untuk lainnya

Piccolo berpendapat bahwa jika seri tidak stasioner, perlu dilakukan *differencing* sehingga menjadi stasioner. Jika data seri memiliki musim, maka musimannya harus dihapus sebelum melakukan analisis lebih lanjut (Montero & Vilar, 2014).

4. Menghitung ukuran ketidakmiripan dengan menggunakan jarak Maharaj.
   Untuk model ARMA yang *invertible* dan stasioner, Maharaj memperkenalkan uji statistik untuk menentukan jarak antar data deret waktu dengan formula

$$d_{MAH}(X_T, Y_T) = \sqrt{T}\left(\widehat{\prod} X_T - \widehat{\prod} Y_T\right)'\hat{V}^{-1}\left(\widehat{\prod} X_T - \widehat{\prod} Y_T\right) \quad (2)$$

   dengan $\widehat{\prod X_T}$ dan $\widehat{\prod Y_T}$ masing-masing adalah estimasi parameter AR (k) untuk deret X_T dan Y_T, dan $V = \sigma^2_{X_T} R^{-1}_{X_T}(k) + \sigma^2_{Y_T} R^{-1}_{Y_T}(k)$ dengan σ^2_{XT} dan σ^2_{YT} yang menunjukkan ragam dari proses *white noise*, dan R_{X_T} dan R_{Y_T} adalah matriks koragam dari kedua deret. Hipotesis nol dari uji ini adalah bahwa tidak ada perbedaan antara proses pembuatan dua data deret waktu yang stasioner. Dua data deret waktu dikelompokkan bersama jika nilai p-*value* lebih besar dari tingkat signifikansi yang ditentukan sebelumnya (Montero & Vilar, 2014).
5. Melakukan penggerombolan hirarki metode Ward dengan menggunakan jarak Piccolo dan jarak Maharaj.
6. Menentukan jumlah gerombol optimum berdasarkan 30 indeks validitas yang dikeluarkan oleh *packages* NbClust. Penentuan gerombol optimum dilihat melalui jumlah gerombol yang diberikan oleh mayoritas indeks validitas (Hermawati, 2018).
7. Tahap 2 sampai 6 dilakukan pengulangan sebanyak 100 kali.
8. Membandingkan hasil penggerombolan data deret waktu dengan menggunakan jarak Piccolo dan jarak Maharaj. Hasil penggerombolan dievaluasi dengan ukuran kemiripan. Ukuran kemiripan yang terbaik adalah hasil penggerombolan yang memiliki persentase kesesuaian objek paling tinggi dari gerombol asalnya.

b. Kajian Aplikasi

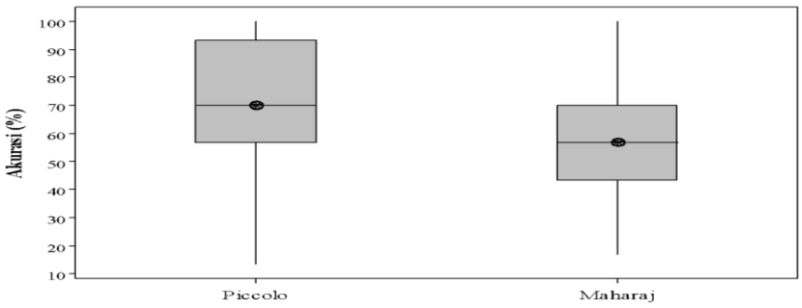
Langkah-langkah metode analisis pada data aplikasi adalah sebagai berikut.

1. Melakukan eksplorasi data curah hujan pada 19 stasiun di Provinsi Banten. Langkah awal ini penting dilakukan untuk mengetahui karakteristik data curah hujan di Provinsi Banten.
2. Melakukan pemodelan ARIMA pada level individu terhadap data curah hujan.
3. Menentukan model ARIMA terbaik berdasarkan dengan prosedur :
   a. Mengecek asumsi stasioner pada rataan dengan uji ACF, PACF, dan ADF
   b. Pendugaan parameter model dengan metode *Maximum Likelihood*. Penentuan model terbaik adalah dengan membandingkan nilai *Akaike's Information Criterion* (AIC).
   c. Pemeriksaan residual (sisaan) dilakukan dengan plot sisaan atau uji Ljung-Box
   d. Menentukan model terbaik
4. Melakukan penggerombolan hirarki metode Ward dengan menggunakan jarak Piccolo dan menentukan jumlah gerombol optimum seperti pada langkah kajian simulasi
5. Melakukan pemodelan pada level gerombol dengan menggunakan ukuran ketidakmiripan Piccolo dengan prosedur sebagai berikut :
   a. Sebelum melakukan pemodelan ditentukan terlebih dahulu satu data deret waktu yang mewakili setiap gerombol dengan menggunakan pendekatan nilai rata-rata (Aghabozorgi et al., 2015). Kemudian dilakukan pemeriksaan kestasioneran data deret waktu pada level gerombol.
   b. Seleksi model terbaik dengan memilih model yang mempunyai (AIC) minimum, dan pendugaan parameternya signifikan.
   c. Evaluasi model yaitu dengan melakukan pengujian terhadap residual model yang diperoleh. Model yang baik adalah model yang memiliki residual yang bersifat random. Pengujian signifikansi ACF dan PACF residual dilakukan dengan uji Ljung-Box.
6. Peramalan curah hujan pada level gerombol.

