ISSN 1225-5343

2000. 8 Vol. 43

Land Use and Land Cover Change

The Contribution of Geography

Journal of Geography Education

Special Issue for the IGU Study Group on
Land Use and Land Cover Change,
29th International Geographical Congress
Seoul, Korea
August 14-18, 2000.

Editors
Manik Hwang
Yukio Himiyama

Department of Geography Education

College of Education
Seoul National University
Seoul, Korea

地理教育論集，43，122-136，2000．8．

Contiguous Spatial Classification: A new approach on quantitative zoning method

Ernan Rustiadi* and Shintaro Kobayashi**

I. Introduction

Classification is a fundamental aspect of scientific activity and widely used as a descriptive tool, summarizing large data sets in a readily appreciated format. According to Johnston (1976), there are two reasons for classification: (1) reducing large number individuals to a small number of groups, and (2) to facilitate description and illustration.

Classification of geographical data needs a specific approach. A homogeneity-based classification should clearly show more similarity between in-group members rather than inter-group members. In spatial classification, an ideal member of one group should close in term of physical distance instead of in their attribute's values. Accordingly, a member of one spatial group should have high spatial autocorrelation. The spatial autocorrelation is a natural phenomenon of geographical data, therefore, grouping or classification of geographical data should determine the autocorrelation between individuals based on their spatial association. Autocorrelation statistics are basic descriptive statistics for any data that are ordered in a sequence because they are provide basic information about the ordering of the data that is not available from other descriptive statistics such as the mean and variance (Odland, 1988). Autocorrelation statistics are function of the same data values that are used to calculate other descriptive statistics but they are also functions of the arrangement of those values in a sequence. The arrangement is expressed by some function that assigns values to pairs of locations in the sequence in order to represent their location with respect to one another. In fact, most of spatial autocorrelation formulations are not used for the classification purposes.

Analysis on the Modifiable Areal Unit Problem (MAUP) can be seen as a kind of homogeneity-based spatial classification. Openshaw (1977) was one of the first who re-emphasized the importance of aggregation effects on geographical data, after Gekkle and Biehl in 1934 started to show the increase of the correlation coefficient due to increasing scale by standardizing the size of areas. Openshaw (1982) noticed that there are two different types zonal arrangement, there are (1) zoning system: grouping system that incorporates a contiguity constraint (spatial aggregation based on contiguous arrangement of zones), and (2) grouping system: a non-contiguous grouping system. Classification method based on attributes values do not guaranty contiguity which make spatial classification often became fragmented, spatially, and make it became difficult to get a general spatial pat-

*Faculty of Agriculture, Bogor Agricultural University, Indonesia.
**Graduate School of Agriculture, Kyoto University, Japan.

tern. In regional classification, a contiguous regionalization is required to make the classification more efficient and applicable on regional policy and management. In regional management, it will be very valuable when data for one set of areal units are progressively aggregated into fewer and larger units. Most of geographical studies have employed spatial aggregation based on contiguous management of zones, something referred to as a zoning system or regionalization.

Region as a geographical term is a specific term of spatial classification. Johnston (1976) recognized two types of regions: (1) formal, comprising places with similar characteristics, and (2) functional, or nodal, emphasizing places with similar linkage patterns to other places. From another point of view, Johnston also subdivided region term into (1) regional types and (2) contiguous region. Regional type comprises place, which are similar on certain pre-determined characteristics (landscape, population structure, etc.). The contiguous regions involve a contiguity constraint — a region must comprise spatially conterminous unit. Therefore, contiguous regions are defined not only on category homogeneity but also on all parts being in direct contact with one another. Zoning system provides a major simplification of real-world complexity that is naturally geographical.

The recent discussion on spatial analysis has lead to the view that a key area for geographical attention concerns is how to analyze spatial information aggregated to zones. Spatial analysis now are challenged to discover methods of analysis that are appropriate for spatial zonal data which are modifiable due to its nature. Openshaw (1996) saw this challenge as the most important of all the GIS-relevant spatial analysis tasks that still need to be handled.

This study aims to apply several methods on spatial classification or zoning methods and then try to determine the characteristic of every class or zones based on their position and attributes. The study is focused on one of suburban regions of Jakarta City, Bekasi District. Bekasi District's will be grouped to be several zones by their determinant factors and geographical positions in attempt to describe the process of suburbanization in the study area.

II. Method

The purposes of classification processes consist of determining the criteria of classification and then allocating the object measured to classes. Considering the number of attributes to be determined, classification process can be distinguished between single and multi variables classification. To develop a contiguous classification system, a measurement of contiguity among spatial units is needed. It is possible to include geographical position variables to be determined by a common multivariate classification procedure such as cluster analysis method.

Two approaches of contiguous classification method will be examined. The first approach is developing a contiguity approach in single and multi variables classification method using cluster analysis methods. The second approach is by developing a new formula of spatial contiguous classification to manipulate the original attribute values in attempt to contrast the different between spatial zones.

1. Non-contiguous spatial classification

The term cluster analysis actually encompasses a number of different classification algorithms. This method has been applied to a wide variety of research problems. Clustering method uses the dissimilarities or distances between object when forming the clusters. These distances can be based on a single or multiple dimensions. The most straightforward way of computing distances among objects in multi-dimensional space is to compute Euclidean distance, which is computed as follow:

$$D_{ij} = \sqrt{(Z_i - Z_j)^2}$$

for a single variable case (1)

and

$$D_{ij} = \sqrt{(Z'_{1i} - Z'_{1j})^2 + (Z'_{2i} - Z'_{2j})^2 + \ldots + (Z'_{mi} - Z'_{mj})^2}$$

for a multivariable case (2)

where, D_{ij} is Euclidean distance between i and j, Z_{mi} and Z_{mj} are the attribute values of i and j for the Z_m variable, and Z'_{mi} and Z'_{mj} are the standardized values of Z_{mi} and Z_{mj}.

Ward's method (Ward, 1963) attempts to minimize the sum of squares (SS) of individual members to group centroids that can be formed at each step. In other words, the variance of the distance is to be minimized as follow:

$$\sigma_k^2 = \frac{1}{n_k} \left\{ \sum_{i=1}^{n_k} (D_{ik})^2 \right\} \quad (3)$$

where σ_k^2 is the variance of the group k, D_{ik} is the distance between place i and the centroid of group k, assuming that i is a member of the group k, and n_k is the number of members of group k.

