Indonesian Journal of Electrical Engineering and Computer Science
Vol. 3, No. 3, September 2016, pp. 714 ~ 720
DOI: 10.11591/ijeecs.v3.i2.pp714-720

Density Based Clustering of Hotspots in Peatland with Road and River as Physical Obstacles

Prima Trie Wijaya*[1,2], Imas Sukaesih Sitanggang[1], Lailan Syaufina[3]
[1]Department of Computer Science, Faculty of Natural Science and Mathematics,
Bogor Agricultural University, Indonesia
[2]Center for Technology of Energy Resources and Chemical Industry,
Agency For The Assessment and Application of Technology, Indonesia
[3]Department of Silviculture, Faculty of Forestry, Bogor Agricultural University, Indonesia
*Corresponding author, email: prima.trie@bppt.go.id

Abstract

Indonesia has the largest peatland area among tropical countries, covering about 21 milions ha, which spread mainly in Sumatera, Kalimantan, and Papua. Land and forest fires occur almost every year in peatland areas in Indonesia. One of indicators for forest and land fires is hotspot. The objective of this study is to group hotspots with road and river as obstacles using the CPO-WCC (Clustering in Presence of Obstacles with Computed number of Cells) algorithm. Clusters of hotspot data were analyzed based on peatland area distribution. This study also evaluates the results of clustering on peatlands in order to obtain the best clusters. Clustering using CPO-WCC algorithm produces three clusters of hotspot. The area of dense cluster is 10202.10 km^2 with number of hotspots per km^2 is 0.985. The higest number of hotspots occurrence is found in peatland with type of Hemists /Saprists (60/40) and depth greater than 400 cm.

Keywords: *cpo-wcc, hotspot, petland, spatial clustering*

1. Introduction

Peatlands in Indonesia spread in some large islands, such as Sumatra, Kalimantan, and Papua [1]. The main peatland area in Sumatra is in Riau Province (about 60%), South Sumatra and Jambi [1]. Indonesia faces serious problem of forest and land fire every year [2]. Forest and land fire in Indonesia have increased every year in which burn area increase, from 2612.09 ha in 2011 to 44411.36 ha in 2014 [3]. Riau is one of the districts that suffers due to forest and land fire where more hotspot found compared to other regions. According to the factors causing forest and land fires in Indonesia, natural factor plays only a very small role compared to the human factor which caused almost 100% of the forest and land fires either intentionally or unintentionally [1]. Land and forest fires affect environment, social, economic and health aspect [4]. Hotspots (active fires) indicate spatial distribution of forest and land fires [5]. Hotspot is an indicator of forest fires [4]. By By knowing the spread of hotspots, it will be easier in the process of preventing and extinguishing forest and landfires.

Data mining is suitable to beapplied in order to analyze distribution of large hotspot data [6]. This technique converts the large amount of data into useful information [7]. One of methods that can be used in data mining techniques to group spatial data is spatial clustering. Clustering is the process of grouping a collection of objects into clusters, so that the objects in one cluster have high level of similarity but they are not similar to the objects of another cluster [7]. Density based clustering is one of the primary methods for spatial clustering [8]. Density Based Spatial Clustering Algorithm with Noise (DBSCAN) as one of the spatial clustering algorithm defines cluster based on their density, cardinal value, and radius of neighbor object. It also provides the best partition value of each input given to this algorithm [9]. DBSCAN is efficiently utilized for processing large spatial database [10].

Moreover, the physical obstacles such as river, lake, and road in the real world affect the process of clustering [11]. Spatial data clustering can be done with constraint-based clustering algorithm, due to the physical condition problems in the real environment.CPO-WCC (Clustering in Presence of Obstacles with Computed number of cells) is one of constraint-based

Received April 2, 2016; Revised July 25, 2016; Accepted August 10, 2016

clustering algorithms [9]. The purpose of this study is to group the hotspot data on peatland in Riau using CPO-WCC Algorithm with rivers and roads as its constraints. Clustering result then will be analyzed based on the characteristics of peatland.The clustering results on hotspots as indicators of forest and peatland fires are expected later to be utilized by related parties involved in controlling forest and land fires in Riau.

