Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2020, 2020:52

https://doi.org/10.28919/cmbn/4778

ISSN: 2052-2541

A NEW APPROACH TO CLUSTERING TIME SERIES DATA USING THE ARIMA MODEL UNCERTAINTY

TRIYANI HENDRAWATI [1,2,*], AJI HAMIM WIGENA[2], I. MADE SUMERTAJAYA[2], BAGUS SARTONO[2]

[1]Department of Statistics, Padjadjaran University, Bandung, Indonesia

[2]Department of Statistics, IPB University, Bogor, Indonesia

Abstract: The piccolo distance is a simplified approach in clustering time-series data, and it needs the data analysts to determine the ARIMA model for each series. Some problems may arise in the modeling step because different criteria may lead to different orders of the best models. This current paper is discussing how to handle this model uncertainty problem by borrowing the concept of an ensemble in the area of data science. Instead of using a single criterion to identify the best model, we proposed to generate the best models using several different criteria. Each series was characterized by the average of the estimates of model parameters obtained. In the clustering process, we employed a hierarchical approached where the optimal number of clusters identified using the Silhouette coefficient. An extensive simulation was completed within this research, and we revealed that our proposed methodology could increase the correct cluster membership by more than 10%. We also implemented our methodology to identify clusters of areas in Indonesia (i.e., Province of Banten) based on the pattern of rainfall level and found an impressive result.

Keywords: model-based clustering; uncertainty model; time series clustering; rainfall data.

2010 AMS Subject Classification: 37M10, 91B84.

*Corresponding author

E-mail address: triyani.hendrawati@unpad.ac.id

Received June 18, 2020

1. INTRODUCTION

The clustering method of time series data is intensively developed, in line with application needs in various fields. One of the advantages of using time series data clustering is to simplify data analysis time by organizing data into similar groups [1]. Time series data are generally large, making it not easy to analyze directly, by conducting clustering first makes it easier, simpler, and more efficient for the next analysis step.

Triacca [25] applied clustering to temperature anomaly data. One of the applications in the field of economy is the annually clustering Gross Domestic Product (GDP) from several countries in the world [26]. In the health sector, Gullo et al. [12] grouped mass spectrometry data, which used as early detection of certain diseases. In climate data applications, time-series data clustering use as the first step before making a model. Rainfall data in several regions grouped first, then modeled according to groups of those regions [24]. There are several methods used to calculate the distance between time series include Piccolo distance [19], autocorrelation function (ACF) distance [11], dynamic time warping (DTW) distance [23], cepstral-based distance [14], Maharaj distance [16], periodogram-based distance [6], copula-based distance [26], and others.

The Piccolo distance is one of the popular methods used to clustering time-series data [19]. We obtained several ARIMA models from a sequence of time-series data, but only one model is selected based on one criterion. However, differences in the selection of model criteria can produce different models, and it can produce different cluster results. It raises the problem of uncertainty of which model will be chosen. There is much debate about which criteria are best for model selection. Some researchers disagree with only selecting the best model based on specific information criteria without considering other models [3], [4], [5].

In this paper, we propose a method for clustering time series data using several criteria to identify the best model. One of the advantages of this proposed method is to improve the correct cluster membership. The systematics of this paper is that in section 2 discusses the proposed method, section 3 presents Simulation, section 4 presents application to Clustering Rainfall Data. Finally section 5 contains conclusions.

2. PROPOSED METHOD

The use of multi models or average models is an alternative so that it does not only use one model that is not necessarily true [3], [5]. Averaging model is one of the ways to determine the model selection. The basic idea of the average model is that if we get several estimators for the same population, then why not combine these estimators to improve estimation. Suppose the value to be estimated is the parameter π, there are several different predictors for π, namely $\hat{\pi}_{\text{Model 1}}$, $\hat{\pi}_{\text{Model2}}$, …, $\hat{\pi}_{\text{Model } m}$ then we can make weight for the estimator:

$$\hat{\pi}_{\text{ Average Model}} = \sum_{i=1}^{m} w_i \, \hat{\pi}_{\text{Model } i} \qquad (1)$$

In this case, what is done is averaging parameter estimators, not averaging the models [8]. The averaging model approach gives an idea for developing the Piccolo distance.