2. Contiguous spatial classifications

Many types of spatial analysis require knowledge of the degree of spatial association in the data. DeMers (1997) defined contiguity as a measure of the degree of wholeness within a region or of the degree to which polygons are in contact with one another. Many scholars have been trying to develop procedures for contiguous regions. Johnston (1970) suggested that the usual classification procedures should be adopted and then tests should be made to see if they also form contiguity regions. Furthermore Johnston (1976) introduced a contiguity constrain to be involved in the grouping procedure.

(1) Contiguous spatial classification using cluster analysis with geographical variables (NC Procedure)

In an attempt to identify contiguous regions, Johnston (1976) introduced a contiguity constraint to the hierarchical clustering with centroid replacement procedures. In this procedure, units were grouped only if they were contiguous (adjacent).

A contiguity constraint can be expressed in many approaches. It can be expressed by introducing geographical variables, the geographical position coordinate (easting X and northing Y). Accordingly, the Euclidean distance on grouping procedure (equation 1 and 2) should be modified to:

$$D_{ij} = \sqrt{(Z'_i - Z'_j)^2 + \beta\left\{(X'_i - X'_j)^2 + (Y'_i - Y'_j)^2\right\}}$$

for a single variable cases (4), and

$$D_{ij} = \sqrt{(Z'_{1i} - Z'_{1j})^2 + (Z'_{2i} - Z'_{2j})^2 + \Lambda + (Z'_{mi} - Z'_{mj})^2 + \beta\left\{(X'_i - X'_j)^2 + (Y'_i - Y'_j)^2\right\}}$$

for a multivariable cases (5)

where β is the weight of contiguity. The contiguity factor will be strengthened if β is set to be less than one ($\beta < 1$) and becomes weaker if β set bigger than one ($\beta > 1$). When β is set to be equal to one ($\beta = 1$), the weight of contiguity will decrease as the number of variable (m) increase. The complete procedure for this spatial contiguous classification is described by flowchart in Fig. 1.

Theoretically, the spatial contiguity among spatial units should be determined in a single

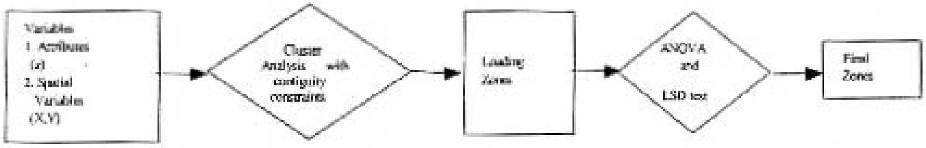

Fig. 1. Flowchart of spatial non-contiguous classification (NC procedure)

variable but in this procedure, the two geographical position variables (X and Y) will work independently to each other. Nevertheless, this procedure is easier and more practical compare to some other procedures. In the Thompson's contiguity constraint procedure, a contiguity matrix among all spatial units should be prepared and be used as a constraint matrix. This procedure is quite complex and time consuming.

The groups resulted from cluster analysis with contiguity constraints (loading zones) are then tested by statistical Least Significant Difference (LSD) test procedure. The final zones should have significant differences among each other.

(2) Contiguous spatial classification using aggregation formula (C′ and C″ procedures)

Basically, spatial autocorrelation is a measurement of general spatial aggregation of a data set but not in individual measurement and not for classification procedure purpose.

The basic factor of zoning system is contiguity among adjacent units. Therefore, a method to measure aggregation among spatial units is important to be developed. The spatial aggregation or spatial association (F_i) level of object i with its surrounding sites is the function of its attribute values (Z) and spatial or spatial association (F_j) with its surrounding region j, or mathematically it is described as

$$F_i = f(Z_i, Z_j) \tag{6}$$

The alternative formulations of F_i depend on the assumption of the association relation between i and j. Spatial association can be identified in a number of ways. Recently, a number of statistics called local statistics identify the association between single z_i and its neighbors up to a specified distance from i. Nass and Garfinkle (1992) used the localized autocorrelation diagnostic statistic, LADS to determine the presence of a contiguous bloc of geographic entities

Anselin (1995) formulated a class of Local Indicators of Spatial Association (LISA) as follows:

$$\Gamma_i = \sum_j w_{ij} y_{ij} \tag{7}$$

where w_{ij} represents the "spatial association" between sites i and j, and yij represents the "association of values (attributes)" of random variable at site i with values at the other sites. The y_{ij} can be in the form of Moran's type statistics, I_i, than $y_{ij} = (z_i - \bar{z})(z_j - \bar{z})$ or in the form of Geary's type statistics, C_i, then $y_i = (z_i - z_j)^2$.

Rustiadi and Kitamura (1998) have defined that attractiveness of a region is the function of population density and as the result of spatial aggregation among an associated spatial units, and formulated as follow:

$$F_i = \frac{P_i + \sum_j^m W_{ij} P_j}{A_i + \sum_j^m W_{ij} A_j}, \quad i \neq j \tag{8}$$

where F_i is an attractiveness index of region i, P_i and P_j are populations of i and j, A_i and A_j

are the total area of i and j. W_{ij} is weight for spatial association between i and j. It can be expressed as a continuous value, ranged from 0 to 1, or as a discrete value, between 0 and 1. In a discrete system, $W_{ij} = 1$ if i and j are adjacent regions and $W_{ij} = 0$ if i and j are not adjacent.

