CO-360

Analyzing the effects of scale and land use pattern metrics on land use database generalization indices

LIU Y., JIAO L., LIU Y.
Wuhan University, WUHAN CITY, CHINA

ABSTRACT

The generalization index system is one of the critical issues for computer-aided land use database generalization. This paper studies the scale and land use pattern effects on land use database generalization indices and estimates the thresholds of these indices based on a typical land use database sample. The index system of land use database generalization is discussed and constructed from macro and micro perspectives. Six land use pattern metrics, namely, land use diversity index, land use dominance index, land use homogeneity index, land use fragmentation index, the index of land use type dominance, and the index of land use type fragmentation, are designed to characterize land use patterns and are introduced into the analysis of land use pattern effect on land use database indices. The analysis framework of the scale and land use pattern effects on the land use database indices are proposed by employing statistical techniques. Based on the land use database samples at multiple spatial scales collected in various land use regions across China, the study generates rules for both scale and land use pattern effects on the indices, including map area proportion of land use types, total map load, parcel map load, and minimum parcel area. The thresholds of these indices in land use database generalization are produced at the scales of 1:50,000, 1:100,000, 1:250,000, and 1:500,000. An experimental generalization at county level demonstrates how to determine the generalization index values considering scale and land use pattern, and how to evaluate the generalization results using our macro indices.

KEYWORDS

Land use database; Landscape pattern; Indices of land use database generalization; Land use database generalization

1. INTRODUCTION

In the generalization of land use database, there are some parameters or thresholds that need to be set properly to control the generalization operations, such as minimum parcel area in parcels generalization, minimum bend diameter in line simplification, etc. These parameters and thresholds are termed as land use data generalization indices. The indices dominate the level of detail of data, and contribute fundamentally to the generalization results and their quality. They represent an important aspect of land use database generalization knowledge (Liu, 2005; Gao et al., 2004), and they are critical for generating multi-scale land use maps and databases (Liu, 2002). A part of the difficulties of land use data generalization are due to the fact that the generalization problem is not specified formally enough (Van Oosterom, 2009), and generalization process often includes subjective aspects such as taste, resulting into artistic solutions. Constructing the generalization indices will help to formalizing the land use data generalization.

Land use database generalization indices were usually set to experiential values in previous generalization practice. In this paper we discuss the effects of map scale and land use pattern on the indices of land use database generalization, and try to determine the values of the indices more rationally by formulating the relationship between the indices and scale and land use pattern metrics.

Many studies on spatial data generalization have focused on general indices for terrain mapping (Butterfield and McMaster, 1991; Muller and Wang, 1992; Oxenstierna, 1997; Lee, 2001; Qi and Jiang, 2001); however, these researches seldom consider indicators for land use database generalization. Some related studies emphasize the operations of graphic generalization of land use data (Jaakkola, 1997; Harma et al., 2004), but only a few of them take semantic constraints into consideration (Ai and Liu, 2002; Gao et al., 2004). Liu (2002) and Liu et al. (2003) proposed a framework for land use database generalization based on models and rules, and provided some basic criteria, such as the maintenance of area proportion of land use types. Ai and Wu (2000) and Ai et al. (2001, 2002) studied the operators of parcel generalization, and discussed parcel merging based on neighborhood analysis. Su et al. (2007) proposed a new skeleton algorithm to assist the generalization of cultivated land polygons in hilly areas. Jaakkola (1997) and Harma et al. (2004) presented raster-based methods of land use database generalization with a given resolution threshold. Martinez-Casasnovas (2000) used a class hierarchy structure for mapping unit definition at different scales in land use mapping and generalization from remotely sensed data. Cheng et al. (2008) introduced a multi-way GAP-tree and its application in land use database generalization. Haunert et al.

(2009) employed a constrained set-up of the tGAP structure to progressively generalize land-cover data. These methods facilitate the graphic generalization of a land use database, while semantic constraints in land use database generalization need to be studied further, such as the importance of land use types and their impact on generalization indices and maintenance of area proportion of land use types.

The micro generalization indices, such as minimum parcel area, were involved in generalization knowledge in several studies, while the macro indices need to be constructed and investigated further. Gao et al. (2004) derived the knowledge base for land use data generalization in the form of production rules, including general knowledge and thematic knowledge. Li et al. (2007) integrated MAPL-tree and thematic knowledge rules, for example, minimum parcel area, in land use data generalization. Liu (2005) conducted a practical study by establishing land use database generalization rules, including minimum parcel area, minimum bend diameter, minimum distance between parcels, and semantic rules. Macro indices, such as land use area proportion change and suitable map load, are also needed to evaluate the generalized results and to calibrate generalization processes.

Landscape metrics were widely used to describe the characteristics of landscape and characterize land-related progress (Rayburn and Schulte, 2009; Wang, et al., 2009a; Wang, et al., 2009b), but were rarely utilized in land use data generalization. Gallant (2009) discussed the nature of landscape variables derived from various land characteristics data sources in wildlife applications. Zhang et al. (2009) employed a landscape pattern index to analyze the map information of different scales and then to determine the appropriate scale to meet the needs of land use data generalization. Several studies on land use data generalization that considered the generalization indicators, such as minimum parcel area in the land use map, were limited to local areas and did not take the regional landscape pattern into consideration in the determination of land use data generalization indices (Liu, 2005; Chen, 2005; Zhang, 2006).

China is conducting its second nationwide land investigation, producing land use maps at county level (1:10,000), which are then generalized into a series of land use maps and databases at smaller spatial scales. These smaller spatial scales range from 1:50,000 to 1:500,000. However, there are not nationwide criteria for land use data generalization.

Previous literature suffers from two major drawbacks. Firstly, the micro indices are emphasized while the macro indices are often overlooked in land use database generalization. The second problem is general ignorance of the land use pattern effect on land use database generalization indices, while the indices for multi-scale land use database generalization over a large area, such as a nation, need to be established, considering both scale and land use pattern effects. This study will examine the relationship between land use database indices and scale, and then formulate the relationship between the indices and land use pattern metrics to identify the influence of land use pattern on the indices. In the remainder of this paper, we first develop the index system of multi-scale land use database generalization from macro and micro aspects and propose the land use pattern metrics. We next estimate the scale and land use pattern effects on the land use data generalization indices based on typical sample datasets. This is followed by an experiment of county-level land use database generalization. We close with brief concluding remarks and discussion.

