Regular Research Article

Correlation of Climate Variability and Burned Area in Borneo using Clustering Methods

Ishardina C. Hidayati[1], Novinda Nalaratih[1], Ayu Shabrina[2,*], Intan N. Wahyuni[2], Arnida L. Latifah[2]

[1] Matematika, Universitas Airlangga
[2] Research Center for Informatics, Indonesian Institute of Sciences
* Correspondence author: ayu.shabrina@lipi.go.id

Abstract: The island of Borneo has faced seasonal forest fires for decades. This phenomenon is worsening during dry seasons, especially when droughts are concurrent with the El Niño-Southern Oscillation (ENSO) phenomenon. Climate is therefore one of the drivers of the fire phenomenon. This paper studies the relationship between climate variables, namely temperature, precipitation, relative humidity, and wind speed, and the occurrence of forest fire using two clustering methods, K-means and Fuzzy C-means (FCM) clustering methods. Borneo is clustered into four areas based on burned area data obtained from Global Fire Emission Data (GFED). It is also clustered according to the combinations of climate variables. Both methods reach the highest correlation between the climate variable and the burned area clusters in September. The K-means method gives a correlation of -0.54 while the FCM gives -0.55. In August until October, relative humidity provides the dominant correlation affecting burned area, even though an additional precipitation or wind variable slightly increases the correlation in the FCM method. In November, temperature largely contributed to the burned area by a positive correlation of 0.31 in K-means and 0.33 in FCM. The evaluation performance of the methods is conducted by an internal validation called the Silhouette index. Both methods have positive index values ranging from 0.39 to 0.69 and the maximum value is influenced by the wind cluster. This indicates that the clustering methods applied in this paper can identify one or a combination of climate variables into dense and well-separated clusters.

Keywords: Forest fire; fuzzy c-means; k-means; clustering; climate; Borneo

---

1. Introduction

Forest and land fires have become a concern of the global community, especially since the El Niño Southern Oscillation (ENSO) in 1982-1983, 1994, and 1997-1998 also resulted in high incidence and intensity of fires (Kita, Fujiwara, & Kawakami, 2000). The El Niño phenomenon causes smaller amounts of rainfall throughout the year, resulting in a longer dry season (Fadholi, 2013). Other climatic factors such as temperature, humidity, and wind also influence forest fires (Aryadi, Satriadi, & Syam'ani, 2018). Forest and land fires cause a decrease in the quality of natural ecosystems, such as forest and vegetation damage (Margono et al., 2012). Uncontrolled fires are a major factor contributing to the loss of forests and land in the tropics (Siegert & Hoffmann, 2000). Forest fires are also considered a threat to sustainable development because of their contribution to the increase of carbon emissions (Tacconi, Moore, & Kaimowitz, 2007).

Barber (2000) stated that forest fires have occurred since the 17th century in Borneo. It is caused by many factors, such as climate, environment, and also social-economic factors (Barber & Schweithelm, 2000). When a drought caused by the El Niño weather phenomenon occurred like in 2015 and 2019, the forest fires are pronounced. This suggests that climate plays an important role in the fires in Borneo. Based on several previous studies (Aflahah, Hidayati, Hidayat, & Alfahmi, 2018; Latifah, Shabrina, Wahyuni, & Sadikin, 2019; Mareta, Hidayat, Hidayati, & Latifah, 2019), there are several climate variables affecting fires in Borneo. Based on climatic information and burned areas in the past, precise information can be generated to help prevent and reduce the risk of fires in the future.

Clustering methods, such as K-means and Fuzzy C-means (FCM) methods have been widely used and established to cluster forest fires to analyze fire risk. Previous studies (Pramesti, Lahan, Tanzil Furqon, & Dewi, 2017; Shoolihah, Furqon, & Widodo, 2017; Sukamto, Id, & Angraini, 2018) implemented methods to cluster the fires phenomenon over Indonesia based on the MODIS

Forest and Society. Vol. 4(2): 280-293, November 2020

281

satellite-based fire hotspot data. Their studies highlight that the K-means clustering method can analyze fire data and capture potential fires at a high, medium or weak level. Another study (Wang, Liu, & Cui, 2017) showed that the K-means clustering method can quickly and accurately map the satellite images of fires and recognize images of burning land in experiments with large amounts of data. Ganesan (2016) compared the K-means and FCM clustering methods in detecting fire zones in Oregon, United States of America. His study concluded that the combination of the fuzzy means clustering method with the genetic algorithm as a segmentation method was more effective than K-means clustering methods for choosing the best pixel photos. Furthermore, the K-means clustering method was also applied to support the identification process of forest fires automatically using spatial digital photos (Prasad & Ramakrishna, 2008). Mustofa (2014) predicted the size of forest fires with climate variables and the forest fire weather index using the FCM clustering method that can successfully classify fire levels into three categories namely low, light and heavily burned area (Shidik & Mustofa, 2014).

Jafarzadeh (2017) studied the relationship between climate and geographical variables and the fire events in Ilam Province, West of Iran. FCM was used to map all of the variables into 3 clusters, namely low, medium and high. Their study concluded that there was a strong relationship between the forest fire clusters and the cluster of 8 input variables, including: distance from the settlements, population density, distance from a road, slope, standing dead oak trees, temperature, land cover and distance from farmland (Jafarzadeh, Mahdavi, & Jafarzadeh, 2017). Differing from the aforementioned studies, in this paper we apply global satellite data that contains dense spatial information about climate and burned areas to cluster areas over Borneo. This is also motivated by the limited available local data in Borneo.

In this paper, we focus on studying only the climate variables, namely temperature, precipitation, relative humidity, and wind speed. This paper clusters Borneo based on the climate characteristics then relates the climate variables to the occurrence of the forest fires that are obtained from the Global Fire Emissions Database (GFED). This study aims to investigate the correlation between the cluster of forest fires and the cluster of climate variables over Borneo Island. Mapping in Borneo can provide information to predict and to prevent the damage of forests and land in the region.