When the aggregation is determined as a multiplication function between the attributes of i and j, the following aggregation formula is proposed :

$$T_i = z_i \sum_{j}^{m_i} W_{ij} z_j \tag{9}$$

where T_i is absolute aggregation value, and z_i is attribute value of object or site i and z_j is attribute value for the other sites. W_{ij} is weight for spatial association between i and j. In a non-stationary system number of mi will vary due to irregularity object or polyons shape and size. In attemp to netralized this irregularity, the following formulation is proposed :

$$\overline{T_i} = \frac{1}{\sum_{j}^{m} W_{ij}} z_i \sum_{j}^{m_i} W_{ij} z_j. \tag{10}$$

For comparison and classification purposes, a standardized and dimensionless measurement is needed. For a data set consists of n objects or sites, a standardized value of spatial aggregation (F_i) is proposed. Therefore, T_i is standardized by $\frac{1}{n}\sum_{i}^{n} z_i^2$. If $D = \frac{1}{n}\sum_{i}^{n} z_i^2$, then

$$F_i = \frac{\overline{T}}{D} = \frac{1}{\sum_{j}^{m} W_{ij}} \frac{z_i \sum_{j}^{m_i} W_{ij} z_j}{\left(\frac{1}{n}\sum_{i}^{n} z_i^2\right)}, \quad i \neq j \tag{11}$$

$F_i < 1$: less spatialy aggregated than the average, $F_i = 1$: same than the average $F_i > 1$: more agregated than the average. $F_i = 1$, if z_i is equal to 0 or if summaration of the attribute's values of its

surrounding (z_j) are equal to 0 or $\sum_{j}^{m_i} W_{ij} z_j = 0$

The F_i measurement is useful to compare the zonal pattern among different attributes or different map themes, nevertheless, it can not be applied for a temporal changes analysis. In order to get more effective method of contiguous classification without standardizing, a square root of $\overline{T_i}$ can be used as a new corrected attribute value of z, or called as z_i''.

$$z_i'' = \sqrt{\overline{T_i}} = \sqrt{\frac{1}{\sum_{j}^{m} W_{ij}} z_i \sum_{j}^{m} W_{ij} z_j} \tag{12}$$

The procedure for this multivariate spatial contiguous classification is described by flowchart in Fig. 2.

III. Data and study area

Suburbanization process in Bekasi District as one of Jakarta Suburb's areas has been passed at least three suburbanization stages, those are pre-suburbanization stage, the first and second stages of suburbanization (Rustiadi, *et al.*, 1998). The pre-suburbanization stage of Bekasi District was characterized by relatively low-dense and dispersed population and low level of both urban and rice field

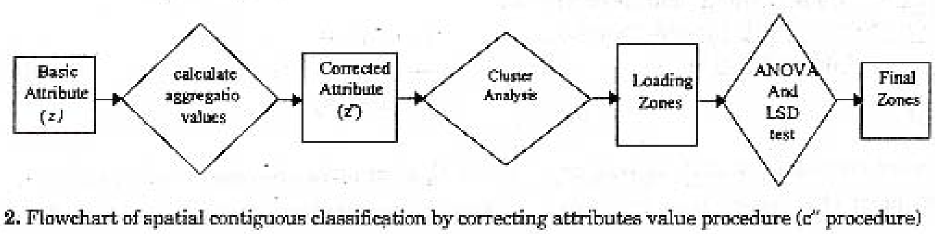

Fig. 2. Flowchart of spatial contiguous classification by correcting attributes value procedure (c'' procedure)

Table 1. Variables groups and zoning indicator parameters

Variable/parameters			unit	Notation
Variables	Basic attributes	1. urban land ratio	people/km^2	μ
		2. population density		ρ
		3. rice field land ratio		τ
	Spatial variables	Longitude coordinate	m (UTM)	X
		Latitude coordinate	m (UTM)	Y
	Aggregation index of region i.	1. absolute aggregation value		T_i
		2. standardized aggregation index		F_I
		3. corrected attributes value		z_I'
Zoning parameters	Coefficient of Variation within group (zone)		%	CV
	Unfragmented spatial unit			k
	Number of significant zone			Q

land ratios. The first stage of suburbanization characterized by two major increases in rice field and urban land ratios implicate the increase of land use mixture between them. The latest stage characterized by: (1) the continuing population agglomeration in the nearest area to Jakarta City, (2) the emerging of less-dense new local urban concentration in several areas, which are relatively far from Jakarta, and (3) the decreasing on rice field land ratio due to the conversion process to urban activities. Despite of these general stages, subregions in Bekasi District have been developed in various rates of growth. The suburbanization intensity vary due to their local conditions, accessibility to the city center, and other factors. A suitable classification of suburb to be some subregions is useful to determine the variation in suburbanization stages.

In this study, three variables will be used as the attributes for spatial classification, namely, population density (ρ), urban land ratio (μ), and rice field land ratio (τ). Table 1 describes the proposed variables groups for characterizing the subregions in the study area.

IV. Analysis results

Table 2 shows the result of single variable classification using three approaches of classification using cluster analysis procedures. The first approach, NC (non contiguous) procedure is employed for three single standardized attributes variables, namely population density (ρ), urban land ratio (μ) and rice field land ratio (τ). The C procedure is a contiguous procedure where as cluster method employed to the previous similar three variables (ρ, μ dan τ) and two variables of geographical position, X and Y. The C'' procedure is cluster analysis procedure apply to a manipulated (corrected) attribute values ρ'', μ'', and τ'' as the result of the calculation of aggregation formula (equation 12). Each year data sets are clustered three groups, L (the lowest group in the year), M (medium) and H (the highest group in the year). A statistical Least Significant Different (LSD) test procedure is conducted to test the different among the group's means.

Table 3 shows the result of multi variables classification procedure, whereas each year data sets are clustered to be three groups (1, 2 and 3). Each new groups are interpreted as a

Table 2. Statistical parameters of the groups resulting from univariate spatial classification for population distribution

