LAND USE DATA GENERALIZATION INDICES BASED ON SCALE AND LANDSCAPE PATTERN

Y.L. Liu [a, b, *], L.M. Jiao [a, b], Y.F. Liu [a, b]

[a] School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan, 430079 China - (yaolin610, lmjiao027, yfliu610)@163.com
[b] Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan, 430079 China - (yaolin610, lmjiao027, yfliu610)@163.com

KEY WORDS: Land use database, Landscape pattern, Land use data generalization, Indices of land use data generalization, Scale

ABSTRACT:

This paper studies the index system of land parcel generalization which is crucial in land use data generalization. We discuss the macro and micro indices of land use data generalization with consideration of spatial scales and landscape pattern. To quantitatively relate the indices and scale and landscape pattern metrics, land use data samples have been collected at multiple spatial scales in various land use regions across China. Based on statistic analysis, we then generate both macro and micro control rules for land use data generalization at various spatial scales and patterns. Finally, we prove the proposed method to be effective with sample data at county level.

1. INTRODUCTION

Indices for parcels generalization are critical for generating multi-scale land use maps and databases. China is conducting its second nationwide land investigation. The land investigation produces 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 studies on spatial data generalization usually focused on general threshold values for terrain mapping (Butterfield and McMaster, 1991; Muller and Wang, 1992; Oxenstierna, 1997; Lee, 2001), whereas these research generally do not consider indicators for land parcels generalization. Liu (2002) and Liu et al. (2003) proposed a framework of 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 parcels generalization, and discussed parcel merging based on neighbourhood analysis. Gao et al. (2004) derived certain thematic knowledge for land use data generalization in the form of production rules. Several studies concentrated on indicators of land use data generalization, such as minimum parcel area in land use map, but were limited in local area (Liu, 2005; Chen, 2005; Zhang, 2006). Nevertheless, these works do not take regional landscape pattern in consideration. Previous literatures suffer from two major setbacks. First, determination of threshold values for multi-scale land use data generalization in a large area, such as a nation, remains subjective. Second, there is a general ignorance of landscape pattern in land use data generalization.

This paper develops the index system of multi-scale land use database generalization from macro and micro aspects, whereby estimates the relationship between indices and scale and the relationship between indices and landscape pattern metrics based on typical samples dataset.

* Corresponding author.

2. DATA AND METHODOLOGY

2.1 Data sampling and preprocessing

Two datasets are used in this paper. A nationwide dataset is used to derive model estimates, while a local dataset serves as a case study. The local dataset is land use data of Zigui county in Hubei province, China, at 1:10,000 scale.

The nationwide land use dataset should be representative in terms of both land use pattern and spatial scale. Land resource in China is zoned into twelve land use regions (Li, 2000). These land use regions can be further divided according to different geomorphologic properties (Resource zoning committee of Chinese academy of science 1959). Therefore, we derive our samples based on two principles as follows: a) The amount of samples in each land use region is proportional to the region area. b) Each geomorphologic zone in a land use region should have at least one sample. Finally we selected land use maps of 51 counties at the scale of 1:10,000, 1:50,000, 1:100,000, and 1:250,000. We also incorporate another 23 subdivided maps for land use at the scale of 1:500,000 in our sample dataset. Our sample counties, which distribute in different land use regions across China, are shown in Figure 1.

We conduct data preprocessing because that our samples are stored in different formats with different reference systems. These data preprocessing includes verification of database, transformation of data formats, transformation of coordinate system, and normalization of coding system for land use types.