Piccolo [19] introduces measurements of the distance between the time series using the ARIMA model. The distance of the two series is expressed by measuring the structural similarity between the two invertible ARIMA processes [6], [9], [19], [20]. The autoregressive expansion contains all useful information about the stochastic structures of a time series. The distance between two-time series X_t and Y_t calculates uses the model parameters:

$$d(X_t, Y_t) = [(\hat{\pi}_x - \hat{\pi}_y)'(\hat{\pi}_x - \hat{\pi}_y)]^{\frac{1}{2}} = \sqrt{\sum_{j=1}^{\infty} \left(\hat{\pi}_{j,x} - \hat{\pi}_{j,y}\right)^2} \qquad (2)$$

where: $\hat{\boldsymbol{\pi}}_x = (\hat{\pi}_{1,x}, \hat{\pi}_{2,x}, \ldots, \hat{\pi}_{j,x}, \ldots)$ and $\hat{\boldsymbol{\pi}}_y = (\hat{\pi}_{1,y}, \hat{\pi}_{2,y}, \ldots, \hat{\pi}_{j,y}, \ldots)$ are parameters of a autoregressive model (AR(∞)) for time series X_t and Y_t. For simplicity of the implementations, AR (p) is used as an approach for AR (∞), where p = 1, 2, …, p [17].

It differs from Piccolo [19] that only uses one model; the proposed method uses five models (multi model). The following are some criteria for choosing the best model that was used in the proposed method. The Akaike's Information Criterion (AIC) [2] formula written as

$$AIC = -2 \log(\hat{\sigma}_e^2) + 2k \qquad (3)$$

where k is the number of parameters, $\hat{\sigma}_e^2$ is the maximum likelihood estimator for the error variance. From this formula, we can see that the more parameters in the model will cause the AIC value to increase so that the best model tends to come from a model with a smaller number of

parameters; in other words, it will be choosing a simpler model. The formula for Bayesian Information Criterion (BIC) [22] is

$$BIC = -2\log(\hat{\sigma}_e^2) + k\log(n) \tag{4}$$

where n is the number of observations, k is the number of parameters, $\hat{\sigma}_e^2$ is the maximum likelihood estimator for the variance of the error. Notice the first term in the BIC formula is the same as the AIC formula; they only differ in the second term. In the second term, the BIC value is influenced by the number of sample sizes and parameters in the model. Now consider the Akaike's information Criterion Bias Corrected (AICc) formula which written as

$$AICc = AIC + \frac{2(k+1)(k+2)}{n-k-2} \tag{5}$$

AICc is a bias parameter corrected AIC. While *Mean Absolute Percentage Error* (MAPE) formulated as

$$\text{MAPE} = \frac{\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\text{x}100\right|}{n} \tag{6}$$

The equation for *Root Mean Squared Error* (RMSE) is

$$\text{RMSE} = \left(\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}\right)^{1/2} \tag{7}$$

where y_t is the observation at time-*t*, and $\hat{y}_t$ is the value of the fit of observation at time-*t* [10], [18].

For instance, model 1, model 2, model 3, model 4, and model 5 are the best models based on the criteria of AIC, BIC, AICc, RMSE, and MAPE, respectively. The parameters estimator are $\hat{\boldsymbol{\pi}}_{AIC}$, $\hat{\boldsymbol{\pi}}_{BIC}$, $\hat{\boldsymbol{\pi}}_{AICc}$, $\hat{\boldsymbol{\pi}}_{RMSE}$, dan $\hat{\boldsymbol{\pi}}_{MAPE}$. Therefore, the average parameters estimator is

