Clustering Pollutants Concentration from Sumatra Peat Fire using Fuzzy C-Means Algorithm

Hanif Rahmat¹, Imas Sukaesih Sitanggang†*

Department of Computer Science, IPB University, Bogor Indonesia
*Corresponding author: imas.sitanggang@ipb.ac.id

Abstract. Indonesia has large peatland areas which can significantly impact both regional and global environments. The forest and land fire in 2015 produced high level of air pollutants concentration in Sumatra, especially in South Sumatra, Riau, and Jambi provinces. This study aims to cluster pollutants concentration from peat fires in Sumatra during July until November 2015 using fuzzy c-means algorithm. The dataset used in this study is concentration of pollutants including CH₄, CO, CO₂, NH₃ and PM₂.₅. Based on the experimental result, pollutants concentration data are grouped into 8 clusters. The result of this study indicates that the highest pollutants concentration is found in the cluster which has only one member with concentration 14,500 µg/m³ of CH₄, 145,980 µg/m³ of CO, 1,183,831 µg/m³ of CO₂, 13,800 µg/m³ of NH₃ and 6,290 µg/m³ of PM₂.₅. Another cluster containing high concentration has 6 objects representing haze locations with the average concentration 8,265 µg/m³ of CH₄, 83,427 µg/m³ of CO, 676,558 µg/m³ of CO₂, 7,913 µg/m³ of NH₃, and 3,597 µg/m³ of PM₂.₅. The largest cluster with 590,816 haze locations contains the lowest concentration of pollutants. Although its average concentration levels are low, yet the maximum concentration are still in moderate or high enough levels. In general, high concentration of pollutants was distributed mostly in South Sumatra, Jambi, and Riau.

1. Introduction
The Indonesia’s peatlands area is estimated about 13.2 to 26.5 million ha, distributed mostly in three major areas those are Sumatra, Kalimantan, and Papua [1]. Furthermore, Indonesia has the largest tropical peatland in the world which the area reached approximately 14.9 ha [2]. With large areas of peatland, Indonesia’s recurrent peatland fires have significant impacts to both regional and global environments. Toxic haze generated by the fires causes severe impacts on public health and economy not only in Indonesia but also within neighboring countries such as Malaysia and Singapore [3]. The catastrophic fires of 2015 made six Indonesian provinces those are Riau, Jambi, South Sumatra, West Kalimantan, Central Kalimantan, and South Kalimantan declaring state of emergency as haze from wildfires on Sumatra and Kalimantan [4]. The 2015 fire event had generated multiple and severe impacts including environmental damages, public health, and economic losses. For example, smoke haze in Riau caused mostly upper respiratory tract infection (83.92%) followed by skin diseases (6.07%), eye diseases (4.83%), asthma (3.83%), and pneumonia (1.34%) [5]. The haze caused by peatland fires contains various pollutants which are dangerous for health to be inhaled. Clustering on pollutants concentration can give information about groups containing similar pollutants concentration level. Furthermore, visualization to the clustering result shows the information about the distribution of pollutants concentration level in the targeted areas.
Clustering is an unsupervised learning approach for classifying data into groups [6]. It is an essential data mining tool for analyzing and exploring air pollution data [7]. Clustering can be divided into hard clustering and soft clustering. In hard clustering, such as k-means clustering, clusters are separated by sharp boundaries; whereas in soft clustering, such as fuzzy c-means (FCM) clustering, there is overlap between clusters [8]. In k-means, each data object is the member of only one cluster, whereas in FCM, a data object is the member of all clusters with varying degrees of fuzzy membership between 0 and 1. Hence, the membership degrees of data points closer to the centers are higher than those of objects scattered in the borders of clusters [9].

Research on clustering of pollutants concentration from Sumatra peat fire haze has been done by Fanni [10]. The study clustered 5 pollutants concentration those are CO₂, CO, CH₄, NH₃, and PM₂.₅ using k-means clustering algorithm. The result of the study showed that the algorithm produced 5 clusters of pollutants concentration. Visualization of each cluster was then compared to hotspots distribution. The comparison showed that high level of pollutants concentration are influenced by heavy haze which potentially caused by peatland fires [10].