2. Materials and Methods

2.1 Data and sources

This paper takes the whole Borneo Island as shown in Figure 1 as the study case without being limited to the administrative areas of the state, province, or district. Geographically, Borneo is directly bordered by the South China Sea to the north, the Sulawesi Sea to the east, the Java Sea to the south, and the Karimata Strait to the west, and is on the equator. It is the third largest island in the world and has diverse climate variations. Most of the land in Borneo is covered by primary and secondary forest that is highly vulnerable to fires. In the simulation, the area is gridded into 40x44 horizontal grids with a size of 0.25 degrees. By neglecting the ocean, the gridded area becomes only 1058 grids.

Forest and Society. Vol. 4(2): 280-293, November 2020

282
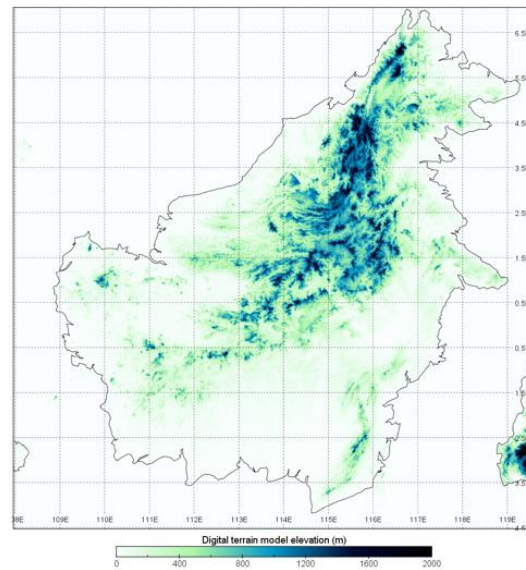
Figure 1. Borneo map based on topographic data in meters

This study uses monthly burned area and climate data in the period of 1 January 1998 − 1 December 2015. The data of the burned area is obtained from the Global Fire Emission Data (GFED) that shows data of globally mapped burned area data available since 1997 (Giglio, Randerson, & Van Der Werf, 2013). It combines the satellite information on fire activity and vegetation productivity to estimate gridded monthly burned area and fire emissions. The climate data consisting of the monthly data of precipitation (PCP), air temperature (TEMP), relative humidity (RH) and wind speed (WIND) are obtained from the reanalysis data of the ERA-Interim. Since the forest fires in Borneo occur seasonally starting in August and ending approximately in the beginning of December, the monthly burned area and the climate data for August until November in 1998-2015 are analyzed.

2.2 Clustering methods

2.2.1 K-means method

K-means method is a grouping algorithm. It is one of the most popular and oldest among practitioners because of the ease of implementation and the speed of processing (Xu, Chiang, Liu, & Tan, 2017). This learning algorithm is included in unsupervised learning methods (Suyanto, 2018). Unsupervised learning techniques can organize data even if they do not have class labels. The K-means grouping algorithm classifies the points given into k groups so that the distance between members of the same group is minimized (Natingga, 2017). The k-means clustering algorithm determines the initial k-centroid (the points are in the middle of the cluster) - one for each cluster. Then each feature is classified into clusters whose centroids are closest to that feature. After classifying all the features, we formed the initial k cluster. For each cluster, we recalculate the centroid to be the average of points in that cluster. After we moved the centroid, we recalculated the class again. Features can change classes. Then we recalculated the centroid again. If the centroid does not move anymore, the k-means clustering algorithm ends.

The shortest distance from one data to all centroids is calculated to group the data. The formula for calculating distances is shown in Equation (1).

$$d(x_i, x_j) = (|x_{i1} - x_{j1}|^g + |x_{i2} - x_{j2}|^g + \ldots + |x_{ip} - x_{jp}|^g)^{1/g} \tag{1}$$

Forest and Society. Vol. 4(2): 280-293, November 2020

283

where g is 1 to calculate Manhattan distance, g is 2 to calculate Euclidean distance, and g is ∞ to calculate Chebychev distance. x_i and x_j are two data that will be calculated the distance and p is the dimension of data (Maimon & Rokach, 2010).

An object is classified into a k cluster such that the Sum of Square Error (SSE) of the object to the k centroid is minimized. SSE is the total of the squared distances between all dimensions of each cluster member with the corresponding cluster centroid. Then, updating a centroid can be calculated from the average of the data objects for each dimension with the following formula

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \tag{2}$$

where μ_k is the centroid point of the k cluster, N_k is the amount of data in the k cluster, and x_q is q data in the k cluster.

2.2.2 Fuzzy c-means method

Fuzzy C-means (FCM) clustering method was introduced by Dunn (1973) and developed by Bezdek (1981) (Bezdek, 1981; Dunn, 1973). FCM method is the extent of K-means or hard c-mean clustering method. FCM is an unsupervised clustering algorithm used to analyze the distance between data points (Ghosh & Kumar, 2013). FCM has been widely applied to problems related to feature analysis and clustering in various fields such as agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, shape analysis, target recognition (Yong, Chongxun, & Pan, 2004).

In contrast to the k-means method, the data set is grouped into c clusters so that every data point is related to every cluster with a different degree of belonging. This degree indicates the level of data points that each cluster has. The data points that lie far from the center of a cluster have a low degree of belonging. This method creates an optimal partition by minimizing the objective function.

$$J_{FCM} = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q d^2(x_k, v_i) \tag{3}$$

with $u_{ik} \in [0,1]$, $\sum_{i=1}^{c} u_{ik} = 1 \forall k$, $0 < \sum_{k=1}^{N} u_{ik} < N \ \forall i$, where X is the data set p- dimension. n is the number of data points, c is the number of clusters with $2 \leq c < n$, u_{ik} is the degree of belonging of x_k in i-cluster , q is a weighting exponent on each fuzzy membership to determine the amount of fuzziness of the clustering. v_i is the center of cluster i and $d^2(x_k, v_i)$ is the distance between a data point and the cluster center. The solution of objective function J_{FCM} can be obtained by performing an iteration process with an algorithm described in (Bezdek, Ehrlich, & Full, 1984).

