VOL. 2 NO. 1, JUNI 2009 ISSN 1979 - 4991

JURNAL AGRIBISNIS TERPADU

VOL. 2	NO. 1	Hal 1 - 76	Serang, Juni 2009	ISSN 1979 - 4991

Jurnal Agribisnis Terpadu ISSN 1979 - 4991

DAFTAR ISI

LAND CONSERVATION ADVICE IN WATERSHED BY CLUSTER ANALYSIS

(A case of Serang Watershed, Kulon Progo, Yogyakarta)

Weksi Budiaji [1], *Bambang Sumantri* [2], *Utami Dyah S.* [2], *R. Gatot Nursinggih H.* [3]

[1] Staf Pengajar Jurusan Agribisnis Fakultas Pertanian
Universitas Sultan Ageng Tirtayasa, *weksi_budiaji@yahoo.com*
[2] Staf Pengajar Departemen Statistika FMIPA Institut Pertanian Bogor
[3] Staf Divisi Perencanaan Dinas Kehutanan Provinsi Daerah Istimewa Yogyakarta

ABSTRAK

Guna mendukung langkah-langkah yang tepat dalam penggunaan lahan perlu adanya gambaran yang jelas mengenai batasan-batasan atau arahan penggunaan lahan dalam mendukung perlindungan ekosistem dan pelestarian lingkungan. Arahan fungsi pemanfaatan lahan sangat diperlukan sebagai dasar pertimbangan pengelolaan Daerah Aliran Sungai (DAS). Arahan fungsi pemanfaatan lahan sangat beragam, untuk menyederhanakan arahan tersebut digunakan analisis gerombol (cluster analysis). Lima metode penggerombolan berhierarki yang berturut-turut (agglomerative) dalam analisis gerombol digunakan untuk melihat karakteristik gerombol yang dihasilkan yaitu single linkage, complete linkage, average linkage, centroid linkage dan ward linkage. Metode penggerombolan yang menghasilkan karakteristik gerombol yang dianggap paling cocok adalah metode centroid linkage dengan jumlah gerombol sembilan. Arahan konservasi yang diperoleh sebanyak sembilan arahan konservasi yaitu agroforestry, rumput bede, kebun campuran, kebun talun, hutan serasah banyak, perkebunan penutupan tanah sempurna, perumputan penutupan tanah sempurna, ladang dan alang-alang murni subur.

Keyword: cluster, linkage, hierarki, agglomerative, daerah aliran sungai

INTRODUCTION

Background

Flood, land slide, drought, erosion, decreasing of land productivity, water, and environment quality are the main problem in watershed management. These conditions are affected by land use mismanagement. To support land use management both inside and outside forest needs distinct description so that it can be used as basic consideration for ecosystem protection and preservation. Land use advice is necessary for watershed management judgment.

Land use advice used to land management judgment will have a great variation. In order to practical usage, these variations can be simplified. One of the statistics analyses to simplify objects is cluster analysis. There are some methods to merge objects in its analysis process. Clusters meaning that are constructed depends on its linkage methods. Five common linkage methods will be treated in all objects so that they result in practical clusters.

Objectives

The main objectives of this paper are to merge land use to gain best practical clusters and choose among five common linkage methods that is the most suitable for this case.

METHODS

Data

The data used in this study is secondary data cite from Hastuti (2004) observation in Serang Watershed, Kulon Progo, Jogjakarta. Land units used in this paper are 72 land units classified as bad and very bad erosion class.

Land unit width, dry rainfall, wet rainfall, land sensitivity, length of slope, and land utilization and conservation is the six variables used to merge 72 objects. Wet rainfall and dry rainfall, moreover, are resulted from divided one year rainfall of Hastuti (2004) rainfall observation. Wet rainfall emerged from May to September and dry rainfall appeared from October to April. Additional notes can be seen in Table 1.

Table 1. Variables Used in Cluster Analysis

Variable	Unit
X_1	Land Unit Width (Ha)
X_2	Dry Rainfall (mm)
X_3	Wet Rainfall (mm)
X_4	Land Sensitivity to Erosion
X_5	Length of Slope
X_6	Land Utilization and Conservation (0 – 1)

Data Analysis

Cluster analysis unifies objects that are similar into one group/ cluster. Different objects, on the contrary, will be separated from them. High similarity among cluster member (high intra-class similarity) and low similarity with other clusters (low inter-class similarity) are expected from cluster analysis (Han and Kamber, 2005).