$$\bar{\hat{\boldsymbol{\pi}}} = \begin{bmatrix}\bar{\hat{\boldsymbol{\pi}}}_1 \\ \dots \\ \bar{\hat{\boldsymbol{\pi}}}_p\end{bmatrix} = \begin{bmatrix}\frac{1}{5}\left(\hat{\pi}_{1,AIC} + \hat{\pi}_{1,BIC} + \hat{\pi}_{1,AICc} + \hat{\pi}_{1,RMSE} + \hat{\pi}_{1,MAPE}\right) \\ \dots \\ \frac{1}{5}\left(\hat{\pi}_{p,AIC} + \hat{\pi}_{p,BIC} + \hat{\pi}_{p,AICc} + \hat{\pi}_{p,RMSE} + \hat{\pi}_{p,MAPE}\right)\end{bmatrix} \tag{8}$$

The average parameters estimator of the model is compiled into new data representing the time series data of an object then used to compute the Piccolo distance. The distance between the X_t and Y_t time series in the proposed method is:

$$d(X_t, Y_t)_{new} = [(\bar{\bar{\boldsymbol{\pi}}}_{\boldsymbol{x}} - \bar{\bar{\boldsymbol{\pi}}}_{\boldsymbol{y}})'(\bar{\bar{\boldsymbol{\pi}}}_{\boldsymbol{x}} - \bar{\bar{\boldsymbol{\pi}}}_{\boldsymbol{y}})]^{\frac{1}{2}} = \sqrt{\sum_{j=1}^{p}\left(\bar{\bar{\pi}}_{j,x} - \bar{\bar{\pi}}_{j,y}\right)^2} \quad (9)$$

where $\bar{\bar{\boldsymbol{\pi}}}_x = \left(\bar{\bar{\pi}}_{1,x}, \bar{\bar{\pi}}_{2,x}, \ldots, \bar{\bar{\pi}}_{p,x}\right)$ *and* $\bar{\bar{\boldsymbol{\pi}}}_y = \left(\bar{\bar{\pi}}_{1,y}, \bar{\bar{\pi}}_{2,y}, \ldots, \bar{\bar{\pi}}_{p,y}\right)$ is the average parameter estimator of the autoregressive model (AR(p)) for time series data X_t and Y_t.

Let $\mathcal{M}$ is the class of ARMA invertible models, and lets $d(X_t, Y_t)_{new}$ be a metric on $\mathcal{M}$. The measure denotes $d(X_t, Y_t)_{new}$ satisfies the properties of a distance:

1. non-negativity

$$d(X_t, Y_t)_{new} \geq 0 \; \forall \; X_t, Y_t \in \mathcal{M}$$

2. symmetry

$$d(X_t, Y_t)_{new} = d(Y_t, X_t)_{new} \forall X_t, Y_t \in \mathcal{M}$$

3. triangularity

$$d(X_t, Y_t)_{new} \leq d(X_t, Z_t)_{new} + d(Z_t, Y_t)_{new} \; \forall \; X_t, Y_t, Z_t \in \mathcal{M}.$$

3. SIMULATION

In this study, we generated 30 time-series data from three clusters with the Autoregressive (AR (p)) model. The cluster is called group *A, B*, and *C*. In each group, ten rows are generated, so there are 30 rows of generated data. The parameter values from the generated data are as follows: model *A* is AR (0.2, 0.1), model *B* is AR (0.4, 0.5), model *C* is AR (0.6, 0.2). The values of the parameters used for this simulation have met stationary conditions [10], and these parameters also used in the study of Kumar and Patel [15]. Each data generation has a length of the observation period (t). There are six periods used: 50, 75, 100, 150, 200, and 300. Generated data for each group has two kinds of variation in groups. The group with small variations has ε_{it} distribution N ~ (0, 0.01), and group with large variations has a distribution of ε_{it} is N ~ (0, 1).