The initial step is to choose the number of clusters c and the weight exponent q with $2 \leq c < n$, then initialize the partition matrix U^0. Set the counter loop for each step as $r = 0,1,2, \dots$
1. Calculate the vector of center clusters

$$v_{ij} = \frac{\sum_{k=1}^{n} (u_{ik})^q x_{kj}}{\sum_{k=1}^{n} (u_{ij})^q} \tag{4}$$

2. Calculate the distance matrix $D_{[c,n]}$

$$D_{ij} = \left(\sum_{j=1}^{q} (x_{kj} - v_{ij})^{1/2} \right) \tag{5}$$

3. Update the partition matrix for r^{th} step, U^r as

$$u_{ij} = \left(\frac{1}{\sum_{j=1}^{c} \left(d_{ik}^r / d_{jk}^r \right)^{\frac{2}{m-1}}} \right) \tag{6}$$

If $\|U^{k+1} - U^k\| < \delta$ then the process stops otherwise it returns to step 2 by updating the cluster centers and the degree of belonging iteratively (Rao & Vidyavathi, 2010).

2.3 Processing and clustering data

This study uses the climate data type netCDF and the burned area data type HDF5. Climate data includes monthly data for precipitation, temperature, and wind speed. Each climate data is a dimensionless array with a size of 40 x 44 x 18 for each month. This data consists of 40 latitude points, 44 longitude points, and a period of 18 years from 1998 to 2015. Furthermore, the GFED burned area data for each month is adjusted based on the latitude and longitude of the island of Borneo and averaged monthly from 1998-2015. To avoid unnecessary computation in the ocean area, we define the climate data as NaN value for the point at the ocean. All processed data are then clustered using the K-Means and the FCM algorithms using the clustering toolbox in Matlab software.

2.4 Silhouette Index

In this paper, we validate clustering results using the Silhouette index proposed by Rousseeuw (1987), an internal validation or accuracy assessment for clustering methods. The Silhouette is a widely used index for assessing the fit of individual objects in the classification, as well as the quality of clusters and it combines two clustering criteria, compactness and separation (Lengyel & Botta-Dukát, 2019). Therefore, it measures how well the data has been clustered. The index value ranges from -1 to +1. The higher the Silhouette index value, the better the clustered data. If the value is very low or negative, the configuration of the clustering is less appropriate, and it could be that too many or too few numbers form the clusters.

3. Results and Discussions

3.1 Cluster of burned area

Using the K-means and FCM clustering methods, the burned area data in August until November in the period 1998-2015 are grouped into four levels. This is based on the Regulation of the Minister of Environment and Forestry Republic of Indonesia about the technical criteria for status of alert and emergency forest and land fire. There are four status types, namely normal, standby, emergency alert, and emergency (Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia, 2018). The upper plot in Figure 2 shows the K-means result, while the lower plot shows the FCM result. The cluster-level in Figure 2 is shown in blue, light blue, yellow, and red. Both methods show similar results, that the area is dominated by cluster 1 (blue) that is centered at almost zero value and with few areas highly burned (red). Cluster 1 represents no fire. Small fires are grouped in cluster 2 (light blue) which mostly occur in the southern, eastern and central part of Borneo. Larger fires shown in clusters 3 (yellow) mostly appear in September and October. The cluster 3 represents the area where approximately 1% area is burning. The largest fire shown in cluster 4 occurs only over a few grid areas in the southern part of Borneo, especially in September.

Forest and Society. Vol. 4(2): 280-293, November 2020

285

In the months of fire susceptibility, it can be concluded that the clusters with small to large burned areas are located mostly in Central, South and East of Borneo. The results of the cluster process using burned area data from the satellite have similarities with the results of previous studies using hotspots as a representation of forest fire data. Sari (2020) concluded that Berau, East Kutai and Kutai Kartanegara are the sites with the highest hotspots in Eastern Borneo classified by Nearest Neighbor Analysis (Sari, Rachmita, & Manessa, 2020). This result is also shown by the light blue area (cluster 2) in Eastern Borneo in Figure 2. The data from the Ministry of Environment and Forestry in the Forestry and Land Monitoring System shows that Central Kalimantan or Southern Borneo is one of the provinces in Indonesia with a high level of forest fires. In addition, Pratamasari (2020) shows that the highest hotspot area in Central Kalimantan or Southern Borneo is located in East Kotawaringin, Palangkaraya City, and Pulang Pisau based on Kernel Density analysis (Pratamasari, Permatasari, Pramudiyasari, Manessa, & Supriatna, 2020). This corresponds to the red area or cluster 4 in September and November in Figure 2.
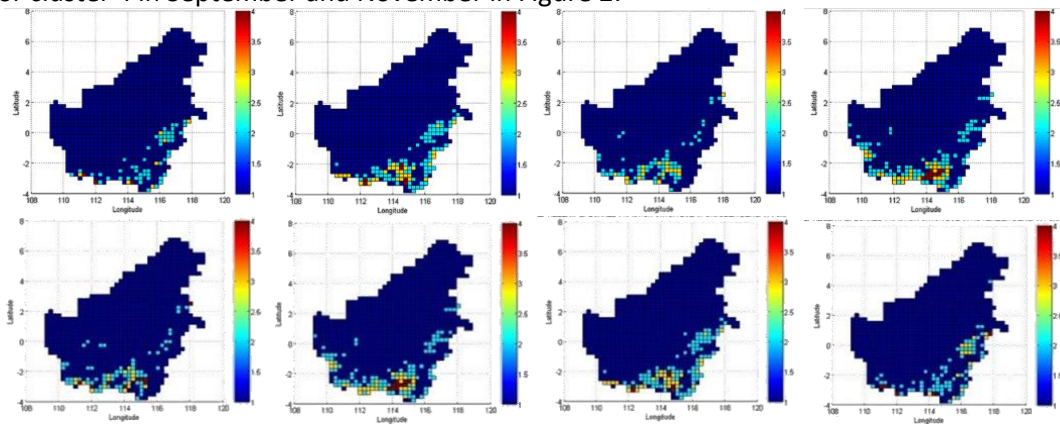
Figure 2. Cluster of the burned area in August to November (from left to right) computed by K-means (upper) and by FCM method (lower). The colors represent cluster divisions: cluster 1 (blue), 2 (light blue), 3 (yellow), and 4 (red)

