Chapter 4
Fuzzy Set Theory in Geospatial Analysis

Yaolong Zhao

4.1 Introduction

Geographical information systems (GIS) are designed to store, retrieve, manipulate, analyze, and map geographical data. Since the 1960s when R.F. Tomlinson first presented the GIS, this field has mainly focused on the construction of the systems, the improvement of system functions, and the extension of its application to other disciplines. The research contents have played an important role in providing spatial decision-making support for both governments and the public, and have also promoted the formation and development of the discipline of Geographic Information Science (Goodchild 1992). However, with the extension and deepening of applications, users began to doubt the results of spatial analysis using GIS (Doucette and Paresi 2000; Morrison 1995; Östman 1997; Stefanakis et al. 1999). The raw material for GIS (i.e., the original data imported into GIS) inevitably always contains errors (Shi et al. 2002). Data models used in GIS to describe the real world are just approximations to objective reality. In addition, all kinds of spatial operations and processing approaches may bring new errors and uncertainties into the production of spatial analysis. Most existing designs of GIS software are based on the hypothesis that no errors exist in geographic entities and their spatial relationships. Generally, GIS can only deal with determinate spatial entities and their relationships. However, using a GIS designed to deal with determinate data for uncertain data will bring about problems, and the results cannot satisfy the users' needs (Shi et al. 2002). As the outputs of GIS play an important role in spatial decision-making support, users began to be

Y. Zhao (✉)
School of Geography, South China Normal University,
Guangzhou, People's Republic of China

Formerly in Division of Spatial Information Science, University of Tsukuba,
Tsukuba, Ibaraki, Japan
e-mail: yaolong@scnu.edu.cn

Y. Murayama (ed.), *Progress in Geospatial Analysis*,
DOI 10.1007/978-4-431-54000-7_4, © Springer Japan 2012

concerned about the quality of spatial data in GIS. This undoubtedly made many scholars think about the field of GIS spatial data quality control (Mowrer and Congalton 2000; Östman 1997; Shi et al. 2002).

The acquisition of spatial data in GIS primarily relies on surveying and geographical investigation. Matured surveying errors and data processing methods (namely surveying adjustments) which are based on probability and statistical theories have been introduced into the field of spatial data quality control. Thus, a set of theoretical systems on spatial data quality control methods gradually came into existence (Goodchild and Dubuc 1987; Goodchild and Gopal 1989; Heuvelink et al. 1989; Shi et al. 2002). However, handling errors of spatial data are somewhat different from the processing of conventional surveying data, as the sources of spatial data are diverse and complex. In addition, operations with spatial data are also complex, and are different from surveying adjustment methods, among which there are strict geometric conditions. Therefore, as well as traditional probability and statistics theory, other theoretical supports are required according to the intrinsic features of spatial data (Burrough and Frank 1996; Burrough et al. 1997; Fisher 1999; Goodchild and Jeansoulin 1998). Fuzzy set theory provides an important approach to dealing with spatial data, and has sporadically been adopted in the field of GIS (Cheng et al. 2001; Fisher 2000).

The connection between fuzzy set theory and spatial data quality control needs to explore the relationship between spatial data error and uncertainty theory. The uncertainty is the deficiency in the degree of knowledge about surveying data, and also the degree of unlikelihood and doubt about the validity of the survey results. Standard deviations, or multiples of these, are always used to express uncertainty. The surveying error is the difference between the measured value and the true value, and this is caused by imperfect processing of the survey data or unsatisfactory surveying conditions. As the true value is generally unknown, the actual value of the error is difficult to find. In theory, surveying adjustment methods allow surveying data to be closer to the true value. In the international "Guide to the Expression of Uncertainty in Measurement" (International Organization for Standardization 1995), it was stressed that surveying error and uncertainty are deemed to be two different concepts which should not be confused. Surveying error is really different from uncertainty by definition. However, both of them are used to express the relationship between the survey data and the true value. They play the same role in describing the confidence level of the survey data. In other words, surveying error is one of the descriptive methods of indicating surveying uncertainty. Goodchild (1999) argued that uncertainty is the difference between the true value and its expression in GIS. If the true value can be determined, error and precision are used to describe the uncertainty (Goodchild 1999). However, the true value is generally unknown, especially when it is related to the cognition of humans.

Randomness and fuzziness are the two conditions that result in the uncertainty of spatial data (Burrough and Heuvelink 1992). Randomness is the uncertainty of cognition generated by the inadequacy of the observation conditions. Fuzziness refers to the uncertainty of differentiation caused by the intermediary transitivity of objective differences (Burrough and Frank 1996; Fisher 2000; Zadeh 1965).