procedure	Statistical parameter	Zones									Average
		1969/70			1981/82			1993			
		L	M	H	L	M	H	L	M	H	
	n_ρ	63	20	7	145	14	3	187	37	8	
	$\bar{\rho}$	445.0^{a}	809.3^{b}	1275.8^{b}	995^{b}	4624^{c}	10069^{d}	1102.2^{b}	4647.5^{c}	11676.8^{d}	
NC	σ_ρ	124.5	111.6	201.21	558	1221	1467	842.5	1428.1	2133.3	
	CV_ρ	27.98	13.79	15.77	56.08	26.41	14.57	76.44	30.73	18.27	31.1
	k/Q		11/2			8/3			7/3		3.3
$\beta = 0.5$	n_ρ	31	37	22	13	70	79	108	96	28	
	$\bar{\rho}$	453.1^{a}	467.7^{a}	990.8^{b}	911.3^{b}	1153.9^{b}	5926.4^{c}	1031.9^{b}	1452.4^{b}	7498.8^{d}	
	σ_ρ	120.3	143.9	236.0	568.0	610.4	2421.3	559.1	1078.7	3062.9	
	CV_ρ	26.55	30.77	23.82	62.33	52.90	40.86	54.18	74.27	40.85	45.2
	k/Q		9/2			5/2			6/2		3.3
$\beta = 1$	n_ρ	33	38	19	76	72	14	76	113	43	
	$\bar{\rho}$	469.3^{a}	483.1^{a}	1016.2^{b}	901^{b}	1186^{b}	6122^{c}	987.3^{b}	1104.6^{b}	6069.6^{c}	
C	σ_ρ	169.91	131.5	245.7	568	655	2386	582.1	653.2	3162.3	
	CV_ρ	36.20	27.22	24.18	63.04	55.23	38.97	58.96	59.13	52.10	46.1
	k/Q		8/2			4/2			5/2		2.8
$\beta = 2$	n_ρ	23	36	31	71	15	76	68	87	77	
	$\bar{\rho}$	463.2^{a}	587.7^{a}	$703.6a^{b}$	851.6^{b}	1249.2^{c}	6316.7^{e}	$934.3b^{c}$	1065.8^{c}	3955.8^{d}	
	σ_ρ	213.9	298.8	281.7	499.8	721.7	2365.4	570.4	880.3	3320.1	
	CV_ρ	46.18	50.84	40.04	58.69	57.77	37.45	61.05	82.60	83.93	53.0
	k/Q		2/2			5/3			4/3		1.4
W_{ij}	n_ρ'	40	38	12	108	42	12	183	35	14	
	$\bar{\rho}'$	$404\ 2^{a}$	652.7^{b}	891.2^{c}	820.4^{c}	1934.4^{e}	4570.2^{f}	1035.4^{d}	4153.0^{f}	7765.9^{g}	
	σ_ρ'	83.3	54.6	120.3	260.7	461.9	784.2	482	938.2	1540.7	
	CV_ρ'	20.61	8.37	13.50	31.78	23.88	17.16	46.55	22.59	19.84	22.7
C′	k/Q		10/3			5/3			6/2		2.6
g_{ij}	n_ρ'	38	41	11	95	55	12	183	36	13	
	$\bar{\rho}'$	374.8^{a}	649.5^{b}	1116.4^{c}	746.2^{b}	1698.6^{d}	6271.2^{f}	1012.9^{c}	4226.2^{e}	9488.9^{g}	
	σ_ρ'	106.1	158.6	253.6	348.5	905.4	2639.9	561.3	1710.1	3270.8	
	CV_ρ'	28.31	24.42	22.72	46.70	53.30	42.10	55.42	40.46	34.47	38.7
	k/Q		7/3			16/3			14/3		2.7

Notes: NC: non-contiguous procedure, C: contiguous classification with spatial variables, C′:contiguous classification with aggregation formula, ρ: population density, n_ρ: number of zone members for the classification of attribute ρ, $\bar{\rho}$: the mean of the zone of attribute ρ, σ_ρ: standard deviation of the group for attribute ρ, L: lowest group in the year, M: medium group in the year, H: highest group in the year, k: number of contiguous (unfragmented) aggregated spatial units (desas), Q: number of zones resulted from clustering prosedure which significantly different ($\alpha = 0.05$), a, b, ..., g: the means with similar marks are not significantly different according to LSD test, β: contiguity weight, W_{ij}: contiguity constraints for spatial association is based on zero/one adjacency, g_{ij}: contiguity constraints for spatial association is based on new association formulated in equation (15), CV: coefficient of variation

Table 3. Statistical parameters of the groups resulting from univariate spatial classification for urban land ratio.

procedure		Statistical parameter	Zones 1969/70 L	M	H	1981/82 L	M	H	1993 L	M	H	Average
NC		n_ρ	50	38	2	105	44	13	167	58	7	
		$\bar{\rho}$	0.054^{a}	0.217^{c}	0.510^{e}	0.089^{b}	0.281^{d}	0.626^{e}	0.087^{b}	0.311^{d}	0.762^{f}	
		σ_ρ	0.041	0.059	0.160	0.047	0.064	0.115	0.048	0.096	0.297	
		CV_ρ	75.93	27.19	31.37	52.81	22.78	18.37	55.17	30.87	38.48	39.3
		k/Q		9/3			22/3			18/3		5.4
C	$\beta = 0.5$	n_ρ	27	48	15	73	57	32	168	57	7	
		$\bar{\rho}$	0.044^{a}	0.127^{bc}	0.313^{d}	0.107^{b}	0.133^{c}	0.456^{e}	0.094^{b}	0.295^{d}	0.762^{f}	
		σ_ρ	0.051	0.059	0.099	0.067	0.079	0.163	0.061	0.114	0.163	
		CV_ρ	115.91	46.46	10.00	62.62	59.40	35.75	64.90	38.64	21.39	50.6
		k/Q		7/3			11/3			19/3		4.1
	$\beta = 1$	n_ρ	27	41	22	68	42	52	83	111	38	
		$\bar{\rho}$	0.050^{a}	0.113^{b}	0.273^{c}	0.099^{b}	0.106^{b}	0.359^{d}	0.105^{b}	0.110^{b}	0.447^{e}	
		σ_ρ	0.063	0.060	0.010	0.065	0.071	0.180	0.069	0.072	0.184	
		CV_ρ	126	53.10	3.66	65.66	66.98	50.14	63.90	65.45	41.16	59.6
		k/Q		5/3			4/2			6/2		1.7
	$\beta = 2$	n_ρ	25	40	25	47	56	59	72	89	71	
		$\bar{\rho}$	0.047^{a}	0.126^{c}	0.230^{d}	0.127^{c}	0.192^{c}	0.225^{cd}	0.100^{b}	0.116^{c}	0.292^{e}	
		σ_ρ	0.064	0.073	0.124	0.124	0.193	0.167	0.066	0.086	0.216	
		CV_ρ	136.17	57.94	53.91	57.64	100.52	74.22	66.67	74.14	73.97	27.2
		k/Q		7/3			5/3			12/3		2.7
C'	W_{ij}	n_ρ'	39	40	11	107	39	16	124	74	34	
		$\bar{\rho}'$	0.048^{a}	0.156^{d}	0.312^{f}	0.104^{c}	0.259^{e}	0.433^{h}	0.075^{b}	0.188^{e}	0.385^{g}	
		σ_ρ'	0.033	0.035	0.050	0.042	0.042	0.055	0.029	0.042	0.076	
		CV_ρ'	68.75	22.44	16.03	40.38	16.22	12.70	38.67	22.34	19.74	28.6
		k/Q		7/3			5/3			12/3		2.7
	g_{ij}	n_ρ'	32	36	22	84	52	26	78	34	120	
		$\bar{\rho}'$	0.033^{a}	0.136^{c}	0.275^{e}	0.080^{b}	0.205^{d}	0.480^{f}	0.180^{d}	0.470^{f}	0.680^{g}	
		σ_ρ'	0.030	0.054	0.099	0.048	0.074	0.175	0.060	0.180	0.380	
		CV_ρ'	90.91	39.71	36.00	60.00	36.10	36.46	33.33	38.30	55.88	51.7
		k/Q		7/3			16/3			14/3		4.1