Table 1. Center value of burned area cluster using K-Means and FCM

Month	K-Means				FCM			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 1	Cluster 2	Cluster 3	Cluster 4
August	0.0010	0.0223	0.0670	0.3064	0.0008	0. 0188	0.0522	0.0901
September	0.0018	0.0344	0.1215	0.2592	0.0015	0.0334	0.1235	0.2625
October	0.0006	0.0145	0.0386	0.0906	0.0004	0.0135	0.0381	0.0945
November	0.0001	0.0021	0.0075	0.0145	0.0000	0.0013	0.0042	0.0128

3.2 Cluster of climate variables and the relationship with burned area cluster

We cluster Borneo into four areas based on single or multiple climate variables using both K-means and FCM clustering methods. The clustering results are then correlated to the burned area cluster. Figure 3(a) shows the correlation between the burned area cluster with the climate variable cluster computed by the K-means clustering method. In this method, the variable of relative humidity gives the largest correlation and it has a negative correlation with the burned area. This agrees with the condition that with a drier climate, larger fires will occur. The relative humidity has the highest correlation value of 37.86% in August, 54.8% in September, and 54.28% in October. Using the FCM method, an additional precipitation variable to the relative humidity gives the highest correlation; 38.52% in August, 55.52% in September. In October, additional wind speed variables to the relative humidity give the highest correlation at 50.89%. In November, both methods reach the

Forest and Society. Vol. 4(2): 280-293, November 2020

286

highest correlation from a single climate variable, namely temperature. The temperature has a positive correlation with the burned area.

The discussion above shows that relative humidity is a very influential variable on the occurrence of forest fires based on the K-means method, whereas with the FCM method, by adding precipitation variables, the correlation value with the burned area variable is better. The results of this study are in line with previous studies. In Wu (2018), it is concluded that wind speed and relative humidity are the dominant weather parameters in measuring the severity of burns. The size of the burned area is negatively correlated with relative humidity; complexity and aggregation in the burned area are positively correlated with wind speed in northeast China's boreal forests (Wu, He, Fang, Liang, & Parsons, 2018). Krystyna (2019) conducted a study that focused on the information of low humidity in the selected area to determine the risk of fire in the forests of Central Poland. The study shows that the lack of relative humidity is a conducive condition for forest fires. Low air humidity also has a large correlation with forest fires in April - August (Konca-Kędzierska & Pianko-Kluczyńska, 2019).
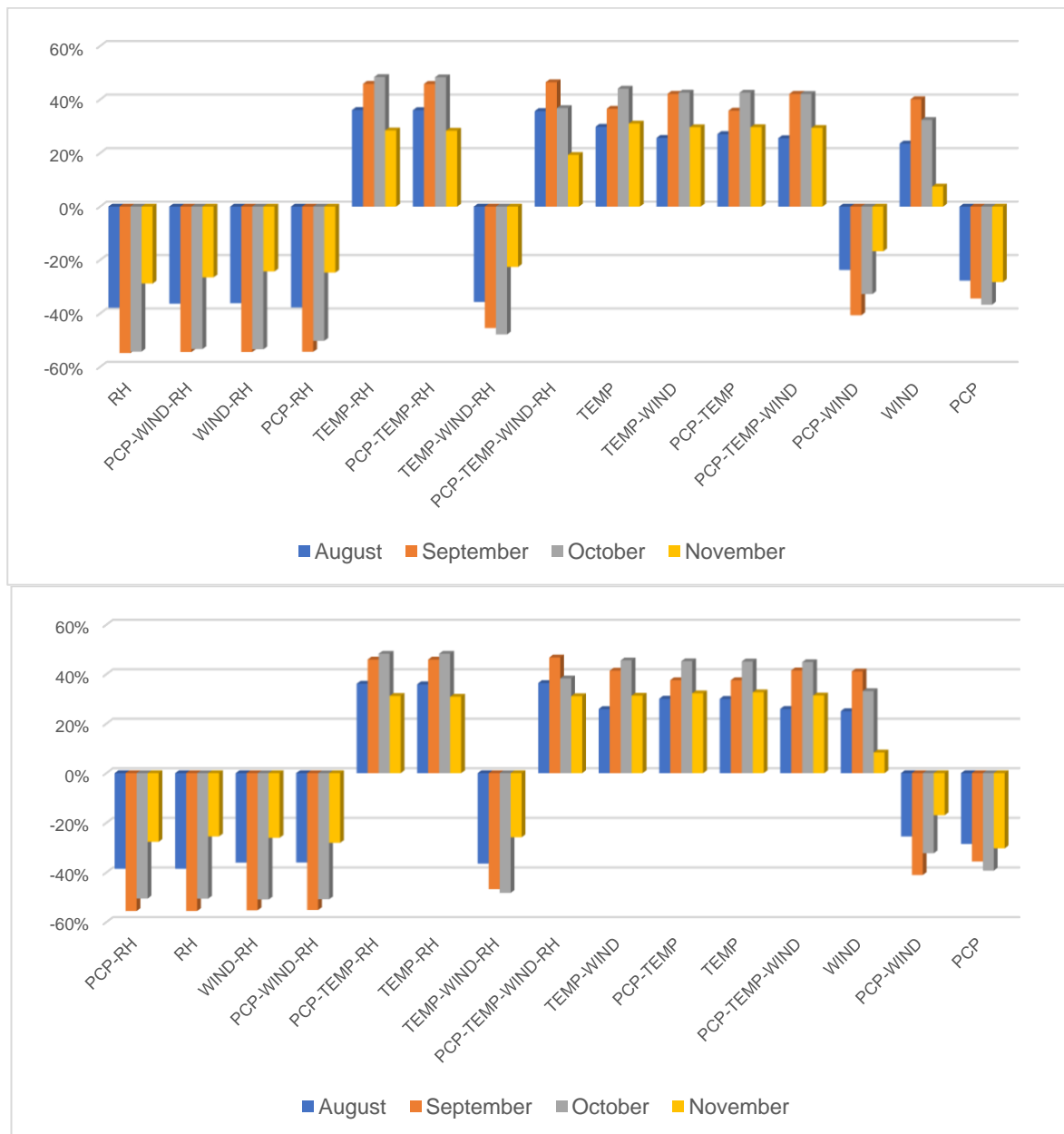
Table 2. Correlation between the burned area cluster and the climate variables cluster using K-Means and FCM