The basic purpose of cluster analysis is finding natural clusters among objects. The first step to measure similarity among objects in cluster analysis is quantitative measurement development (Johnson and Wichern, 1998). One of methods to measure similarity is distance measurement (Hair et al., 1998). Distance that is common used is Euclidean distance. Formula to measure Euclidean distance is

$$d(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})'(\mathbf{x}-\mathbf{y})}$$

$$= \sqrt{\sum_{i=1}^{p}(\mathbf{x}_i - \mathbf{y}_i)^2}\text{, where}$$

x_i = value x in the i-th variable

y_i = value y in the i-th variable

Distance measure that is important suggested by Hair et al. (1998) is squared Euclidean distance. It is suggested if there are correlation among variables and linkage method usage are centroid linkage and ward linkage. Squared Euclidean distance is Euclidean distance without root square.

Similarity between two clusters, in the contrary, is measured by formula

$$s_{(uv)} = 100\left(1 - \frac{d_{(uv)}}{d_{(\max)}}\right),$$

where $d_{(uv)}$ is distance between cluster U and V, $d_{(max)}$ is maximum value of original distance matrix, **D** (Minitab, 2006).

Agglomerative Hierarchical Clustering Algorithm

One of methods to join objects is agglomerative hierarchical method (Hair et al., 1998). Johnson and Wichern (1998) say that agglomerative hierarchical method results from sequence joining objects (agglomerative). First, the number of clusters is similar to the number of objects. Objects that is the most similar will be unified. Then, along with the decresing of similarity, all of objects will be merge into one cluster. The most important characteristics that is mentioned by Hair et al. (1998) is dendogram or three graph. It emerges in hierarchical method because its first step is nested to the next steps.

The following are the steps in agglomerative hierarchical clustering algoritm for grouping N objects cited from Jonson and Wichern (1998):

1. Start with N clusters, each containing a single entity and an N x N symmetric matrix of distance $\mathbf{D} = \{d_{ik}\}$.
2. Search the distance matrix for the nearest (most similar) pair of clusters. Let the distance between "most similar" clusters U and V be d_{uv}.
3. Merge clusters U and V. Label the newly formed cluster (UV). Update the entries in the distance matrix by (a) deleting the rows and columns corresponding to clusters U and V and (b) adding a row and column giving the distances between cluster (UV) and the remain clusters.
4. Repeat step 2 and 3 a total of $N - 1$ times. (All objects will be in single cluster after the algorithm terminates.) Record the identity of cluster that are merged and the levels (distances) at which the mergers take place.

According to Hair et al. (1998), there are five popular agglomerative methods. These are single linkage, complete linkage, average linkage, centroid linkage, and ward linkage. All of these definitions are taken from Minitab (2006).

1. Single Linkage

   In single linkage, or "nearest neighbor," the distance between two clusters is the minimum distance between an observation in one cluster and an observation in the other cluster. When observations lie close together, single linkage tends to identify long chain-like clusters that can have a relatively large distance separating observations at either end of the chain. In terms of the distance matrix,

   $$d_{(uv)w} = \min\{ d_{uw}, d_{vw} \},$$

   where d_{uw} is distance between cluster U and W, and d_{vw} is distance between cluster V and W.

2. Complete Linkage

   In complete linkage, or "furthest neighbor," the distance between two clusters is the maximum distance between an observation in one cluster and an observation in the other cluster. In terms of the distance matrix,

   $$d_{(uv)w} = \max\{d_{uw}, d_{vw}\},$$

   where d_{uw} is distance between cluster U and W, and d_{vw} is distance between cluster V and W.

3. Average Linkage

   In average linkage, the distance between two clusters is the average distance between an observation in one cluster and an observation in the other cluster. In terms of the distance matrix,

   $$d_{(uv)w} = \frac{N_u d_{uw} + N_v d_{vw}}{N_{(uv)}},$$

   where N_u, N_v, and $N_{(uv)}$ are the number of objects in cluster U, V, and UV. Whereas d_{uw} is distance between cluster U and W, and d_{vw} is distance between cluster V and W.

