MEASURES FOR SEMANTIC QUALITY AFTER POLYGON GENERALISATION

Tao CHENG and Zhilin LI
Department of Land Surveying and Geo-Informatics
The Hong Kong Polytechnic University
Hung Hom, Kowloon,
Hong Kong
Email: {lstc; lszlli}@polyu.edu.hk

ABSTRACT

Map generalization changes the semantics and geometry of map objects according to the context defined by users. How to evaluate and ensure the quality of generalization has become a major issue in contemporary digital cartography. The semantic change after generalization has been seldom studied compared with the other two aspects, i.e. geometry and topology. This research investigates the effect of generalization operations on the semantics of maps objects. A set of quantitative measures for semantic change is put forward. A case study of a land use map is carried to illustrate the practical usefulness of these proposed measures, with merging as an example for polygon generalization. The results indicate that these measures are not only sound in theory but also meaningful in practice

KEY WORDS
semantics; generalization; quality assessment, measures

1 INTRODUCTION

Map Generalization is a process of selection and simplification of details of a map according to the scale and/or the purpose of the map [1]. Such a process is used to derive small-scale maps from large-scale maps. Many operations are applied to achieve the aim of generalization. An operation defines the geometric transformation that is to be achieved; and a generalization algorithm is then used to implement the particular transformation. There are many operations used in generalization and the basic ones are *simplification, selection/elimination, merging, aggregation, symbolization, collapse, exaggeration, displaceme,t, typification* and *smoothing* [2, 3, 4].

In order to derive the simplified model (or representation) of reality at a smaller scale, these operations essentially modify the geometry, topology and/or semantics of the objects from a high resolution to a low resolution. For example, *simplification* and *exaggeration* only modify the metric aspects of the objects; *selection/elimination, merging, symbolization* and *aggregation* essentially modify the topological aspects; and *displacement* is primarily metric, but can also cause topological changes in some cases [5]. The changes in topology and geometry may result in semantic changes, and *vice versa.* The topological changes indirectly imply metric changes as well, but not *vice versa.* For instance, where the real world distance between a lake and a village is too small to be displayed graphically at a given scale, it is certain that the railway and road between the lake and village also cannot be represented, resulting in a spatial conflict in representation. *Displacement* is a solution, but leads to a loss in positional accuracy [6]. Further, due to the removal of the railway/road, the semantics of the map is then changed, which leads to loss in the semantic accuracy.

How to evaluate and ensure the quality of generalization is becoming a major issue in contemporary digital cartography. The assessment of automated generalization results has so far largely relied on visual and qualitative methods that are specific for particular procedures [7,8,9,10], particular aspects of the spatial objects (attribute accuracy) [11,12], and particular shape changes (only for linear objects) [13]. Very few quantitative assessment methods are available when multiple objects are involved or the entire map needs to be characterized [4, p.150]. Recently a method has been proposed by Bard [14] for evaluating generalization quality by comparing the difference between the observed result (after generalization) and the theoretical result (reference or initial characterization of the object). In Bard's approach, three geometric properties are taken into consideration, i.e. position, concavity ratio and size. Galanda [15] modeled the constraints for polygon generalization, which could serve as a guideline for general quality control. But semantic issue is rarely touched in such work.

Indeed, the change in semantics after generalization has been seldom studied compared with the other two aspects, geometry and topology. This research is to investigate the effect of generalization operations on the semantics of polygon objects on thematic maps, as semantics plays a fundamental role in the modeling and representation of geographic objects [16, p.559]. A set of quantitative measures will be developed to measure such effects.

The next section defines a strategy for the evaluation of the quality of semantics after generalization. A set of quantitative measures, i.e. accuracy, consistency and completeness, is described in Section 3. The effect of generalization on semantics is analyzed in Section 4, where merging is taken as a case for deep analysis. An experimental test is reported in Section 5. A discussion on the practical usefulness of these measures is given in Section 6. Some concluding remarks are given in Section 7.

2 ASSESSMENT OF SEMANTIC QUALITY: A STRATEGY

2.1 MERGING AS A CASE FOR STUDY

For the generalization of polygons on thematic maps, the two basic operations are *merging* and *aggregation.* Here *merging* refers to a process to eliminate small areas or sub-polygons. This is also referred to as *coarsen* in some literature. After the merging process, the original objects cease to exist (see Figures 1 and 2). *Aggregation* refers to the process that deletes edges between similar objects and forms a composite object. The semantics of the original objects are then transferred to the new composite object, but the original objects do not cease to exist [17]. *Merging* is usually done with priority given to the neighbor that shares the longest border or the largest area. These two approaches to *merging* have been implemented in Arc/Info as *elimination* [18]. Indeed, *Merging* is actually a geometry-driven approach. On the other hand, *aggregation* is usually carried out because of common thematic characteristics with the neighbors, thus it is actually a class-driven approach. The semantic change in aggregation is not as obvious as in merging. Therefore, this paper will concentrate on the semantic change due to merging operation. Although merging can be applied to point, line and area objects, we mainly consider the area objects (i.e. polygons) in thematic maps.

2.2 3 ELEMENTS FOR DATA QUALITY

For evaluation of semantic quality, attribute accuracy, completeness, consistency and currency are the elements to be considered [19].

Accuracy is the probability of correctly assigning a value. It describes the stochastic errors of observations on attributes and is a measure that indicates the probability that an assigned value is acceptable.

Completeness refers to the symmetric difference between the perceived reality and the database at a given moment.

Consistency is the validation of semantic constrains. It is the result of the validation of semantic constrains of the objects, attributes or relations.

Currency measure changes over time and describes the semantic quality at a given time, say T.

Here accuracy, completeness and consistency of the semantics after generalization are discussed as currency of the semantics involves the temporal change of the data.