Variable	K-Means				Variable	FCM			
	Aug	Sept	Oct	Nov		Aug	Sept	Oct	Nov
RH	-37.86%	-54.80%	-54.28%	-28.79%	PCP-RH	-38.52%	-55.52%	-50.52%	-27.66%
PCP-WIND-RH	-36.33%	-54.42%	-53.33%	-26.47%	RH	-38.52%	-55.52%	-50.53%	-25.48%
WIND-RH	-36.17%	-54.41%	-53.40%	-24.29%	WIND-RH	-36.03%	-55.22%	-50.89%	-26.05%
PCP-RH	-37.72%	-54.35%	-50.24%	-24.72%	PCP-WIND-RH	-35.98%	-55.11%	-50.79%	-28.06%
TEMP-RH	36.10%	45.89%	48.41%	28.47%	PCP-TEMP-RH	36.15%	45.92%	48.27%	31.21%
PCP-TEMP-RH	36.05%	45.84%	48.31%	28.37%	TEMP-RH	35.91%	45.92%	48.27%	30.89%
TEMP-WIND-RH	-35.69%	-45.48%	-47.83%	-22.60%	TEMP-WIND-RH	-36.45%	-46.71%	-48.23%	-25.82%
PCP-TEMP-WIND-RH	35.71%	46.48%	36.82%	19.31%	PCP-TEMP-WIND-RH	36.42%	46.75%	38.28%	31.11%
TEMP	29.84%	36.57%	44.08%	31.05%	TEMP-WIND	25.92%	41.45%	45.59%	31.33%
TEMP-WIND	25.68%	42.14%	42.60%	29.66%	PCP-TEMP	30.10%	37.57%	45.24%	32.27%
PCP-TEMP	27.10%	35.87%	42.58%	29.68%	TEMP	30.03%	37.55%	45.14%	32.67%
PCP-TEMP-WIND	25.56%	42.14%	42.10%	29.42%	PCP-TEMP-WIND	25.95%	41.53%	44.89%	31.41%
PCP-WIND	-23.74%	-40.67%	-32.73%	-16.74%	WIND	25.04%	41.11%	33.16%	8.34%
WIND	23.55%	40.10%	32.36%	7.51%	PCP-WIND	-25.48%	-41.03%	-32.26%	-16.92%
PCP	-27.70%	-34.37%	-36.70%	-28.25%	PCP	-28.51%	-35.55%	-39.25%	-30.29%

The result in section 3.1 shows that large scale fires in cluster 3 and 4 occurred in several areas in central Kalimantan, especially in September. According to data from the Ministry of Environment and Forestry, the area in central Kalimantan or southern Borneo is covered by 80% of forest and 20% of peatland areas. When the amount of rainfall is lower than usual, climate conditions become drier, making peatlands drier, therefore maintaining conditions where ignition can easily occur and make them prone to fire (Erianto I Putra, Vetrita, & Graham, 2016; Erianto Indra Putra & Hayasaka, 2011). A low level of relative humidity combined with a reduced amount of rainfall makes the peatland area flammable. Usup et al. (2004) found that during the dry season, degraded peatlands

experienced a decrease in surface water, making them more flammable. Shrubs in degraded areas can easily dry out and become flammable materials. The presence of fires combined with dry conditions, especially in combustible materials can cause forest fires difficult to extinguish and can last longer (Usup, Hashimoto, Takahashi, & Hayasaka, 2004).

Syaufina & Sukmana (2008) stated that precipitation is one of the climate elements that has a high correlation with the occurrence of forest fires. This contradicts our finding that a single variable of precipitation (PCP) has a relatively low correlation compared to the other variables in both methods. Nevertheless, this paper agrees closely with the results of previous studies. Aflahah et al. (2018) and Pratamasari et al. (2020) show a low correlation between precipitation and hotspot data. When the precipitation level is low, the visibility is low, the temperature is high, the number of hotspots increases so that the potential for forest and land fires is greater (Erianto Indra Putra, Hayasaka, Takahashi, & Usup, 2008).


Figure 3. Correlation between the burned area cluster and the climate variables cluster using (a) K-means (top) and (b) FCM (bottom) method

Forest and Society. Vol. 4(2): 280-293, November 2020

288

Table 3. Center value of the precipitation and relative humidity cluster using FCM and the relative humidity cluster using K-Means

