Evaporation Climate Zonation Based on Observation Data in Indonesia Using Cluster Analysis

I D G A Putra, R B Perdana, A Sopaheluwakan, Suryadi

1 Center for Research and Development, Indonesian Agency for Meteorological Climatological and Geophysics (BMKG), Jakarta, Indonesia
2 Center for Database, Indonesian Agency for Meteorological Climatological and Geophysics (BMKG), Jakarta, Indonesia
3 Center for Applied Climate Service, Indonesian Agency for Meteorology Climatology and Geophysics (BMKG), Jakarta, Indonesia
4 Berau Meteorological Station, Indonesian Agency for Meteorological Climatological and Geophysics (BMKG), North Kalimantan Province, Indonesia

Abstract. Identification of regional evaporation zones is very important because information on the classification of evaporation patterns in an area can be used for various purposes, one of which is knowing the amount of water in the reservoir that is lost due to evaporation. In this study, the evaporation zone in Indonesia was created using cluster analysis. The evaporation data from 127 BMKG observation stations for 6 years from 2014 to 2019 were used after the normal monthly calculation was then standardized with mean = 0 and standard deviation = 1 to ensure that all variables were given the same weight so that it became the 127 x 12 matrix used in the analysis. Hierarchical cluster analysis was chosen to regionalize evaporation. Five different techniques are applied to initially determine the most suitable method for the area. Cluster stability is also tested. It was decided that the Ward method is the most likely to produce acceptable results in this evaporation variable. 3 different classifications of evaporation zones are found in Indonesia. Evaporation zone 1 forms a seasonal cycle with one peak evaporation that occurs with the highest peak occurring in August and the minimum evaporation occurring in February in areas in the northern and central regions of Sumatra island, eastern Kalimantan, parts of Java island, North Sulawesi and West Papua. Evaporation zone 2 forms a seasonal cycle with two peaks that occur in December and March and minimum evaporation occurs in June which is in the southern region of the island of Sumatra, almost the entire island of Kalimantan, a small part of the island of Java, central Sulawesi, the archipelago. Maluku as well as the northern island of Papua. Evaporation zone 3 forms a seasonal cycle with one peak evaporation that occurs in October and the minimum evaporation occurs in June which is in a small area of the island of Central Sumatra, South Kalimantan, the west coast of Sulawesi, and all parts of the island of Nusa Tenggara.

1. Introduction
Knowledge of the characteristics of climate variables in an area based on zoning is very important to provide the information needed for various needs such as regional development planning and analysis of economic potential as well as disaster adaptation and mitigation. In 2015 Das and Ghosh zoned...
using a new approach based on the Cross Detrended Multifractal Correlation Analysis (MF-DXA) for several climate variables in the Eastern and North-Eastern region of India. In this approach, the spatiotemporal pattern of each site, as determined by multifractal correlation studies, has been exploited by K-means-based clustering techniques, which can accurately detect various climatic zones over a large area [1]. One of the techniques for creating climate zoning is using the hierarchical clustering analysis method with two grouping criteria (K-means and Ward methods) as done by Martínez and Carbajal (2017) for the monthly average maximum and minimum temperatures and accumulated monthly rainfall in Mexico [2]. Research conducted by Suryanto in 2017 for grouping 15 daily rainfall in the province of Yogyakarta concluded that the K-means method is better than the fuzzy c-means method [3]. Sutikno (2019) uses cluster analysis based on Dynamic Time Warping (DTW) distance and autocorrelation function (ACF) to determine the characteristics of the season zones in the entire Mojokerto region and map the pattern of season zones in Mojokerto based on rainfall data [4].

One of the climate variables observed at stations belonging to the Indonesian Agency for Meteorology Climatology and Geophysics (BMKG) is the variable evaporation. Analysis of evaporation in Indonesia is still quite rare. Meteorologically, the value of evaporation is related to the hydrological cycle of an area. Evaporation information can also be used to support information about the discharge analysis of a reservoir and is used for the industry supporting salt drying. The purpose of this study is to create zoning for evaporation in all regions of Indonesia based on observational data from BMKG stations so that it is hoped that the characteristics of the monthly evaporation intensity in an area can be identified as well as seasonal pattern cycles that can be used to see the economic potential of an area as well as disaster adaptation and mitigation.