Notes: NC: non-contiguous procedure, C: contiguous classification with spatial variables, C': contiguous classification with aggregation formula, μ: urban land ratio, n_μ: number of zone members for the classification of attribute μ, $\bar{\mu}$: the mean of the zone of attribute μ, σ_μ: standard deviation of the group for attribute μ, L: lowest group in the ear, M: medium group in the ear, H: highest group in the ear, k: number of contiguous (unfragmented) aggregated spatial units, Q: number of zones resulted from clustering procedure, which significantl different ($\alpha = 0.05$), a, b, ..., g: the means with similar marks are not significantl different, β: contiguit weight, W_{ij}: contiguit constraints for spatial association is based on zero/one adjacenc, g_{ij}: contiguit constraints for spatial association is based on a new association formulated in equation ( ), CV: coefficient of variation

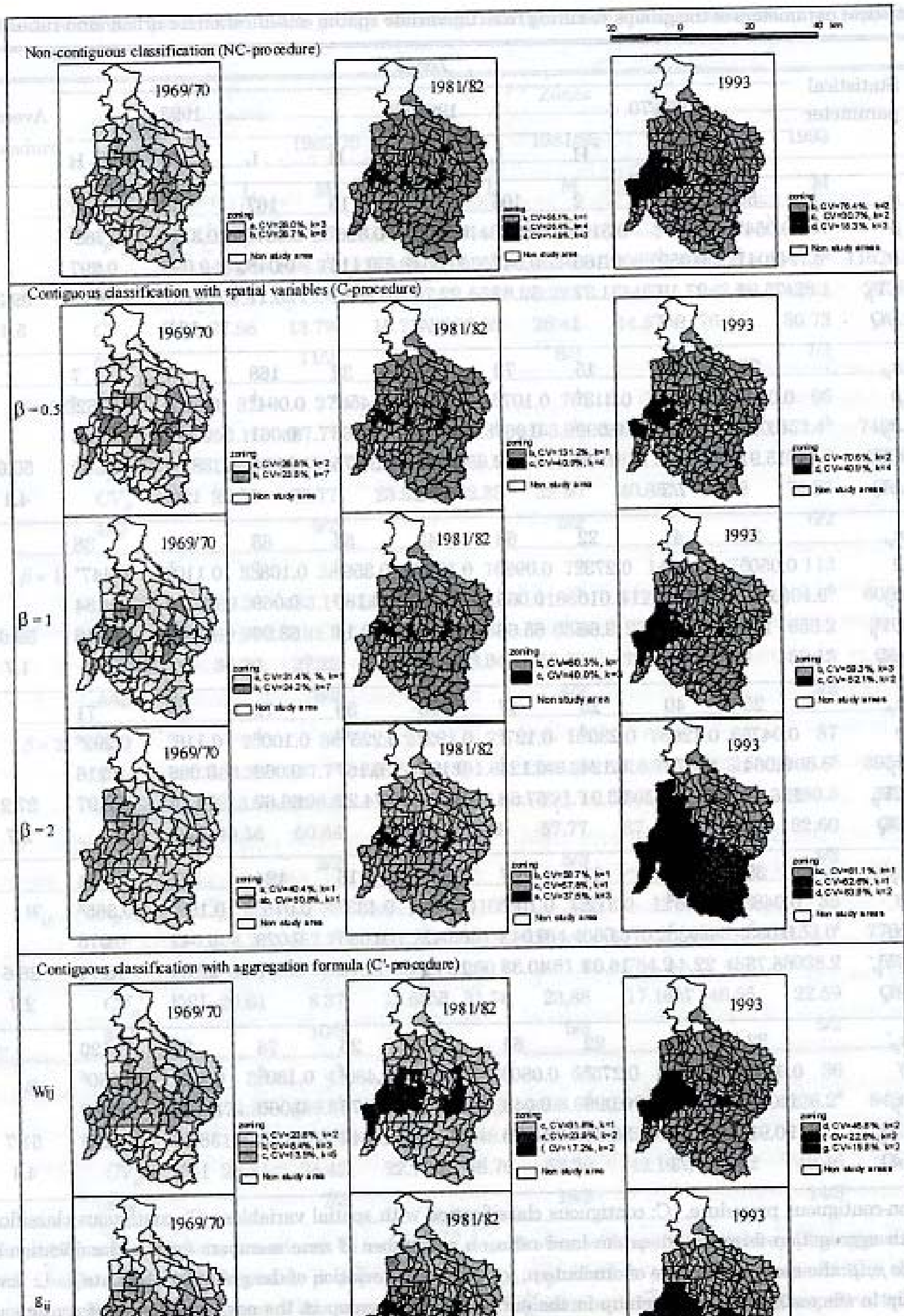

Fig. 3. Non-contiguous and contiguous classification of population density, using cluster analysis.