Furthermore, research on clustering of CO and CO₂ concentration especially from Sumatra peat fire haze has been done by Ni’am et al. [11]. In addition, Luo & Chen has conducted a study on potential sources and transportation path of PM₂.₅ in Shanghai using HYSPLIT modelling, k-means clustering algorithm, and Concentration-Weighted Trajectory (CWT) [11], [12].

This study aims to cluster pollutants concentration caused by peatland fires in Sumatra using FCM algorithm. The pollutants studied are Methane (CH₄), Carbon Monoxide (CO), Carbon Dioxide (CO₂), Ammonia (NH₃), and Particulate Matter PM₂.₅. The pollutants concentration data used are those in the period of July until November 2015. It is expected that the results can help in understanding the characteristics of pollutants caused by peatland fires. It will be useful in mitigation program. Furthermore, from the view-point of ecologist, the results can be explored to relate it with environmental conditions, e.g. land cover etc.

2. Data and Study Area
The data used in this study are pollutants concentration data obtained from European Centre for Medium-range Weather Forecasts (ECMWF) data server. Data were observed from satellite-based sensors assimilated by The Global Fire Assimilation System (GFAS) [13]. The global data then were selected for the area of Sumatra.

The data are daily pollutants concentration data in Sumatra from July to November 2015. Selection of the period is due to quite high number of fire points in that period [5]. Pollutants concentration studied are methane (CH₄), carbon monoxide (CO), carbon dioxide (CO₂), ammonia (NH₃), and PM₂.₅.

3. Methods
3.1. Data Pre-processing
The first step in data pre-processing is data conversion from matrix data to tabular data. The data obtained from ECMWF data server are stored as netCDF file. Each data that have matrix data structure was converted into tabular. The data were then filtered into the domain in which can cover Sumatra area. After that, the data were saved as comma-separated values (CSV) file and then it was combined as daily pollutants data from July 1 until November 30, 2015.

The next step of data pre-processing is clipping data. The data were selected using clipping technique to Sumatra polygon (shapefile) in PostgreSQL with PostGIS extension. Finally, the pollutants concentration unit was converted from unit kg/m²s into µg/m²s.

3.2. Data Visualization
Visualization of data was used to give overview about the data generally. The visualization was conducted using ‘openair’ package in R. The data of pollutants concentration was visualized in summary plot, scatter plot, and calendar plot. Summary plot was used to provide summary statistics with graphical display of the data, while scatter plot was used to see relationship between data objects.
Furthermore, calendar plot displayed daily pollutants concentration data in calendar-based graphics with scale of the concentration level.

3.3. Clustering Using Fuzzy C-Means

Fuzzy c-means (FCM) clustering is a soft clustering technique for classifying data into groups. In fuzzy clustering, each data point belongs to all the clusters with varying memberships and these membership values range between zero and one. For each data point, its membership value represents how strongly it belongs to its cluster. Therefore, the membership value of a data point will be closer to zero if it is farther from a cluster compared with the distances between that point to remaining clusters. The sum of membership values of a data point to all the clusters will be equal to 1. The FCM clustering algorithm is shown in Table 1 [8].

| Table 1. Fuzzy c-means algorithm |
|----------------------------------|
| a. Fix number of clusters ($c$) and tolerance value ($tol$) |
| b. Generate initial membership matrix of the data with ‘c’ clusters |
| c. Let $r$ be the iteration index. $r = 1, 2, 3, ...$ |
| d. Update cluster centers ($\theta$) and then membership values ($U$) by using |
| $\theta^r_j = \frac{\sum_{j=1}^{N} \left( \frac{u^r_j}{(r-1)^{q}} \right)^{q-1} X_j}{\sum_{j=1}^{N} \left( \frac{u^r_j}{(r-1)^{q}} \right)^{q}}$ |
| $u^r_j = \frac{1}{\Sigma_{k=1}^{N} \left( \frac{d^r_k}{d^r_k} \right)^{q-1}}$ |
| e. Repeat the above step till $\left| U^r - U^{(r-1)} \right| \geq t$ |

In this study, clustering was conducted using ‘fcm’ function available in ‘ppclust’ package in R. The clustering result was then visualized using ‘leaflet’ package. The visualization was aimed to see the distribution of each cluster of pollutants concentration in Sumatra.