2. DATA AND METHODOLOGY

2.1 Data sampling and preprocessing

The data source of the study is a multi-scale land use dataset of China. The Chinese land use database originates from countrywide land use investigations and the regular renewal of land use data. Land use investigation and renewal are implemented by incorporating remote sensing image interpretation and field surveys at county level. The original datasets (usually at 1:10,000 scale) are then generalized to a series of datasets at smaller scales, including 1:50,000, 1:100,000, 1:250,000, 1:500,000, etc. In particular mapping areas, some other scales are also employed for convenience, such as 1:75,000, 1:150,000, 1:200,000, etc.

Data samples were processed from our whole dataset. The nationwide land use data samples should be representative in terms of both land use pattern and spatial scale. Li (2000) zoned land resources in China into 12 land use regions, which can be further divided according to different geomorphologic properties (Resource Zoning Committee of Chinese Academy of Science, 1959). A land use zoning schema considering land use characteristics and landscape patterns was generated (see Table 1).

Table 1. Land use regions and samples.

Land use region	Sub-region of land use	Number of samples
Middle China	The Plain of Middle and Lower Reaches of Changjiang River	7
	Hilly area in Middle China	6
	Mountainous area in Middle China	5
North China	Hilly and mountainous area in North China	2
	Plain area in North China	1
Northeast China	Mountainous area in Northeast China	3
	Plain area in Northeast China	1
Northwest China	North Sinkiang	2
	South Sinkiang	1
	Loess Plateau	3
	Grass land in Inner-Mongolia	2
Southwest China	Yungui Plateau	6
	Chuanshan Basin	3
	Qingzang Plateau	1
Southeast China	Hilly area in Southeast China	6
	Southeast Plain	2

The samples were derived under two principles as follows: a) The number of samples in each land use region is positively correlative with the diversity of land use pattern in the region, i.e., more diverse of the region, more samples; and b) Each geomorphologic zone in a land use region should have at least one sample. Finally, we selected 51 counties (see Table 1) as sample areas, and multi-scale land use data of these counties were collected. Our sample counties, which are distributed in different land use regions across China, are shown in Figure 1.

Figure 1. Spatial distribution of land use data sample.

Data preprocessing was implemented since the samples were stored in different formats with different geo-referencing systems. The data preprocessing includes verification of the database, transformation of data formats, transformation of coordinate systems, and normalization of the coding system for land use types.

2.2 Index system of land use database generalization

Previous studies usually employed the indicators for land use data generalization at micro level, such as minimum parcel area, and minimum distance between parcels. Nevertheless, we wished to incorporate the indices at macro level, such as area proportion of land use types and suitable map load. These macro indices can be used to control generalization operations and to evaluate the generalization result. The index system of land use database generalization was built from both macro and micro perspectives (see Table 2).

Macro indices for land use database generalization include map load, area proportion of different land use types, and semantic characteristics. Map load describes the amount of map content from a macro perspective. There are at least three kinds of map loads in land use maps: maximum map load, optimum map load, and features map load. The area proportion of land use types serves as an important threshold in land use database generalization because the generalization of parcels will lead to changes in area proportion. The characteristics of spatial distribution of land use types also needs to be maintained since land use maps are primarily used to express the spatial pattern of land use types. Additionally, semantic characteristics or a hierarchy of land use classes are used to evaluate the semantic distance between adjacent polygons that is important for parcel merging control.

Micro indices for land use database generalization include minimum parcel area, minimum distance between parcels, and minimum bend diameter. Minimum parcel area reflects the importance of land use types and land use patterns; hence different land use types can have different area thresholds at the same spatial scale.

Table 2. Index system of land use data generalization.

Types	Thresholds	Description
Macro indices	Hierarchy of land use types	A hierarchy of classification system of land use according to socio-economic purposes.
	Map load	Maximum map load, optimum map load, features map load (subject features and basic features).
	Area proportion of land use types	Percentage of each land type in the case study area.
	Spatial contrast of distribution	Density of scattered parcels for some land use type, land use pattern index.
Micro indices	Minimum parcel area	Minimum parcel area for each land use type.
	Minimum distance between parcels	Distance between closest island parcels.
	Minimum bend diameter	Diameter of the minimum bend of parcel boundary.

Only a portion of the indices in Table 2 will be discussed further in this paper. The hierarchy of land use types is usually consistent during generalization except for conditions in which one or more land use types occupy very small percentages. The spatial contrast of distribution should be maintained during generalization. Thus for macro indices, we will focus on the other two indices, i.e., map load and area proportion of land use types. For micro indices, minimum distance between parcels and minimum bend diameter are mainly determined by human visual discrimination ability, whereas minimum parcel area changes with scale and landscape pattern. The scale effect and landscape effect on minimum parcel area are discussed in this paper.

2.3 Method of scale effect analysis

Map scale and land use pattern exert influence on indices of land use database generalization at different strengths. Map scale has a determinant influence on the value of generalization indices; thus maps with different scales of the same area will have significantly different index values. Land use pattern has a relatively smaller impact than map scale, and its effects can be perceived in a large region with various land use patterns.

Statistics charts and tables are used to demonstrate the scale effect on land use area proportion change, total map load, and parcels map load. The total map load is calculated as the average of the total map area of symbols (linear symbols and markers) and annotation in per unit map area. The parcels map load is the average of the map area of parcel boundaries in per unit map area. We employ exploratory statistics and non-linear regression models to evaluate the relationship between minimum parcel area and map scale. For convenience, map scale was replaced by the scale's denominator. A logarithm function was chosen as a non-linear regression function type according to the dataset.

2.4 Method of land use pattern effect analysis

Patch-based indices are developed to quantify landscape characteristics in landscape ecology, such as the index of landscape diversity, index of landscape dominance, index of landscape homogeneity, and index of landscape fragmentation. Comparing land use parcels with ecology patches, we can define land use pattern indices. In addition to the aforementioned indices employed in landscape ecology, we proposed two new indices to describe the characteristics of a specific land use type: dominance index of land use type and fragmentation index of land use type.

(1) Index of land use diversity (*H*)

This index is modified from landscape diversity index (Wang, 2003), and describes the degree of diversity of land use types,

[redacted] (1)

where H is the index of land use diversity, m is the number of land use types, and p_k is the percentage of land use type k. If m=1, then H=0 , the minimum. H increases as m increases or proportional distribution of land use types increases.