2.3 GENERALIZATION AT 3 SPATIAL LEVELS

It has been pointed out that map generalization could be carried out at one the three spatial levels 15, 20], i.e.

- Map level,
- Class (group) level and
- individual feature level.

For polygon generalization, an individual feature means a polygon. Such a three-level decomposition is also referred to as macro, meso and micro levels by other researchers [21,22]. Quality evaluation of generalization at these three levels has also been conducted [14].

The elements of semantic quality for different spatial levels are shown in Table 1. It can be seen that the accuracy can be mapped to a polygon, a class and a map; consistency and completeness can only mapped to a class and a map since it don't make sense for a polygon.

Table 1. An evaluation matrix for semantic assessment

Data Quality	Spatial levels		
Component	Polygon	Class	Map
Accuracy	+	+	+
Completeness		+	+
Consistency		+	+

3 MEASURES FOR SEMANTIC QUALITY

In the following sections, a set of quantitative measures for the semantic quality will be developed. The discussion is mainly related to category maps such as a landuse map that has discrete (or nominal) value attributes.

3.1 ACCURACY

Accuracy can be assessed at three levels. At polygon (i.e. feature) level, the attribute accuracy can be described by the certainty index for the new area objects created after generalization. The accuracy of the new area objects created after generalization will be discussed in Section 4.4 and can be calculated by Equations 13–15.

At class level, the accuracy for Class K can be calculated as follows

$$\mu^{C_K} = \frac{\sum_{i=1}^{N} \mu_i^{C_K} * Area(i)_{C_K}}{\sum_{i=1}^{N} Area(i)_{C_K}} \quad (1)$$

where, i is the objects which belong to Class K after generalization; $\mu_i^{C_K}$ is the certainty of Area i belonging to Class K, which can be derived based upon Equations 13-15; N is the number of objects belonging to Class K after generalization; $Area(i)_{C_K}$ is the area of Polygon i which belongs to Class C_K.

At map level, the accuracy should be evaluated for the whole map as follows:

$$\mu^{M} = \frac{\sum_{j=1}^{M}\sum_{i=1}^{N_j} \mu_i^{C_j} * Area(i)_{C_j}}{\sum_{j=1}^{M}\sum_{i=1}^{N_j} Area(i)_{C_j}} \quad (2)$$

where M is the number of classes after generalization; N_j is the number of objects belonging to Class j after generalization; $\mu_i^{C_j}$ is the certainty of Area i belonging to Class C_j, which can be derived based upon Equations 13-15; $Area(i)_{C_j}$ is the area of Polygon i belonging to Class C_j.

It can be seen that valuations of accuracy at three-levels are mainly related to the object and classes after generalization. They all have a value within 0 to 1. The higher the value is, the better the accuracy is.

3.2 CONSISTENCY

Consistency for a given constrains is given by the ratio between the number of violations and the number of checks of constraint, which was defined by Salgé [19] as,

$$\rho = \frac{N^V}{N} \quad (3)$$

where N^V is the number of violation and N is the number of checks of the constraint.

The obvious violation in semantics after generalization is the change of class types for the objects. Therefore, Equation 3 is modified to describe the attribute change at the class level and the map level.

At class level, the consistency can be defined as

$$\rho_O^{C_K} = \frac{N_{O_K}^{\ V}}{N_{C_K}} \quad (4)$$

where $N_{O_K}^{\ V}$ is the number of area objects which belonged to Class C_K (before generalization) and changed their class type after generalization; N_{C_K} is the number of area objects belonging to Class K (before generalization).

As for the map level, the consistency can be defined as

$$\rho_O^{\ M} = \frac{N_O^{\ V}}{N} = \frac{\sum_{j=1}^{M} N_{O_j}^{\ V}}{\sum_{j=1}^{M} N_{C_j}} \quad (5)$$

where M is the number of classes (before generalization); $N_{O_j}^{\ V}$ is the number of area objects which belonged to Class C_K (before generalization) and changed their class type after generalization; N_{C_j} is the number of area objects belonging to Class j (before generalization).

In addition to number of objects that change their class types after generalization, the consistency of semantics can be represented by the attribute change in terms of areas of the objects. Bregt and Bulens [23] defined the attribute change index as follows:

$$\textit{attribute change index} = \frac{\textit{sum of absolute area differences per class}}{\textit{total surface}} \quad (6)$$

In Equation 6, the area difference is that of the original with generalization per class.

Equation 6 can be considered as another measure for the consistency at the map level. In order to have a same format, Equation 6 is re-written as Equation 7:

$$\rho_{Area}^{\ M} = \frac{Area_O^{\ V}}{Area} = \frac{\sum_{j=1}^{M} Area_{O_j}^{\ V}}{\sum_{j=1}^{M}\sum_{i=1}^{N} Area(i)_{C_j}} \quad (7)$$

where M is the number of classes (before generalization); $Area_{O_j}^{\ V}$ is the absolute area difference of objects belonging to Class j before generalization and after; $Area(i)_{C_j}$ is the area of Object i which belongs to Class j before generalization.

At class level, the consistency in area can be defined as Equation 8.

$$\rho_{Area}^{\ C_K} = \frac{Area_O^{\ V}}{Area_{C_K}} = \frac{\sum_{i=1}^{N} Area(i)_{C_K} - \sum_{i=1}^{N_0} Area(i)_{C_K}}{\sum_{i=1}^{N_0} Area(i)_{C_K}} \quad (8)$$

where N is the number of objects belonging to Class C_K after generalization; N_0 is the number of objects belonging to Class C_K before generalization, $Area(i)_{C_K}$ is the area of Polygon i which belongs to Class C_K before generalization.