4. Results and Discussion

4.1. Pre-processing Data

From one netCDF file, it can be obtained daily data of a pollutant in a month with location variables. The daily data of pollutants each month were combined and saved as CSV file as daily pollutants data from July 1 until November 30, 2015. The result example after combining data set is shown in Table 2.

| Table 2. Combined daily pollutants data |
|----------------------------------------|
| longitude  | latitude  | ch4july1 | ch4july2 | ch4july3 | ch4july4 | ch4july5 | ch4july6 | ch4july7 |
|------------|-----------|----------|----------|----------|----------|----------|----------|----------|
| 95.050003  | 5.750000  | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 |
| 95.150002  | 5.650000  | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 |
| 95.250000  | 5.850000  | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 |
| 95.250000  | 5.450000  | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 |
| 95.250000  | 5.350000  | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 |
| 95.250000  | 5.250000  | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 |
| 95.550003  | 5.550000  | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 | 2.12E-22 |

The next step of data pre-processing is data selection. Data selection was conducted to obtain pollutants concentration data in area of Sumatra. The data were selected using clipping technique to Sumatra polygon (shapefile) in PostgreSQL with PostGIS extension. After selection, the data of pollutants concentration was converted from unit kg/m’s into µg/m’s. This data pre-processing process resulted in 592,416 rows of data. The example data set after pre-processing is shown in Table 3.
Table 3. Example data set of pollutants

| Date     | Longitude | Latitude | CH₄ (µg/m³) | CO (µg/m³) | CO₂ (µg/m³) | NH₃ (µg/m³) | PM₂.₅ (µg/m³) |
|----------|-----------|----------|-------------|------------|-------------|-------------|---------------|
| 21/10/2015 | 104.8499985 | -3.849999905 | 0           | 0          | -1.08E-10   | 0           | 0             |
| 21/10/2015 | 104.9499969  | -3.849999905 | 0           | 0          | -1.08E-10   | 0           | 0             |
| 21/10/2015 | 105.0500031  | -3.849999905 | 0           | 0          | -1.08E-10   | 0           | 0             |
| 21/10/2015 | 105.1500015  | -3.849999905 | 6.18        | 62.4       | 506          | 5.92        | 2.69          |
| 21/10/2015 | 105.25       | -3.849999905 | 21.8        | 221        | 1790         | 20.9        | 9.51          |
| 21/10/2015 | 105.3499985  | -3.849999905 | 0           | 0          | -1.08E-10   | 0           | 0             |
| 21/10/2015 | 105.4499969  | -3.849999905 | 2.21        | 22.3       | 181          | 2.11        | 0.96          |
| 21/10/2015 | 105.5500031  | -3.849999905 | 3.31        | 33.4       | 271          | 3.17        | 1.44          |
| 21/10/2015 | 105.6500015  | -3.849999905 | 106         | 1070       | 8710         | 102         | 46.3          |
| 21/10/2015 | 105.75       | -3.849999905 | 3.53        | 35.6       | 289          | 3.38        | 1.54          |
| 21/10/2015 | 105.8499985  | -3.849999905 | 115         | 1160       | 9410         | 110         | 50            |
| 21/10/2015 | 105.9499969  | -3.849999905 | 0           | 0          | -1.08E-10   | 0           | 0             |
| 21/10/2015 | 106.5500031  | -3.150000095  | 0           | 0          | -1.08E-10   | 0           | 0             |
| 21/10/2015 | 101.6500015  | -3.150000095  | 0           | 0          | -1.08E-10   | 0           | 0             |
| 21/10/2015 | 101.75       | -3.150000095  | 0           | 0          | -1.08E-10   | 0           | 0             |

4.2. Visualization

Summary plot of pollutants concentration as shown in Figure 1 indicates that the highest concentration of pollutants through July-November 2015 was in the middle of October 2015. CO₂ had the highest concentration level since the average of it was 104.4 µg/m³'s while the concentration level of PM₂.₅ was the lowest as the average of it was 0.6 µg/m³'s. In addition, plot of concentration shows similar pattern over the period. That indicates positive relationship between the pollutants concentration level.