(2) Index of land use dominance (*D*)

The index is modified from landscape dominance index (Wang, 2003). The index of land use dominance indicates to what extent the land use is dominated by one or two types. The index is defined as:

[redacted] (2)

where D refers to the index of land use dominance, H_{max} represents the maximum of the index of land use diversity, and p_k is the percentage of land use type k.

(3) Index of land use homogeneity (*E*)

The index is modified from landscape homogeneity index (Wang, 2003). The index of land use homogeneity describes the homogeneity of land use types in a land use pattern, which is given by:

[redacted] (3)

(4) Index of land use fragmentation (*C*)

The index is modified from landscape fragmentation index (Wang, 2003). The index of land use fragmentation describes the degree of fragmentation of land use pattern,

[redacted] (4)

where N and A are the number of land use parcels and the total area (in map scale, mm^2) of the studied region, respectively.

(5) Dominance index of land use type(D_t)

This index measures how much the land use is dominated by one or several large parcels. Compared with land use dominance index that indicates the degree of land use is dominated by a few of land use types, the dominance index of land use type emphasizes the dominance level of several parcels in a land use type. The index is expressed as:

[redacted] (5)

where D_t is the dominance index of land use type t, n is the number of parcels belonging to land use type t, H_{tmax} is the maximum of the index of diversity of land use type t, and p_i is the area percentage of parcel i of land use type t.

(6) Fragmentation index of land use type (C_t)

This index captures the fragmentation degree of the distribution of a land use type,

[redacted] (6)

where N represents the number of land use parcels belonging to land use type t, and A_t is the total area (in map scale, mm^2) of the land use type t in the studied region.

We employed statistical methods to analyze the effects of the land use pattern metrics on the indices of land use database generalization based on our sample maps. We used correlation analysis to identify the land use pattern metrics that are correlated with the land use generalization indices. Then we implemented regression analysis to build the regression functions between the land use indices and the correlated land use pattern metrics.

3. RESULT AND ANALYSIS

3.1 Scale effect on indices of land use database generalization in China

（1）Scale effect of macro indices

There are four driving forces for changes of map area proportions of land use types in generalization. The first cause is collapse of parcels. For example, some polygonal residential areas are simplified as points, or polygonal roads or rivers are simplified as lines during generalization. The second cause is boundary simplification of parcels. The third cause is the reorganization of parcels, such as elimination or merging and the fourth is exaggeration of some important parcels. Table 3 shows the land use area proportions at different scales in the Middle and Lower Reaches of Changjiang River. Figure 2 illustrates the change in area proportion of land use types at different scales in this area.

Table 3. Area proportion of land use types in Middle and Lower Reaches of Changjiang River at different spatial scales.

Code	01	02	03	04	10	11	20
(1) Plain region							
10,000	56.22%	0.96%	6.87%	1.86%	0.71%	22.81%	22.94%
50,000	61.78%	0.88%	7.69%	1.67%	0.22%	17.71%	8.76%
100,000	63.21%	0.88%	8.03%	1.58%	0.04%	16.12%	6.93%
250,000	65.54%	0.89%	8.02%	1.47%	0.01%	15.19%	4.58%
500,000	70.96%	0.95%	8.37%	1.38%	0.00%	14.69%	2.80%
(2) Hilly and mountainous region							
10,000	49.97%	0.72%	24.66%	3.20%	0.23%	12.40%	17.97%
50,000	50.71%	0.66%	24.43%	2.90%	0.14%	11.08%	9.32%
100,000	53.40%	0.61%	25.73%	3.05%	0.01%	7.07%	5.05%
250,000	55.06%	0.65%	28.00%	3.07%	0.00%	6.52%	1.75%
500,000	56.04%	0.60%	31.06%	2.99%	0.00%	3.92%	1.13%

Code: 01 Cultivated land; 02 Orchard; 03 Forest land; 04 Grass land; 10 Transportation land; 11 Water bodies and related facilities; 20 cities, towns, villages and isolated industrial districts.

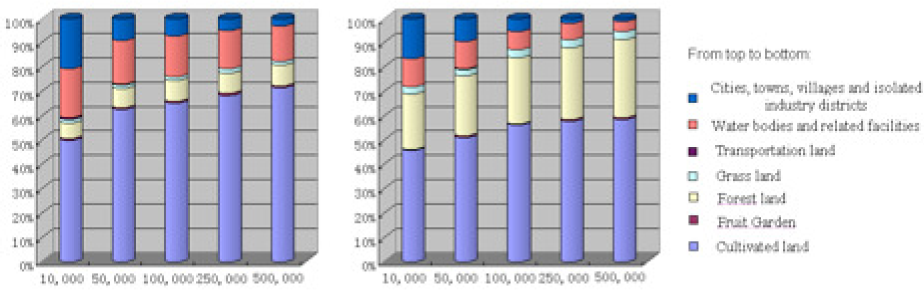

Figure 2. Multi-scale changes of area proportion of land use types in Middle and Lower Reaches of Changjiang River.

Figure 2 shows that the map area of cities, towns, villages, isolated industrial districts and the area of water bodies and water resource facilities decrease considerably in both plain regions and hilly and mountainous regions. The area of transportation land decreases as well. These decreases are caused mainly by polygon collapse. However, the decrease of the area of transportation land is not obvious in the figure, since the total area of transportation land is so small. On the contrary, the area of cultivated land in plain regions and the area of forest land in hilly regions increase. The reason is that these land use types serve as the background of generalization operations, especially polygon merging, polygon elimination, and polygon collapse. We estimate the degree of area proportion changing caused by the generalization operations for some land use types, including cultivated land, orchards, forest land, and grass land. The other land use types are excluded from this estimation since their area change is unavoidable and mainly caused by polygon collapse. The rate of area proportion change of a land use type is calculated by

[redacted] (7)

where r is the rate of area proportion change, p_0 and p_1 are the area proportions of the land use type in generalized map and original map, respectively. The maximum changing rate of area proportion of land use types between scales was calculated based on our nationwide samples, and the result is summarized in Table 4.

Table 4. Changing rate of area proportion of some land use types in generalization.