Cluster	Var	K-Means				Var	FCM			
		Aug	Sep	Oct	Nov		Aug	Sep	Oct	Nov
1	RH	92.5789	90.1568	89.9090	93.3438	PCP	0.0432	0.0091	0.0616	0.2609
						RH	92.6006	90.0252	89.7568	93.2711
2	RH	94.5659	92.6536	93.6460	95.5099	PCP	0.0797	0.0277	0.0976	0.3580
						RH	94.5817	92.5636	93.7054	95.5189
3	RH	95.9963	95.0200	95.8775	97.0093	PCP	0.1789	0.0892	0.1976	0.4860
						RH	96.0128	95.0216	96.0063	97.0878
4	RH	97.1213	96.7641	96.8671	97.6525	PCP	0.2766	0.1933	0.2790	0.5903
						RH	97.1555	96.7938	96.9359	97.6655
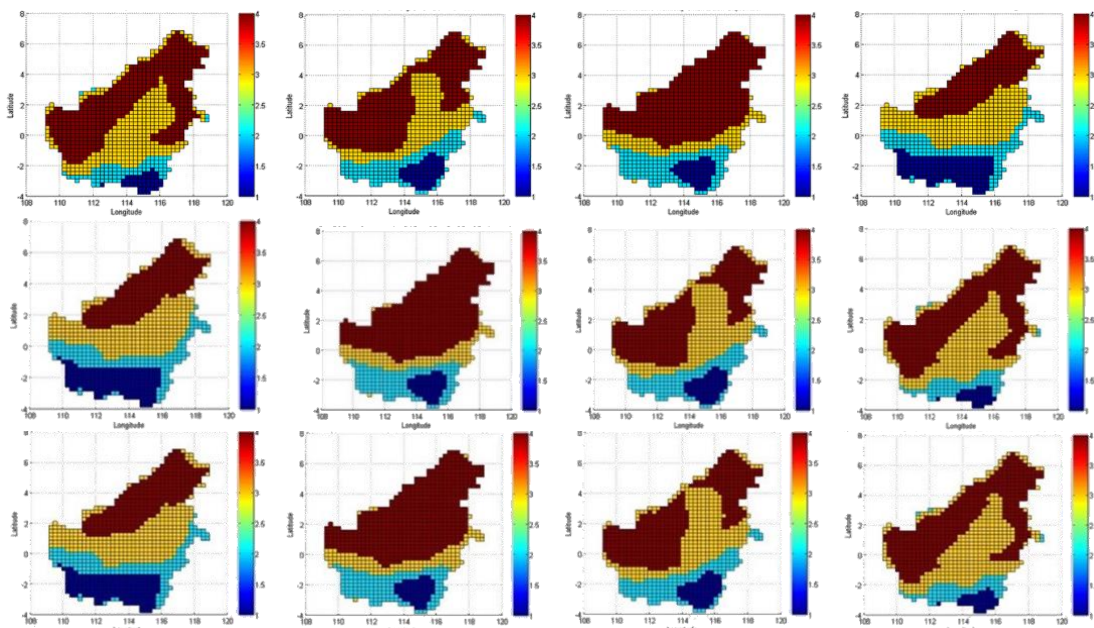
Figure 4. Cluster of relative humidity using the K-means method, relative humidity, and precipitation - relative humidity using the FCM method (top to bottom) in August to November (from left to right). The colors represent cluster divisions: cluster 1 (blue), 2 (light blue), 3 (yellow), and 4 (red)

The upper plots of Figure 4 show the cluster of the relative humidity in August-November using the K-means clustering method, while the lower plots show the cluster of the relative humidity and precipitation using the FCM method. In each month, both methods give similar cluster areas. The cluster 1 to 4 sequentially located from Southern to Northern Borneo. Compared with Figure 4, the distribution of these cluster locations has a slightly opposite pattern with the burned area cluster. The largest fire illustrated by cluster 4 (red) in Figure 4 is correlated with cluster 1 (blue) in Figure 4. Cluster 1 in Figure 4 is centered in (92.6006; 90.0252; 89.7568; 93.2711) for relative humidity cluster and (0.0432; 0.0091; 0.0616; 0.2609) for precipitation cluster using the FCM method, and in (92.5789; 90.1568; 89.9090; 93.3438) for relative humidity cluster using the K-means method. This is the lowest level of the climate variables in the selected months. This can be interpreted that the

low level of precipitation and relative humidity is correlated with the vast fire area. It means the low level of precipitation and relative humidity can increase the risk of forest fires.

3.3 Evaluation of the clustering method

Table 4. Validation measurement using Silhouette Index for K-Means and FCM

Variable	K-Means				Variable	FCM			
	Aug	Sept	Oct	Nov		Aug	Sept	Oct	Nov
WIND	0.6359	0.6392	0.6663	0.6886	WIND	0.6372	0.6356	0.6634	0.6878
PCP-WIND	0.6169	0.6171	0.6209	0.6583	PCP-WIND	0.6183	0.6124	0.6187	0.6568
TEMP	0.5880	0.5745	0.6128	0.5967	RH	0.5454	0.6023	0.5580	0.5902
RH	0.5460	0.6033	0.5630	0.5911	TEMP	0.5874	0.5725	0.6116	0.5881
PCP-TEMP	0.5704	0.5513	0.5843	0.5779	PCP-TEMP	0.5689	0.5495	0.5833	0.5664
PCP	0.5662	0.5464	0.5435	0.5473	PCP	0.5659	0.5421	0.5424	0.5463
TEMP-RH	0.4290	0.4710	0.5249	0.5538	TEMP-RH	0.4223	0.4620	0.5253	0.5534
PCP-RH	0.5064	0.5794	0.5104	0.5510	PCP-TEMP-RH	0.4195	0.4560	0.5151	0.5455
PCP-TEMP-RH	0.4258	0.4661	0.5146	0.5458	PCP-RH	0.5061	0.5744	0.5059	0.5445
TEMP-WiND	0.4919	0.5063	0.5311	0.5254	TEMP-WIND	0.4908	0.5065	0.5309	0.5254
PCP-TEMP-WIND	0.4872	0.5001	0.5225	0.5184	PCP-TEMP-WIND	0.4864	0.5004	0.5220	0.5185
WIND-RH	0.4579	0.4886	0.5124	0.5154	WIND-RH	0.4495	0.4781	0.5112	0.5140
PCP-WIND-RH	0.4523	0.4829	0.5017	0.5037	PCP-WIND-RH	0.4439	0.4722	0.5004	0.5017
TEMP-WIND-RH	0.4013	0.4127	0.4636	0.4825	TEMP-WIND-RH	0.3987	0.4082	0.4516	0.4789
PCP-TEMP-WIND-RH	0.3992	0.4102	0.4587	0.4796	PCP-TEMP-WIND-RH	0.3965	0.4056	0.4465	0.4746

For data that does not have the ground truth labels as a reference, the Silhouette index can be used to evaluate the model. In table 4, the results for the K-means and FCM methods have values ranging from 0.39 to 0.69 for the four months specified. This shows that all climate variations cluster models have good dense and well separated clusters. In the K-means and FCM methods, a single wind variable is the best-defined clusters with a Silhouette index above 0.6 in August until November. While the climate variables with the best correlation to burned area (see the grey shaded in Table 4), namely RH for the K-means method and PCP-RH for the FCM method has a Silhouette value between 0.5 - 0.6. This value indicates that the two methods with the variables that have the best correlation also have a good evaluation value, although not the best.