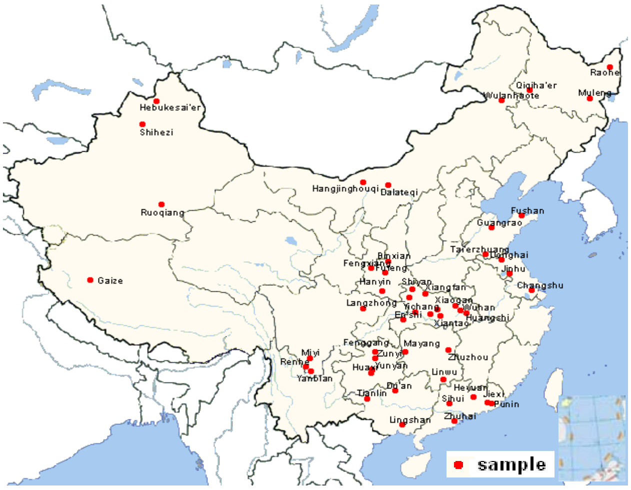


Figure 1 Map of land use sample dataset

2.2 Index system of land use data generalization

Traditionally, researchers employ indicators for land use data generalization at micro scale, such as minimum parcel area, and minimum distance between parcels. Nevertheless we should incorporate certain indices at macro level, such as area proportion of land use types, and spatial distribution characteristics. These macro indices can be used to control generalization operations and to evaluate the generalization result. Hence we build the index system of land use data generalization from both macro-perspective and micro-perspective.

Macro indices for land use data generalization include map load, area proportion of different land use types, and semantic characteristics. Map load describes the 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 data generalization because that generalization of parcels will lead to change in area proportion. Spatial contrast 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 are important for parcel merging control in land use data generalization. The hierarchy of land use types is often used to analyze the similarity between two land use types.

Micro thresholds for land use data generalization include minimum parcel area, minimum distance between parcels, and minimum bend diameter. Minimum parcel area reflects the importance of land use type and landscape pattern. Hence different land use types can have different area thresholds at the same spatial scale.

2.3 Scale effect on indices of land use data generalization

Map scale and land use pattern exert influence on indices of land use data generalization at different strength. Map scale has determinant influence on generalization indices, thus the maps with different scales in the same area will have significantly different indices. Land use pattern has relative smaller impacts than map scale, and its effects can be perceived in a large region with various terrains at the same scale.

We employ exploratory statistics and non-linear regression model to evaluate the relationship between generalization indices and map scale. For convenience, we replace map scale with the scale's denominator. A logarithm function was chosen as a non-linear regression function type according to the dataset.

2.4 Land use pattern effect on indices of land use data generalization

In landscape ecology, patch-based indices are developed to quantify landscape characteristics, such as index of landscape diversity, index of landscape dominance, index of landscape homogeneity, and index of landscape fragmentation. Comparing land use parcels with ecology patches, land use pattern indices are defined as follows.

2.4.1 Design of the land use pattern indices

We define land use indices according to the comparison between land use parcels and landscape patches. In addition to aforementioned indices employed in landscape ecology, we propose two new indices to describe a specific land use type: dominance index of land use type and fragmentation index of land use type.

①Index of land use diversity (H)

This index describes the diversity of land use types based on informatics.

$$H = -\sum_{k=1}^{m} P_k Iog_2(P_k) \qquad (1)$$

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. When m goes to infinity, H reaches its maximum.

②Index of land use dominance (D)

The index of land use dominance measures the degree of how the land use was dominated by one or two types.

$$\begin{aligned} D &= H_{\max} + \sum_{k=1}^{m} P_k Iog_2(P_k) \\ &= \log_2(m) + \sum_{k=1}^{m} P_k Iog_2(P_k) \end{aligned} \qquad (2)$$

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 .

③Index of land use homogeneity (E)

This index describes the homogeneity of land use types in a land use pattern (Wang, 2003), which is given by:

$$E = H/H_{\max} = -\lg\left[\sum_{k=1}^{m}(P_k)^2\right] \Big/ \lg(m) \qquad (3)$$

④Index of land use fragmentation(C)

This index describes the degree of fragmentation of land use pattern,

$$C = N / A \qquad (4)$$

where N and A are the number of land use parcels and the total area of the studied region, respectively.