2. Methodology

2.1 Data

This study uses historical data on daily evaporation variables for 7 years starting from January 2014 to December 2019 in 127 BMKG observation stations spread throughout Indonesia (details of research locations can be seen in Figure 1 above). The data on the evaporation at the stations were observed using an evaporation pan tool which was carried out simultaneously at 00.00 UTC at all stations and inputted into the megasoft software which was sent to the central storage server for the BMKG database in Jakarta. The evaporation data was downloaded in the megasoft application in CSV format.
2.2 Method
To calculate the value of the evaporation variable and visualize the results in this study as a whole using the R programming software which is available free to use. The method used in this research is to add up the daily intake to the accumulated monthly intake data for the time series period from January 2014 to 2019 in all research stations. The monthly time series data are then averaged over multi-years in the same month so that they become normal monthly evaporation data from January to December for all stations so that a 127 x 12 matrix is formed (all stations x normal monthly evaporation). Monthly normal evaporation matrix data is then standardized with mean = 0 and standard deviation = 1, to ensure that all variables have the same weight that will be used as input in the next cluster analysis [2]. Several researchers previously used cluster analysis in making climate zoning. Five commonly used hierarchical cluster clustering methods are hierarchical cluster analysis using the single linkage method, complete linkage, centroid linkage, Ward, and the average linkage [5][6][7]. The strengths and limitations of some clustering methods must be understood to make informed decisions about which method is most suitable for a particular study. Therefore, several hierarchical clustering procedures of single linkage, complete linkage, average linkage, centroid linkage, and Ward's method were carried out in this study, and the results of climate regionalization were compared to determine the most suitable clustering algorithm for the evaporation variables in the Indonesian region. The cluster description and algorithm can be seen in Table 1 below.

| Cluster methods     | Description                                                                 | Cluster algorithm                      |
|---------------------|----------------------------------------------------------------------------|----------------------------------------|
| single linkage      | This method uses the distance between the closest members of the two clusters. | \( D_{12} = \min d(x_i, y_i) \)        |
| complete linkage    | This method uses the distance between the members that is the most distant (most different) | \( D_{12} = \max d(x_i, y_i) \)        |
| average linkage     | This method involves looking at the distances between all pairs and averaging all of these distances. This is also called UPGMA- Unweighted Pair Group Mean Averaging. | \( D_{12} = \frac{1}{k l} \sum_{i=1}^{k} \sum_{j=1}^{l} d(x_i, y_i) \) |
| centroid method     | This method involves finding the mean vector location for each of the clusters and taking the distance between the two centroids. | \( D_{12} = d(\bar{x}, \bar{y}) \)     |
| ward method         | This method minimizes the total in-cluster variance. The clusters are combined which combine to produce a minimum loss of information. | \( D_{12} = \frac{2 \cdot |k| \cdot |l|}{|k| + |l|} \cdot |\bar{x} - \bar{y}|} \) |

The results of the clustering algorithm trial will separate stations based on different climatic characteristics based on zones, while stations that have almost the same characteristics will enter into one climate group. The results of evaporation zoning will be mapped spatially and temporally to determine the detailed characteristics of evaporation in Indonesian territory.
3. Result and discussion

3.1 Climatic zoning formed based on hierarchical cluster analysis

Table 2. The number of stations in each clustering method is carried out.

| Cluster Method  | Zone 1 | Zone 2 | Zone 3 |
|-----------------|--------|--------|--------|
| Single linkage  | 125    | 1      | 1      |
| Complete linkage| 45     | 58     | 24     |
| Average linkage | 80     | 43     | 4      |
| Centroid linkage| 125    | 1      | 1      |
| Ward method     | 39     | 59     | 29     |