This type of intermediary transitivity may be due to the fuzziness of the ruler used to describe objective things, or the inevitability of unclear cognition about those things. GIS can model the real world. The model is based on the cognition and abstraction of the real world, and is the approximate reflection of the real world. The randomness and fuzziness of reality cognition bring uncertainty about the spatial data into GIS, and thus affect the spatial data quality and the results of spatial analysis.

The uncertainty of spatial data derived from the concept of fuzziness can be separated into several aspects, as follows.

1. Uncertainty comes from the vagueness of the distribution of geographical phenomena or the concept of a geographical entity

   Variation and fuzziness are two intrinsic attributes in nature which affect the accuracy of spatial data representation. For instance, the range of grassland is not always determinate; somewhere grassland always moves gradually toward forest or desert areas, or else there exists a transition area which reflects a smooth transition state from grassland to forest or desert. The boundary of soil units and the classification of vegetation type are usually fuzzy, and different operators often draw different classification maps, and so forth. If such fuzzy information does not undergo appropriate processing, the data imported into GIS will definitely have fuzzy characteristics, and result in uncertainty.

   This kind of fuzziness can be reflected in both graphics and the attribute of accuracy in the quality of the contents of spatial data. The intuitionistic reflection in the accuracy of graphics is an error. However, the real position of a boundary is difficult to determine owing to the vagueness of the concept. Therefore, the fuzziness of a concept may be reflected in the accuracy of the attribute. For example, the vagueness of vegetation classifications will result in mistakes when describing parcels of vegetation.
2. Uncertainty derived from spatial relationships

   In a qualitative description of a spatial relationship, there ubiquitously exist inaccurate terms. For example: what is "*nearby*" the village; land to the "*south*" (or "*north*") of the river is suitable for arable farming. In buffer analysis, the proximity to a river is usually described in terms like "in the area "*about*" 5 km to the "*north*" (or "*south*") of the river"; in visual interpretations of images, a ground feature may be described as residential houses, factory buildings, or other type of architecture. There are often inaccuracies when describing geographical properties. For instance, the descriptions of boundaries always are fuzzy, as multiple feature boundary lines often coincide with each other. The mixed pixels generated in the processing of remote sensing data and the overlapping in mode identification are also fuzzy. These emerge in the process of finding specified descriptions of geographical phenomena.

   This kind of vagueness in specific descriptions of spatial relationships is mainly reflected in attribute accuracy, logistic consistency, and the integrity and temporal accuracy of the quality and contents of spatial data. If the description of geographic phenomena is not clear, the accuracy of the attribute will always be affected, as well as the logistic consistency and integrity of the data. During a set time interval, spatial data may fail to accurately reflect the situation at that time. Therefore, the temporal accuracy will also be influenced.

3. Uncertainty derived from spatial analysis and spatial reasoning operations

   Spatial analysis is a reasoning process about knowledge, the results of which can provide spatial decision-making support to users. The language of human beings is similar to fuzzy semantic expressions, for example, we want to find the "*largest*" area affected when a reservoir bursts, or to consider the villages which are at a distance of "*about*" 200 m from the reservoir, and so forth. Sometimes we need to consider whether an area is "*suitable*" for a certain kind of crop. Here, "*suitable*" can be divided into fuzzy terms like "*totally unsuitable*," "*not very suitable*," "*suitable*," "*comparatively suitable*," and "*totally suitable*."

   Uncertainty generally exists in spatial data, and originates from the vagueness of the concept of spatial entities and spatial relationships. This uncertainty influences the quality of spatial data, and thus affects the results of GIS applications. Traditional Boolean set theory can deal with spatial entities with determinate boundaries and concepts in cognition. However, for those spatial entities with fuzzy boundaries and concepts, Boolean set theory fails to reflect the vagueness among them. In traditional approaches, such spatial entities would be modeled approximately, and accordingly, the approximate model would result in a loss of information. Such uncertainty needs fuzzy set theory.

4.2 Fuzzy Set Theory

Fuzzy set theory was first presented in 1965 by the famous cybernetics expert L.A. Zadeh in his ground-breaking paper *Fuzzy Sets* (Zadeh 1965). In his research on human thinking and judgment of the modeling process, he built up a theoretical system using rigorous mathematical methods to describe fuzzy phenomena. Fuzzy set theory is an extension of the traditional classic set theory. The aim of the extension is to overcome the accurate "either–or" bi-value logic of classic set theory. Thus, there is a smooth transition between elements and non-elements of a set, so that one element can partially belong to a set, but not completely belong or completely not belong to the set. The difference between a fuzzy set and a classic set is that the fuzzy set has explicitly put forward the terms of a membership function through which the degree of each element belonging to a set can be calculated. Set operations like intersection and union in classic set theory are still applicable in fuzzy sets.