2. Research Method

2.1. Study Area and Forest Fires Data

The data used in this research wereforest fire data in 2014 that were occurred in peatland in Riau. The data were taken from the Directorate of Forest Fire (DPKH) Ministry of Environment and Forestry. Road and river data weretaken from BAKOSURTANAL (currently Geospatial Information Agency). Furthermore, peatland data were obtained from *Wetlands International Indonesia Program*. Total peatland in Indonesia is about 20.6 million hectares in which 35% of peatland are located in Sumatera [12].

2.2. CPO-WCC Algorithm

The CPO – WCC algorithm firstly divides the area into rectangular cells with the same length and width, and counts number of points within the cell. Then the algorithm labels each cell with solid and not solid. The label solid represents dense area whereas the label not solid represents non dense area Next, each solid cells passed by obstacles is divided into rectangular with the same size. The process will be repeated until there is no solid cell passed by obstacles. Furthermore, the CPO – WCC algorithm will give name of area per cluster with name area cluster 1, cluster 2, and etc. The CPO-WCC algorithm is defined as follows [13].

2.2.1. Algorithm CPO – WCC

Input : A set of N objects (points) and set of polygon obstacles in a spatial area *S*.
Output : The relatively *dense* region in *S*.
Method :

1. Let la and lo be the dimensions of the spatial area. Determine two numbers, x, y, such that $\frac{x}{la} = \frac{y}{lo}$ and the average number of points in each cell, t, given by $\frac{\mathrm{N}}{\left(\frac{\mathrm{la}}{\mathrm{x}}\right)*\left(\frac{y}{lo}\right)}$ ranges from several dozes to several thousand.

2. Divide the spatial area into $\left(\frac{\mathrm{la}}{\mathrm{x}}\right) * \left(\frac{y}{lo}\right)$ rectangular cells that have equal areas by dividing the longitude and latitude of the spatial area into $\left(\frac{\mathrm{la}}{\mathrm{x}}\right)$ $\left(= \left(\frac{lo}{y}\right)\right)$ equal segments.

3. For each cell, c, determine the following parameters:
n_{c} : the number of the objects in the cell.
m_{c} : the mean of points in the cell.

4. For each cell, c, if $n_{\mathrm{c}} \geq t$, then label c as $dense$, otherwise label c as $non - dense$.

5. For each obstacle, O, all cells that intersect O are labeled as $obstructed$.

6. For each $obstructed$ cell, c, divide the cell c into a number of small pieces of equal areas in the same way as the spatial area is divided such that the average number of the points in each piece (smaller than the average number of points in a cell inside the spatial area) is in the range from several dozens to several hundreds.

7. For each small piece, p, label p as either $obstructed$ (i.e. intersects any obstacle) or $non - obstructed$.

8. For each $non - obstructed$ piece, p, that is not marked before in this step, the area constituted from p and its $non - obstructed$ neighbors is marked and all $non - obstructed$ neighbors that has just examined are put into a queue. Each time, one pieceis taken from the queue and repeat the same procedure except that those $non - obstructed$ piece that are not marked before are enqueued. When the queue is empty, one sub-cellis identified.

9. For each sub-cell, sc, obtained, the following parameters are determined:
n_{sc} : the number of the objects in the sub-cell
m_{c} : the mean of points in the sub-cell
p_{sc} : the percentage of the area covered by sc from c.

10. For each sub-cell, sc, obtained in step 9, if $\frac{n_{sc}}{t} \geq p_{sc}$, then label sc as dense. Otherwise label scas $non - dense$.

11. For each $dense$, $non - obstructed$ cell that is not, previously, processed in the current step, or dense sub-cell that also is not processed before in the current step, its neighboring $non - obstructed$ cells or sub-cells is examined to see if the average number of points in a cell within this small area is greater than or equal t. If so, this area is marked and all $dense$ cells or sub-cells that are just examined are put into a queue. Remove from the queue each $dense$ cell or sub-cell that has been examined before in a previous iteration. Each time one cell or sub-cell is takenfrom the queue and repeat the same procedure. When the queue is empty, one regionis identified.

3. Results and Analysis

The results of data preprocessing using QGIS and Microsoft Excel showed that there were 18292 hotspots in peatland in Riau, whereas there are only 18246 hotspots in peatland. Types of peatland in Riau are categorized based on the level of maturity into Fibrists, Hemists, Saprists, Saprists/Hemists, Hemists/Saprists, Fibrists/Saprists. Figure 1 shows the distribution of hotspots based on peat types in Riau, whereas Figure 2 shows the distribution of hotspots based on the depth of peat in Riau.