3.3 COMPLETENESS

The completeness can be given by two figures: the rate of over-completeness (or commission) and the rate of missing data (omission). Completeness applies for the objects of a class, attributes and relations between objects [19, p.147]. Since objects and classes will be lost after generalization, here we only discuss the situation of omission. Omission is defined by Salgé [19] as follows,

$$\tau^{-} = \frac{N^{-}}{\max(N, N^{0})} \quad (9)$$

where N^0 is the number of occurrence in the perceived reality; N is the number of occurrence in the sample; N^- is the number of occurrence in the perceived reality which does not exist in the sample.

Here we assume the original map before the generalization to be the perceived reality, and the map after generalization to be the sample. In such a case, $N^0 \geq N$ and $\max(N, N^0) = N^0$. Therefore Equation 9 is modified as Equation 10 to evaluate semantic completeness at the class level in terms of object omission for each class.

$$\tau_O^{-C_K} = \frac{N_{O_{C_K}}^{-}}{N_{O_{C_K}}^{0}} \quad (10)$$

where $\tau_O^{-C_K}$ is the ratio of object lost in Class C_k; $N_{O_{C_K}}^{-}$ is the number of objects lost for Class C_K (omitted) after generalization; $N_{O_{C_K}}^{0}$ is the number of objects belonging to Class C_K before generalization.

Equations 11 and 12 are the semantic completeness at map level in terms of object omission (the loss of objects for the whole map after generalization) and class omission (the loss of thematic classes after generalization), respectively.

$$\tau_O^{-M} = \frac{N_O^{-}}{N_O^{0}} \quad (11)$$

where τ_O^{-M} is the ratio of object lost for the whole map N_O^- is the number of objects lost for the whole map after generalization; N_O^0 is the number of objects for the whole map before generalization.

$$\tau_C^{-M} = \frac{N_C^{-}}{N_C^{0}} \quad (12)$$

where τ_C^{-M} is the ratio of class lost for the whole map N_C^- is the number of class lost (omitted) after generalization; N_C^0 is the number of class types before generalization.

3 EFFECT Of GENRALIZATION ON SEMANTCS: MERGING AS A CASE

Here an example is used to illustrate the effect of generalization operation on the semantic change, with merging as an example.

4.1 SITUATION 1: COARSEN

The first situation (see Figure 1) is that the small area D is contained in (or isolated by) a large area A (as background). After generalization, D is merged into A. We use A' to represent the area after merging, although in the database it is still represented as A.

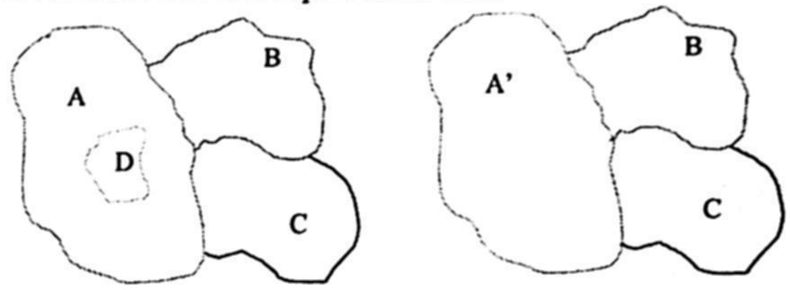

Figure 1. The case of merging (1) - Small area D is merged into background area A.

4.2 SITUATION 2: AREA- AND BORDER-BASED

The second situation involves a small area D adjacent to several larger areas (Figure 2A). The merging will depend upon some geometric criteria defined. In general, there are two ways to merge Area D. The first way is to merge D with Area A due to its biggest size (Figure 2B). This approaches of merging is referred to as area-based merging in this paper. The second way is to merge D with Area C due to its longest common boarder with D (Figure 2C). This approach is referred to as border-based merging in this paper. These two ways are geometry-based approaches and they are implemented GIS software such as ArcInfo (see [18]).

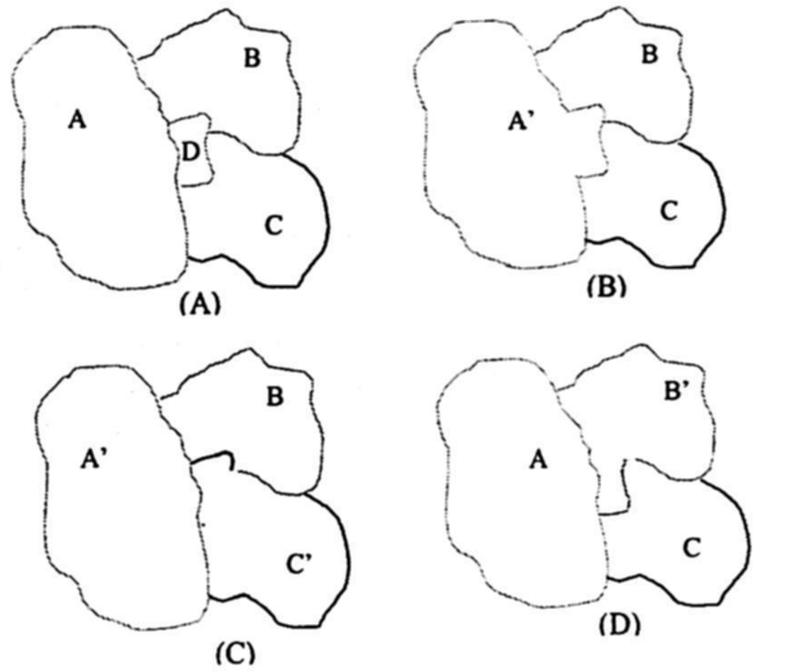

Figure 2. The case of merge (2) - Small area D is merged into one of its adjacent areas.
(A) Small area D and its adjacent areas; (B) Small area D is merged into A, (C) Small area D is merged into C; (D) Small area D is merged into B

4.3 SITUATION 3: SEMANTICS-BASED

The third way can be a semantics-based or knowledge-based approach, i.e., to merge D with Area B due to its closest semantic similarity with D (Figure 2D). Semantic similarity of two area objects has been discussed in several papers. For examples, the sharpness of a boundary for a polygon was defined based upon the purity of polygons [24]; the sharpness of boundaries for regions in raster format was defined based upon the fuzzy membership value [25, p.64]; a semantic similarity evaluation matrix was proposed in [26].