Figure 1. Summary plot of pollutants concentration July-November 2015
Figure 2. Scatter plots of CH₄ (a), CO (b), and CO₂ (c) concentration on July 1, 2015 until November 30, 2015

Figure 3. Calendar plots of CH₄ (a), CO (b), and CO₂ (c) concentration on July 1, 2015 until November 30, 2015
Figure 2 shows scatter plots of pollutants concentration data. Each scatter plot indicates a positive relationship between pollutants concentration. Moreover, the highest concentration data point of pollutants was far away from the others. It could be potentially an outlier.

Calendar plot is used to plot the concentration of pollutants daily from July to November 2015. The darker color means the higher concentration. Figure 3 shows calendar plots of pollutants concentration CH₄, CO, and CO₂ showing that the highest concentration of pollutants was in September and October 2015. Furthermore, the most frequent of high concentration occurrence was in October, and the highest concentration was on 19 and 21.

Figure 4 shows calendar plots of CO₂ concentration in Riau, Jambi, and South Sumatra as forest and peatland fires occurred in those areas during July until November 2015 [4][10]. The calendar plots show that the highest concentration level of CO₂ in Riau was on July 25, 2015 while in Jambi was on September 1, 2015. Furthermore, the pollutants concentration in South Sumatra was the most influential to the pollutants concentration level in Sumatra. The highest concentration of CO₂ in South Sumatra was on September 12 and October 19 and 21, 2015. The pollutants concentration level in South Sumatra had the same pattern as that in Sumatra overall.

Figure 4. Calendar plots of CO₂ concentration in Riau (a), Jambi (b), and South Sumatra (c) on July 1, 2015 until November 30, 2015

4.3. Clustering Pollutants Concentration
Clustering was performed using the C-Means module on Probabilistic and Possibilistic Cluster Analysis (ppclust) package that is available in R. Variables included in clustering are concentration of pollutants those are CH₄, CO, CO₂, NH₃, and PM₂.5. Number of clusters (c) was selected from 2 to 10 and the clustering results were evaluated by the Sum of Squared Errors (SSE). Figure 5 shows the SSE value and cluster quality for each cluster.
The cluster number was selected such that percentage of cluster quality is high enough to represent the clustering results with the smallest value of SSE. Based on the experimental results, the cluster number of 8 was selected to cluster the pollutants concentration. The quality of clustering reaches 97% while the SSE is 2.97E+11. Table 4 shows the Fuzzy C-Means clustering results of CH₄, CO, CO₂, NH₃, and PM₂.₅ with the number of cluster is 8.

**Table 4. Fuzzy C-Means clustering result with 8 clusters**

| Cluster | Total members | Percentage of cluster members | Descriptive Statistic | Pollutants Concentration |
|---------|---------------|-------------------------------|-----------------------|--------------------------|
|         |               |                               | CH₄ (µg/m³) | CO (µg/m³) | CO₂ (µg/m³) | NH₃ (µg/m³) | PM₂.₅ (µg/m³) |
| 1       | 590,816       | 99.723                        | Min 0        | 0          | 0           | 0          | 0           |
|         |               |                               | Max 58.1     | 586        | 4,750       | 55.6       | 26.3       |
|         |               |                               | Mean 0.18    | 1.88       | 16.73       | 0.16       | 0.09       |
| 2       | 1             | 0.0001                        | -            | 14,500     | 145,980     | 1,183,831  | 13,800      | 6,290       |
|         | 93            | 0.016                         | Min 785      | 7,930      | 64,300      | 752        | 342        |
|         |               |                               | Max 1,630    | 16,500     | 133,769     | 1,560      | 711        |
|         |               |                               | Mean 1,092   | 11,025     | 89,418      | 1,046      | 475.4      |
| 4       | 1,154         | 0.195                         | Min 19.8     | 302        | 4,770       | 2.79       | 25.4       |
|         |               |                               | Max 281      | 2,830      | 23,000      | 269        | 125        |
|         |               |                               | Mean 120.2   | 1,217      | 9,970       | 114.4      | 53.06      |
| 5       | 271           | 0.046                         | Min 282      | 2,850      | 23,100      | 270        | 123        |
|         |               |                               | Max 782      | 7,890      | 64,000      | 749        | 340        |
|         |               |                               | Mean 442.2   | 4,464      | 36,200      | 423.5      | 192.5      |
| 6       | 50            | 0.008                         | Min 1,680    | 16,900     | 137,273     | 1,610      | 730        |
|         |               |                               | Max 2,880    | 29,100     | 235,752     | 2,760      | 1,250      |
|         |               |                               | Mean 2,153   | 21,736     | 176,245     | 2,061      | 937        |
| 7       | 25            | 0.004                         | Min 3,200    | 32,300     | 26,2245     | 3,070      | 1,390      |
|         |               |                               | Max 5,750    | 58,100     | 471,126     | 5,510      | 1,757      |
|         |               |                               | Mean 4,035   | 40,724     | 330,304     | 3,864      | 2,500      |
| 8       | 6             | 0.001                         | Min 6,520    | 65,800     | 533,628     | 6,240      | 2,840      |
|         |               |                               | Max 10,300   | 103,863    | 842,282     | 9,850      | 4,480      |
|         |               |                               | Mean 8,265   | 83,427     | 676,558     | 7,913      | 3,597      |