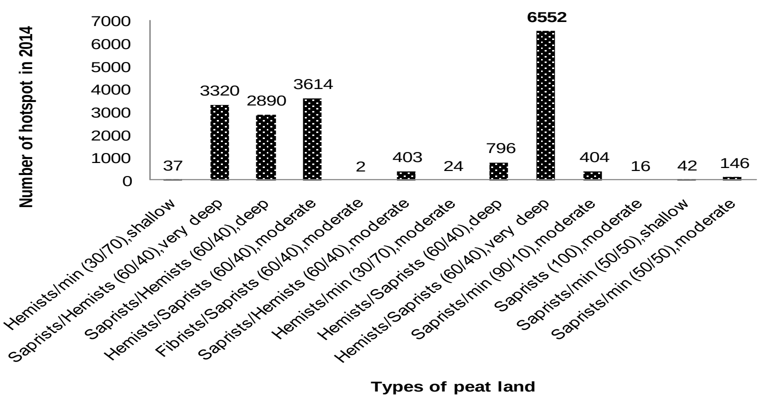


Figure 1. Hotspot Distribution based on Peat Types in Riau in 2014

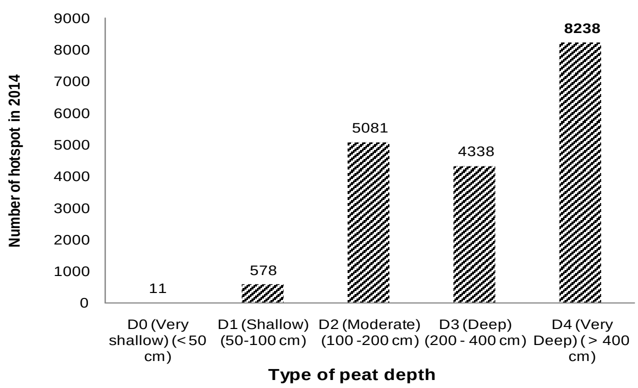


Figure 2. Number of Hotspots in 2014 based Peatland Depth

In Figure 1, the label "*Hemist /Saprists*" (60/40), very deep" means that the portion of Hemists and Sapristis 60% and 40%, respectively. *Very deep* means the depth of peat greated than 400 cm. Moreover, Figure 1 and 2 show that the most hotspot distribution was on

Hemist/Saprist which is 6552 hotspots. In addition, based on peatland depth, the largest number of hotspots were found in the peat with depth greater than 400 cm.

There are 13 land cover types ofpeatlands in Riau. The number of hotspots based on land cover is shown in Figure 3. It shows the largest number of hotspots in the peat swamp forest which is 12825 hotspots.

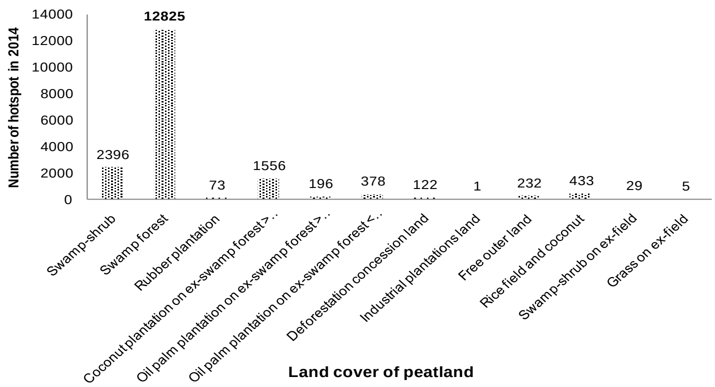
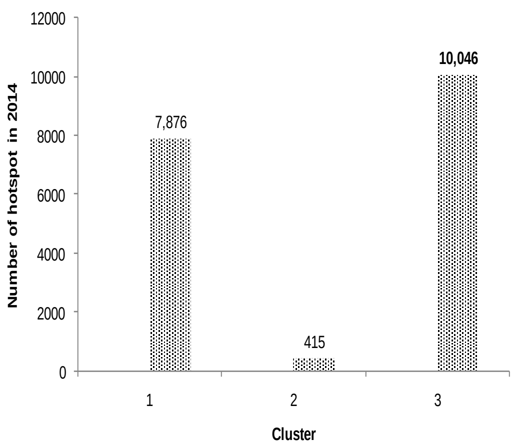