⑤Dominance index of land use type（D_t）

This index measures how much the land use is dominated by one or several large parcels, and is expressed as:

$$D_t = H_{t\max} + \sum_{i=1}^{n} P_i Iog_2(P_i) = \log_2(n) + \sum_{i=1}^{n} P_i Iog_2(P_i) \tag{5}$$

where D_t is the dominance index of land use type t, n is the number of parcels belonging to land use type t, $H_{t\max}$ 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.

⑥Fragmentation index of land use type（C_t）

This index captures the fragmentation degree of the distribution of land use types,

$$C_t = N_t / A_t \tag{6}$$

where N_t represents the number of land use parcels belonging to land use type t, and A_t is the total area of the land use type t in the studied region.

2.4.2 Analysis of the effect of land use pattern metrics on the indices of land use data generalization

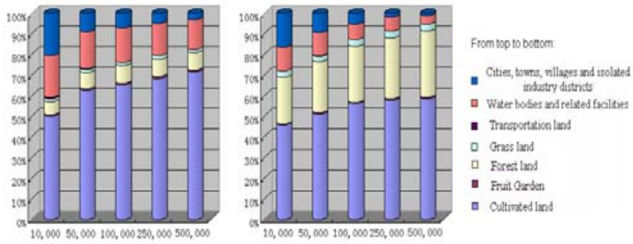
We use correlation analysis to find the major explanatory factors of the thresholds for land use data generalization. Consequently we can set up the regression model between the threshold and major explanatory variables. We can then determine the indices for land use data generalization appropriately based on model estimates.

3. RESULT AND ANALYSIS

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

3.1.1 Scale effect of macro thresholds

There are three driving-forces for changes in area proportions of land use types in generalization. The first cause is collapse of parcels. For example, polygonal residential area is simplified as points, polygonal roads or rivers are simplified as lines during generalization. The second is boundary simplification of parcels whereas a third cause is the elimination, aggregation, amalgamation or exaggeration of parcels. Figure 2 illustrates the changing of area proportion of land use types in Middle and Lower Reaches of Changjiang River at different spatial scales.



(1) Plain area (2) Hilly and mountainous area

Figure 2 Multi-scale changing 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 the map objects of transportation land decreases as well. However, the decrease is not obvious in the figure since the total area of transportation land is relatively small. On the contrary, the area of the map objects of cultivated land in plain region and the area of the ones of forest land in hilly region increase. We produce the ranges of area proportion changes in land use data generalization based on our nationwide samples. The result is summarized in Table 2.

Table 1 Changing rate of area proportion of land use types in generalization (%)

Scales of before and after generalization	Changing rate of area proportion of land use types
1:10k~1:50k	12~20
1:50k~1:100k	4~6
1:100k~1:250k	4~7
1:250k~1:500k	4~6
1:10k~1:500k (accumulative)	15~30

Map load is related to map scale and land use pattern. When map scale decreases, map area and map content increase, and hence the map load increases. Even at the same scale, Map load is larger in regions with more land use types and fragmentary land use distribution. Therefore map load is correlated with land use fragmentation index at the same spatial scale. Land use fragmentation index increases along with the decrease of map scale, whereas the ratios among land use fragmentation index in different areas are almost constant. At the scale of 1:10000, land use fragmentation index can be categorized into three classes: low-level fragmentation (>0.5), medium-level fragmentation (0.3~0.4), and high-level fragmentation (<0.3). We analyze the scale effect of map load in three subdivisions of land use fragmentation. The appropriate total map load of land use maps and map load of parcel features in land use maps are shown in Table 3 and Table 4 respectively. These ranges and averages can be used as benchmarks of area proportion control and map load control in generalization of land use maps.