We grouped the generated data using the Piccolo distance and the proposed method, then compared it. The generated data modeled by using the Autoregressive (AR (p)) approach with p = 1, 2, ..., 5. Each object in the data will be selecting the best model; the model is chosen based on

the smallest AIC value. We used the model parameters' values to calculate distances using the Piccolo distance formula, as in equation (2). All distances of each object were calculated and formed the distance matrix. The method for clustering used Ward method. Each step in the Ward method combines the object pairs with the smallest increase in error sum of squares (ESS), with formula

$$\text{ESS} = \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})'(\boldsymbol{x}_j - \overline{\boldsymbol{x}}) \tag{10}$$

where $\boldsymbol{x}_j$ is *a j-th* object, and $\overline{\boldsymbol{x}}$ is an average of objects [13]. The optimal number of clusters is using an average value of the Silhouette index. The formula is

$$\text{Silhouette} = \frac{\sum_{i=1}^{n} S(i)}{n} \tag{11}$$

$$S(i) = \frac{b(i)-a(i)}{max\{a(i);b(i)\}} \tag{12}$$

where the value $a(i)$ is the average dissimilarity of object *i-th* to others in cluster C_r; the d_{iC_s} is the average dissimilarity of object *i-th* to others in clusters C_s; the $b(i)$ is $\min_{s \neq r}\{d_{iC_s}\}$; and *n* is the number of observations [3], [7], [21]. Repeat the clustering process as above, but use different criteria for model selection; based on the smallest BIC, AICc, RMSE, and MAPE values. This process was repeated 100 times.

For the proposed method, the generated data obtained were each modeled with the Autoregressive (AR (*p*)) approach with *p* = 1, 2, ..., 5. Each object in the data was selected for the five best models based on the smallest of AIC, BIC, AICc, RMSE, and MAPE. The average parameter estimator of the model is calculated by using the formula in equation (8). It is used to calculate distances with the proposed method, as seen in equation (9). The distance of all objects was calculated to form the matrix of distance. Then the distance matrix was grouped by the Ward method in equation (10). This process was repeated 100 times.

The correct cluster membership was evaluated for both Piccolo and the proposed method. Groups originating from the same generated data will be grouped into the same group as well. If the data comes from the same group but grouped differently, there has been an error of grouping

or missing class error. The percentage of correct cluster membership (CC) calculated by the formula:

$$\mathrm{CC} = \frac{number\ of\ object - number\ of\ miss\ class}{number\ of\ object} \times 100\% \qquad (13)$$

The average of correct cluster membership for 100 replications calculated. We compared the average of correct cluster membership of Piccolo with the proposed method. We also explored the proposed method for different order p in the autoregressive (AR (p)) model. Two kinds of autoregressive (AR (p)) model approach scenarios created, with p maximum of five and p maximum of 10.

Table 1 shows the results of Piccolo compared to the proposed method for generated data with variance 0.01. As the length of the observation period (t) increases, the percentage of correct cluster membership increases for both the proposed method and the Piccolo method. The use of different model selection criteria results in different correct cluster membership. In Table 1, the lowest and highest correct cluster membership values marked with bold numbers. The Piccolo method (with RMSE criterion) has the lowest correct cluster membership than other methods for t = 50. Otherwise, for t = 75, 100, 150, 200, and 300, the percentage of correct cluster membership of the Piccolo method (with MAPE criterion) has the lowest correct cluster membership compared to other methods.

Table 1. Percentage of correct cluster membership of the Piccolo and the proposed method for generating data with σ_e^2 = 0.01.

t	Piccolo Method					Proposed
	AIC	BIC	AICc	RMSE	MAPE	Method
50	78.03%	83.8%	79.53%	**72.47%**	74.27%	**84.53%**
75	82.67%	86.4%	83%	79.97%	**78.23%**	**88.8%**
100	90.27%	88.1%	90.67%	89.13%	**81.43%**	**92.83%**
150	94.77%	93.33%	94.93%	94.5%	**85.3%**	**96.3%**
200	96.5%	97.03%	96.33%	96.8%	**86.27%**	**97.3%**
300	98.87%	98.93%	99%	98.57%	**87.8%**	**99.2%**

The proposed method increases the correct cluster membership by 12.06% compared to the Piccolo method with RMSE criterion for t = 50. Compared to the Piccolo method with MAPE criteria, the proposed method succeeded in increasing the correct cluster membership by 10.57%, 11. 4%, 11%, 11.03%, 11.4% for t = 75, 100, 150, 200, and 300 respectively. The Piccolo method with AIC, BIC, and AICc criterion also provides lower correct cluster membership than the proposed method.