Other semantics, such as priority given to "wood into forest" rather than "wood into lake", may also be considered as criteria for merging. Indeed, it can be considered as knowledge-based approach. The use of the prior-knowledge of thematic maps has been proved an effective approach to landuse map generalization [27].

4.4 THEMATIC ACCURACY (CERTAINTY) OF INDIVIDUAL POLYGON AFTER MERGING OPERATION

Although these three ways has different criteria for merging, the thematic certainties of the new created area *A', B' or C'* can be calculated as follow (although in the database they are still represented as *A*, *B*, *C*, respectively):

$$u_{A'}^{C_A} = \frac{Area(A)\cdot u_A^{C_A} + Area(D)\cdot u_D^{C_A}}{Area(A)+Area(D)} \tag{13}$$

$$u_{B'}^{C_B} = \frac{Area(B)\cdot u_B^{C_B} + Area(D)\cdot u_D^{C_B}}{Area(B)+Area(D)} \tag{14}$$

$$u_{C'}^{C_C} = \frac{Area(C)\cdot u_C^{C_C} + Area(D)\cdot u_D^{C_C}}{Area(C)+Area(D)} \tag{15}$$

These values are used for computation of thematic accuracy at class and map levels.
Here we assume Areas *A, B, C* and *D* are fuzzy objects [28] and they belong to Classes *A, B, C* and *D* with membership function values as $\{u_A^{C_A}, u_A^{C_B}, u_A^{C_C}, u_A^{C_D}\}^T$, $\{u_B^{C_A}, u_B^{C_B}, u_B^{C_C}, u_B^{C_D}\}^T$, $\{u_C^{C_A}, u_C^{C_B}, u_C^{C_C}, u_C^{C_D}\}^T$ and $\{u_D^{C_A}, u_D^{C_B}, u_D^{C_C}, u_D^{C_D}\}^T$, respectively. In Equations 13-15, $u_A^{C_A}, u_B^{C_B}, u_C^{C_C}$ are the membership function values of Area *A*, *B* or *C* belonging to Class *A*, *B* or *C*, respectively; $u_D^{C_A}, u_D^{C_B}, u_D^{C_C}$ are the membership function values of Area *D* belonging to Class *A*, *B or C*, respectively.

If Areas *A* and *D* are crisp objects which belong to Class *A* and *D*, respectively and certainly, i.e. $u_A^{C_A}=1$ and $u_D^{C_A}=0$, Equation 13 becomes Equation 16.

$$u_{A'}^{C_A} = \frac{Area(A)}{Area(A)+Area(D)} \tag{16}$$

If Areas *A* and *D* are thematically similar, i.e. $u_A^{C_A}=1$ and $u_D^{C_A}=1$, then $u_{A'}^{C_A}=1$. It means there is no uncertainty created by merging if *A* and *D* fully belong to a same class.

If the value of Equation 13, 14 or 15 is close to 1, it indicates the change in semantics is small and the quality of generalization is good; on the other hand, if its value is close to 0, it indicates the change in semantics is a lot and the quality of generalization is low.

3 CASE STUDY

A land use map of a sub-area of Wuhan, China (Figure 3) is used for test in this study. It is assumed that this land map is free of uncertainty and is used as source data for generalization.

Figure 3. A detailed land use map (Baoqin Area, Hankou, Wuhan, China)

At the polygon level, the merging operation is applied to eliminate the small areas with highest priority given to the metric constraint of minimal area. Here the two algorithms available in ArcInfo (i.e. longest shared boundary (border-based) and largest area (area-based)) are used to compare their difference in effect on semantic quality after generalization. In order to check the influence of metric constraint (i.e., size of the minimum area, or minimal mapping unit - MMU), three threshold values (1000, 2000 and 4000) are applied to both algorithms. The different values of the MMU correspond to different target map scale. Figures 4A and 4B show the merged results based upon largest area and border respectively, with the MMU= 2000.

5.1 ACCURACY

Figure 4C represents the semantic accuracy of Figure 4B (which are calculated based upon Equations 13-15). It represents the thematic accuracy at polygon level.

At class level, the semantic accuracy (μ^C) for the border-based algorithm with MMU=2000 is listed in Column 2 of Table 2, which are calculated based upon Equation 1.

Based upon the accuracy at the class level, the semantic accuracy at the map level is calculated based upon Equation 2 and the measure of $\mu^M = 0.979$.