Table 2 Range of suitable total map load of land use maps (%)

Fragm-entation	1:10k	1:50k	1:100k	1:250k	1:500k
Low	3~7	13~17	17~21	22~26	26~30
Medium	5~9	15~19	20~24	26~30	31~35
High	5~9	17~21	23~27	30~34	35~39

Table 3 Average of suitable map load of parcel features (%)

Fragm-entation	1:10k	1:50k	1:100k	1:250k	1:500k
Low	1.100	2.227	2.712	3.353	3.838
Medium	1.200	2.327	2.812	3.453	3.938
High	1.700	2.666	3.082	3.631	4.047

3.1.2 Scale effect of microscopic thresholds

There are four factors influencing the minimum parcel area in generalization. The first one is the precision required by mapping purpose. The second is the resolution determined by map scale. The third is the importance of land use types, while the fourth is the spatial pattern of land use. We employ non-linear regression to fit the relationship between minimum parcel area and map scale. Taking cultivated land as an example, the scatterplot of the samples is shown in Figure 3. The samples shown in the figure are tidied by eliminating the outliers with two times of variance method. Logarithm function is selected to fit the relationship between minimum parcel area and scale.

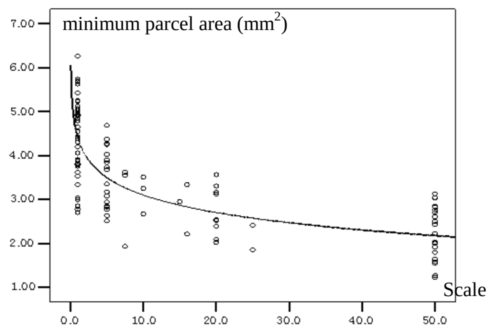


Figure 3 Regression analyses between minimum parcel area of cultivated land and map scale

The regression function is

$$Y=-0.574*\ln(X)+4.424 \qquad (7)$$

where Y is the minimum parcel area of cultivated land, X is the denominator of map scale. The independent variable explains 61.1 of the variations of minimum parcel area ($R^2 = 0.611$). The F-ratio of 155.310,indicates that the model is well-fitted. Moreover, we generate the regression functions for other land use types with the same routine, and the model estimates and associated statistics are presented in Table 5.

Table 4 Regression results between minimum parcel area and map scale

Land use type	Regression (Y =)	R	F
Cultivated land	-0.574*ln(X)+4.424	0.781	155.310
Fruit Garden	-0.875*ln(X)+6.300	0.760	102.530
Forest land	-1.240*ln(X)+8.977	0.800	159.526
Grass land	-1.187*ln(X)+8.860	0.761	109.999
Transportation land	-2.151*ln(X)+7.571	0.806	89.084
Water bodies	-1.626*ln(X)+8.979	0.766	140.662
Others	-1.670*ln(X)+9.860	0.719	77.208
cities, towns, villages, industry districts	-0.626*ln(X)+4.035	0.802	176.121

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

3.2.1 Land use pattern effect of macro indices

The changing of 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 as for Table 2, we should employ the lower limit in regions with a lower fragmentation, and the upper limit in regions with a higher fragmentation. Map load is influenced primarily by the index of land use fragmentation, see Table 4.

3.2.2 Land use pattern effect of microscopic indices

Minimum parcel area correlates with not only map scale but also land use pattern. We employ correlation analysis to explore the land use pattern indices which influence the minimum parcel area. Then we use regression analysis to quantify these influences. We take minimum parcel area of cultivated land in 1:50,000 map as an example to describe the analytical procedure, and the correlation analysis results are demonstrated in Table 6.