Table 2 shows the results of Piccolo compared to the proposed method for the group of generated data with variance 1. For t = 50, the proposed method increases the correct cluster membership by 12.56% compared to the Piccolo method with RMSE criterion. Proposed methods compared to Piccolo method with MAPE criterion increases the correct cluster membership by 11.06%, 10.93%, 10.63%, 11.8%, 11.36% for t = 75, 100, 150, 200, and 300 respectively. Table 1 and Table 2 indicate that the clustering using the proposed method gives better results than Piccolo.

Table 2. Percentage of correct cluster membership of the Piccolo and the proposed method for generating data with σ_e^2 = 1.

t	Piccolo Method					Proposed
	AIC	BIC	AICc	RMSE	MAPE	Method
50	78.03%	83.37%	78.67%	**71.57%**	73.73%	**84.13%**
75	85.17%	86.1%	83.23%	82.8%	**78.77%**	**89.83%**
100	89.43%	86.8%	90.47%	87.93%	**81.77%**	**92.7%**
150	95.17%	92.43%	94.37%	94.43%	**86.17%**	**96.8%**
200	97.5%	96.4%	97.23%	96.87%	**86.23%**	**98.03%**
300	98.37%	98.53%	98.63%	98.5%	**87.67%**	**99.03%**

The correct cluster membership of the proposed method evaluated with a different order of p in AR (p). The order, namely p maximum five and p maximum 10. Table 3 shows the correct cluster membership of the proposed method for p maximum 5 and 10 on generation data with variance 0.01. Table 4 shows the correct cluster membership of generation data with variance 1.

Table 3 shows that for t = 50, 100, and 300, the correct cluster membership with p maximum 10 shows better than p maximum five even though the difference is small. While for t = 150, p

maximum five is slightly above p maximum 10. Table 4 shows at t = 50, 100, and 150, the percentage of correct cluster membership for p maximum 10 lower than p maximum five. However, there is a little difference at t = 300, p maximum 10 shows higher than p maximum five.

Table 3. Percentage of the correct cluster membership of the proposed method for the AR (p) approach with p maximum five and ten on generation data σ_e^2 = 0.01.

t	p maximum 5	p maximum 10
50	84.53%	**85.37%**
100	92.83%	**93.27%**
150	**96.30%**	96.07%
300	99.20%	**99.47%**

Table 4. Percentage of the correct cluster membership of the proposed method for the AR (p) approach with p maximum five and ten on generation data σ_e^2 = 1.

t	p maximum 5	p maximum 10
50	**84.13%**	83.60%
100	**92.70%**	92.63%
150	**96.80%**	96.06%
300	99.03%	**99.43**%

We use the mean difference test at the 0.05 significance level to test the similarity of the results obtained from the two kinds of maximum p used in the proposed method. Based on the mean difference test, there was no significant difference in the AR (p) approach between p maximum five and maximum 10. In other words based on these results, there is no difference between using the p approach maximum of five or 10 on AR (p).

4. CLUSTERING RAINFALL DATA

The application of the proposed method used in this study is the intensity of monthly rainfall obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG) Province of Banten, Indonesia. The data contains the average value of rainfall intensity (mm) from 15 rainfall

stations located in Banten from 1998 to 2014. The names of the rainfall stations are (1) Baros, (2) Carenang, (3) Cinangka, (4) Ciomas, (5) Padarincang, (6) Pamarayan, (7) Regas Hilir, (8) Cibaliung, (9) Cimanuk, (10) Jiput, (11) Labuhan, (12) Menes, (13) Pandeglang, (14) Sepatan, and (15) Ciliman.