Scales before and after generalization	1:10,000~ 1:50,000	1:50,000~ 1:100,000	1:100,000~ 1:250,000	1:250,000~ 1:500,000
Maximum changing rate of area proportion of land use types	12~15%	5~7%	7~9%	9~11%

The area proportions of land use types change sharply when the land use database is generalized from 1:10,000 to 1:50,000 due to the large scale gap.

Map load is related to map scale and land use pattern. When the map scale decreases, the map load increases with the increase in map content density. Even at the same scale, map load is larger in the regions with many land use types and fragmentary land use distribution. Therefore map load is correlated with the land use fragmentation index at the same spatial scale. At the scale of 1:10000, the land use fragmentation index can be categorized into three classes: low level fragmentation (<0.3), medium level fragmentation (0.3~0.4), and high level fragmentation (>0.5). The scale effects on map load were analyzed in three subdivisions of land use fragmentation. The suitable total map load and parcels map load of land use maps are shown in Table 5 and Table 6, respectively. These ranges and averages can be used as benchmarks to control area proportion change and map load in the generalization of the land use database.

Table 5. Range of suitable total map load of land use maps.

Subdivisions	1:10,000	1:50,000	1:100,000	1:250,000	1:500,000
Low fragmentation	3%–7%	13%–17%	17%–21%	22%–26%	26%–30%
Medium fragmentation	5%–9%	15%–19%	20%–24%	26%–30%	31%–35%
High fragmentation	5%–9%	17%–21%	23%–27%	30%–34%	35%–39%

Table 6. Average of suitable map load of parcel features.

Subdivisions	1:10,000	1:50,000	1:100,000	1:250,000	1:500,000
Low fragmentation	1.100%	2.227%	2.712%	3.353%	3.838%
Medium fragmentation	1.200%	2.327%	2.812%	3.453%	3.938%
High fragmentation	1.700%	2.666%	3.082%	3.631%	4.047%

Total map load increases against map scale during generalization. The map load in fragmented areas is larger than in less fragmented areas at the same scale. The suitable total map loads increase greatly when the land use databases are generalized from 1:10,000 to 1: 50,000, and then increase gradually at the rate of about 4 percent in low fragmentation areas, about 5 percent in medium fragmentation areas, and about 6 percent in high fragmentation areas.

The suitable map load of parcel features obviously increases when the map scale changes from 1:10,000 to 1:50,000. The suitable parcels map loads apparently differ between the areas with different fragmentation at 1:10,000 and 1:50,000, and the average suitable map load of parcels increases by about 0.1 percent when the fragmentation changes from low to medium or from medium to high at scales of 1:10,000, 1:250,000, and 1:50,000.

（2）Scale effect of micro indices

There are four factors influencing the minimum parcel area in generalization. The first one is the precision required by the mapping purpose. The second is the map resolution determined by map scale. The third is the importance of land use types, while the fourth is the spatial pattern of land use. Regression analysis was employed to fit the relationship between minimum parcel area and map scale. Taking cultivated land as an example, a total of 114 samples were selected. The samples were tidied by eliminating 12 outliers with the two times standard deviation method. The scatter plot of the tidied samples is shown in Figure 3. A logarithm function is selected to fit the relationship between minimum parcel area and scale.

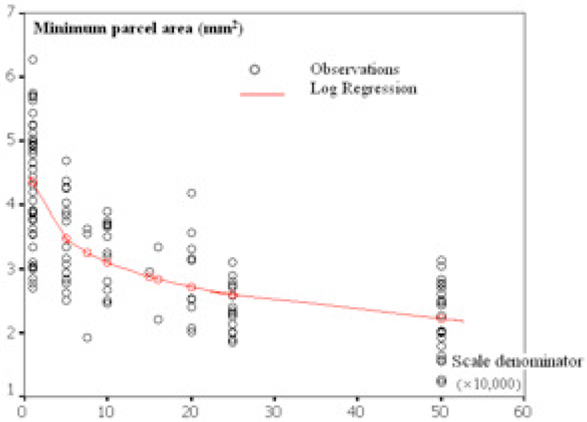

Figure 3. Regression analyses between minimum parcel area of cultivated land and map scale.

The regression function is:

$$Y=-0.544*\ln(X)+4.352 \quad (8)$$

where Y is the minimum parcel area of cultivated land, X is the denominator of map scale. The independent variable explains 57.6% of the variations of minimum parcel area (R2 = 0.576) and the F-ratio is 170.887. Moreover, the regression functions for other land use types were generated with the same routine. The model estimates and associated statistics are presented in Table 7. Figure 4 illustrates the changes of minimum parcel area against scale according to the results.

Table 7. Regression results between minimum parcel area and map scale.

Land use type	Regression function	*R*	*F*	*p*
Cultivated land	$Y=-0.544*\ln(X)+4.352$	0.759	170.887	.000
Orchard	$Y=-0.875*\ln(X)+6.300$	0.760	102.530	.000
Forest land	$Y=-1.240*\ln(X)+8.977$	0.800	159.526	.000
Grass land	$Y=-1.187*\ln(X)+8.860$	0.761	109.999	.000
Transportation land	$Y=-2.151*\ln(X)+7.571$	0.806	89.084	.000
Water bodies and related facilities	$Y=-1.626*\ln(X)+8.979$	0.766	140.662	.000
Others	$Y=-1.670*\ln(X)+9.860$	0.719	77.208	.000
Cities, towns, villages and isolated industrial and isolated industrial districts	$Y=-0.626*\ln(X)+4.035$ $Y=-0.626*\ln(X)+4.035$	0.802 0.802	176.121 176.121	.000 .000

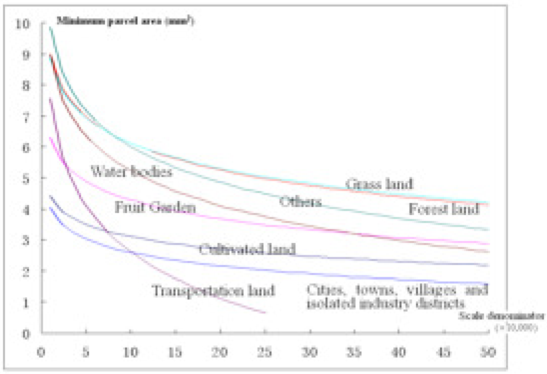

Figure 4. Changes of minimum parcel area of land use types against scale.