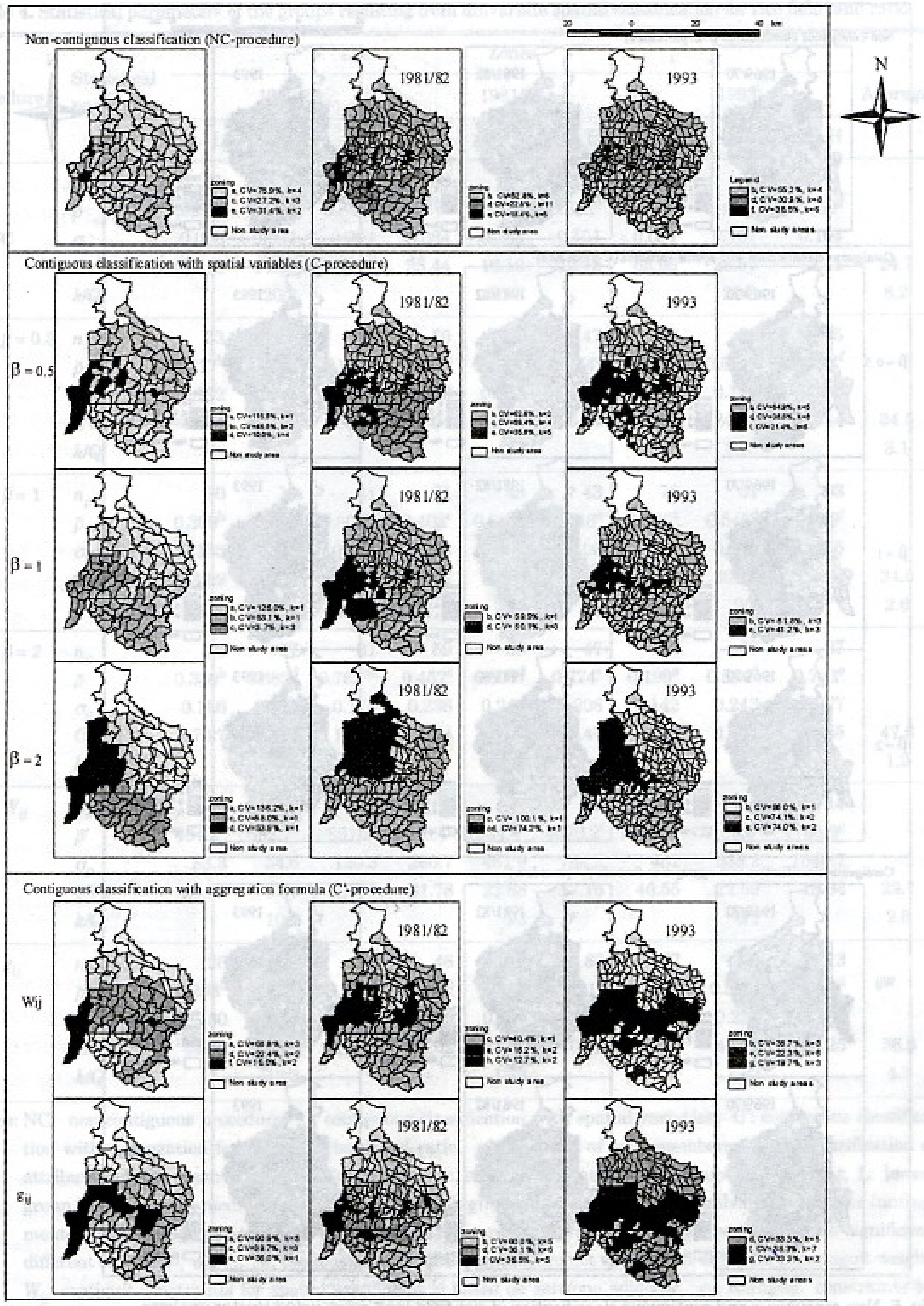

Fig. 4. Non-contiguous and contiguous classification of urban land ratio, using cluster analysis.

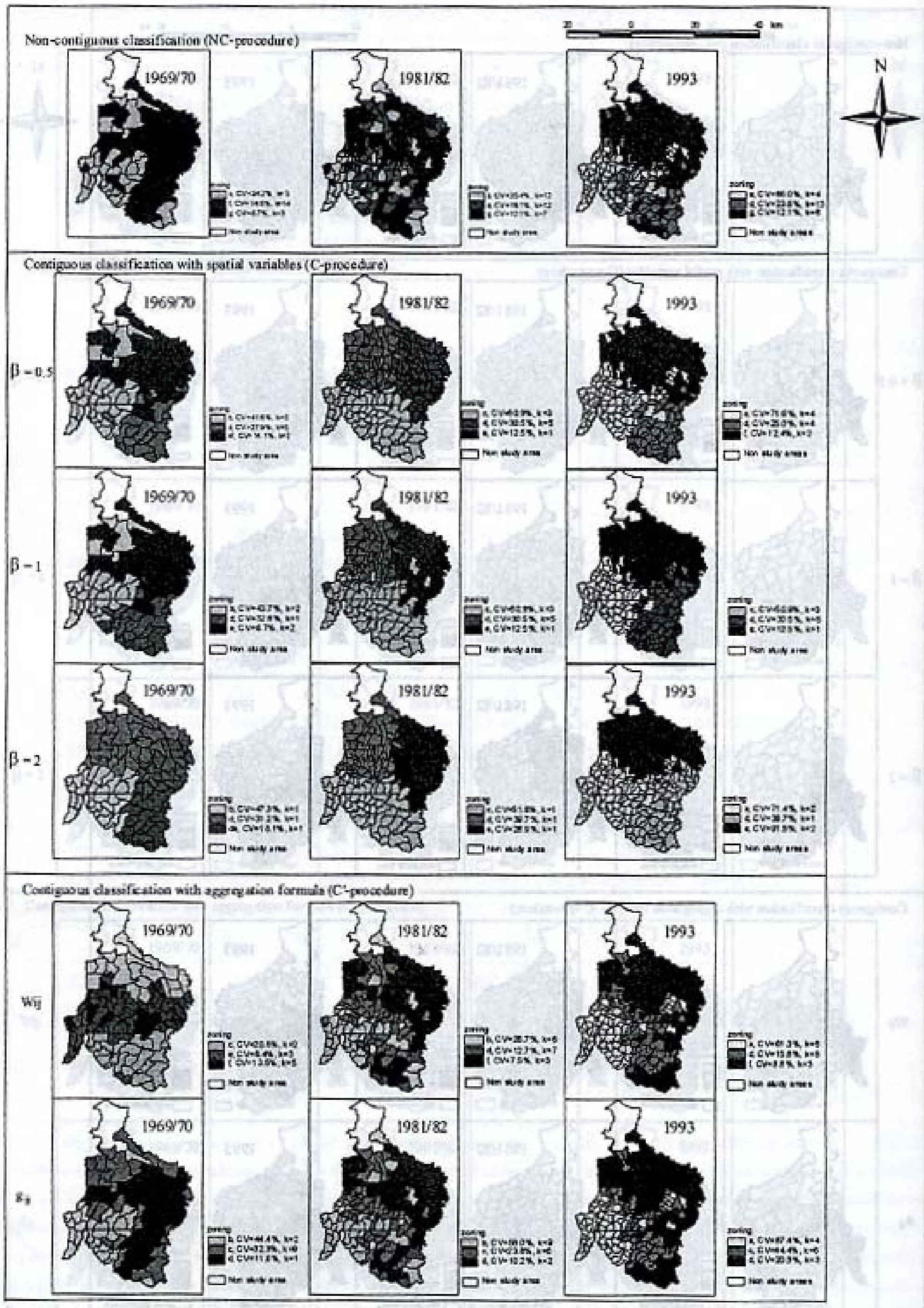

Fig. 5. Non-contiguous and contiguous classification of rice field land ratio, using cluster analysis.