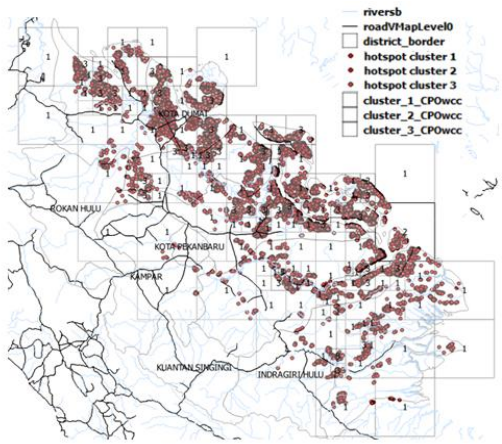
Figure 3. Number of Hotspots in 2014 based on Land Cover of Peatland

3.1. Clustering Hotspot using CPO-WCC Algorithm

In this study, hotspot clusters were created using CPO-WCC algorithm. The algorithm results hotspot clusters in five iterations with different length and width cell size on each iteration. The cell size is 0.5 degree, 0.25 degree, 0.125 degree, 0.0625 degree, and 0.03125 degree respectively in the first, second, third, fourth and fifth iteration. Iterations in CPO-WCC algorithm will continue to repeat if the area had a solid density and the areas are passed by the physical constraints such as only road, only riveror road and river. The results of CPO-WCC algorithm are shown in Figure 4.



a) Results of Clustering



b) Distribution of Hotspots Overlaid with Road and River

Figure 4. Results of CPO-WCC Algorithm

Figure 4(a) shows that CPO-WCC algorithm results three clusters where the largest number of hotspots is in *cluster* 3 and followed by *cluster* 1 and *cluster* 2. Next, Figure 4(b) shows the hotspots spread from Bengkalis, Indragiri Hilir, Indragiri Hulu, Kampar, Dumai, Pekanbaru, Pelalawan, Rokan Hilir, Rokan Hulu, and Siak in Riau province. The minimal number of hotspots so that grid was broken down is greater than or equal to 255 hotspots, 219, 146, 44 and 5 hotspots respectively in the first, second, third, fourth and fifth iteration.Those grid was broken down if the number of hotpots in solid grid is passed by river, road or both.

Figure 4 shows that the densest cluster is in cluster 3 and Table 1 shows the areas of all cluster as the results of CPO-WCC algorithm.

Table1. Cluster area in Riau Province

Cluster	Number of Hotspot	Area		Hotspot / km^2
		km^2	Percentage	
1	7849	81919,36	86%	0,096
2	397	3098,15	3%	0,128
3	10046	10202,10	11%	0,985

The area that hasthe highest density is in *cluster* 3 which every 1 km^2 has 0.985 hotspots (Table 1) area in cluster 3 located in the district bengkalis, Indragiri hilir, Indragiri hulu, dumai city, pelalawan, rokan hilir, and siak

Based on the extent of the area, the largest area was in *cluster* 1 which is 81919.36 km^2. The difference of these areas occurred because every area having the densest hotspots and the areas passed by only road or river or both of them would be cut into smaller cell as shown in Figure 4(b).

3.2. Hotpot Cluster Based on Peatland Characteristic

The cluster produces area having the most hotspots in *cluster* 3 (Table 1) with 10046 hotspots.The distribution of hotspots in cluster 3 based on the characteristic of peatland is shown in Figure 5.

According to Figure 5 and 6, the most hotspots in peatland of cluster 3 are in peatland with type of Hemists/Saprists (60/40) and depth greater than 400 cm. Those 4 257 hotspots were spread in some villages. There were 852 hotspots in Tanjung Leban village, 539 hotspots in Basilam Baru village, and 2866 hotspots were spread in 46 villages. Figure 7 shows the spread of hotspot based on the peat depth in cluster 3.