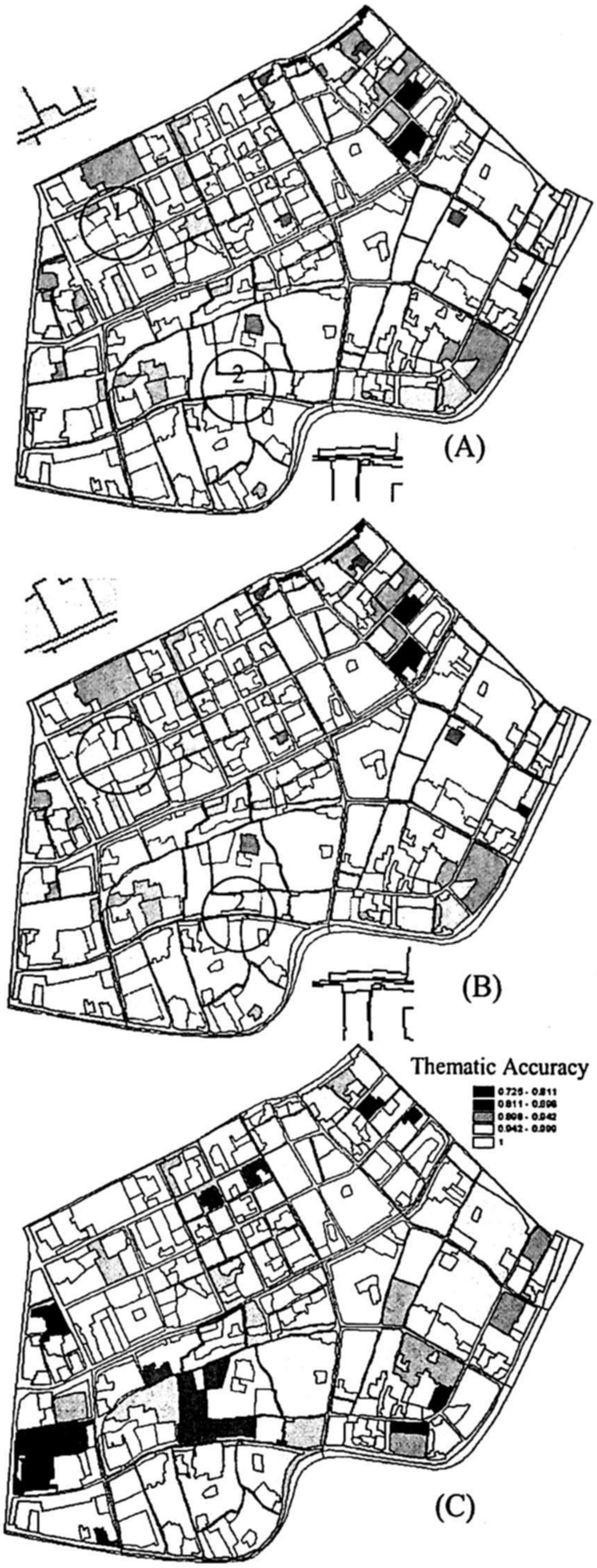

Figure 4. Generalization results by Merging.

A. The merged results based upon the largest area.
B. The merged results based upon the largest border.
C. Thematic accuracy of Figure B. Circle 1 indicates an area in *a* doesn't exist in b; Circle 2 indicates an area in *B* doesn't exist in A.

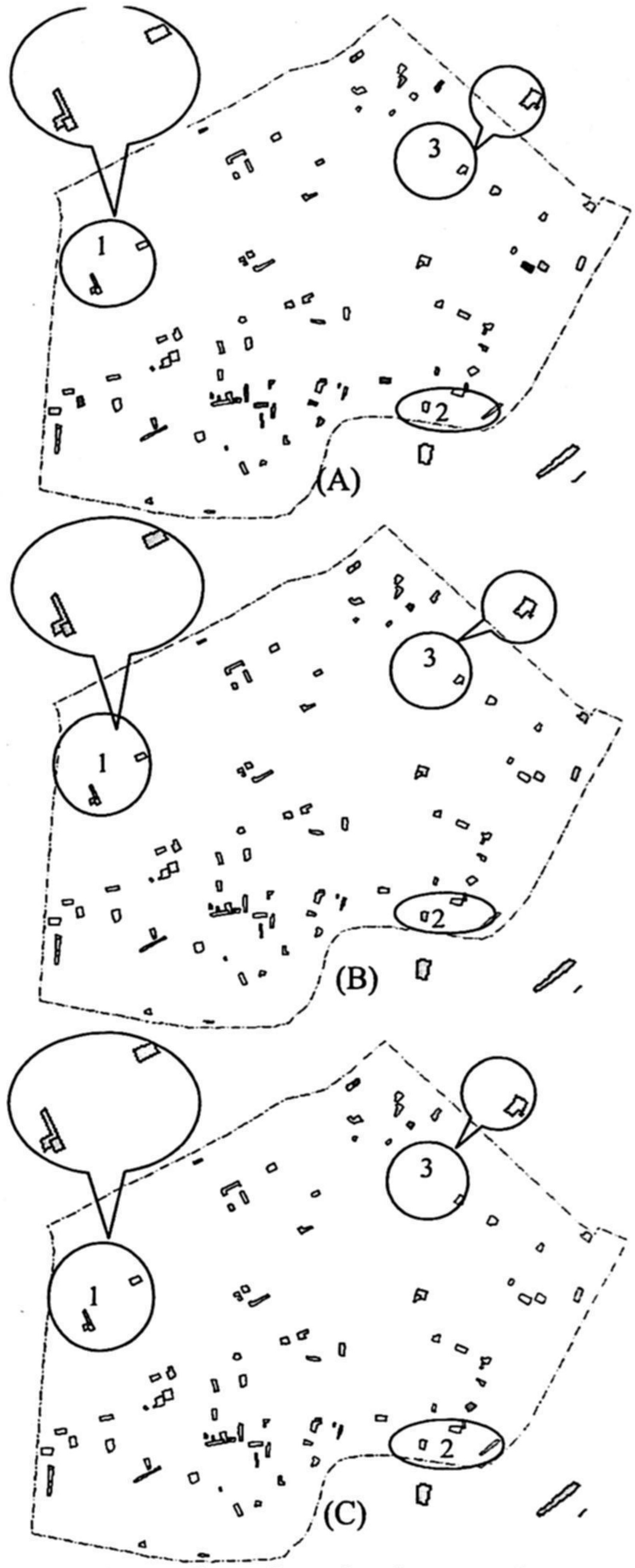

Figure 5. The small areas have been merged out.

