

# Empirical Study on Location Indeterminacy of Localities

Sungsoon Hwang, Jean-Claude Thill

Department of Geography, State University of New York at Buffalo, 105 Wilkeson Quad, Buffalo NY 14261 U.S.A.

## 1. Introduction

Humans perceive localities on a daily basis but perception is likely to be vague, particularly when it comes to boundaries of localities. In contrast to vague human perception of localities, data and operations used in a geographic information system (GIS) are dominantly based on crisp sets. Due to the discrete nature of computing environments, vague concepts prevalent in spatial objects (e.g. nearness and other qualitative relations) are usually forced into discrete constructs. For example, a vaguely stated phrase such as “near Chicago” is usually only georeferenced *either* in or outside of Chicago in GIS, although it can be better seen as confidence interval. Accordingly, if georeferencing is broadly understood as pinpointing the most plausible location intended by users, the reliability of georeferencing depends on mental maps of spatial entities and of their mutual spatial relationships. The problem is that such mental maps lack sharp boundaries of localities so that the question comes down to how to best model such indeterminate boundaries of localities given the crisp nature of common spatial datasets.

There are two possible approaches to modeling indeterminate boundaries of localities (Robinson 1988). The first approach is purely empirical. In short, empirical evidence can be directly collected from those who perceive localities and weaved together to compose boundaries of localities, provided that a certain degree of agreement (i.e. consistently overlapping boundaries) is met (Montello et al 2003). Second, a hypothetical model can be adopted where it is assumed that the location indeterminacy of locality is governed by some generally accepted rules (e.g. spatial autocorrelation). In this study, the second approach is chosen over the first one because the first approach is not feasible when considering the purpose of this study –

examining location indeterminacy across multiple localities. In this study, fuzzy logic (Zadeh 1965) is applied to geographic databases where fuzzy set membership of a locality is determined by (1) absolute distance measures, (2) qualitative spatial relations, and (3) scale associated with the locality.

The model of fuzzy regions allows a vaguely-conveyed locality to be georeferenced while it could be considered inadequate to georeferencing by a conventional method. When traffic crash data are georeferenced on the basis of a fuzzy locality model, the georeferenced data can be used to examine the gap between what is perceived and what physically exists. In other words, it can help us delineate possible locality boundaries represented in mental maps. The location indeterminacy of locality is measured as the ratio of locality entities that are georeferenced outside of their actual boundaries relative to records believed to be *in* the boundary. The variation of location indeterminacy may be associated with some characteristics of localities. Analysis of variance examines if location indeterminacy can vary by different neighborhood types such as urban versus rural area.

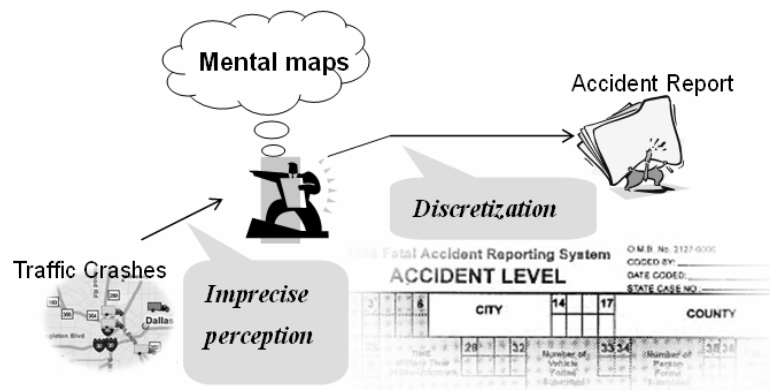
This study can help address the following related research questions: Can we represent the perceptual environments (e.g. mental maps) in a crisp-set based computing environment (e.g. GIS)? Is a fuzzy-set approach promising in accomplishing this task? What kind of factors do we have to take into account to represent mental maps in GIS? How can it be implemented in GIS? To what extent are humans confident in identifying the boundary of a locality? Does the level of confidence vary with certain characteristics of localities (such as urban versus rural)?