3. Hasil dan Pembahasan

3.1 Evaluasi Keakuratan Metode jarak Maharaj dan Piccolo

Kajian simulasi dilakukan pada data yang dibangkitkan dari 3 model deret waktu yaitu AR(1), AR(2), dan AR(3). Parameter yang digunakan pada ketiga model tersebut adalah hasil bangkitan dengan batasan yang berbeda-beda. Keakuratan jarak dilihat dari ketepatannya dalam mengklasifikasikan objek dalam gerombol asalnya. Nilai rata-rata salah klasifikasi pada metode 1 adalah lebih kecil dari metode 2. Metode 1 memiliki ragam yang lebih besar, artinya kesalahan klasifikasi pada metode Piccolo adalah cukup beragam dari 100 ulangan. Nilai minimum dari metode 1 adalah 0, artinya dari 100 ulangan metode Piccolo mampu mengklasifikasikan objek data sesuai dengan gerombol awalnya. Besar nilai maksimum kesalahan klasifikasi dari kedua metode adalah sama.



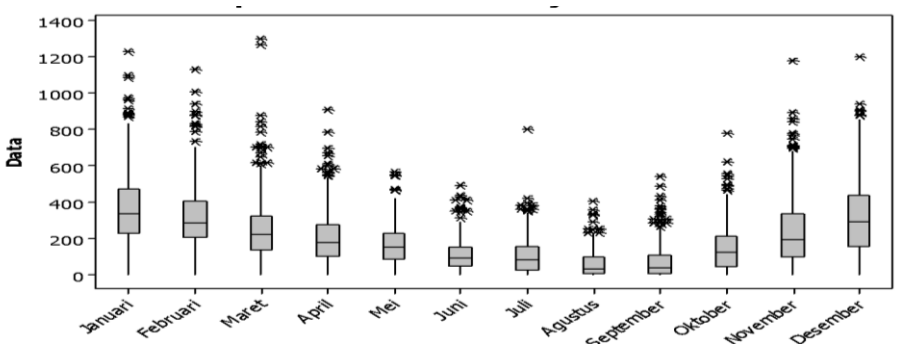
Gambar 1: Diagram kotak garis nilai akurasi evaluasi penggerombolan pada metode Piccolo dan metode Maharaj.

Secara umum pada metode Piccolo dan metode Maharaj yang diterapkan, kedua metode memiliki nilai keakuratan yang tinggi dalam melakukan penggerombolan. Metode Piccolo memiliki nilai rata-rata akurasi yaitu sebesar 70.1 persen (Gambar 1). Sementara metode Maharaj memiliki nilai rata-rata akurasi sebesar 56 persen. Artinya metode penggerombolan hirarki dengan menggunakan pendekatan jarak Piccolo memiliki nilai akurasi yang lebih besar dibandingkan dengan pendekatan jarak Maharaj. Akibatnya, berdasarkan kajian simulasi ini metode Piccolo yang diusulkan dapat digunakan sebagai alternatif metode jarak yang sangat baik untuk proses penggerombolan data deret waktu pada data kasus.

3.2 Penggerombolan Data Deret Waktu pada Data Curah Hujan Bulanan

Eksplorasi terhadap data curah hujan di Provinsi Banten dilakukan untuk melihat gambaran umum dari data curah hujan bulanan. Gambaran umum tersebut menunjukkan gambaran curah hujan pada setiap periode waktu. Gambaran secara umum data curah hujan bulanan dalam satuan milimeter ditampilkan pada Gambar 2.