A. represents the original class types

B. represents the class types after generalization based upon largest area;

C. represents the class types after generalization based upon largest border;

Three circles in each figure represent three typical situations. Circle 1 represents an area in C which changes its class type to different class types in A and B; Circle 2 represents an area in C which changes its class type to a same class type in A and B; and Circle 3 represents an area in C which doesn't change its class type in both *A* and *B*.

Table 2. Semantic quality at the class level

Class Type	Accuracy (μ^C)	Consistency (ρ^C_{Area})						Completeness (τ^{-c}_O)		
	2B	1A	1B	2A	2B	3A	3B	1	2	3
99	0.983	0.001	0.003	-0.002	0.007	-0.025	0.003	0.029	0.100	0.200
C11	0.964	-0.028	-0.028	-0.070	-0.036	-0.149	-0.081	0.250	0.400	0.600
C12	1.000	0.000	0.000	0.000	0.000	-0.258	-0.193	0.000	0.000	0.571
C21	1.000	-0.009	-0.009	-0.024	-0.024	-0.139	-0.107	0.095	0.143	0.381
C22	1.000	0.000	0.000	-0.406	-0.406	-1.000	-1.000	0.000	0.500	1.000
C23	1.000	0.000	0.000	-0.189	-0.189	-0.344	-0.344	0.000	0.400	0.600
C25	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C26	1.000	0.000	0.000	-0.052	-0.052	-0.276	-0.276	0.000	0.167	0.583
C32	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C35	1.000	0.000	0.000	0.000	0.000	-1.000	-1.000	0.000	0.000	1.000
C36	1.000	0.000	0.000	-0.130	-0.130	-0.130	-0.001	0.000	0.667	0.667
C41	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C51	0.982	-0.014	-0.014	0.004	0.004	-0.040	-0.008	0.286	0.286	0.429
C52		-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000
C61	1.000	0.000	0.000	0.000	0.000	0.000	0.048	0.000	0.000	0.000
C62	1.000	0.000	0.000	-0.071	-0.071	-0.349	-0.349	0.000	0.167	0.500
C63	1.000	0.000	0.000	-0.239	-0.239	-1.000	-1.000	0.000	0.500	1.000
C65	1.000	0.000	0.000	0.000	0.000	-1.000	-1.000	0.000	0.000	1.000
C7		-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000
D3	0.841	0.000	0.190	0.000	0.190	-1.000	-1.000	0.000	0.000	1.000
G11	0.975	0.000	0.026	0.000	0.026	0.000	0.438	0.000	0.000	0.000
M1	1.000	-0.193	-0.193	-0.472	-0.040	-1.000	-1.000	0.500	0.800	1.000
M13	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
M2	0.990	0.000	0.000	-0.049	-0.040	-0.107	-0.097	0.000	0.250	0.438
M21		0.000	0.000	-1.000	-1.000	-1.000	-1.000	0.000	1.000	1.000
M3	1.000	0.000	0.000	0.000	0.000	-0.098	0.143	0.000	0.000	0.250
R11	1.000	0.000	0.000	0.000	0.000	0.002	0.002	0.000	0.000	0.333
R12		0.000	0.000	-1.000	-1.000	-1.000	-1.000	0.000	1.000	1.000
R21	0.982	0.003	0.007	-0.021	-0.010	-0.106	-0.070	0.034	0.172	0.414
R22	1.000	0.008	0.000	-0.086	-0.095	-0.201	-0.225	0.000	0.333	0.667
R3	0.725	0.000	0.000	0.000	0.380	0.000	0.380	0.000	0.000	0.000
R31	0.981	0.000	0.003	0.002	0.011	0.008	0.042	0.087	0.159	0.275
R32	0.991	-0.012	-0.012	-0.100	-0.092	-0.429	-0.421	0.056	0.278	0.722
R41	0.962	0.007	0.005	0.042	0.038	0.057	0.057	0.043	0.043	0.217
R42	0.977	0.006	0.006	-0.012	-0.018	0.004	-0.011	0.250	0.450	0.650
S1	0.970	0.012	0.003	0.070	0.031	0.324	0.158	0.000	0.000	0.000
T1	1.000	0.000	0.000	0.000	0.000	-0.382	-0.382	0.000	0.000	0.500
T23		-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000
T4	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
T42	1.000	0.000	0.000	-0.963	-0.963	0.000	0.000	0.000	0.750	0.000
U3	1.000	0.000	0.000	0.000	0.000	-1.000	-1.000	0.000	0.000	1.000
U9	1.000	-0.068	-0.068	-0.068	-0.068	-0.068	-0.068	0.500	0.500	0.500
W1	1.000	-0.038	-0.038	-0.067	-0.067	-0.144	-0.072	0.333	0.444	0.556
average	0.982	-0.077	-0.073	-0.184	-0.159	-0.345	-0.312	0.127	0.291	0.513
max	1.000	0.012	0.190	0.070	0.380	0.324	0.438	1.000	1.000	1.000
min	0.725	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	0.000	0.000	0.000
stdev	0.051	0.258	0.261	0.346	0.356	0.432	0.462	0.274	0.343	0.378

5.2 CONSISTENCY

Semantic consistency in area at the class level is summarized in Column 3 (ρ^{C}_{Area}) of Table 2, where negative value means the area for the class is reduced; and value "-1" means that there is no objects in this class (i.e. the class is omitted). It is calculated based upon Equation 8. It can be seen that the consistency index ρ^{C}_{Area} increases with the MMU at the class level. It implies that the inconsistency in semantics at the class level increase with an increase in MMU.