The research objectives of this study are twofold: One is to model indeterminate boundaries of localities by applying the concept of fuzzy sets to geographic databases; the second is to examine if the location indeterminacy of localities significantly varies by the urban versus rural settings. The rest of this paper is organized as follows: Sect. 2 reviews background issues for the problems in hand. In Sect. 3, we will discuss how to model indeterminate boundaries of localities in GIS. Sect. 4 presents the results of an empirical analysis to compare the mean of location indeterminacy between urban and rural areas. In Sect. 5, analysis results are interpreted and the project is summarized.

## **2. Background**

Let us consider traffic crash sites where police officers compile required information in an accident report. The typical way of recording accident

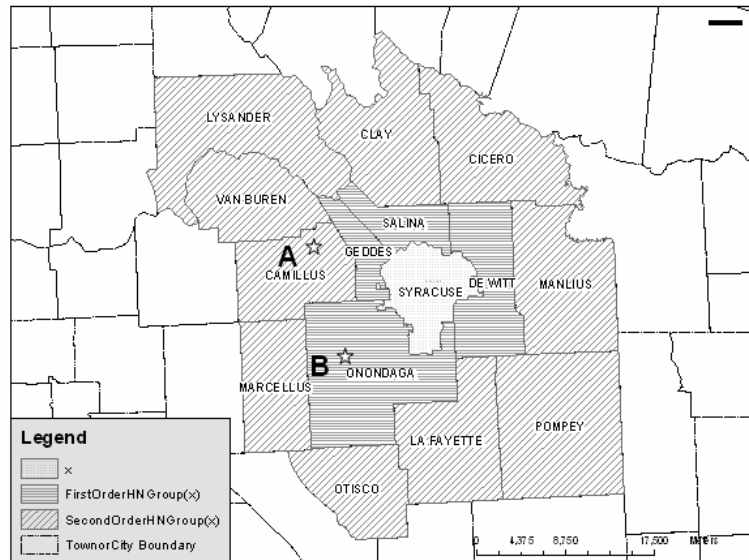
location is either by means of a linear referencing system (LRS) for highway accidents (which is not considered in this study), or by reporting the names of the roadway and locality where the crash occurred. As illustrated in Fig. 1, an accident report does not necessarily reflect what is out there *as it actually is*; In particular, accident location is captured through the cognitive and perceptual filter of filing officers. Moreover, police officers are often forced to make crisp judgment on vague perception and cognition in response to rigid requirements of coding forms. This study is particularly focused on the vague perception and cognition of locality boundaries. Given the spatial resolution of data used in this study, the type of locality we address is confined to local jurisdiction (e.g. City of Buffalo). The central issue of this study is depicted in Fig. 1. In the lower right corner of this figure, the CITY column suggests a spatial relation *in*, but it can be argued that *near* also applies under uncertainty.



**Fig. 1.** How accident location is recorded: losing certainty  
The example of accident coding form is adapted from NHTSA (1995)

Mental maps can be seen from different levels of abstraction – idiosyncrasy versus general principles. As for idiosyncrasy, mental maps vary with individuals – their unique experiences, preferences, and the level of familiarity with areas (Gould and White 1986, Thill and Sui 1993). Thus, it is hardly amenable to generalization. Similarly, mental maps may be affected by unique characteristics of surroundings. For instance, people may find it easier to identify the boundary of localities surrounded by a water body (e.g. island, peninsular) or high mountains (e.g. river basin). Besides, it is easy to identify a locality with salient characteristics (e.g. Paris with Eiffel Tower). All these exemplify *specifics* of cognitive mapping of locality.

In contrast, properties inherent to a geospatial object determine the degree to which its boundary is perceived with vagueness. First, the degree of belonging to a certain locality declines as the distance to the locality increases (similar to Tobler's First Law of Geography). That is, it is a function of *Euclidean distance or some other measure of spatial separation* (Yao and Thill 2004). This is widely accepted, but this alone cannot explain irregular or asymmetric form of indeterminate boundaries. Second, it is affected by *spatial qualitative relation*. In Fig. 2, suppose that A and B are within the indeterminate boundaries of the locality of Syracuse. Location A may be perceived less near Syracuse than B due to the intervening locality of Geddes between A and Syracuse (Ullman 1956). Third, it is *scale-dependent* (Goodchild 2001). The size or hierarchy of referents matters (Gahegan 1995). For example, Buffalo, NY is near Syracuse, NY at a regional scale but the two localities cannot be considered near at the local scale. The three properties mentioned above are key determinants of location indeterminacy of locality.