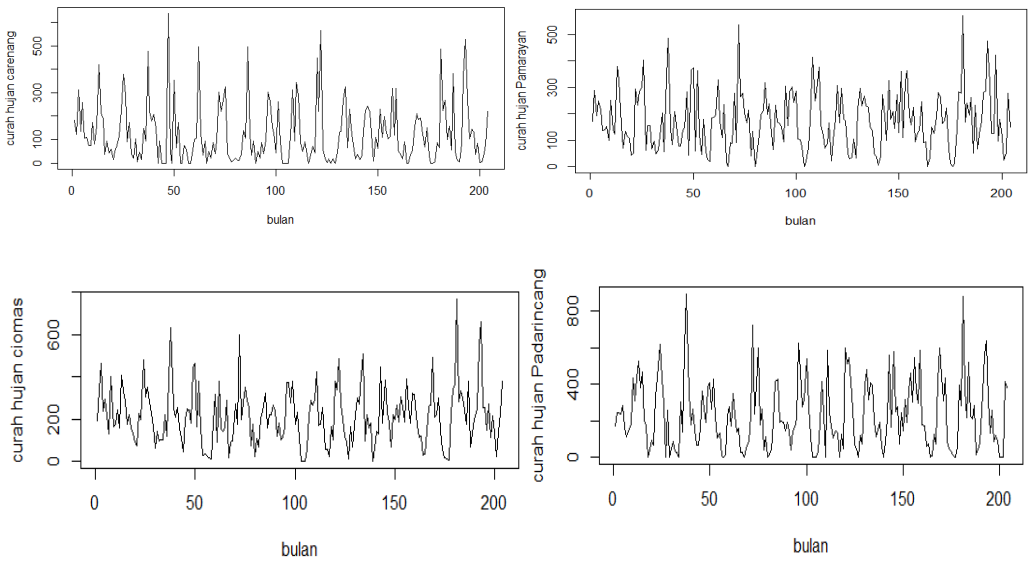
Deskripsi data pada Gambar 2 menunjukkan bahwa curah hujan mempunyai sebaran data yang bervariasi. Dalam kurun waktu bulan Januari tahun 1998 hingga bulan Desember tahun 2018, intensitas curah hujan paling rendah adalah 0 mm/bulan. Artinya tidak terjadi hujan selama 1 bulan pada kurun waktu 1998 hingga 2018 di Provinsi Banten.



Gambar 2: Gambaran umum curah hujan di Provinsi Banten.

Nilai rata-rata lebih besar dibandingkan nilai mediannya pada semua lokasi stasiun curah hujan. Hal ini menggambarkan bahwa pola sebaran data curah hujan menjulur ke kanan. Kemenjuluran kurva ke kanan menunjukkan bahwa terdapat pencilan positif. Stasiun Cibaliung, Jiput, dan Menes mempunyai variasi nilai yang lebih tinggi dibandingkan lokasi stasiun lainnya.

Stasiun curah hujan di Provinsi Banten yang menjadi kajian dalam penelitian ini terdiri dari 19 stasiun yaitu Baros, Carenang, Cinangka, Ciomas, Padarincang, Pamarayan, Petir, Pontang, Regas Hilir, Tirtayasa, Cibaliung, Cimanuk, Jiput, Labuhan, Menes, Pandeglang, Sepatan, Banjar Irigasi, dan Ciliman. Kestasioneran data deret waktu secara deskriptif dapat dilihat dari plot data terhadap waktu, koefisien otokorelasi maupun *correlogram* yang disajikan pada Gambar 3. Koefisien otokorelasi menunjukkan keeratan hubungan secara linear antara nilai peubah yang sama dengan periode waktu yang berbeda.



Gambar 3: Plot data intensitas curah hujan bulanan beberapa stasiun di Provinsi Banten.

Pemilihan model ARIMA dari setiap peubah adalah berdasarkan nilai AIC terkecil, sehingga model ARIMA untuk masing-masing stasiun secara berurutan adalah ARMA (3,0,3), ARMA (3,0,2), ARMA (3,0,2), ARMA (3,0,2), ARMA (2,0,3), ARMA (2,0,2), ARMA (2,0,3), ARMA (2,0,1), ARMA (3,0,2), ARMA (2,0,3), ARMA (3,0,2), ARMA (5,0,3), ARMA (4,0,1), ARMA (2,0,2), ARMA (3,0,2), ARMA (3,0,2), ARMA (2,0,1), ARMA (1,0,1), dan ARMA (2,0,1). Dari 19 model ARMA terdapat beberapa stasiun yang memiliki model deret waktu yang sama. Sehingga diasumsikan dapat dilakukan penggerombolan dengan tujuan untuk mendapatkan model yang lebih efisien.