4. Conclusions

This paper studied the correlation between the occurrence of forest fires and climate variables over Borneo using clustering methods, namely K-means and FCM clustering methods. A single or multiple combination of the climate variables are grouped into four cluster areas as well as the burned area. The K-means and FCM clustering methods show that the relative humidity contributes most to the burned area. In the K-mean clustering method, a single climate variable (relative

humidity) has shown a strong negative correlation with the burned area in most of the dry season. In FCM clustering methods, the combination of the relative humidity and precipitation gives the largest negative correlation to the burned area. Different from other months, both methods show that the temperature is more likely to affect fires in November. The clustering results are evaluated using the Silhouette index. Based on the Silhouette index value, all the clustered climate variables are in a good position, dense and well separated.

Acknowledgments

Author's Contributions

All authors contributed equally to this work.

Conflicts of interest

Author declares no conflict of interest.

References

Aflahah, E., Hidayati, R., Hidayat, R., & Alfahmi, F. (2018). Pendugaan hotspot sebagai indikator kebakaran hutan di Kalimantan berdasarkan faktor iklim Hotspot asummption as a forest fire indicator in Kalimantan based on climate factor. *Journal of Natural Resources and Environmental Management*, *9*(2), 405–418. doi: https://doi.org/https://doi.org/10.29244/jpsl.9.2.405-418

Aryadi, M., Satriadi, T., & Syam'ani, S. (2018). KECENDERUNGAN KEBAKARAN HUTAN DAN LAHAN DAN ALTERNATIF PENGENDALIAN BERBASIS KEMITRAAN DI PT. INHUTANI II KOTABARU Trend of Forest and Land Fire and Control Alternative Based on Partnership at PT. Inhutani II Kotabaru. *Jurnal Hutan Tropis*, *5*(3), 222. doi: https://doi.org/10.20527/jht.v5i3.4789

Barber, C. V., & Schweithelm, J. (2000). *Trial by fire: Forest fires and forestry policy in Indonesia's era of crisis and reform. Washington D.C, USA, World Resources Institute (WRI), Forest Frontiers Initiative. In collaboration with WWF-Indonesia and Telapak Indonesia Foundation.*

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. doi: https://doi.org/10.1007/978-1-4757-0450-1

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, *10*(2–3), 191–203. doi: https://doi.org/10.1016/0098-3004(84)90020-7

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, *3*(3), 32–57. doi: https://doi.org/10.1080/01969727308546046

Fadholi, A. (2013). Studi Dampak El Nino Dan Indian Ocean Dipole (Iod) Terhadap Curah Hujan Di Pangkalpinang. *Jurnal Ilmu Lingkungan*, *12*(2), 43. doi:https://doi.org/10.14710/jil.11.1.43-50

Ganesan, P., Sathish, B. S., & Sajiv, G. (2016). A comparative approach of identification and segmentation of forest fire region in high resolution satellite images. *IEEE WCTFTR 2016 - Proceedings of 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare*, 3–8. doi: https://doi.org/10.1109/STARTUP.2016.7583959

Ghosh, S., & Kumar, S. (2013). Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. *International Journal of Advanced Computer Science and Applications*, *4*(4), 35–39. doi: https://doi.org/10.14569/ijacsa.2013.040406

Giglio, L., Randerson, J. T., & Van Der Werf, G. R. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *Journal of*

Forest and Society. Vol. 4(2): 280-293, November 2020

291

Geophysical Research: Biogeosciences, *118*(1), 317–328. doi: https://doi.org/10.1002/jgrg.20042

Jafarzadeh, A. A., Mahdavi, A., & Jafarzadeh, H. (2017). Evaluation of forest fire risk using the Apriori algorithm and fuzzy c-means clustering. *Journal of Forest Science*, *63*(8), 370–380. doi: https://doi.org/10.17221/7/2017-JFS

Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia. (2018). Berita Negara. In *Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor P.9/MENLHK/SETJEN/KUM.1/3/2018 Tentang Kriteria Teknis Status Kesiagaan dan Darurat Kebakaran Hutan dan Lahan*. Retrieved from http://ditjenpp.kemenkumham.go.id/arsip/bn/2018/bn375-2018.pdf

Kita, K., Fujiwara, M., & Kawakami, S. (2000). Total ozone increase associated with forest fires over the Indonesian region and its relation to the El Nino-Southern oscillation. *Atmospheric Environment*, *34*(17), 2681–2690. doi: https://doi.org/10.1016/S1352-2310(99)00522-1

Konca-Kędzierska, K., & Pianko-Kluczyńska, K. (2019). The influence of relative humidity on fires in forests of Central Poland. *Forest Research Papers*, *79*(3), 269–279. doi: https://doi.org/10.2478/frp-2018-0027

Latifah, A. L., Shabrina, A., Wahyuni, I. N., & Sadikin, R. (2019). Evaluation of Random Forest model for forest fire prediction based on climatology over Borneo. *2019 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, 4–8. doi: https://doi.org/10.1109/IC3INA48034.2019.8949588

Lengyel, A., & Botta-Dukát, Z. (2019). Silhouette width using generalized mean—A flexible method for assessing clustering efficiency. *Ecology and Evolution*, *9*(23), 13231–13243. doi: https://doi.org/10.1002/ece3.5774

Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. doi: https://doi.org/10.1007/978-0-387-09823-4

Mareta, L., Hidayat, R., Hidayati, R., & Latifah, A. L. (2019). Pengaruh Faktor Alami dan Antropogenik Terhadap Luas Kebakaran Hutan dan Lahan di Kalimantan. *Jurnal Tanah Dan Iklim*, *43*(2), 143–151. doi: https://doi.org/10.2017/jti.v43i2.10456

Margono, B. A., Turubanova, S., Zhuravleva, I., Potapov, P., Tyukavina, A., Baccini, A., … Hansen, M. C. (2012). Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010. *Environmental Research Letters*, *7*(3). doi: https://doi.org/10.1088/1748-9326/7/3/034010