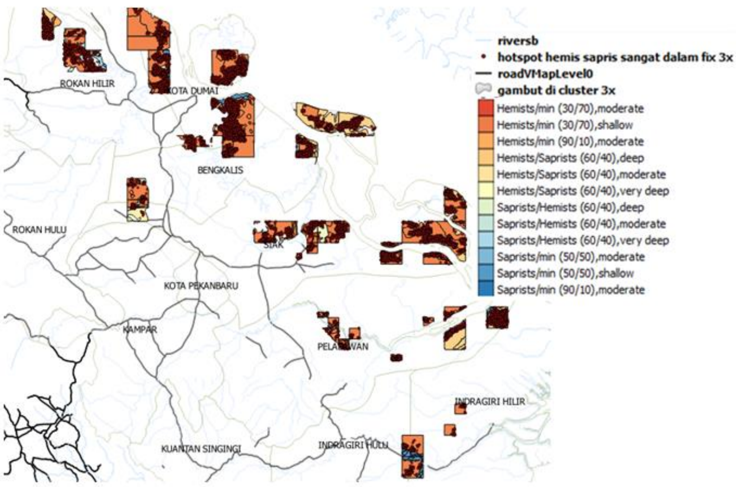


Figure 5. Distribution of Hotspots in 2014 based on Peatland Types In Cluster 3

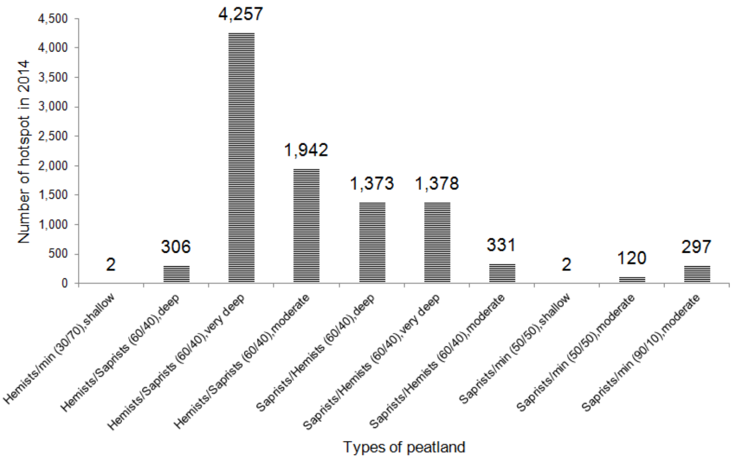
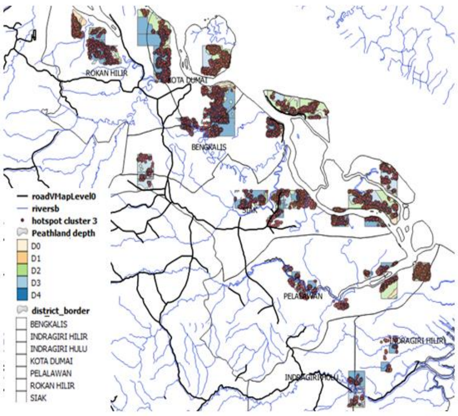
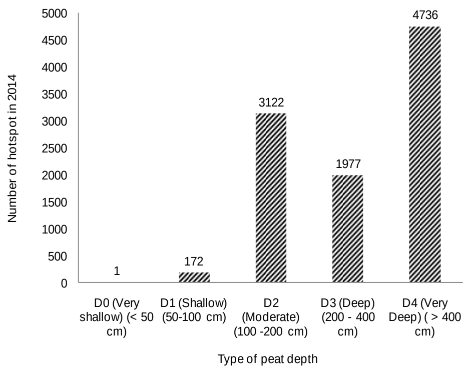


Figure 6. Number of Hotspots based on Peatland Types in Cluster 3 Algorithm



a) Hotspot Distribution



b) Number of Hotspots

Figure 7. Spread of Hotspots And Number Of Hotspots in 2014 based on Peatland Depth in Cluster 3