Semantic consistency at the map level is reported in Table 3. Both consistency indexes in terms of area (ρ^{M}_{Area}) and object (ρ^{M}_{O}) violation increases with an increase in MMU at the map level. The differences between the indices of the border-based and the area-based are not obvious at three spatial levels. This is also approved by the results that an average 98.2% of merged areas based upon the largest area and 97.2% of merged areas based upon the largest border change their class types after merging (see Figure 5). It implies that the inconsistency in semantics at the map level increases with an increase in MMU.

Table 3. Thematic consistency at the map level

MMU	ρ^{M}_{Area}		ρ^{M}_{O}	
	A	*B*	A	B
1000	0.006	0.006	0.089	0.087
2000	0.031	0.027	0.214	0.212
4000	0.123	0.087	0.416	0.409

(*A* and *B* represents the area-based and border-based merging, respectively)

In general, the consistency decreases with an increase in MMU at the class level and the map level. There is not much difference with the area-based and the border-based algorithm.

5.3 COMPLETENESS

Column 3 (τ^{-C}_{O}) of Table 2 summarizes the semantic completeness at the class level. Generally speaking, the completeness indicator increases with an increase in MMU at the class level. It implies that the completeness in semantics at the class level decreases with an increase in MMU. Since the number of objects lost in both algorithms are the same, the completeness for both algorithms are the same at the class level.

Table 4 summarizes the semantic completeness at the map level. It can be seen that the number of objects and classes after generalization decrease with an increase in MMU. In another word, the semantic completeness index at map level increases with the MMU. Since the number of objects lost and the number of classes lost in both algorithms are the same, the completeness for both algorithms are the same at the map level.

Table 4. Semantic completeness at the map level

MMU	*No. of Objects Merged*	τ^{-M}_{O}	*No.of Classes Lost*		τ^{-M}_{C} (A,B)
			A	B	
1000	39	0.068	4	4	0.092
2000	92	0.114	5	5	0.216
4000	179	0.274	12	12	0.421

(Number of original objects is 425; Number of original classes is 44.)

It implies that completeness of semantics decreases with an increase in MMU. For the whole map, the reduction of object number is around 6.8% to 27%; the reduction of class number is around 10% to 40%.

In general, the completeness indicators increase with an increase in MMU at the class level and the map level. It implies that the quality of completeness of semantics decreases with an increase in MMU at the class level and the map level. There is no difference with the area-based and border-based algorithm.

6 PRACTICAL USEFULNESS OF THE QUALITY MEASURES: A DISCUSSION

The three quality measures of the semantics at three levels are discussed in above sections. What is their practical usefulness? Here we try to interpret these quality measures from the practical viewpoint. The accuracy of the semantics represents the purity of the polygon, the class or the map. At the polygon level, the purity of each polygon is calibrated so that any area has inclusion is clear and cautious can be taken for further analysis. At the class level, if we are especially interested in a class (such as a special landuse), we should consider if the algorithms could satisfy the requirement of the semantics for that class. At the map level, we may pay attention to the quality for the whole map to check if the map has sufficient semantic quality for further usage.

The consistency in area represents the mis-matches of areas for different classes after generalization. The area sizes of a class of polygons before and after generalization could be quite different. If this mis-match is too big to be accepted, the generalization operation should then be avoided. In other words, alternative generalization operations should be applied.

The completeness represents the omission of objects and classes at the class and map levels. Since the omission of objects might imply the omission of the owners of a parcel (a polygon in the landuse map), the indicator of completeness of object omission provides a further constraint for quality control of generalization. If a complete class of polygons is omitted after generalization, and if this class type is very important for the application, omission of that class should then be avoided. It also implies that the particular generalization operation shouldn't be applied in such a case.

Generally speaking, the quality measures offer a means for quality control of generalization. It provides quality information of maps after generalization to map user and help them decide if the generalized map is qualified for further analysis. The quality information can also be used to facilitate the cartographers in choosing proper operations and algorithms for map generalization.

7 CONCLUSIONS

This paper picks up a particular type of maps, polygon maps, for discussion. The aim is to evaluate the effect of generalization on the semantic changes. Special attention has been given to the uncertainty in semantics created in the generalization operation, with merging operation as a case analysis. A set of quantitative measures, i.e. accuracy, consistency and completeness, has been developed for semantic quality assessment. A set of real-life data (landuse map) has been used to evaluate these measures. In such a test, different MMU values are used, which correspond to different target map scales.

It is found that the semantic quality in terms of accuracy, consistency and completeness decreases with an increase in the MMU (minimal mapping unit) value. It means that the semantic quality of the thematic map become poorer at a smaller scale. This is in accordance with our common sense. Such is also the case for the quality of topographic maps. Therefore, the set of quantitative measures developed in this project is theoretically sound and meaningful in practice.

It should be noted here that, although the set of quality measures for semantic change is only applied to merging operation in this study, it is applicable to other operations for polygon maps.

In this study, only the situation in which the original source data have no uncertainty has been analyzed. It is desirable to study the situation where there is uncertainty in the source data in further research. Also it would be of great interest to systematically analyze the changes of semantics after generalization with MMU and then to find out the correspondence between MMU and map scale.

REFERENCES

[1] ICA (International Cartographic Association), Multilingual dictionary of technical terms in cartography, Wiesbaden, Franz Steiner, 1973.

[2] J. Keates, *Cartographic Design and Production*, (2nd edition). Longman Scientific, London,1989, 261pp.

[3] R. B. McMaster and K.S. Shea, *Generalization in Digital Cartography*, Association of American Geographers, 1992, 134p.