Table 5 Correlation analysis between minimum parcel area of cultivated land and land use pattern indices (Scale 1:50,000)

Indices	Pearson coef.	Sig. (2-tailed)
Diversity index (H)	−0.510	0.052
Dominance index (D)	0.440	0.101
Homogeneity index (E)	−0.487	0.065
Fragmentation index (C)	−0.678	0.024
Dominance index of land use type (D_t)	0.325	0.237
Fragmentation index of land use type (C_t)	−0.799	0.018

Table 6 reveals that fragmentation index of cultivated land and general index of land use fragmentation have the most significant influences on the minimum parcel area of cultivated land at 1:50,000 scale. In contrast, the diversity index, predominant index, homogeneity index and dominance index of land use type are not significant. Therefore fragmentation index of cultivated land is adopted as explanatory variable in the regression analysis at the second stage, and the linear regression takes the following form:

$$Y=-18.843*X+4.532 \qquad (8)$$

Where Y is the minimum parcel area threshold for cultivated land, X is the fragmentation index of cultivated land. The Correlation coefficient R^2, F-ratio, and p-ratio are 0.638. 5.112, and 0.008 respectively, all of which are statistically significant and show that the model is well-fitted. The model estimates reveals that one percent change in the fragmentation index has a marginal effects of 0.19mm^2 in 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 mm^2 to 4.5 mm^2.

Table 6 changing of minimum parcel area with the fragmentation index of land use type

Land use type	Decrease of minimum parcel area for each percent increase of the fragmentation index of land use type (mm2) / The average of fragmentation index of land use type (%)			
	1:50k	1:10 0k	1:250k	1:500k
Cultivated land	0.19/ 1.62	0.17/ 2.51	0.14/ 3.41	0.12/ 5.45
Fruit Garden	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	0.20/ 2.03	0.20/ 2.44	0.18/ 2.94	0.15/ 3.24
Cities, towns, villages and 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

3.3 Experiment of land use data generalization

We use a local land use database at 1:10,000 scale, as mentioned before, in our case study. The study area has a relative large fragmentation index of land use pattern (0.53%). The control range of area proportion changing of land use types are set to upper values in Table 2. The minimum parcel area of cultivated land for 1:50,000 map is 4.0mm2 according to the regression functions in Table 3. The number of cultivated parcels in 1:50,000 map is estimated by fractal selection method, and the fragmentation index of cultivated land (3.43%) is computed using equation 6. Therefore the minimum cultivated parcel area in 1:50,000 map is adjusted to 3.7mm2 according to Table 5. The other thresholds of minimum parcel area for other land use types and other map scales can be estimated in the same way. We present the minimum parcel area for different land use types at various spatial scales in Table 8.

Table 7 The minimum parcel area in multi-scale land use maps (mm^2)

Land use type	1:50k	1:100k	1:250k	1:500k
Cultivated land	3.7	3.0	2.5	2.2
Fruit Garden	4.3	3.2	2.7	2.4
Forest	8.0	7.2	6.5	6.0
Grass	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 industry districts	3.3	2.8	2.3	2.0
Others	8.5	7.7	7.1	6.5

The land use data generalization in the study area is implemented based on these indices. Some results of a part of the area are shown in Figure 4. Macroscopic indices prior to and after generalization are shown in Table 9.

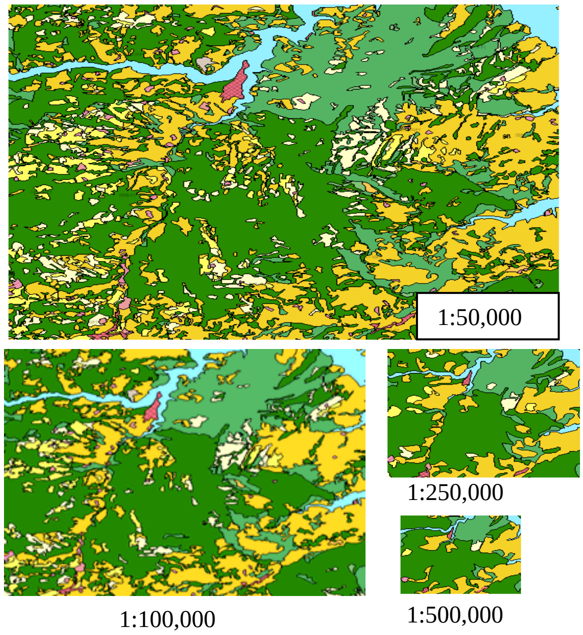


Figure 4 Results of land use data generalization

Table 8 Changing of macroscopic indices of land use data generalization

Index	1:10k	1:50k	1:100 k	1:250 k	1:500 k
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
Maximum extent of the range of area proportion change	—	16%	6%	6%	5%

The comparison of the observations in the experiment and theoretical indices are shown in Figure 5 to 7.