Table 2 above is the result of making three clusters formed from 127 stations that were tested based on five methods, namely the single linkage procedure, complete linkage, average linkage, centroid linkage, and Ward's method. Based on the calculation results of the five clusters method above, it can be seen that four methods, namely single linkage, complete linkage, average linkage, and centroid linkage produce unstable station distribution, this can be seen from only 1 station which is included in zone 3 in the centroid linkage and single methods, linkage and 4 stations in zone 3 on the average method. The only method that produces a stable distribution of the number of clusters is Ward's method, this is consistent with research conducted by Unal et. al in turkey [9]. Ward's minimum variant method has become the most widely used clustering technique in climate zoning research based on its stability, so this study uses cluster results from the ward method for further interpretation [10]. Based on the results of the ward method cluster calculation for the evaporation variable, 3 evaporation zones are produced spatially and temporally, the results will be displayed based on the discussion of the next sub-chapter.
3.2 Monthly evaporation pattern based on zoning

Figure 2. The station's monthly average evaporation cycle pattern based on zoning

(A) Zone 1 monthly evaporation pattern
(B) Zone 2 monthly evaporation pattern
(C) Zone 3 monthly evaporation pattern

Figure 2.A is the normal monthly evaporation and the average of the stations that fall into zone 1 based on the calculation of the ward's method. Based on the boxplot, it can be seen that the characteristics of zone 1 evaporation pattern have a relatively low fluctuating monthly pattern within one year with a range of average monthly evaporation values ranging from 75 mm/month to 110 mm/month. Evaporation in zone 1 forms a seasonal cycle with the evaporation of one peak occurring in July, August, and September with the highest peak occurring in August with an average value reaching 110 mm/month. The minimum evaporation in zoning area 1 occurs in February with a value of about 75 mm/month. Figure 2.B is the normal monthly evaporation and the average of the stations that fall into zone 2 based on the calculation of the ward's method. Based on the boxplot, it can be seen that the characteristics of zone 2 evaporation pattern have a monthly pattern which is also relatively less fluctuating in one year with a range of average monthly evaporation values ranging from 80 mm/month to 120 mm/month but has a characteristic higher evaporation value than evaporation in zone 1. Evaporation in zone 2 forms a seasonal cycle with the evaporation of two peaks that occurred in December and March with an average value reaching 120 mm/month. The minimum evaporation in zoning area 2 occurs in June with a value of about 80 mm/month. Figure 2.C is the normal monthly evaporation and the average of the stations that fall into zone 3 based on the calculation of the ward's method. Based on the boxplot, it can be seen that the characteristics of the zone 3 evaporation pattern have a relatively fluctuating monthly pattern within one year with a range of average monthly evaporation values ranging from 80 mm/month to 170 mm/month. Evaporation in zone 3 forms a...
seasonal cycle with one peak of evaporation that occurs in August, September, October, and November. The maximum evaporation peak occurs in October with an average value of reaching 170 mm/month. The minimum evaporation in zone 3 occurs in June with a value of about 80 mm/month.