**Fig. 2.** The effect of qualitative spatial relation on nearness

It can be noted that this study addresses localities with official recognition in contrast to localities with informal recognition such as vernacular regions (e.g. Midwest) (Zelinsky 1980). Moreover, accident reports are filled in by police officers who are presumably familiar with surrounding areas, so that the variation of mental maps on an individual basis is rea-

sonably controlled for. In that sense, the results of this study should not simply be seen as representative of a statistical sample, but as conservative when it comes to location indeterminacy of locality.

### 3. Modeling Location Indeterminacy of Locality

Let us consider locality  $l$  to be a kind of fuzzy region. As such, it is denoted as  $\tilde{A}_l$ .  $\tilde{A}_l$  is composed of the following three parts: *Core*, *Boundary*, and *Exterior* (also known as egg-yolk model; Cohn and Gotts 1996). These parts are defined as crisp regions (regular closed sets) denoted by  $reg_c$ . Also let  $\mathcal{R}^2$  denote the two-dimensional geographic space, and let the fuzzy-set membership function of  $\tilde{A}_l$  be  $\mu_{\tilde{A}_l}$ .  $\tilde{A}_l$  and  $\mu_{\tilde{A}_l}$  are defined as follows:

$$\begin{aligned}\tilde{A}_l &= Core(\tilde{A}_l) \vee Boundary(\tilde{A}_l) \vee Exterior(\tilde{A}_l) \\ Core(\tilde{A}_l) &= reg_c(\{(x,y) \in \mathcal{R}^2 \mid \mu_{\tilde{A}_l}(x,y) = 1\}) \\ Exterior(\tilde{A}_l) &= reg_c(\{(x,y) \in \mathcal{R}^2 \mid \mu_{\tilde{A}_l}(x,y) = 0\}) \\ Boundary(\tilde{A}_l) &= reg_c(\{(x,y) \in \mathcal{R}^2 \mid 0 < \mu_{\tilde{A}_l}(x,y) < 1\})\end{aligned}$$

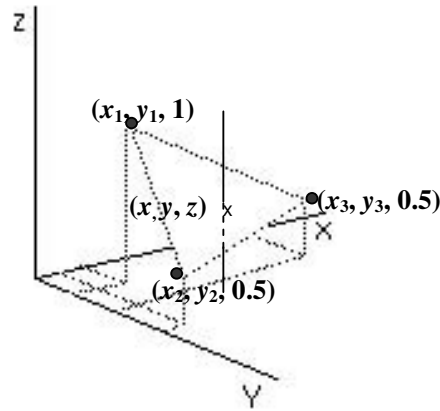
The *core* identifies the part of the region that definitely belongs to  $\tilde{A}_l$ . The *exterior* determines the part that definitely does not belong to  $\tilde{A}_l$ . The indeterminate character of  $\tilde{A}_l$  is summarized in the *boundary* of  $\tilde{A}_l$  in a unified and simplified manner (Erwig and Schneider 1997; Schneider 1999). The *core* and *boundary* can be adjacent with a common border, and *core* and/or *boundary* can be empty. When the *boundary* is an empty set,  $\tilde{A}_l$  becomes a crisp region. Thus, a crisp region is a special case of a fuzzy region.

The question boils down to delineating the nonempty set *boundary*, which is illustrated in Fig. 2. Where the function named *FirstOrderHNGroup*( $x$ ) is defined on the set of neighbors which “meet” locality  $x$ , where each neighbor has the same spatial resolution level as  $x$  (HN is the abbreviation of HorizontalNeighbor in the sense that the function results in a group with the same resolution level as  $x$ ). For example,  $x$  and its considered neighbors should belong to the same level of administrative boundary (e.g. city, county, state). *SecondOrderHNGroup*( $x$ ) is defined on the set of neighbors which “meet” *FirstOrderHNGroup*( $x$ ) at the exclusion of  $x$ . The term “meet” is one of the spatial relation predicates defined in Egenhofer and Franzosa (1991) to extend Allen’s (1983) temporal logic to the spatial domain. The exterior of *FirstOrderHNGroup*( $x$ ) becomes the 0.5-cut

boundary. Accordingly, the fuzzy set membership value for *boundary* can be redefined as follows:

$$\begin{aligned}\mu_{\tilde{A}_I}(x,y) &= [0, 0.5] && \text{if } (x,y) \text{ is in } \textit{SecondOrderHNGroup}(I) \\ \mu_{\tilde{A}_I}(x,y) &= [0.5, 1] && \text{if } (x,y) \text{ is in } \textit{FirstOrderHNGroup}(I)\end{aligned}$$