The rainfall data were clustered based on the proposed method. Figure 1 shows the results of clustering presented with a dendrogram. However, we have not had the option to decide the number of groups. The method used to determine many groups is the Silhouette index.

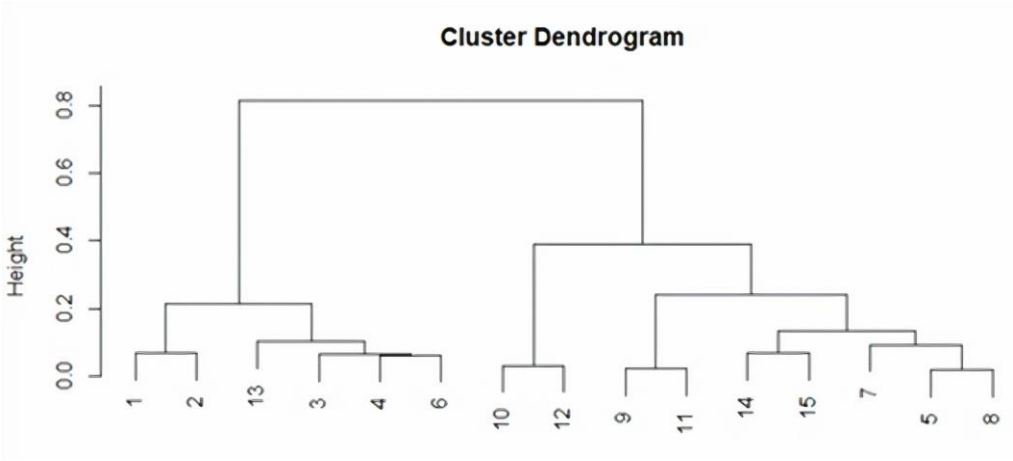


Figure 1. Cluster dendrogram of rainfall data based on the proposed method.

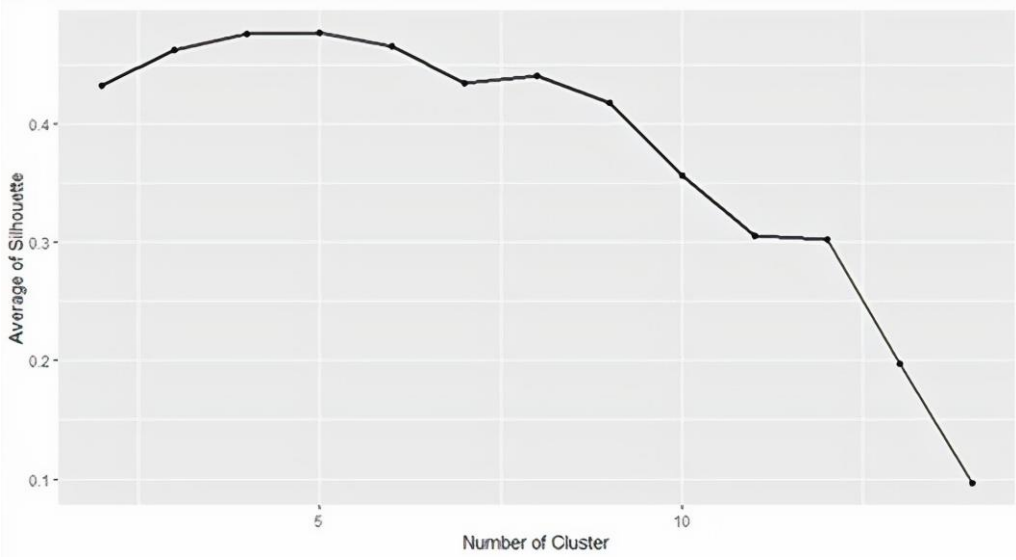
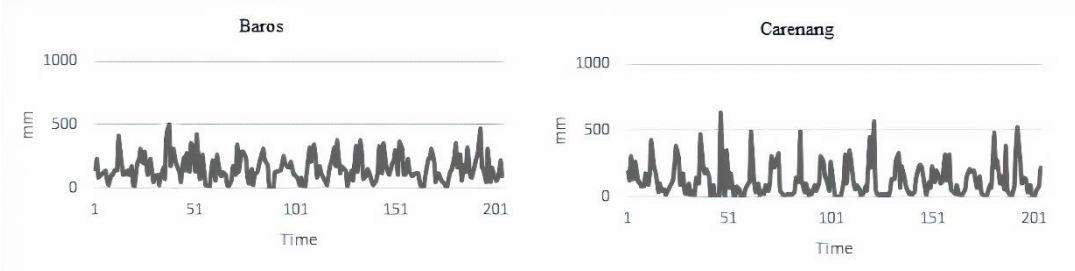