The minimum parcel area thresholds of grass land and forest land are larger than the thresholds of the other land use types at the same land use database. The ones of construction land (cities, towns, villages and isolated industrial districts) and cultivated land are the two smallest. The parcel area threshold of transportation land decreases sharply against map scale, and reaches almost zero at 1:250,000 since little transportation land can be represented as polygons at this scale.

3.2 Land use pattern effect on indices of land use database generalization

(1) Land use pattern effect of macro indices

The changes in area proportion of land use types is influenced by land use pattern. Regions with higher fragmentation values usually have larger changes in the area proportion. Thus in Table 4, we should employ the lower limit in regions with a lower fragmentation, and the upper limit in regions with a higher fragmentation. Map load in fragmented areas is larger than in ordinary areas, so it is influenced by the index of land use fragmentation (see Table 5).

(2) Land use pattern effect of micro indices

The minimum parcel area correlates not only with map scale but also with land use pattern. Correlation analysis was employed to explore the land use pattern metrics, which influence the minimum parcel area. Then regression analysis was used to quantify these influences. Minimum parcel area of cultivated land in the 1:50,000 map is used as an example to describe the analytical procedure, and the correlation analysis results are presented in Table 8.

Table 8. Correlation analysis between minimum parcel area of cultivated land and land use pattern indices (scale 1:50,000).

	Diversity index (H)	Dominance index (D)	Homogeneity index (E)	Fragmentation index (C)	Dominance index of land use type (D_i)	Fragmentation index of land use type (C_i)
Pearson coef.	-0.510	0.440	-0.487	-0.501	0.325	-0.799*
Sig.(2-tailed)	0.052	0.101	0.065	0.057	0.237	0.018

* Significant at the confidence level of 95%

Table 8 demonstrates that the fragmentation index of cultivated land is the only land use pattern index that is correlated significantly with the minimum parcel area of cultivated land at 1:50,000 scale. The fragmentation index of cultivated land is adopted as an explanatory variable in the regression analysis at the second stage. The fitted linear regression function takes the following form:

$$Y=-18.843*X+4.532 \quad (9)$$

where Y is the minimum parcel area threshold for cultivated land, and X is the fragmentation index of cultivated land. The correlation coefficient R2, F-ratio, and p-ratio are 0.638, 5.112, and 0.008, respectively. The model reveals that one percent change in the fragmentation index has a marginal effect of 0.19 mm2 on the minimum parcel area threshold. The range of the fragmentation index of cultivated land is between 0.25 and 5.00, and therefore the theoretical range of the minimum parcel area threshold for cultivated land is from 3.6 mm2 to 4.5 mm2.

By the same routine, we formulate the relationship between minimum parcel area and the land use type fragmentation at different scales for other land use types (see Table 9).

Table 9. Change of minimum parcel area against the fragmentation index of land use type.

Land use type	Decrease of minimum parcel area for each percentage increase in the fragmentation index of land use type (mm^2) / The average of fragmentation index of land use type			
	1:50,000	1:100,000	1:250,000	1:500,000
Cultivated land	0.19/1.62	0.17/2.51	0.14/3.41	0.12/5.45
Orchard	0.20/1.86	0.17/2.56	0.13/3.53	0.10/4.76
Forest	0.11/1.08	0.10/2.02	0.09/2.72	0.06/3.24
Grass	0.13/1.04	0.12/2.01	0.10/2.71	0.07/3.12
Transportation	0.24/0.65	0.22/1.05	0.19/1.65	0.17/1.72
Water bodies	0.20/2.03	0.20/2.44	0.18/2.94	0.15/3.24
Cities, towns, villages and isolated industry districts	0.25/2.15	0.23/2.65	0.21/3.40	0.19/3.87
Others	0.10/0.98	0.10/1.68	0.08/2.38	0.06/2.43

The fragmentation index of land use types increases along with the decrease of map scale. At the same scale level, the minimum parcel area decreases against the land use type fragmentation index. The reduction of minimum parcel area for each percent increase of land use type fragmentation index decreases gradually from 1:50,000 to 1:500,000. Construction land and transportation land lead the reduction of minimum parcel area against land use type fragmentation, and water bodies, cultivated land and orchards follow.

3.3 Experiment of land use database generalization

The local land use database (at 1:10,000 scale) of Zigui county, Hubei province was used to demonstrate how to determine the generalization index values considering scale and land use pattern, and how to evaluate the generalization results using our macro indices. The study area has a relatively high land use fragmentation (0.53). The control range of area proportion changes in land use types are set to upper values in Table 4. The minimum parcel area of cultivated land for the 1:50,000 map is 4.0 mm^2 according to the regression functions in Table 3. The number of cultivated parcels in the 1:50,000 map is estimated by the fractal selection method (Wang and Wu, 1996), and the fragnentation index of cultivated land (3.43) is computed using equation 6. Therefore, the minimum cultivated parcel area in the 1:50,000 map is adjusted to 3.7mm2 according to Table 9 (mm2). The other thresholds of minimum parcel area for other land use types and other map scales can be estimated in the same way. The minimum parcel area for different land use types at various spatial scales are presented in Table 10.

Table 10. The minimum parcel area in multi-scale land use maps (mm^2).

Land use type	1:50,000	1:100,000	1:250,000	1:500,000
Cultivated land	3.5	3.1	2.6	2.2
Orchard	4.3	3.2	2.7	2.4
Forest	8.0	7.2	6.5	6.0
Grass land	7.8	6.5	6.1	5.8
Transportation	4.5	3.5	—	—
Water	7.0	6.1	5.3	4.7
Cities, towns, villages and isolated industry districts	3.3	2.8	2.3	2.0
Others	8.5	7.7	7.1	6.5

We implemented the generalization of the test land use database in ArcGIS 9.0. Most generalization tasks were completed by flexibly using the functions or commands whose parameters were set based on the indices described above. Some errors in automated generalization results were corrected interactively. Some results for a part of the area are shown in Figure 5.