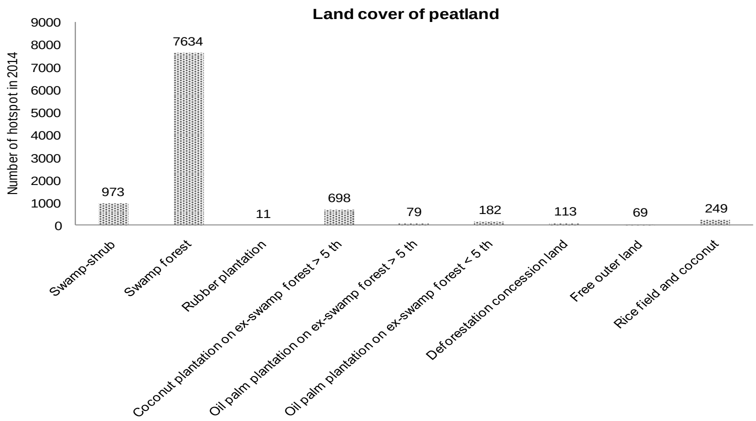


Figure 8. Number of Hotspots in 2014 based on Land Cover in Cluster 3

Figure 7(a) and 7(b) show that majority hotspots that reached 4736 hotspots were located in peatland with very deep/very dense (D4) level (> 400 cm). From those, there were

851 hotspots scattered in Tanjung Leban village, 489 hotspots in Basilam Baru village, and 3396 hotspots spread in 76 villages. Figure 8 shows the land cover of peatland in cluster 3.

The distributionof hotspots were dominated by swamp forest with 7634 hotspots (Figure 8). Those hotspots were scattered in Tanjung Leban village that have 975 hotspots and 554 hotspots in Basilam Baru, and the rest hotspots were scattered in 109 villages in Riau Province.

4. Conclusion

Provide a statement that what is expected, as stated in the "Introduction" chapter can ultimately result in "Results and Discussion" chapter, so there is compatibility. Moreover, it can also be added the prospect of the development of research results and application prospects of further studies into the next (based on result and discussion).

References

[1] Suratmo FG, Husaeni EA, Jaya NS, Sahardjo BA, Rachmatsjah O, Hirorki I, Nocolas MVJ, Ismunandar S, Prabowo D, Soedarmo, Sumatri, Prakoso JH. Pengetahuan Dasar Pengendalian Kebakaran Hutan. Bogor: Fakultas Kehutanan IPB. 2003.

[2] Kirana AP, Sitanggang IS, Syaufina L. Poisson Clustering Process on Hotspot in Peatland Area in Sumatera. *TELKOMNIKA Telecommunication, Computing, Electronics and Control*. 2015; 13(4): 1376-1383.

[3] Ministry of the Environment and Forestry: The data area of the fires. 2016. http://sipongi.menlhk.go.id/hotspot/luas_kebakaran.

[4] Agus F, Subiksa IGM. LahanGambut: Potensi untuk Pertanian dan Aspek Lingkungan. Bogor: Balai Penelitian Tanah dan World Agroforestry Centre (ICRAF). 2008.

[5] Sitanggang IS, Yaakob R, Mustapha N, Ainuddin AN. A Decision Tree Based on Spatial Relationship for Predicting Hotspots in Peatlands. *TELKOMNIKA Telecommunication, Computing, Electronics and Control*. 2014; 12(2): 511-518.

[6] Usman M, Sitanggang IS, Syaufina L. *Hotspot Distribution Analyses based on Peat Characteristics using Density-Based Spatial Clustering*. The 1st International Symposium on LAPAN-IPB Satellite for Food Security and Enviromental Monitoring. Procedia Environmental Sciences. 2015; 24: 132-140.

[7] Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. San Francisco (USA): Morgan-Kaufmann Publisher. 2012.

[8] Parimala M, Lopez D, Senthikumar NC. A Survey on Density Based Clustering Algorithms for Mining Large. *International Journal of Advanced Science and Technology*. 2011; 31: 59-66.

[9] Wahyunto, Ritung S, Suparto, Subagjo H. Sebaran Gambut dan Kandungan di Sumatera dan Kalimantan. Bogor: Wetlands International IP. 2005.

[10] Ester M, Kriegel HP, Sander J, Xu X. *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.* KDD-96 Proceedings. 1996: 226-231.

[11] Deborah LJ, Baskaran R, Kannan A. A survey on Internal Validity Measure for Cluster Validation. *International Journal of Computer Science & Engineering Survey (IJCSES)*. 2010; 2(2): 85-102.

[12] Syaufina L. Kebakaran Hutan danLahan di Indonesia. Malang: Bayumedia Publishing. 2008.

[13] El-Zawawy MA, El-Sharkawi ME. Clustering with O bstaclessIn Spatial Databases. 2009.