Figure 2. The plot of the average silhouette coefficient.

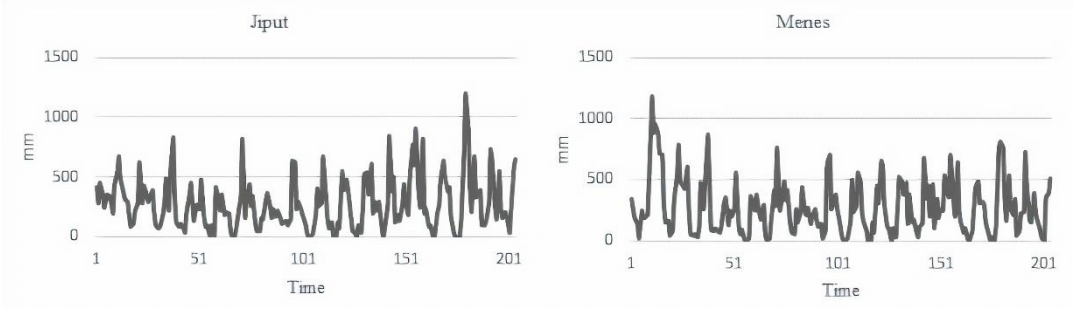
The maximum average Silhouette of the data used to determine the optimal number of groups. The average Silhouette coefficient was calculated for the clusters. The maximum value is 0.47656014, so that the optimal number of groups is five (Figure 2). The clusters formed can be seen in Table 5.

Table 5. Rainfall stations based on the cluster.

Cluster	Rainfall stations
1	Baros and Carenang
2	Cinangka, Ciomas, Pamarayan, and Pandeglang
3	Padarincang, Regas Hilir, Cibaliung, Sepatan, and Ciliman
4	Cimanuk and Labuhan
5	Jiput and Menes



a



b

Figure 3. The plot of rainfall for cluster one (a) and five (b).

Figure 3 shows the rainfalls plot members for cluster one and five. The plot of rainfall in one cluster has similar patterns. Baros and Carenang have low rainfall values, most of the values in cluster one are less than 500 mm and the maximum value is not more than 700 mm. Jiput and Menes have high rainfall values, some values are more than 500 mm, but the maximum values are not more than 1,200 mm.

5. CONCLUSIONS

This study was clustering time-series data used several criteria to identify the best model. Based on the simulation, each model selection criteria results in a different percentage of correct cluster membership. The proposed method that used five criteria is better than the primary Piccolo distance that used one criterion. The proposed method increases the correct cluster membership by more than 10% compared to the Piccolo distance method. There were not different groups between generating data with variations 0.01 and variations 1. For the proposed method, the order p approaches in AR (p) maximum of five and 10 do not give different results. We have applied the proposed method to clustering of rainfall stations in Province of Banten (Indonesia). The results obtained five groups which the rainfall in one group have the same pattern. Further research is needed for clustering using different model selection criteria.

ACKNOWLEDGEMENTS

We would like to thank Dian Kusumaningrum for first proofread.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] S. Aghabozorgi, S. Shirkhorshidi, T.Y. Wah, Time-series clustering - A decade review, Inform. Syst. 53 (2015), 16–38.