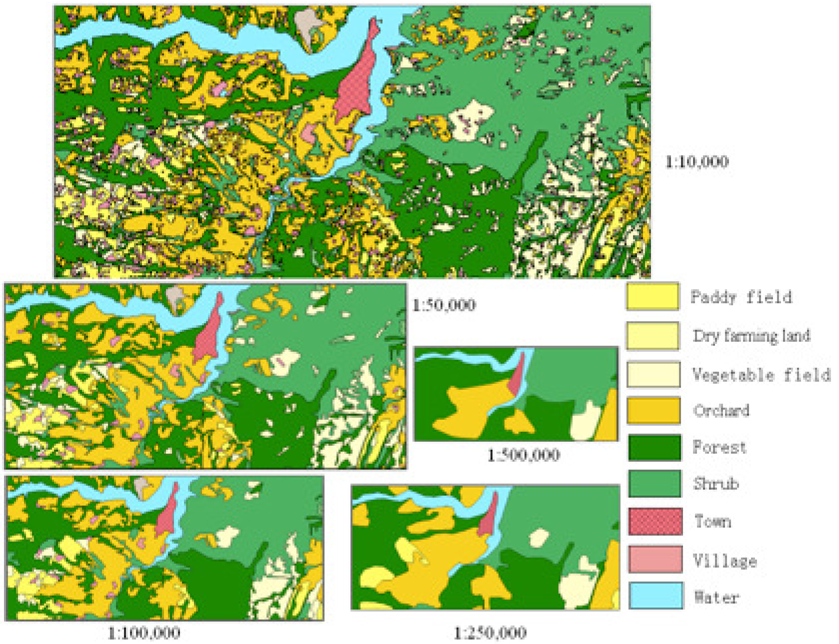

Figure 5. Results of land use data generalization.

A visual evaluation of generalized maps is often implemented by surveying cartographic experts (Stoter, et al., 2009). We invited seven experts with extensive cartographic experience to participate our visual evaluation. A questionnaire was designed to collect the opinions of the experts. The questionnaire has four questions. Each question asked experts to evaluate our generalized maps according to a specific aspect and give a score between 0 and 10 (0 for very poor, 3 for poor, 5 for fair, 7 for good, and 10 for excellent). Question A was designed to ask whether the generalized maps are legible and have proper map loads. Question B asked whether the change of area proportions of land uses is acceptable. Question C asked the experts to evaluate how well the characteristics of land uses were maintained in the generalized maps. Question D asked whether the shapes of land use parcels were simplified properly. In addition, the experts were also asked to give brief reasons for their scores and additional comments if they had. The average values of the scores for question A, B, C, and D were 8.7, 8.0, 8.6, and 8.0, respectively. The visual evaluation of the generalized maps showed that the experts generally regard the maps as between good and excellent. The test area is in a mountainous region. Forest coves the majority of the test area, and agricultural land is scattered. The experts agreed that the generalized maps maintained the land use characteristics, typically, dominance of forest, and high fragmentation of cultivated land (including paddy field, dry farming land and vegetable field) and orchard.

Area proportions of cultivated land, orchards, forest, grass land and water bodies at different scales are computed and shown in Table 11. The maximum rate of area proportion change is calculated and shown as the last row in Table 11. Figure 6 illustrates the comparison of land use area proportion between different scales.

Table 11. Area proportions of land use types in land use data at different scales.

Land use type	1:10,000	1:50,000	1:100,000	1:250,000	1:500,000
Cultivated land	14.62%	13.55%	12.69%	11.88%	11.09%
Orchard	10.74%	10.73%	10.09%	9.65%	9.04%
Forest	66.74%	68.87%	70.51%	71.91%	73.39%
Grass land	0.31%	0.27%	0.25%	0.24%	0.23%
Water bodies	4.34%	4.05%	4.06%	4.07%	4.14%
Maximum ratio of the area proportion change	—	14.0%	6.4%	6.4%	6.7%

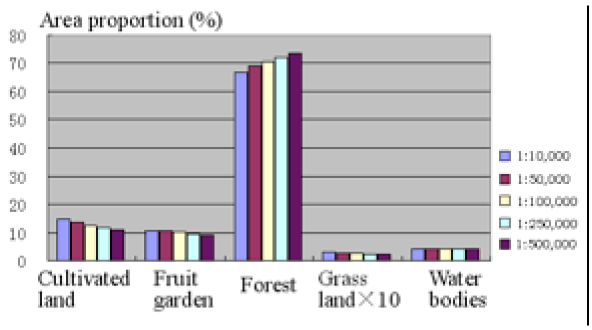

Figure 6. Comparison of land use area proportion between different scales.

There are no siginificant changes in the area proportions of cultivated land, orchards, grass land, and water bodies. The area proportion of forest land increases gradually because it is the major land use type and represents the background of generalization operations, such as polygon elimation and collapse of fragmented land use types. The maximun ratio of land use proportion change accords with our thresholds in Table 4. There is a comparatively large change of area proportion of land use types when the land use database is generalized from 1:10,000 to 1:50,000, and orchard has the maximum changing rate of area proportion, 14.0%. In contrast, the changes of area proportion in generalization between other scales ranges from about 6% to 7%. The area proportions of construction land and transportation land decrease sharply during generalization due to polygon collapse and are not included in Table 11.

Table 12 presents other macro indices for different scales, including total map load and map load of parcels. The comparison of the observations in the experiment and theoretical indices are shown in Figures 7 and 8.

Table 12. Change of macro indices of land use data generalization (percentage).

Index	1:10,000	1:50,000	1:100,000	1:250,000	1:500,000
Total map load (%)	7.9	19.8	25.6	31.0	37.2
Map load of parcels (%)	1.5	2.6	3.1	3.7	4.1

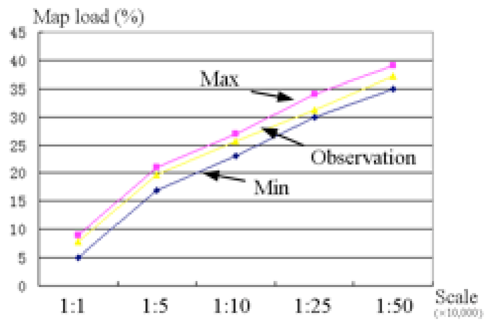

Figure 7. Total map load.

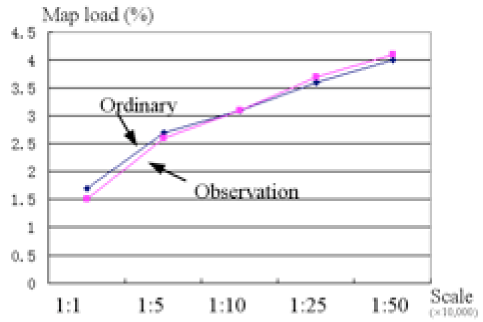

Figure 8. Map load of parcels.