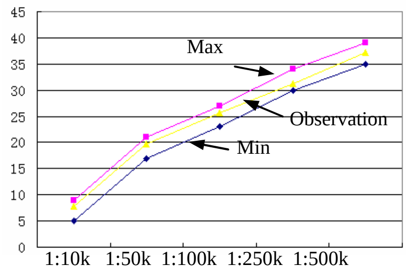


Figure 5 Total map load

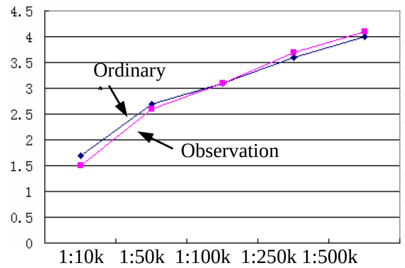


Figure 6 Map load of parcels

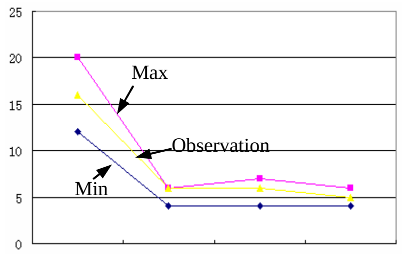


Figure 7 Change of area proportion of land use types

Both the total map load and the parcels map load increase as the scale increases, however the increase rate decreases gradually. There is a significant change of area proportion of land use types when land use data is generalized from 1:10,000 to 1:50,000, which is about 16%. In contrast, the changes of area proportion in generalization among other scales ranges from 5% to 6%. These observations in the experiment coincide with the aforementioned rules in this paper.

4. CONCLUSION

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. 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. 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. The minimum parcel area was observed to decrease with map scale, and the logarithm functions between minimum parcel area and map scale for different land use types were formulated. Land use pattern exerts impact on the land use database generalization indices when at the same map scale. It was observed that higher land use fragmentation increases the land use area proportion change in generalization and map load at the same scale. It was found that the minimum parcel area is correlated significantly with land use type fragmentation index, and the changing rule of minimum parcel area against land use type fragmentation index was generated at each main map scale. 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. The experiment of land use database generalization indicates that our indices are applicable. At the same time, the experiment also proves implicitly the methodology of this paper and shows the results to be reasonable.

Further research should include rules for computer-aided land use data generalization and land use data auto-generalization.

ACKNOWLEDGEMENTS

This study was funded by National Science Foundation, Project number: 40771168

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, pp. 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), pp. 486~492.

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), pp. 426~431.

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

Chen X.W., 2005. Structural models and algorithms in land use database generalization, dissertation of Wuhan University.

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

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

Li Y., 2000. China Land Resource, China Land Press, Beijing.

Liu X.H., 2005. Method and Practice of land use data generalization, dissertation of China Agriculture University.

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.

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

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

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

Resource zoning committee of Chinese academy of science, 1959. China Geomorphology zoning, Science Press, Beijing, China.

Wang Q., Wu H.H., 1996. Fractal method in generalization of polygon groups, Journal of Wuhan Technical University of Surveying And Mapping, 1, pp. 59~63.

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), pp. 649~656.

Zhang W., 2006. Research on methodology of land use map generalization, dissertation of Zhejiang University.

Zhu G.R., 2004. Cartography, Wuhan University Press, Wuhan, China.