[2] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, in: E. Parzen, K. Tanabe, G. Kitagawa (Eds.), Selected Papers of Hirotugu Akaike, Springer New York, New York, NY, 1998: pp. 199–213.

[3] D.R. Anderson, K.P. Burnham, Avoiding Pitfalls When Using Information-Theoretic Methods, J. Wildlife Manage. 66 (2002), 912–918.

[4] M. J. Brewer, A. Butler, S. L. Cooksley, The relative performance of AIC, AICC, and BIC in the presence of unobserved heterogeneity, Methods Ecol. Evol. 7 (2016), 679–692.

[5] K.P. Burnham, D.R. Anderson, Multi model Inference: Understanding AIC and BIC in Model Selection, Soc. Methods Res. 33 (2004), 261–304.

[6] J. Caiado, N. Crato, D. Peña, A periodogram-based metric for time series classification, Comput. Stat. Data Anal. 50 (2006), 2668–2684.

[7] M. Charrad, N. Ghazzali, V. Boiteau, A. N. Maintainer, Determining the Best Number of Clusters in a Data Set, CRAN. (2015), 1–9.

[8] G. Claeskens, Statistical Model Choice, Ann. Rev. Stat. Appl. 3 (2016), 233–256.

[9] M. Corduas, D. Piccolo, Time series clustering and classification by the autoregressive metric, Comput. Stat. Data Anal. 52 (2008), 1860–1872.

[10] J.D. Cryer, K. Chan, Time Series Analysis with Application in R, second edition. Springer, New York, 2008.

[11] P. D'Urso, E.A. Maharaj, Autocorrelation-based fuzzy clustering of time series, Fuzzy Sets Syst. 160 (2009), 3565–3589.

[12] F. Gullo, G. Ponti, A. Tagarelli, G. Tradigo, P. Veltri, A time series approach for clustering mass spectrometry data, J. Comput. Sci. 3 (2012), 344–355.

[13] A.K. Jain, R.C. Dubes, Algorithms for Clustering Data. Prentice Hall, New Jersey, 1988.

[14] K. Kalpakis, D. Gada, V. Puttagunta, Distance measures for effective clustering of ARIMA time-series, in: Proceedings 2001 IEEE International Conference on Data Mining, IEEE Comput. Soc, San Jose, CA, USA, 2001: pp. 273–280.

[15] M. Kumar, N.R. Patel, Clustering Data with Measurement Errors, in: W. Jank, G. Shmueli (Eds.), Statistical Methods in E-Commerce Research, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008: pp. 243–267.

[16] E. A. Maharaj, Clusters of time series, J. Classif. 17 (2000), 297–314.

[17] M.P. Montero, J.A. Vilar, Package TSclust: An R package for time series clustering, J. Stat. Softw. 62 (2015), 1–43.

[18] D.C. Montgomery, C.L. Jennings, M. Kulachi, Introduction to Time Series Analysis and Forecasting. John Wiley & Sons Inc., New Jersey, 2008.

[19] D. Piccolo, A Distance Measure for Classifying Arima Models, J. Time Ser. Anal. 11 (1990), 153–164.

[20] D. Piccolo, The Autoregressive metric for comparing time series models, Statistica. 70 (2010), 459-480.

[21] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987), 53–65.

[22] G. Schwarz. Estimating the dimension of a model, Ann. Stat. 6 (1978), 461-464.

[23] Y. Shou, N. Mamoulis, D. Cheung, Fast and exact warping of time series using adaptive segmental approximations, Mach. Learn. 58 (2005), 231–267.

[24] S. Soltani, R. Modarres, and S.S. Eslamian, The use of time series modeling for the determination of rainfall climates of Iran, Int. J. Climatol. 27 (2007), 819–829.

[25] U. Triacca, Measuring the distance between sets of ARMA models, Econometrics. 4 (2016), 32.

[26] B. Zhang, B. An, Clustering time series based on dependence structure, PLoS ONE. 13 (2018), e0206753.