Table 4. Statistical parameters of the groups resulting from univariate spatial classification for rice field land ratio.

procedure	Statistical parameter	Zones									Average
		1969/70			1981/82			1993			
		L	M	H	L	M	H	L	M	H	
NC	n_ρ	31	20	39	39	46	77	63	58	111	
	$\bar{\rho}$	0.287^{c}	0.598^{f}	0.849^{g}	0.237^{b}	0.516^{e}	0.833^{f}	0.123^{a}	0.454^{d}	0.859^{g}	
	σ_ρ	0.098	0.087	0.074	0.084	0.083	0.101	0.081	0.107	0.104	
	CV_ρ	34.15	14.55	8.72	35.44	16.10	12.12	65.95	23.57	12.11	24.7
	k/Q		20/3			31/3			23/3		8.2
C, $\beta = 0.5$	n_ρ	33	18	39	59	56	47	89	98	45	
	$\bar{\rho}$	0.317^{b}	0.652^{d}	0.814^{e}	0.457^{c}	0.607^{d}	0.774^{e}	0.204^{a}	0.601^{d}	0.868^{ef}	
	σ_ρ	0.132	0.182	0.115	0.236	0.241	0.208	0.146	0.150	0.108	
	CV_ρ	41.64	27.91	14.13	51.64	39.70	26.87	71.57	24.96	12.44	34.5
	k/Q		9/3			9/3			10/3		3.1
C, $\beta = 1$	n_ρ	30	19	41	71	48	43	78	61	93	
	$\bar{\rho}$	0.309^{b}	0.616^{d}	0.804^{ef}	0.403^{c}	0.669^{d}	0.848^{e}	0.196^{a}	0.545^{d}	0.869^{f}	
	σ_ρ	0.135	0.201	0.121	0.205	0.204	0.106	0.147	0.218	0.115	
	CV_ρ	43.69	32.63	15.05	50.87	30.49	12.5	75.00	40.00	13.23	34.8
	k/Q		5/3			9/3			9/3		2.6
C, $\beta = 2$	n_ρ	29	27	31	59	56	47	77	68	87	
	$\bar{\rho}$	0.330^{b}	0.682^{d}	0.762^{de}	0.457^{c}	0.607^{d}	0.774^{e}	0.199^{a}	0.584^{d}	0.762^{e}	
	σ_ρ	0.156	0.213	0.178	0.236	0.241	0.208	0.142	0.243	0.277	
	CV_ρ	47.27	31.23	18.11	51.64	39.70	91.47	71.36	41.61	36.35	47.6
	k/Q		3/3			3/3			5/3		1.2
C′, W_{ij}	n_ρ'	40	38	12	108	42	12	183	35	14	
	$\bar{\rho}'$	404.2^{a}	652.7^{b}	891.2^{c}	820.4^{c}	1934.4^{e}	4570.2^{f}	1035.4^{d}	4153.0^{f}	7765.9^{g}	
	σ_ρ'	83.3	54.6	120.3	260.7	461.9	784.2	482	938.2	1540.7	
	CV_ρ'	20.61	8.37	13.50	31.78	23.88	17.16	46.55	22.59	19.84	22.7
	k/Q		10/3			5/3			6/2		2.6
C′, g_{ij}	n_ρ'	26	26	38	46	53	63	67	52	113	
	$\bar{\rho}'$	0.293^{b}	0.569^{c}	0.830^{d}	0.293^{b}	0.577^{c}	0.851^{d}	0.166^{a}	0.505^{c}	0.815^{d}	
	σ_ρ'	0.130	0.184	0.096	0.170	0.136	0.087	0.145	0.224	0.165	
	CV_ρ'	44.37	32.34	11.57	58.02	23.57	10.22	87.35	44.36	20.25	36.8
	k/Q		12/3			17/3			13/3		4.7

Notes: NC: non-contiguous procedure, C: contiguous classification with spatial variables, C′: contiguous classification with aggregation formula, τ: urban land ratio, n_τ: number of zone members for the classification of attribute τ, $\bar{\tau}$: the mean of the zone of attribute τ, σ_τ: standard deviation of the group for attribute τ, L: lowest group in the ear, M: medium group in the ear, H: highest group in the ear, k: number of contiguous (unfragmented) aggregated spatial units, Q: number of zones resulted from clustering procedure, which significantl different ($\alpha = 0.05$), a, b, ..., g: the means with similar marks are not significantl different, β: contiguit weight, W_{ij}: contiguit constraints for spatial association is based on zero/one adjacenc, g_{ij}: contiguit constraints for spatial association is based on a new association formulated in equation ( ), CV: coefficient of variation