The clustering result was visualized using ‘leaflet’ package that is available in R. Figure 6 shows the distribution of pollutants concentration in Sumatra based on the clustering result. According to Table 3, the highest concentration of pollutants is found in cluster 2 which has only one member. Hence, we can say that such data point is considered to be an outlier because it lies in the abnormal distance from other values. The data point is located in Ogan Komering Ilir district area on October 19, 2015 as shown in Figure 7(b). Furthermore, pollutant in the cluster 2 that had the highest concentration level was CO₂ by 1,183,831 µg/m³, while concentration level of PM₂.₅ was the lowest by 6,290 µg/m³.
Figure 6. Plot of pollutants concentration in all clusters

Cluster 8 with 6 locations as the members have the second highest concentrations of pollutants. These concentrations of pollutants were located in Jambi and South Sumatra areas. It can be considered that such concentrations occurred in September 12 and October 21, 2015. Plot of pollutants concentration in cluster 8 is shown in Figure 7(c).

Figure 7. Plots of pollutants concentration in cluster 1 (a), cluster 2 (b), and cluster 8 (c).
The lowest concentrations are found in cluster 1 with 590,816 locations are members of the cluster as shown in Table 3. The average concentration levels in cluster 1 are low, yet the maximum concentration are still in moderate or high enough levels. The cluster has maximum concentration 58.1 µg/m$^3$ of CH$_4$, 586 µg/m$^3$ of CO, 4,750 µg/m$^3$ of CO$_2$, 55.6 µg/m$^3$ of NH$_3$, and 26.3 µg/m$^3$ of PM$_{2.5}$.

The remaining clusters ordered from the highest concentrations are clusters 7, 6, 3, 5, and 4 with the number of members are 25, 50, 93, 271, and 1,154 respectively. In general, high concentrations of pollutants are distributed mostly in South Sumatra, Jambi, and Riau. Plots of concentrations in each cluster are shown in Figure 8.

![Figure 8](image_url)

Figure 8. Plot of pollutants concentration in cluster 3 (a), cluster 4 (b), cluster 5 (c), cluster 6 (d), and cluster 7 (e).

In this study, FCM algorithm made a good performance in clustering pollutants concentration which resulted in 97% cluster quality. However, the execution time was very high, besides the algorithm had to calculate such very large data. Since FCM executes more fuzzy calculations iteratively, its time complexity is sharply higher than the other partitioning clustering methods such as k-means. Though in many cases FCM can be very stable even in presence of outliers and overlapping, it requires more calculation time [15][16]. Thus, it is considered to cluster very large dataset to use method which results in comparable cluster quality with better runtime performance [17].