To compute the continuous and linear<sup>1</sup> fuzzy-set membership value in  $\textit{Boundary}(\tilde{A}_I)$ , we create a Delaunay Triangulation whose nodes are comprised of any vertices on *core*, *0.5-cut boundary*, and *exterior*. Fig. 3 illustrates two nodes on the *0.5-cut boundary*, and one node on *core*. The membership value is obtained by intersecting a vertical line with the plane defined by the three nodes of the triangle as shown in Fig. 3.



**Fig. 3.** TIN surface created to interpolate fuzzy set membership value

In Fig. 3, the generalized equation for linear interpolation of a point  $(x, y, z)$  in a triangle facet is  $Ax + By + Cz + D = 0$ , where  $A$ ,  $B$ ,  $C$ , and  $D$  are constants determined by the coordinates of the triangle's three nodes. Thus, the fuzzy set membership value can be obtained from the following equation given the  $x$ - and  $y$ -coordinates:  $\mu_{\tilde{A}_I}(x,y) = (-Ax - By - D) / C$ .

As illustrated in Fig. 4, a continuous membership value between 1 and 0 is computed in the indeterminate *boundary* (Wang and Hall 1996; Stefanakis et al 1999). In Figure 4, the locality Buffalo has a full membership in

<sup>1</sup> Some studies on nearness suggest that the relationship between distance and nearness can be approximated by a linear relationship (Gahegan 1995). An s-shaped function has also been proposed (Worboys 2001).

its *core*, a partial membership in its *boundary* (e.g. Amherst), and no membership beyond *exterior* (e.g. Alden).

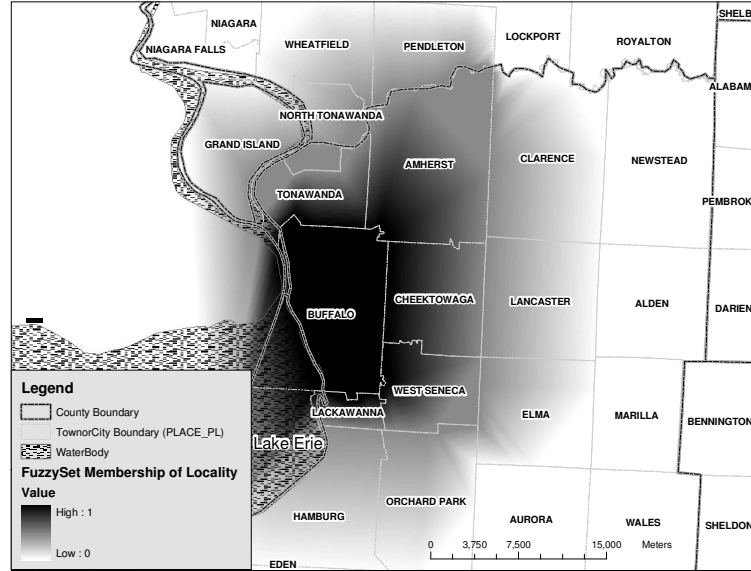


Fig. 4. Fuzzy set membership of locality Buffalo, NY

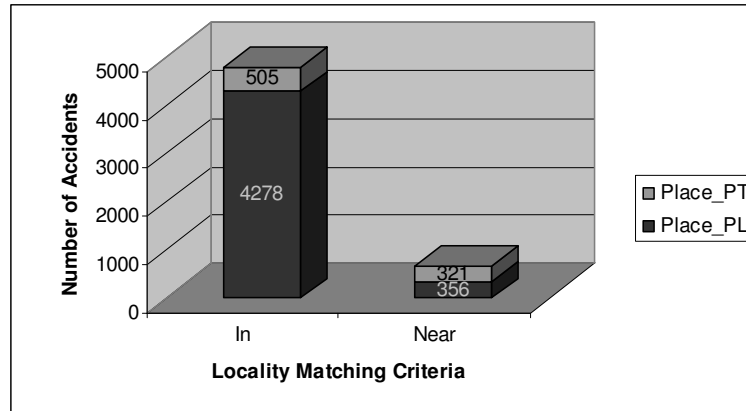
#### 4. Empirical Analysis of Location Indeterminacy of Localities