[4] R. Weibel and G. H. Dutton, Generalising spatial data and dealing with multiple representations, in Longley, P.A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (eds.), *Geographical Information Systems*, Second Edition, Vol. 1, 1999, 125–155.

[5] G. Dettori and E. Puppo, How generalization interacts with the topological and metric structure of maps, *Proceedings of the 7th International Symposium on Spatial Data Handling*, edited by M. J Kraak and M. Molenaar, 1996, P.559–570.

[6] F. Harvey and F. Vauglin, Geometric match processing: applying multiple tolerances, *Proceedings of the 7th International Symposium on Spatial Data Handling*, edited by M. J. Kraak and M. Molenaar, 1996, P.155–171.

[7] Z. Li and S. Openshaw, A comparative study of the performance of manual generalization and automated generalization of line features. *Geographical Information 1992-1993 (AGI Year Book)*, 1992, 501-514.

[8] M. De Berg, M. van Kreveld and S. Schirra, A new approach to sub-division simplification, *ACSM/ASPRS Annual Convention and Exposition 4* (*Proceedings AutoCarto 12*): 1995, 79–88.

[9] F. Brazile, A generalization machine design that incorporates quality assessment. In *Proceedings 8th International Symposium on Spatial Data Handling*, Vancouver, 1998, pp. 349–360.

[10] R. Weibel and G. H. Dutton, Constraint-based automated map generalization, *Proceedings of International Symposium on Spatial Data Handling*, T. K. Poiker and N. Chrisman (eds.), 1998, P.214–224.

[11] E. M. João, The importance of quantifying the effects of generalization, In Müller, J. C., Lagrange, J. P. and Weibel, R. (eds), *GIS* and *Generalization, Methodology and Practice*, London: Taylor & Francis, 1995,183–193.

[12] M. Painho, The effects of generalization on attribute accuracy in natural resource maps, In Müller, J. C., Lagrange, J. P. and Weibel, R. (eds.), *GIS and Generalization, Methodology and Practice*, London: Taylor & Francis, 1995, 194–206.

[13] A. Skopeliti and L. Tsoulos, A methodology for the assessment of generalization quality, paper presented at Fourth Workshop on Progress in Automated Map Generalization, August 2-4, 2001, Beijing, China, http://www.geo.unizh.ch/ICA/docs/beijing2001/papers01.html

[14] S. Bard, Evaluation of generalisation quality, paper presented in the *Fifth Workshop on Progress in Automated Map Generalization*, Paris, 2003, http://www.geo.unizh.ch/ICA/docs/paris2003/papers03.html.

[15] M. Galanda, Modelling constraints for polygon generalization, paper presented in *the Fifth Workshop on Progress in Automated Map Generalization*, Paris, France, 28-30 April 2003, http://www.geo.unizh.ch/ICA/docs/paris2003/papers03.htm

[16] F. Harvey, F. Vauglin and A. Bel Hadj Ali, Geometric matching of areas – comparison measures and associated Links, *Proceedings of the 8th International Symposium on Spatial Data Handling*, T. K. Poiker and N. Chrisman (eds.), 1998, P.557–568.

[17] T. Cheng, Quality assessment of model-oriented generalisation. *In Proceedings of the 4th Workshop ACI on Progress in Automated Map Generalisation.* Beijing, China, 2001, http://www.geo.unizh.ch/ICA/docs/beijing2001/.

[18] D. Lee, Recent Generalization Development and road ahead, paper presented in the *Fifth Workshop on Progress in Automated Map Generalization*, Paris, 2003, http://www.geo.unizh.ch/ICA/docs/paris2003/papers03.html

[19] F. Salgé, Semantic accuracy, in *Elements of Spatial Data Quality* (edited by S.C. Guptill and J. L. Morrison, Exeter: Elsevier Science), 1995, P. 139-152.

[20] Z. Li, Multi-scale representation of spatial data: An Overview. *Proceedings of Mapping Sciences'98*. 24-28 May 1998, Perth 409-416.

[21] A. Ruas, OO-constraint modelling to automate urban generalization process, *Proceedings of the International Symposium on Spatial Data Handling*, Vancouver, Canada, 1998, 225–235.

[22] S. Musti`ere and B. Moulin, What is spatial context in cartographic generalisation? *Proceedings of the Symposium on Geospatial Theorey, Processing and Applications*, Ottawa, Canada, CDROM, 2002.

[23] A. Bregt and J. Bulens, Application-oriented generalization of area objects, in Molenaar, M. (editor), *Methods for the Generalization of Geo-Databases*, Delft: Netherlands Geodetic Commission New Series, Nr. 23, 1996, 57–64.

[24] F. Wang and G. B. Hall, Fuzzy representation of geographical boundaries in GIS, *International Journal of Geographical Information Science, 10(5)*, 1996, 573–590.

[25] T. Cheng, *A Process-oriented Data Model for Fuzzy Spatial Objects*, Ph.D. thesis, Wageningen Agriculture University/ITC, Enschede, 1999, 163pp. (ISBN 90-5808-038-2).

[26] Y.L. Liu, M. Molenaar and M. J. Kraak, Semantic similarity evaluation model in categorical database generalization, in *Proceeding ISPRS Commission IV Symposium*, Ottawa, Canada, CDROM, 2002.

[27] W.X. Gao, J. Gong, and Z. Li, Knowledge-based generalization on land-use data, paper presented in the *Fifth Workshop on Progress in Automated Map Generalization*, Paris, 2003, http://www.geo.unizh.ch/ICA/docs/paris2003/papers03.html.

[28] T. Cheng, M. Molenaar and H. Lin, Formalization and application of fuzzy objects, *International Journal of Geographical Information Science, 15(1)*, 2001, pp.27– 42.