3.3 Penentuan Jumlah Gerombol Optimum

Dalam menentukan jumlah gerombol optimum pada penggerombolan berhirarki juga dapat dilakukan dengan penghitungan indeks validitas. Charrad et al. (2014)

mengumpulkan 30 indeks validitas yang tersedia hanya dalam satu paket data yaitu *packages* NbClust pada program R. Mayoritas dari 30 indeks validitas menunjukkan gerombol optimum yang terbentuk adalah 3 gerombol. Diantara indeks yang mayoritas tersebut adalah indeks Hartigan, Scott, Ratkowsky, dan Ball.

Penentuan gerombol dengan indeks Hartigan, Scott, dan Ratkowsky adalah pada saat perubahan paling besar dari setiap level gerombol. Berdasarkan padaTabel 3 perubahan yang paling signifikan dari ketiga indeks tersebut adalah pada level 3. Sementara pada indeks Ball nilai k optimum adalah yang memiliki nilai indeks yang paling besar. Dengan memperhatikan mayoritas dari 30 indeks validitas maka nilai k optimum gerombol adalah pada level 3. Sehingga data curah hujan Provinsi Banten digerombolkan dalam 3 gerombol. penggerombolan data dengan menggunakan metode Piccolo.

Tabel 3: Indeks gerombol.

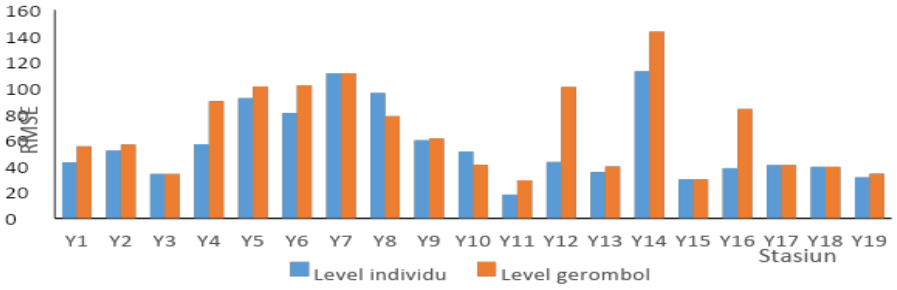
Jumlah Gerombol	Hartigan	Scott	Ratkowsky	Ball
2	552,68	427,47	0,05	30,34
3	5,14	547,09	0,50	4,73
4	7,62	632,39	0,43	3,44
5	47,26	669,38	0,39	2,63
6	280,15	774,62	0,37	1,71
7	30,87	859,29	0,36	0,54
8	21,12	884,49	0,34	0,40
9	52,34	911,07	0,32	0,31
10	95,45	996,88	0,31	0,21

Gerombol 1 terdiri dari stasiun Baros, Pontang, Cimanuk, Jiput, Sepatan, Banjar Irigasi, dan Ciliman. Gerombol 2 terdiri dari stasiun Carenang, Ciomas, Padarincang, Pamarayan, Petir, Regas Hilir, Labuhan, dan Menes. Sedangkan gerombol 3 terdiri dari 4 stasiun lainnya yaitu Cinangka, Tirtayasa, Cibaliung, dan Pandeglang. Metode pemilihan model pada level gerombol adalah seperti yang dilakukan pada level individu. Sehingga model deret waktu untuk masing-masing gerombol secara berurutan adalah ARMA (1,0,1), ARMA (2,0,3), dan ARMA (1,0,1).

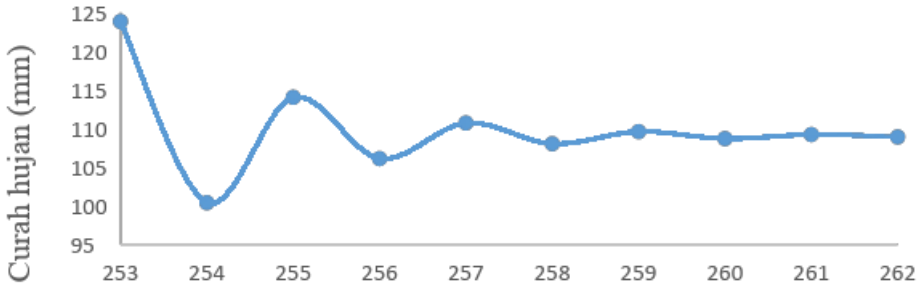
3.4 Evaluasi Pemodelan Level Individu dan Gerombol

Gambar 4 merupakan hasil perbandingan RMSE pemodelan pada level individu dan level gerombol. Garis tegak merupakan sumbu x yang menjelaskan nilai RMSE dari model deret waktu setiap stasiun. Garis horizontal merupakan sumbu y yang menunjukkan peubah dari 19 stasiun.