Both the total map load and the parcels map load increase as the scale increases; however, the rate of increase decreases gradually. These observations in the experiment accord with our thresholds, which can be found in Table 5 and Table 6.

Comparing with subjectively determined indices, our indices were determined based on empirical models considering scale and land use pattern effects. The indices helped to produce rational and identical generalized maps. While the indices were apllied to a large region, the contrast of geographic characteristics between subareas were to be well preserved due to consideration of land use pattern. The marco indices proposed in paper can be used to quantitatively evaluate the generalization results.

4. CONCLUDING REMARKS AND DISCUSSION

4.1 Concluding remarks

Land use database generalization indices are restricted by visual discrimination ability, and influenced by subjective factors, map usage, importance of land use types, etc (Zhu, 2004). It is hard to formulate the relationships between land use database generalization indices and these factors individually. This study formulates the scale and land use pattern effects on land use database generalization indices, and leads to the following conclusions:

(1) The paper introduces land use pattern metrics into the determination of land use generalization indices, and proposes an analysis framework of scale and land use pattern effects on the indices. The paper highlights an empirical method to determine land use database indices by examining the relationships between the indices and map scale and land use pattern metrics.

(2) Taking China as an example, our major findings based on countrywide land use samples include:

Map scale dominates the determination of land use database generalization indices. The map area proportions of cities, towns, villages and isolated industrial districts, water bodies and related facilities, and transportation land decrease significantly with map scale, while the map area of cultivated land in flat regions and forest in hilly and mountainous regions increase gradually. Total map load and parcels map load increase with progressive land use database generalization from the original scale to smaller scales. The minimum parcel area was observed to decrease with decreasing map scale. We formulated the logarithm functions between minimum parcel area and map scale for different land use types.

For the land use maps at the same scale, the value of a specific index for land use database generalization (e.g., minimum parcel area) changes with different land use patterns. It was observed that the change of land use area proportion before and after generalization in the area with high land use fragmentation is more significant than the change in other areas. The map load in the area with high land use fragmentation is larger than the map load in the area with lower land use fragmentation. It was found that the minimum parcel area of a land use class is correlated significantly with the fragmentation index of the land use type.

We built the regression function describing the changing of minimum parcel area against land use type fragmentation index.

As a result, the thresholds of the ratio of land use area proportion change were generated, and the thresholds of total map load and suitable parcels map load during generalization with consideration of different land use fragmentation levels were also produced. The minimum parcel area at each main scale can be calculated based on our quantified rules.

(3) The experiment of land use database generalization demonstrates how to determine the generalization index values according to our findings mentioned above, and how to evaluate the generalization results using our macro indices. At the same time, the experiment also proves implicitly the methodology of this paper.

4.2 Discussion

There are some issues worthy of discussion and clarification.

(1) The different effects of land use fragmentation and land use type fragmentation

It can be inferred from our study that land use fragmentation influences macro indices, while land use type fragmentation impacts micro indices. When the mapping area is fragmented, the map content will be dense and complex. Thus the map load and parcels map load will be larger than the ones in less fragmented areas. In this study, we implemented the statistics of map load according to fragmentation zoning. It is also more difficult to maintain the land use area proportion in generalization due to too many parcels being generalized. This explains why we employ a range to control the area proportion change ratio in Table 2. The minimum parcel area is correlated with land use type fragmentation but not the general fragmentation index. It can be intuitively explained as the minimum parcel area of a land use type is influenced by the fragmentation index of the land use type and is not related to other land use types.

(2) Regionality of land use pattern metrics

Land use pattern metrics are calculated aimed at a specific region, so regionality is an intrinsic attribute of these metrics. Land use pattern metrics will change when the calculation unit shrinks or is enlarged. Our land use database samples are at county level, and land use pattern metrics are computed at the same level. So the changing rules for land use database generalization indices related to land use pattern metrics are applicable only when the land use pattern metrics are computed at county level. Sometimes we implement land use database generalization under county level, while the mapping area has distinct land use patterns, for example, mountainous areas and flat areas in the same county. The minimum parcel area deduced from our results will be used as the average, while two adjusted values could be employed in two different sub-areas. A quantitative method for this calibration is to be investigated and verified in further study.

(3) Interpretation of the difference in minimum parcel areas among land use types

It was found that the minimum parcel areas of different land use types differed from each other, and could be divided into three categories. The thresholds of minimum parcel areas of construction land and cultivated land are much smaller than the others, the ones of orchard and water bodies are moderate, and the ones of forest land, grass land, and other land are the largest (see Table 10 and Figure 4). The minimum parcel area of a land use type in a land use database is dominated by the priority of the land use type. The study was conducted in the Chinese land use context. China has the largest population in the world, and has experienced a comparatively fast socio-economic development in recent decades. Construction land expanded rapidly and cultivated land was encroached upon undesirably. Conservation of cultivated land and the strict supervision of construction land are two major concerns in Chinese land use management. Thus these land use types own the highest priority in land use mapping. In our experimental study, it is found that the three smallest minimum parcel area thresholds are the minimum parcel area threshold of transportation land, the threshold of cities, towns, villages and isolated industrial districts, and the threshold of cultivated land. It means these land use types are illustrated in more detail than the others in most Chinese areas. Nevertheless, there are still outliers where some other land use types should be emphasized and portrayed in fine detail, such as water bodies in XinJiang (a water-scarce area) and islands in coastal areas. In these regions, special demands should be fulfilled but these do not contradict our general results.

(4) Evaluation of generalized map using macro indices and feedback

The micro indices, such as minimum parcel area, minimum distance between parcels, and minimum bend diameter, are used as parameters for generalization operations, whereas the macro indices served as evaluation indices of the generalized map. When the values of macro indices violate the thresholds determined in this study after a series of generalization operations, one or more feedback processes are necessary until the macro indices are reached. For example, the ratio of area proportion change of cultivated land is 8% when a land use dataset is generalized from 1:50,000 to 1:100,000, which is larger than the threshold (7%). Then the generalization operations of cultivated land polygons need to be reviewed and adjusted. To fulfill all the macro indices, one or more iterations may be needed.