5. Conclusion

The result of this study indicates that the highest pollutants concentration is found in the cluster which has only one member with concentration 14,500 µg/m$^3$ of CH$_4$, 145,980 µg/m$^3$ of CO, 1,183,831 µg/m$^3$ of CO$_2$, 13,800 µg/m$^3$ of NH$_3$ and 6,290 µg/m$^3$ of PM$_{2.5}$, followed by cluster 8 which contains 6 objects representing haze locations. Moreover, the largest cluster with 590,816 haze locations contains the lowest concentration of pollutants. Although its average concentration levels are low, yet the maximum concentrations are still in moderate or high enough level. In general, high concentration of pollutants was distributed mostly in South Sumatra, Jambi, and Riau.

In this study, FCM algorithm made a good performance in clustering pollutants concentration, though the execution time was very high since FCM executes more fuzzy calculations iteratively,
besides the dataset used in this study was very large. Hence, it is considered to cluster such large dataset using a method which results in comparable cluster quality with better runtime performance.

References

[1] Wahyunto, Nugroho K, Agus F 2016 Perkembangan Pemetaan dan Distribusi Lahan Gambut on Lahan Gambut Indonesia p. 33-60 (Jakarta:Indonesian Agency of Agricultural Research And Development)

[2] Ritung S, Wahyunto, Nugroho K, Sukarman, Hikmatullah, Suparto, and Tafakresnanto C 2011 Peta Lahan Gambut Indonesia Skala 1:250,000 Indonesian Center for Agricultural Land Resources Research and Development (ICALRRD)

[3] Carmenta R, Zabala A and Phelps J 2015 Indonesian Peatland Fires: Perceptions of Solutions Center for International Forestry Research (CIFOR)

[4] Oktama C I 2015 Six Provinces Declare State of Emergency as Haze Worsens (Online) https://jakartaglobe.id

[5] Ministry of Health of the Republic of Indonesia 2015 Masalah Kesehatan Akibat Kabut Asap Kebakaran Hutan dan Lahan Tahun 2015 (Online) www.depkes.go.id

[6] Jain A K, Murty M N and Flynn P J 1999 Data clustering: A Review ACM Computing Surveys 31 (3) 264–323

[7] Aghabozorgi S, Shirkhorshidi A S, Wah T Y, Soltanian H and Herawan T 2014 Spatial and Temporal Clustering of Air Pollution in Malaysia: A Review International Conference on Agriculture, Environment and Biological Sciences (ICFAE)

[8] Joopudi S, Rathi S S, Narasimhana S and Rengaswamy R 2013 A New Cluster Validity Index for Fuzzy Clustering Mumbai: 10th IFAC International Symposium on Dynamics and Control of Process Systems: The International Federation of Automatic Control : 325-330

[9] Cebeci Z 2017 Partitioning Cluster Analysis Using Fuzzy C-Means (Online) https://cran.r-project.org

[10] Fanni S 2017 Analisis Pengelompokan Data Polutan Kabut Asap Kebakaran Lahan Gambut di Sumatra Menggunakan Algoritme K-Means dan Openair (Bogor: IPB University)

[11] Ni’am M, Sitanggang I S and Nuryanto D E 2017 Clustering of CO and CO₂ Concentration from Sumatra Peat Fire Haze Using HYSPLIT and K-means Algorithm IOP Conf. Series: Earth and Environmental Science

[12] Luo M and Chen C 2015 Potential sources and transport pathways of PM₃.₅ in Shanghai, China 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM) 104-107

[13] European Centre for Medium-range Weather Forecasts (ECMWF) 2019 CAMS Global Fire Assimilation System (GFAS) data documentation (Online) https://confluence.ecmwf.int

[14] Akagi S K, et. al. 2011 Emission factors for open and domestic biomass burning for use in atmospheric models Atmos. Chem. Phys. 11 4039-4072 doi: 10.5194/acp-11-4039-2011

[15] Heil J, Häring V, Marschner B, Stumpe B 2019 Advantages of fuzzy k-means clustering in the classification of diffuse reflectance soil spectra: A case study with West African soils Geoderma 337(2019) 11-21

[16] Mingoti S A, Lima J O 2006 Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms European Journal of Operational Research 174 1742-1759

[17] Cebeci Z, Yildiz F 2015 Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures Journal of Agricultural Informatics 6(3) 13-23