4. Centroid Linkage

   In centroid linkage, the distance between two clusters is the distance between the cluster centroids or means. In terms of the distance matrix,

$$d_{(uv)w} = \frac{N_u d_{uw} + N_v d_{vw}}{N_{(uv)}} - \frac{N_u N_v d_{uv}}{N^2_{(uv)}},$$

where N_u, N_v, and $N_{(uv)}$ are the number of objects in cluster U, V, and UV. Whereas d_{uw} is distance between cluster U and W, d_{vw} is distance between cluster V and W. and d_{uv} is distance between cluster U and V.

5. Ward Linkage

In Ward's linkage, the distance between two clusters is the sum of squared deviations from points to centroids. The objective of Ward's linkage is to minimize the within-cluster sum of squares. In terms of the distance matrix,

$$d_{(uv)w} = \frac{(N_w + N_u)d_{uw} + (N_w + N_v)d_{vw} - N_w d_{uv}}{N_w + N_{(uv)}}$$

where N_w, N_u, N_v, and $N_{(uv)}$ are the number of objects in cluster W, U, V, and UV. Whereas d_{uw} is distance between cluster U and W, d_{vw} is distance between cluster V and W, and d_{uv} is distance between cluster U and V.

RESULTS AND DISCUSSION

Correlation among Variables

Table 2 shows that there are correlations among variables. The correlations emerge between X_2 and X_3, X_3 and X_4, X_3 and X_6, X_5 and X_6. Because correlation among variables is too many indentified, squared Euclidean distance is applied.

Table 2. Correlation among Variables

		X_1	X_2	X_3	X_4	X_5	X_6
X_1	Pearson Correlation	1.00	0.02	0.08	0.09	0.09	-0.15
	Sig. (2-tailed)	.	0.89	0.52	0.44	0.44	0.21
	N	72.00	72.00	72.00	72.00	72.00	72.00
X_2	Pearson Correlation	0.02	1.00	0.77(**)	0.15	0.13	-0.22
	Sig. (2-tailed)	0.89	.	0.00	0.20	0.27	0.06
	N	72.00	72.00	72.00	72.00	72.00	72.00
X_3	Pearson Correlation	0.08	0.77(**)	1.00	0.29(*)	0.15	-0.33(**)

	Sig. (2-tailed)	0.52	0.00	.	0.01	0.22	0.01
	N	72.00	72.00	72.00	72.00	72.00	72.00
X_4	Pearson Correlation	0.09	0.15	0.29(*)	1.00	0.00	-0.18
	Sig. (2-tailed)	0.44	0.20	0.01	.	0.98	0.12
	N	72.00	72.00	72.00	72.00	72.00	72.00
X_5	Pearson Correlation	0.09	0.13	0.15	0.00	1.00	-0.40(**)
	Sig. (2-tailed)	0.44	0.27	0.22	0.98	.	0.00
	N	72.00	72.00	72.00	72.00	72.00	72.00
X_6	Pearson Correlation	-0.15	-0.22	-0.33(**)	-0.18	-0.40(**)	1.00
	Sig. (2-tailed)	0.21	0.06	0.01	0.12	0.00	.
	N	72.00	72.00	72.00	72.00	72.00	72.00