Natingga, D. (2017). *Data Science Algorithms in a Week*. Retrieved from https://www.packtpub.com/big-data-and-business-intelligence/data-science-algorithms-week

Pramesti, D. F., Lahan, Tanzil Furqon, M., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(9), 723–732. doi: https://doi.org/10.1109/EUMC.2008.4751704

Prasad, K. S. N., & Ramakrishna, S. (2008). An autonomous forest fire detection system based on spatial data mining and fuzzy logic. *IJCSNS International Journal of Computer Science and Network Security*, *8*(12), 49–55. Retrieved from http://paper.ijcsns.org/07_book/200812/20081208.pdf

Pratamasari, A., Permatasari, N. K. F., Pramudiyasari, T., Manessa, M. D. M., & Supriatna, S. (2020). Spatial Distribution Patterns Analysis of Hotspot in Central Kalimantan using FIMRS MODIS Data. *Jurnal Geografi Lingkungan Tropik*, *4*(1), 24–34. doi: https://doi.org/10.7454/jglitrop.v4i1.74

Putra, Erianto I, Vetrita, Y., & Graham, L. L. B. (2016). Degraded peatlands, Ground Water Level and severe peat fire occurrences. *International Peat Congress*, (August).

Forest and Society. Vol. 4(2): 280-293, November 2020

292

Putra, Erianto Indra, & Hayasaka, H. (2011). The effect of the precipitation pattern of the dry season on peat fire occurrence in the Mega Rice Project area, Central Kalimantan, Indonesia. *Tropics*, *19*(4), 145–156. doi: https://doi.org/10.3759/tropics.19.145

Putra, Erianto Indra, Hayasaka, H., Takahashi, H., & Usup, A. (2008). Recent Peat Fire Activity in the Mega Rice Project Area, Central Kalimantan, Indonesia. *Journal of Disaster Research*, *3*(5), 334–34. doi: https://doi.org/10.20965/jdr.2008.p0334

Rao, V. S., & Vidyavathi, S. (2010). Comparative Investigations and Performance Analysis of Fcm and Mfpcm Algorithms on Iris Data. *Indian Journal of Computer Science and Engineering*, *1*(2), 145–151. Retrieved from https://www.ijcse.com/docs/IJCSE10-01-02-12.pdf

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(C), 53–65. doi: https://doi.org/10.1016/0377-0427(87)90125-7

Sari, N. M., Rachmita, N., & Manessa, M. D. M. (2020). Hotspot Distribution Analysis In East Kalimantan Province 2017-2019 to Support Forest and Land Fires Mitigation. *Indonesian Journal of Environmental Management and Sustainability*, *4*(1), 28–33. doi: https://doi.org/10.26554/ijems.2020.4.1.28-33

Shidik, G. F., & Mustofa, K. (2014). Predicting Size of Forest Fire. In *Lecture Notes in Computer Science* (pp. 316–327). doi: https://doi.org/10.1007/978-3-642-55032-4_31

Shoolihah, A.-M., Furqon, M. T., & Widodo, A. W. (2017). Implementasi Metode Improved K-Means untuk Mengelompokkan Titik Panas Bumi. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *1*(11), 1270–1276. Retrieved from http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/453

Siegert, F., & Hoffmann, A. A. (2000). The 1998 forest fires in East Kalimantan (Indonesia): A quantitative evaluation using high resolution, multitemporal ERS-2 SAR images and NOOA-AVHRR hotspot data. *Remote Sensing of Environment*, *72*(1), 64–77. doi: https://doi.org/10.1016/S0034-4257(99)00092-9

Sukamto, S., Id, I. D., & Angraini, T. R. (2018). Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means. *JUITA : Jurnal Informatika*, *6*(2), 137. doi: https://doi.org/10.30595/juita.v6i2.3172

Suyanto. (2018). *Machine Learning: Tingkat Dasar dan Lanjut*. Retrieved from http://biobses.com/product/machine-learning-tingkat-dasar-dan-lanjut/

Syaufina, L., & Sukmana, A. (2008). Tinjauan PenyebabUtama Kebakaran Hutan di Daerah Tangkapan Air Danau Toba. In *ITTO PROJECT PD 394/06 REV.1 (F) Centre of Forest and Nature Conservation Research and Development (CFNCRD)*.

Tacconi, L., Moore, P. F., & Kaimowitz, D. (2007). Fires in tropical forests - What is really the problem? Lessons from Indonesia. *Mitigation and Adaptation Strategies for Global Change*, *12*(1), 55–66. doi: https://doi.org/10.1007/s11027-006-9040-y

Usup, A., Hashimoto, Y., Takahashi, H., & Hayasaka, H. (2004). Combustion and thermal characteristics of peat fire in tropical peatland in Central Kalimantan, Indonesia. *Tropics*, *14*(1), 1–19. doi: https://doi.org/10.3759/tropics.14.1

Wang, Z., Liu, P., & Cui, T. (2017). Research on forest flame recognition algorithm based on image feature. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(2W7), 925–928. doi: https://doi.org/10.5194/isprs-archives-XLII-2-W7-925-2017

Wu, Z., He, H. S., Fang, L., Liang, Y., & Parsons, R. A. (2018). Wind speed and relative humidity influence spatial patterns of burn severity in boreal forests of northeastern China. *Annals of Forest Science*, *75*(3). doi: https://doi.org/10.1007/s13595-018-0749-z

Xu, T. S., Chiang, H. D., Liu, G. Y., & Tan, C. W. (2017). Hierarchical K-means Method for Clustering Large-Scale Advanced Metering Infrastructure Data. *IEEE Transactions on Power Delivery*, *32*(2), 609–616. doi: https://doi.org/10.1109/TPWRD.2015.2479941

Forest and Society. Vol. 4(2): 280-293, November 2020

293

Yong, Y., Chongxun, Z., & Pan, L. (2004). A novel fuzzy c-means clustering algorithm for image thresholding. *Measurement Science Review*, *4*(1), 11–19. Retrieved from http://www.measurement.sk/2004/S1/Yong.pdf