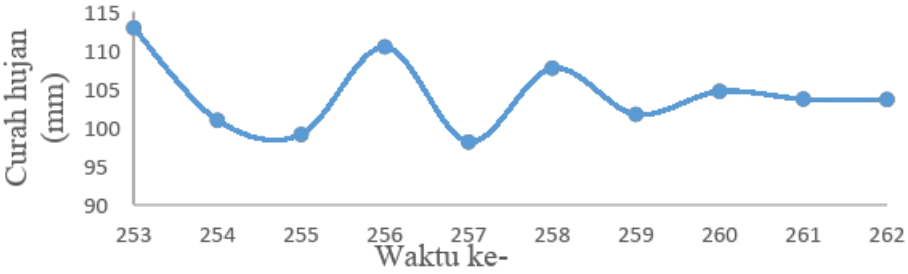
Secara eksploratif terlihat bahwa tidak ada perbedaan yang signifikan antara RMSE level individu dan level gerombol. Hal ini diperkuat dengan uji 2 populasi yang menghasilkan nilai p-value sebesar 0,293. Artinya tidak berbeda nyata antara RMSE pada kedua level. Sehingga dapat dikatakan pemodelan level gerombol layak dan efektif.



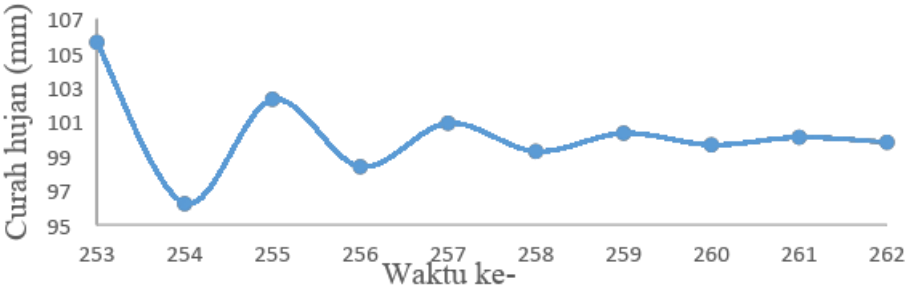
Gambar 4: Perbandingan RMSE level individu dan gerombol.



(a)



(b)



(c)

Gambar 5: Peramalan curah hujan pada level gerombol (a) gerombol 1, (b) gerombol 2, (c) gerombol 3.

3.5 Peramalan

Tujuan dari pemodelan level gerombol adalah untuk melakukan peramalan. Pada tahap peramalan dilakukan sebanyak 10 periode yaitu Januari – Oktober 2019. Gambar 5 menunjukkan hasil peramalan untuk 3 gerombol. Ketiga gerombol memiliki pola yang sama yaitu pola menurun.

4. Simpulan

Pada kajian simulasi yang dilakukan, metode Piccolo dan Maharaj dapat melakukan penggerombolan data deret waktu. Kedua metode pendekatan jarak dapat mengklasifikasikan objek ke dalam gerombol asalnya. Namun kedua metode jarak juga menghasilkan kesalahan klasifikasi. Berdasarkan nilai akurasi, metode jarak Piccolo memberikan nilai akurasi yang lebih tinggi daripada metode jarak Maharaj. Stasiun curah hujan Provinsi Banten memiliki gerombol optimum sebanyak 3 gerombol. Pemodelan level individu dan level gerombol memiliki nilai keakuratan yang tidak jauh berbeda.

Daftar Pustaka

Adinugroho, W. (2018). *Pengembangan model peramalan harga minyak goreng dengan pendekatan clustering time series*. Bogor (ID): IPB.

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering - A decade review. *Information Systems*, *53*(2015): 16–38.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). an R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6): 1–36.

Hermawati, F. (2018). *Perbandingan metode cluster ensemble, two step cluster dan metode gower pada penggerombolan data bertipe campuran (studi kasus pada penggerombolan kecamatan di kabupaten Malang tahun 2017)*. Bogor (ID): IPB.

Liao, T. W. (2005). Clustering of time series data-a survey. *Pattern Recognition*, *38*: 1857–1874.

Montero, P., & Vilar, J. A. (2014). TSclust: An R package for time series clustering. *Journal of Statistical Software*, *62*(1): 1–43.

Piccolo, D. (1990). A distance measure for classifying ARIMA models. *Journal of Time Series Analysis*, *11*(2): 153–164.