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

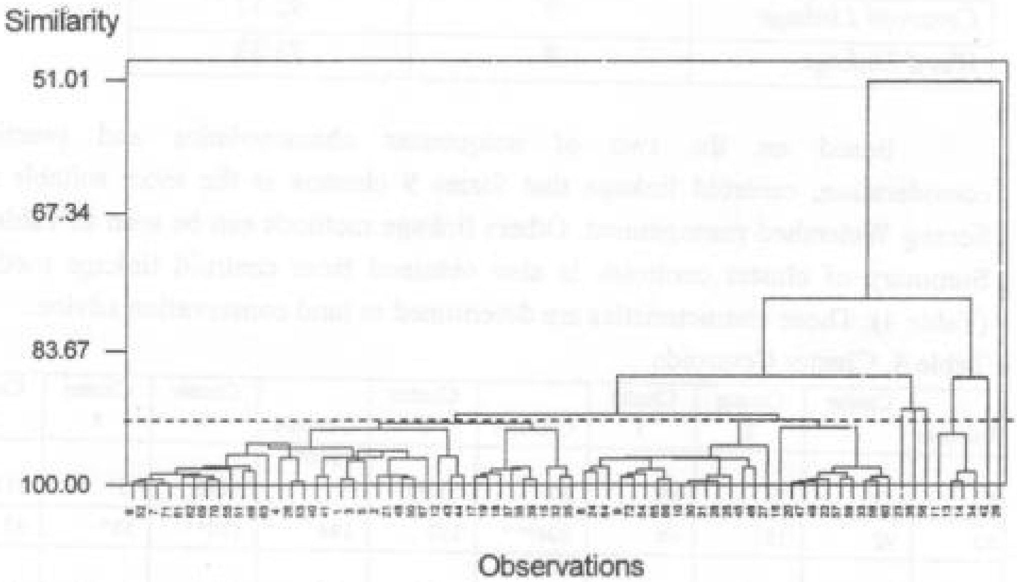

Figure Centroid Linkage Dendogram

Cluster Analysis

Cluster analysis result, moreover, is in a dendogram form. Every linkage has different dendogram shape each other. Dendogram cut in certain similarity or distance will result in a certain number of clusters. The Figure describes hcw centorid linkage dendogram are cut in certain similarity level.

There are two characteristics of cluster members that must be considered. First, X_6 or Land Utilization and Conservation variable scored 1 must become single cluster. Score 1 in this variable means infertile land so that it needs individual solving. Second, X_1 or Land Unit Width that is large should be a cluster alone. It is important because large land unit have great effect to the erosion speed.

Table 3. Cluster Number and Similarity Level of Each Linkage Method

Method	Cluster Number	Similarity Level
Single Linkage	8	96.47
Complete Linkcge	9	83.17
Average Linkage	7	88.79
Centroid Linkage	9	92.52
Ward Linkage	8	73.33

Based on the two of uniqueness characteristics and practical consideration, centroid linkage that forms 9 clusters is the most suitable for Serang Watershed management. Others linkage methods can be seen in Table 3. Summary of cluster centroids is also obtained from centroid linkage method (Table 4). Those characteristics are determined in land conservation advice.

Table 4. Cluster Centroids

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster4	Cluster 5	Cluster6	Cluster 7	Cluster 8	Cluster 9
X1	122	71	49	39	832**	1940***	1297	5*	1118**
X2	92	191	68	204***	195	184	120**	58*	97
X3	1524	2086	1267	2204***	2028	2081	1566	1241*	1782**
X4	0.4	0.45	0.28*	0.37	0.46	0.46	0.3	0.56***	0.43**
X5	5.5**	9.2***	2	2.3	3.1	9.5***	3.1	3.1	9.5***
X6	0.25**	0.21**	1.00***	0.32**	0.20*	0.20*	0.20*	1.00***	0.20*

* : low value
** : average value
*** : high calue

Land Conservation Advice

Cluster 1 is determined by 122 Ha in width average, 92 mm dry rainfall average, 1524 mm wet rainfall average, and 0.4 land sensitivity average score. In addition, 5.5 length of slope average score and 0.25 land utilization and conservation average score also determine this cluster. The main characteristics are little steep to steep classification of slope and village, mix plantation, and dry cultivation land unit utilization and conservation. Because of village existence, land conservation advice will be economically good if its advice is agro forestry. Agro forestry production will help house hold income.

Characteristics obtained from cluster 2 are 71 Ha in width average, 191 mm dry rainfall average, 2086 mm wet rainfall average, 0.45 land sensitivity average score, 9.2 length of slope average score and 0.21 land utilization and conservation average score. The most important characteristics are high steep of slope and village, mix plantation, wet and dry cultivation land unit utilization and conservation. As high steep of slope is present, Bede grass is proper for its slope.