3.3 Spatial evaporation climate zone

Figure 3 shows the results of the three-zone clusters using the ward's method from 127 BMKG observation stations, the same color means that they are in one zoning and have almost similar evaporation characteristics which are plotted spatially. Evaporation zone 1 which is shown by a red box with an evaporation pattern forms a seasonal cycle with the maximum evaporation peak occurring in August and the minimum evaporation occurring in February (see figure 2.A) predominantly in the northern and central regions of Sumatra island, eastern Kalimantan, Part of the island of Java, north of the island of Sulawesi and west of the island of Papua. Evaporation zone 2 which is indicated by a yellow box with an evaporation pattern forms a seasonal cycle with two peaks that occur in December and March, while the minimum evaporation pattern occurs in June (see Figure 2.B) dominantly in the southern region of Sumatra island, almost all of the island of Borneo, a small part of the island of Java, central Sulawesi, the Maluku Islands and the north of the island of Papua. Evaporation zone 3 which is shown by a green box with an evaporation pattern forming a seasonal cycle with one peak that occurs in October, while the minimum evaporation pattern occurs in June (see Figure 2.C) is dominant in a small part of the island of Central Sumatra, South Kalimantan, the west coast of Sulawesi and all parts of the island of Nusa Tenggara. In spatial characteristics, the zoning pattern of evaporation occurs randomly in Indonesia. Although based on the hydrological cycle, evaporation is related to rainfall, but in terms of zoning, different results are obtained from rainfall zoning based on Aldrian and Susanto research in 2003 [11], so it is necessary to study further in more detail to determine the factors that influence the evaporation pattern of an area based on the influence of global phenomena, regional to local factors.
4. Conclusion
Based on the results of cluster analysis based on the ward's method for 7 years from 2014 to 2019, it resulted in 3 evaporation zones in Indonesia. Evaporation zone 1 forms a seasonal cycle with one peak evaporation that occurs with the highest peak occurring in August and the minimum evaporation occurring in February in areas in the northern and central regions of Sumatra island, eastern Kalimantan, parts of Java island, North Sulawesi and West Papua Province. Evaporation zone 2 forms a seasonal cycle with two peaks that occur in December and March and minimum evaporation occurs in June which is in the southern region of the island of Sumatra, almost the entire island of Kalimantan, a small part of the island of Java, and central Sulawesi. Maluku archipelago, as well as the northern island of Papua. Evaporation zone 3 forms a seasonal cycle with one peak evaporation that occurs in October and the minimum evaporation that occurs in June which is located in a small area of the island of Central Sumatra, South Kalimantan, the west coast of Sulawesi, and all parts of the island of Nusa Tenggara.

References
[1] Das M and Ghosh S K, 2015 Detection of climate zones using multifractal detrended cross-correlation analysis: A spatio-temporal data mining approach ICAPR 2015 - 2015 8th Int. Conf. Adv. Pattern Recognit. p. 1–6.
[2] Pineda-Martinez L F and Carbajal N, 2017 Climatic analysis linked to land vegetation cover of Mexico by applying multivariate statistical and clustering analysis Atmosfera 30, 3 p. 233–242.
[3] Suryanto J, 2017 Analisa Perbandingan Pengelompokkan Curah Hujan 15 Hari Provinsi DIY Menggunakan Fuzzy Clustering Dan K-Means Clustering J. AGRIFOR XVI p. 229–242.
[4] Ayundari I and Sutikno S, 2019 Penentuan Zona Musim di Mojokerto Menurut Karakteristik Curah Hujan Dengan Metode Time Series Based Clustering Inferensi 2, 2 p. 63.
[5] DeGaetano A T, 1996, Delineation of mesoscale climate zones in the northeastern United States using a novel approach to cluster analysis, Journal of Climate, 9, 8. p. 1765–1782.
[6] Terassi P M de B and Galvani E, 2017 Identification of homogeneous rainfall regions in the Eastern watersheds of the state of Paraná, Brazil Climate 5, 3.
[7] Juaeni I Yuliani D Ayahbi R Hardjana T and Nurzaman, 2010 Pengelompokan wilayah curah hujan Kalimantan Barat berbasis metode ward dan fuzzy clustering J. Sains Dirgant. 7, 2 p. 82–99.
[8] Islam N, 2019 Defining Homogenous Climate zones of Bangladesh using Cluster Analysis International Journal of Statistics and Mathematics 6, 1 p. 119–129.
[9] Unal Y Kindap T and Karaca M, 2003 Redefining the climate zones of Turkey using cluster analysis Int. J. Climatol. 23, 9 p. 1045–1055.
[10] Kahya E Demirel M C and Bég O A, 2008 Hydrologic homogeneous regions using monthly streamflow in Turkey Earth Sci. Res. J. 12, 2 p. 181–193.
[11] Aldrian E and Dwi Susanto R, 2003 Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature Int. J. Climatol. 23, 12 p. 1435–1452.