This study proposes the index system of land use database generalization and designs the descriptive indices of land use pattern. The scale effects of macro and micro indices for land use database generalization were analyzed. Furthermore, this paper investigated the land use pattern effect on the indices of land use database generalization. The methods and implications of this study are confirmed by a case study in Zigui County, Hubei, China. Further research should include rules for computer-aided land use database generalization and land use database auto-generalization.

ACKNOWLEDGEMENT

This study is funded by National Science Foundation of China (40771168, 40901188).

REFERENCES

Ai, T.H., Guo, R.Z., Chen, X.D., 2001. Simplification and Aggregation of Polygon Object Supported by Delaunay Triangulation Structure. Journal of Image and Graphics (7), 703–709.

Ai, T.H., Liu, Y.L., 2002. Aggregation and Amalgamation in Land-use Data Generalization, Geomatics and Information Science of Wuhan University, 27(5), 486–492 (In Chinese).

Ai, T.H., Wu, H.H., 2000. Consistency Correction of Shared Boundary between Adjacent Polygons. Journal of Wuhan Technical University of Surveying and Mapping (5), 426–431 (In Chinese).

Butterfield, B.P., McMaster, R.B., 1991. Map generalization: making rules for knowledge representation. London, Longman.

Chen, X.W., 2005. Structural models and algorithms in land use database generalization. Dissertation of Wuhan University (In Chinese).

Cheng, C.X., Nium, F.Q., Cai, J., Zhu, Y., 2008. Extensions of GAP-tree and its implementation based on a non-topological data model. International Journal of Geographical Information Science 22(6), 657–673.

Gallant, A.L., 2009. What You Should Know About Land-Cover Data. Journal of Wildlife Management 73(5), 796–805.

Gao, W.X., Gong, J.Y., Li, Z.L., 2004. Thematic knowledge for the generalization of land use data. Cartographic Journal, 41(3), 245–252.

Harma, P., Teiniranta, R., Torrna, M., Repo, R., Jarvenpaa, E., Kallio, M., 2004. Production of CORINE2000 Land cover data using calibrated LANDSAT 7 ETM satellite image mosaics and digital maps in Finland. IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), 2703–2706.

Haunert, J.-H., Dilo, A., van Oosterom, P., Constrained set-up of the tGAP structure for progressive vector data transfer. Computers & Geosciences, 35(11), 2191-2203

Jaakkola, O., 1997. Quality of multiscale land cover data. Geographical Information '97, From Research to Application through Cooperation, 335–344.

Lee, D., 2001. Generalization in the new generation of GIS. Proceedings of ICC, Beijing.

Li, J.L., Cheng, C., Niu, F., Cai, J., Zhu, Y., Zhu, D., 2007. A progressive display prototype system based on MAPL-tree and knowledge rules. Geoinformatics 2007, Geospatial Information Science, Pts 1 and 2, 6753, 75302–75302.

Li, Y., 2000. China Land Resource. Beijing: China Land Press (In Chinese).

Liu, X.H., 2005. Method and Practice of land use data generalization. Dissertation of China Agriculture University (In Chinese).

Liu, Y.L., 2002. Categorical database generalization in GIS. Wageningen University, ITC Dissertation.

Liu, Y.L., Molenaar, M., Ai, T.H., Liu, Y.F. 2003. Categorical database generalization aided by data model. The 21st International Cartographic Conference, Durban.

Martinez-Casasnovas, J.A., 2000. A cartographic and database approach for land cover/use mapping and generalization from remotely sensed data. International Journal of Remote Sensing 21(9), 1825–1842.

Muller J C, Wang Z. Area_patch Generalization: A Competitive Approach. The Cartographic Journal, 1992, 29(2):137~144.

Oxenstierna, A. Generalization rules for database-driven cartography. In: proceedings ICC, Stockholm, 1997.

Qi, Q.W., Jiang, L.L. 2001. Research on the Index System and Knowledge Rules for Geographic-Feature-Oriented Generalization. Progress in Geography 20,1–12 (In Chinese).

Rayburn, A.R., Schulte, L.A., 2009. Landscape change in an agricultural watershed in the US Midwest. Landscape and Urban Planning 93(2), 132–141.

Resource Zoning Committee of Chinese Academy of Science, 1959. China Geomorphology Zoning. Beijing: Science Press (In Chinese).

Stoter, J., Burghardt, D., Duchêne, C., Baella, B., Bakker, N., Blok, C., Pla, M., Regnauld, N., Touya, G., Schmid, S., 2009. Methodology for evaluating automated map generalization in commercial software. Computers, Environment and Urban Systems, 33(5):311-324.

Su, M., Li, M.C., Wang, S.H., Zhang M., 2007. Cartographic generalization of cultivated land in hilly areas based on skeleton algorithm. Geoinformatics 2007, Cartographic Theory and Models, 6751: W7510–W7510.

Van Oosterom, P., 2009. Research and development in geo-information generalisation and multiple representation. Computers, Environment and Urban Systems, 33(5):303-310.

Wang, J., Ju, W.M., Li, M.C., 2009a. Characterizing Urban Growth of Nanjing, China, by Using Multi-stage Remote Sensing Images and Landscape Metrics. Joint Urban Remote Sensing Event, 1–3,123–128.

Wang, Q., Wu, H.H., 1996. Fractal method in generalization of polygon groups. Journal of Wuhan Technical University of Surveying and Mapping, 1, 59–63 (In Chinese).

Wang, S.Y., Ding, C.B. Liu, J.S., 2009b. Landscape evolution in the Yellow River Basin using satellite remote sensing and GIS during the past decade. International Journal of Remote Sensing 30(21), 5573–5591.

Wang S.Y., Zhang Z.X., Zhou Q.B., Liu B., Wang, C.Y., 2003. Analysis of landscape patterns and driving factors of land use in China, Acta Ecologica Sinica 23(4), 649~656. (In Chinese).

Zhang, W., 2006. Research on methodology of land use map generalization. Dissertation of Zhejiang University (In Chinese).

Zhang, X.P., Yan, T.L., Cui X.G., Zhu D.H., 2009. Quantitatively Optimized Selection of Proper Scale in Land Use Cartographic Generalization. Computer and Computing Technologies in Agriculture II, 1, 169–178.

Zhu, G.R., 2004. Cartography. Wuhan: Wuhan University Press (In Chinese).