Cluster 3 is determined by 49 Ha in width average, 68 mm dry rainfall average, 1267 mm wet rainfall average, 0.28 land sensitivity average score, 2 length of slope average score and 1 land utilization and conservation average score. The important characteristics are low of land sensitivity, slightly steep to little steep classification of slope and infertile land of land unit utilization and conservation. Since land sensitivity is low, mix plantation land conservation advice is suitable.

Cluster 4 characteristics are 39 Ha in width average, 204 mm dry rainfall average, 2204 mm wet rainfall average, 0.37 land sensitivity average score, 2.3 length of slope average score and 0.32 land utilization and conservation average score. The main characteristics are high in dry and wet rainfall, mix plantation and dry cultivation of land unit utilization and conservation. Because of high rainfall, the erosion will be very heavy by splashing rainfall. The Combination of short and tall plants such in Talun plantation will reduce speed erosion.

Characteristics obtained from cluster 5 are 832 Ha in width average, 195 mm dry rainfall average, 2028 mm wet rainfall average, 0.46 land sensitivity average score, 3.1 length of slope average score and 0.20 land utilization and conservation average score. The main characteristics are average of land unit width and mix plantation of land unit utilization and conservation. Land conservation advice for this cluster is perfect covered forest.

Cluster 6 is determined by 1940 Ha in width average, 184 mm dry rainfall average, 2081 mm wet rainfall average, and 0.46 land sensitivity average score. Furthermore, 9.5 length of slope average score and 0.2 land utilization and conservation average score also determine this cluster. The main characteristics are large of land unit width, high steep classification of slope and mix plantation land unit utilization and conservation. Perfect land covered of plantation is good for this cluster.

Cluster 7 characteristics are 1297 Ha in width average, 120 mm dry rainfall average, 1566 mm wet rainfall average, 0.30 land sensitivity average score, 3.1 length of slope average score and 0.20 land utilization and conservation average score. The most important characteristics are average in dry rainfall, mix plantation of land unit utilization and conservation. Land conservation advice for it is perfect land covered of grass.

Characteristics obtained from cluster 8 are 5 Ha in width average, 58 mm dry rainfall average, 1241 mm wet rainfall average, 0.56 land sensitivity average score, 3.1 length of slope average score and 1 land utilization and conservation average score. The important characteristics are small of land unit width, low in dry and wet rainfall, high of land sensitivity, and infertile land unit utilization and conservation. Dry cultivation can be treated for this cluster.

Cluster 9 is determined by 1118 Ha in width average, 97 mm dry rainfall average, 1782 mm wet rainfall average, 0.43 land sensitivity average score, 9.5 length of slope average score and 0.20 land utilization and conservation average score. The main characteristics are average of land unit width, average in wet rainfall, average of land sensitivity, high steep classification of slope and mix plantation of land unit utilization and conservation. Land conservation advice for this cluster is fertile coarse grass.

CONCLUSIONS

Squared Euclidean distance is satisfied applied for Serang Watershed case. Moreover, Cluster linkage method that is suitable is centroid !inkage which results in 9 clusters. Land conservation advice which merges from those are agro forestry, Bede grass, mix plantation, Talun plantation, perfect covered forest, perfect land covered of plantation, perfect land covered of grass, dry cultivation, and fertile coarse grass.

Last but not least, for further research, it is recommended that dry and wet rainfall be evaluated because these may be changed due to climate change.

REFERENCES

Hair JFJ, RE Anderson, RL Tattam, WC Black. 1998. *Multivariate Data Analysis 5th ed.* New Jersey: Prentice-Hall.

Han J, M Kamber. Data Mining Concepts and Techniques: Slides for Textbook Chapter 8. http://www.cs.sfu.ca/~han/ bk/8clst.ppt [16 juni 2005]

Hastuti D. 2004. Evaluasi Penggunaan Lahan untuk Arahan Konservasi Lahan Di Daerah Aliran Sungai Serang Kulon Progo Yogyakarta [Tesis]. Program Studi Ilmu Kehutanan Konsentrasi Konservasi Sumberdaya Alam dan Lingkungan. Yogyakarta: Universitas Gadjah Mada.

Johnson RA, DW Wichern. 1998. *Applied Multivariate Statistical Analysis 4th ed.* New Jersey: Prantice-Hall Int.

Minitab Inc. 2006. Minitab User's Help Release 15 for Windows.