This section presents location indeterminacy revealed in a time series of 8631 traffic crashes that occurred between 1996 and 2001 in New York State and compiled at the accident level in the Fatal Accident Reporting System (FARS) (NHTSA 1995). Of these, 5460 cases<sup>2</sup> are considered for examining location indeterminacy. First, crash records in the dataset are georeferenced by taking into account uncertainty (e.g. incompatible name of roadway between FARS and reference database, nearness implicit in identifying locality) (Hwang and Thill 2003). Second, the geographic location of a georeferenced record is compared with locality information given in the original FARS data record. The discrepancy between them may in-

<sup>2</sup> Excluded cases are composed of 2053 cases georeferenced by LRS, 614 cases with invalid code for locality matching (e.g. null or incorrect code), and non-geocodable cases (504).

indicate location indeterminacy of the concerned locality. If a record is georeferenced exactly *in* the locality as given in the original data (i.e. no discrepancy), fuzzy set membership becomes 1. If a record is georeferenced to the indeterminate boundary of the locality (that is, *near* locality), fuzzy set membership becomes any value between 0 and 1 based on the definition of fuzzy set membership function described in sect. 3.

Fig.5. graphs 5460 cases by reference data (Place\_PL or Place\_PT<sup>3</sup>), and locality matching criteria (In or Near). Near-cases account for 12.4% (677/5460). This means that police officers misidentify the locality of car crashes 12.4% of the time. But it should be noted that it is a very conservative estimate because the total count of 5460 usable cases only accounts for cases that are georeferenced with a reasonably good quality (i.e. data in poor quality is more likely to have near-cases).



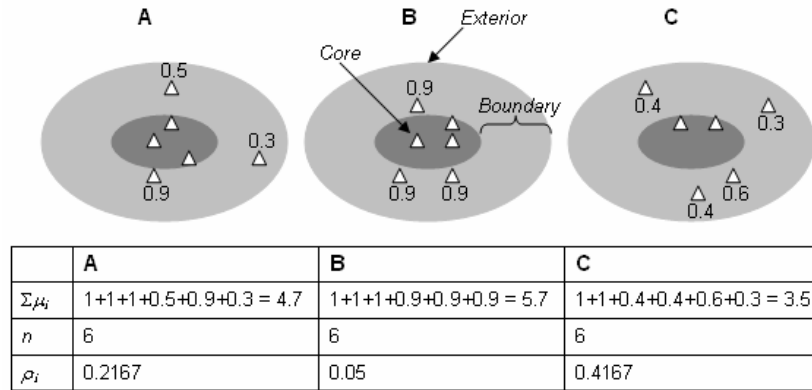
**Fig. 5** Probability of pinpointing localities exactly (in) versus roughly (near)  
Source: FARS accident level data (1996-2001 New York State)

In the third of the analysis, location indeterminacy of locality  $i$  is computed as follows:  $\rho_i = 1 - (\sum \mu_i)/n$  where  $\mu_i$  is the fuzzy set membership value of locality  $i$  for each record, and  $n$  is the total number of records that are reported to occur in locality  $i$ . To illustrate this, consider the three different scenarios in Fig. 6. Fuzzy set membership value 1 is assigned to accidents within *Core*, and  $[0,1]$  is assigned to accidents in *Boundary*. It is

<sup>3</sup> Place\_PL corresponds to County Subdivisions in U.S. Census terms. On the other hand, the polygon boundaries of Place\_PT are usually not well-defined. Thiessen polygons drawn around centroids derived from hardcopy maps are used as a proxy of polygon boundary if unknown.



shown that B has the lowest location indeterminacy 5% while C has the highest location indeterminacy 42%.



**Fig. 6** Illustration of computation of location indeterminacy of locality

Finally, localities are classified as urban or rural<sup>4</sup> in an attempt to examine if location indeterminacy can vary by neighborhood types. The average number of fatal crashes that occurred during the study period in rural areas is 2 (612 accidents / 298 localities), while that of fatal crashes in urban areas is 16 (3822 accidents / 246 localities). The mean value of location indeterminacy for rural localities turns out to be 0.1165, which compares to 0.0966 for urban localities.