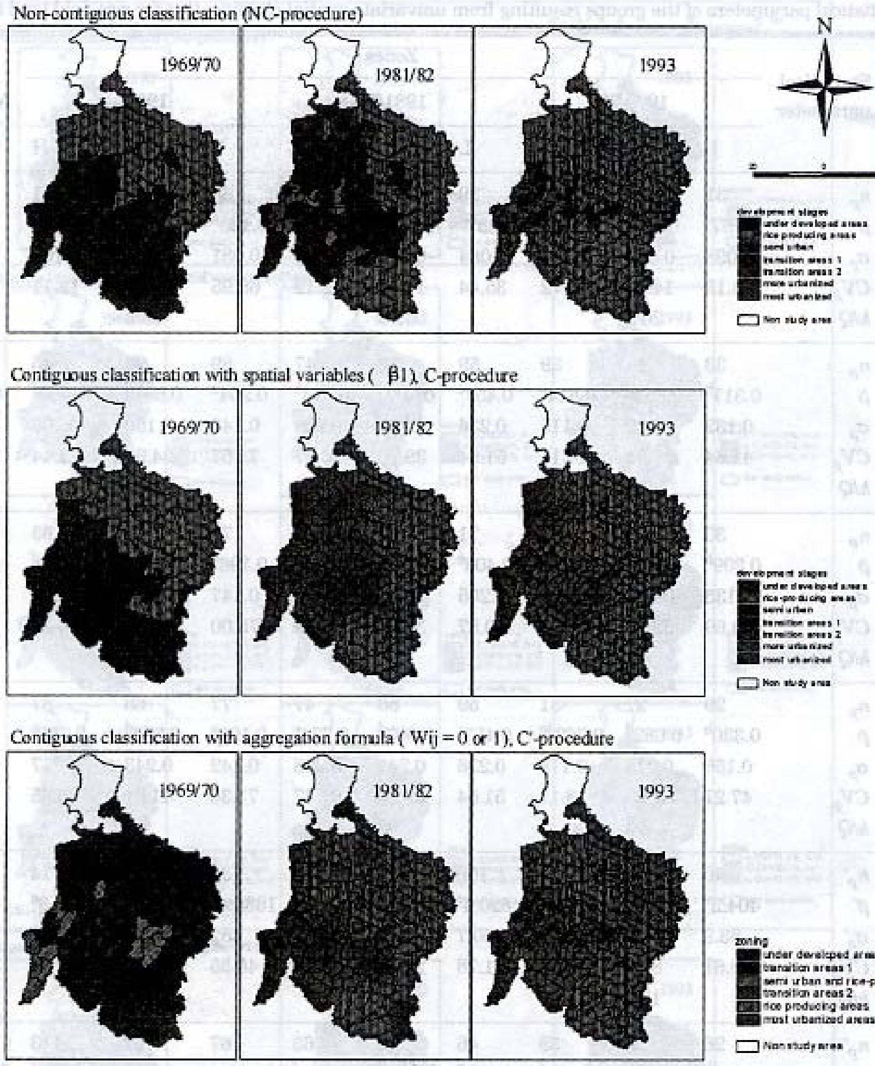

Fig. 6. Multivariate classification

development stages.

By using cluster analysis procedure, for every data set of single attribute variables, the region is splitted into three subregions (zones). For the case of contiguous classification, statistical Less Significant Different (LSD) tests show that for some attributes, the differences among the means of the zone's are not significant (Table 2). The visualization of these results is showed in Fig. 3a, 3b and 3c. The maps in these figures only show zoning which significant differences.

The contiguous classification procedure in all cases of classification has figured out clearer and simpler zoning result. Furthermore, this method could not guaranty to make a perfect unfragmented zoning without any enclaves. The level of fragmentation more depends on data condition, and the weight β (for cluster analysis using spatial variables procedure) and

W_{ij} (for the case of aggregation formula). A negative spatial autocorrelated data set will be fragmented and more difficult to be simplified in only several zones. So far, this method could contrast the spatial aggregation in attempt to summarize the general spatial pattern.

According to the univariate population cluster zoning system results, the pattern of population distribution has changed from separated less-dense population clusters to be two highly concentrated population clusters. One of the clusters is a wide population concentration at the border area adjacent to Jakarta City, and the other one is a small population concentration located far from the border and the most urbanized areas distributed along the main road from west to east (Fig. 3). Urban land use area changed to be farther from Jakarta City (Fig. 4). The regionalization of rice field areas have become clearer which become more concentrated in the northern part of Bekasi District (Fig. 5). The zoning patterns of two proceeding contiguous classifications show significant result.

As shown in Table 2 and 3, compared to the other procedures, the contiguous classification using aggregation formula (C'') tend to form zoning system with small variances among the members (small coefficient of variation). This consequences on resulting highly statistical significant differences among zones.

Multivariate classification needs more complex interpretation. For the case of Bekasi District, it seems that classifying the development of suburbanization to be three zones is too simplify the real phenomenon (Fig. 6). By using multivariate classification, it become clear that the most urbanized areas in the study area are developed originally from the older local urban centers, where the rice production centers were also located. Most of the current rice-producing areas are relatively new production areas.

V. Conclusions

This study described the classification methods applied to spatial data, especially the spatial distribution data of population and land uses. The variation of population distribution, urban land uses and development stages in suburbs area can be spatially visualized by non-contiguous and contiguous spatial classification methods. The contiguous spatial classification methods are able to describe and measure spatial pattern quantitatively. These quantitative analysis were able to show un ambiguous description of changes on spatial pattern of land use and population distribution.

References

Anselin, L., 1995, "Local Indicators of Spatial Association-LISA," *Geographical analysis*, Vol. 27, No. 2, pp 93-115

Getis, A., 1995, Spatial dependence and heterogeneity and proximal data bases, In Spatial Analysis and GIS, Fotheringham, S. and P. Rogerson (Eds), pp. 105-120.

Getis, A and J.K. Ord, 1992, "The analysis of spatial association by use of distance statistics," *Geographical Analysis*, Vol. 24, No. 3, pp. 189-205.

Johnston, R.J., 1976, Classification in geography, CATMOG 6, (Geobooks, Norwich)

Nass C, and D. Garfinkle, 1992, "Localized autocorrelation diagnostic statistic (LADS) for spatial models," *Regional Science and Urban Economics*, Vol. 22, pp. 333-346.

Odland, 1988, Spatial autocorrelation, Scientific geography series, Sage Pub. p. 87.

Openshaw, S., 1977, The modifiable areal unit problem, CATMOG 38, (Geobooks, Norwich)

Openshaw, S., 1996, "Developing GIS-relevant zone-based spatial analysis method," in Longley P. and M. Batty (Eds), Spatial Analysis: Modelling in A GIS Environment, Geoinformation International, pp. 55-73.

Rustiadi, E. Mizuno. K. and S. Kobayashi, 1998, "Measuring spatial pattern of suburbanization process," *Proceeding of annual Rural Planning Association Seminar, Tokyo, April 3-4, 1998*, pp. 33-36.

Rustiadi, E. and T. Kitamura, 1988, "Analysis of land use changes in city suburbs," *Journal of Rural Planning Association*, Vol. 17, No. 1, pp. 20-31.

Ward, J.H. Jr, 1963, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, Vol. 58, pp. 236-244.

DeMers, M.N., 1997, Fundamentals of Geographical Information Systems, p 248, 463.