At first glance, the difference appears to be trivial, but interpretation of the location indeterminacy rate of rural areas requires some caution because of the small number problem (Kennedy 1989). That is, location indeterminacy rate is highly variable when the number of crashes within an area is rather small. To work around the small number problem, we convert observed rates to empirical Bayes estimates in a way that prior distribution is taken into account (Bailey and Gatrell 1995). Consequently, somewhat unreliable values of location indeterminacy are smoothed out while relatively reliable values are expected not to change much. Instead of adjusting for overall distribution, two sets of observed rates (which is urban versus rural) are adjusted for their within-group distributions because of significant difference in their respective overall patterns.

Table 1 presents descriptive statistics of location indeterminacy value for urban and rural areas, after adjustment for the overall pattern of loca-

<sup>4</sup> Depending on the spatial definition of Urbanized Area (U.S. Census Bureau 1999).

tion indeterminacy within each group. It indicates that people are 94% (or somewhere between 93% and 95%) sure in identifying urban localities while they are 88% (or somewhere between 86% and 90%) sure in identifying rural localities. It means that the boundary of rural areas is perceived 6% less accurately than in urban areas.

**Table 1.** Descriptive statistics of Bayesian estimates of location indeterminacy value: comparison between urban locality and rural locality

	N	Mean	Std. Dev	Std. Error	95% Conf. Interval for Mean	
					L Bound	U Bound
Urban	246	<b>0.0597</b>	0.0631	0.0040	0.0518	0.0677
Rural	298	<b>0.1178</b>	0.1759	0.0102	0.0977	0.1378
Total	544	0.0915	0.1399	0.0060	0.0798	0.1033

Analysis of variance (Table 2) conducted on Bayesian estimates of location indeterminacy confirms the difference between urban versus rural locality in terms of location indeterminacy is significant (F-statistics 24.209). Therefore, it can be concluded that the boundary of localities is perceived differently depending on neighborhood types. The interpretation and implication of this empirical analysis will be given in the next section.

**Table 2.** The result of analysis of variance (One-way ANOVA) conducted on 544 localities grouped as urban or rural.

	Sum of Squares	df	Mean Square	F	Sign.
Between Groups	0.454	1	0.454	24.209	0.000
Within Groups	10.170	542	0.019		
Total	10.624	543			

## 5. Conclusions

The result of analysis of variance sheds some light on our initial hypothesis that mental maps on urban settings may be less error-prone than those on rural settings. It is maybe because city provides more landmark or route upon which judgment on indeterminate boundaries of localities can be based. As suggested by Golledge et al (1995), direct experience (e.g. navigation) on geographic space helps humans acquire ultimate spatial knowledge (e.g. survey knowledge). Under the theory of spatial knowledge acquisition, it can be argued that urban settings provide more favorable

conditions for forming consistent mental maps. The empirical study presented in this paper supports this idea.

Moreover, it may be worthwhile considering characteristics of localities as a *fiat* spatial object (Smith 1995). According to Smith, a *fiat* object, unlike other physical environments (e.g. mountain, lake) does not exist in a physical way (thus intangible); rather it is the outcome of human conceptualization. Therefore, any factors considered to facilitate human conceptualization may refine our mental maps. One of them is a scale factor. Identifying localities (here we focus on “unity condition” (Guarino and Welty 2000) – what composes localities as a part-whole relation) becomes easier at a micro-scale. As urban settings are denser, they provide a reasonable scale in which humans can conceptualize localities without much difficulty. Conversely, a highly dispersed pattern of settlements constitutes rather challenging environments in which humans can conceptualize localities as a whole.

In summary, we built a hypothetical model of localities in a way that location indeterminacy is incorporated as a fuzzy set membership value. As evident in a multi-year empirical datasets of traffic crashes in New York state, location indeterminacy of spatial objects seems to prevail, even though we deal with spatial objects with official recognition. Moreover, the confidence level in identifying localities may vary with neighborhood types. That is, people are found to be 6% more confident in identifying urban areas than rural areas. Boundaries of rural areas are perceived more vaguely and variably than urban areas.

## References

- Allen JF (1983) Maintaining knowledge about temporal intervals. *Communications of the ACM* 26(11):832-843
- Bailey TC, Gatrell AC (1995) *Interactive Spatial Data Analysis*. Longman Scientific & Technical, Essex, pp.303-308
- Cohn A, Gotts N (1996) The ‘egg-yolk’ representation of regions with indeterminate boundaries. In: Burrough P, Frank AU (ed) *Geographic Objects with Indeterminate Boundaries*, Taylor & Francis, London, pp 171–187
- Egenhofer M, Franzosa R (1991) Point-set topological spatial relations. *INT J of Geographical Information Systems* 5(2):161-174
- Erwig M, Schneider M (1997) Vague regions. In: 5th international symposium on advances in spatial databases (SSD’97), LNCS Vol 1262, Springer, pp 298-320
- Gahegan M (1995) Proximity operators for qualitative spatial reasoning, In: Frank AU, Kuhn W (ed) *Spatial Information Theory: A Theoretical Basis for GIS*. LNCS Vol 988, Springer, Berlin, Germany, pp 31-44

- Golledge RG, Dougherty V, Bell S (1995) Acquiring spatial knowledge: survey versus route-based knowledge in unfamiliar environments. *Annals of AAG* 85(1):134-158
- Goodchild MF (2001) A geographer looks at spatial information theory, In: Montello DR (ed) *COSIT 2001*, Springer-Verlag, London, pp.1-13
- Gould P, White R (1986) *Mental Maps*. Harmondsworth, Penguin
- Guarino N, Welty C (2000) Ontological analysis of taxonomic relationships. In: Laender A, Storey V (ed) *Proceedings of ER-2000: The International Conference on Conceptual Modeling*, Springer-Verlag
- Hwang S, Thill JC (2003) *Georeferencing Historical FARS Accident Data: A Preliminary Report*, Unpublished document, Department of Geography and NCGIA, State University of New York at Buffalo
- Kennedy, S (1989) The small number problem and the accuracy of spatial databases. In: Goodchild M, Gopal S (ed) *Accuracy of Spatial Databases*, Taylor & Francis, London, pp 187-196
- Montello DR, Goodchild MR, Gottsegen J, and Fohl P (2003) Where's downtown?: behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation*, 3(2&3):185-204
- NHTSA (1995) *FARS 1996 Coding and Validation Manual*. National Center for Statistics and Analysis, National Highway Traffic Safety Administration, Department of Transportation, Washington, D.C.
- Robinson VB (1988) Some implications of fuzzy set theory applied to geographic databases. *Computers, Environment, and Urban Systems* 12:89-97
- Schneider M (1999) Uncertainty management for spatial data in databases: fuzzy spatial data types. In: Gutting RH, Papadias D, Lochovsky F (ed) *SSD'99*, LNCS Vol 1651, Springer-Verlag, Heidelberg, pp 330-351
- Smith B (1995) On drawing lines on a map. In: Frank AU, Kuhn W (ed) *Spatial information theory: A theoretical basis for GIS*. LNCS Vol 988, Springer, Berlin, Germany, pp 475-484
- Stefanakis E, Vazirgiannis M, Sellis T (1999) Incorporating fuzzy set methodologies in a DBMS repository for the application domain of GIS. *INT J Geographical Information Science* 13(7):657-675
- Thill JC, Sui DZ (1993) Mental maps and fuzzy preferences. *Professional Geographer* 45: 264-276
- Ullman EL (1956) *The Role of Transportation and the Bases for Interaction*. In: William TJ (ed) *Man's Role in Changing the Face of the Earth*, University of Chicago Press, pp 862-80.
- Wang F, Hall GB (1996) Fuzzy representation of geographic boundaries in GIS. *INT J of Geographical Information Systems* 10(5):573-590
- Worboys M (2001) Nearness relations in environmental space. *INT J Geographical Information Science* 15(7):633-651
- Yao X, Thill JC (2004) How far is too far? A statistical approach to context-contingent proximity modeling. *Transactions in GIS*, forthcoming
- Zadeh LA (1965) Fuzzy sets. *Information and Control* 8:338-353
- Zelinsky W (1980) North America's vernacular regions. *Annals of AAG* 70(1):1-16