4.2.1 Fuzzy Set

When people consider a specific problem, they always confine the issue within a limited range, which is the so-called universe, and is usually represented by capital letter U, V. The components in the universe are elements which are usually embodied by lowercase x, y. Given a universe U, a group of different elements in the universe is called a set, which is usually represented by A, B and so on. In classic set theory, the relationship of an element x with a set A has only two cases: $x \in A$ or $x \notin A$. However, the existence of vagueness in the objective world makes it impossible for the "either–or" thought in classical set theory to present all the relationships of each element within the set.

In classic set theory, an eigenfunction is used to depict the relationship between elements and a set. Each set A has an eigenfunction $C_A(x)$. If $x \in A$, then $C_A(x) = 1$; if $x \notin A$, then $C_A(x) = 0$.

$$C_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{4.1}$$

Eigenfunction $C_A(x)$ is a mapping from the universe U to a range [0, 1]. Usually it can separate elements in the set A from those outside of the set A. $C_A(x)$ is a binary value function which can only distinguish two situations, to be or not to be, and is applicable to objects with determinate definition. As it cannot distinguish the degree of membership, it is not suitable for fuzzy phenomena.

The basic idea of a fuzzy set makes the absolute affiliation relations in a classic set flexible. In the form of an eigenfunction, the grade of membership is not confined to 0 or 1, but can be any value between 0 and 1. Given a universe U and a membership function, each element x in U can be connected with a value $\mu_A(x)$ in [0, 1]. $\mu_A(x)$ is used to express the grade of membership of element x belonging to the set A. Here, A is a fuzzy set, and $\mu_A(x)$ is equivalent to eigenfunction $C_A(x)$. Its value is no longer confined to 0 and 1, and has expanded to any value between [0, 1].

Definition 4.1 Given a universe U and its mapping μ_A in the closed interval [0, 1]

$$\mu_A : U \to [0,1]$$

$$x \to \mu_A(x), x \in U$$

A fuzzy subset A in the universe U can be determined, and generally be referred to as a fuzzy set. $\mu_A(x)$ is called the grade of membership belonging to the fuzzy set A.

4.2.2 Fuzzy Set Operations

Operations between two fuzzy sets actually operate on the grade of membership point by point.

1. $\supseteq$ indicates inclusion

   Given A, B as two fuzzy sets in the universe U, if there are $\mu_A(x) \le \mu_B(x)$ for any $x \in U$, then B includes A, denoted $B \supseteq A$.

If $\mu_A(x) = \mu_B(x)$, fuzzy set *A* equals *B*, denoted *A*=*B*.

A fuzzy set with all membership at 0 is called a null set or an empty set, denoted Φ.

2. A^c indicates the complementary set of fuzzy set *A*

   Given that *A* is a fuzzy set in the universe *U*, the complementary set A^c can be defined as follows:

$$\mu_{A^c}(x) = 1 - \mu_A(x)$$

3. $A \bigcup B$ indicates the union set of fuzzy sets *A* and *B*

   Given two fuzzy sets *A*, *B* in a universe *U*, a new fuzzy set *C* is the union set of *A* and *B*. For any $x \in U$, the membership of *x* included by *C* can be determined by the larger of $\mu_A(x)$ and $\mu_B(x)$

$$C = A \bigcup B \Leftrightarrow \forall x \in U$$

$$\mu_C(x) = \max(\mu_A(x), \mu_B(x))$$

4. $A \bigcap B$ indicates the intersection of sets *A* and *B*, and can be defined as:

$$D = A \bigcap B \Leftrightarrow \forall x \in U$$

$$\mu_D(x) = \min(\mu_A(x), \mu_B(x))$$

5. Cut the operation of fuzzy set *A*

   Given that *A* is a fuzzy set in a universe *U*, for any real number $\lambda \in [0,1]$, the λ -level cut set of fuzzy set *A* is

$$A_\lambda = \{x \big| \mu_A(x) \geq \lambda, x \in U\}$$

   A_λ is a classic set. A fuzzy set is converted to a common set by the cut operation.

4.2.3 *Fuzzy Relationships*

There are various relationships in the world. The relationship between two objects is usually represented by a trenchant subset, such as terms like *x equals y* or *x is larger than y*, and so on.

Definition 4.2 Element *x* in a set *A* and element *y* in a set *B* can form an ordered pair (*x*, *y*). All these pairs (*x*, *y*) constitute a set which is a direct product of *A* and *B*, denoted $\mathrm{A} \times \mathrm{B}$.

$$\mathrm{A} \times \mathrm{B} = \{(x, y) \,|\, x \in A, y \in B\}$$

Definition 4.3 As for sets *A* and *B*, any subset *R* of their direct product $U \times V$ is called a binary relation between *A* and *B*, or simply referred to as a relation.

Given that both A and B are finite sets, the relation R can be signified as

$$R = \{r_{ij}\}_{m \times n}$$

Here, m stands for the number of elements in set A, n is the number of elements in set B, and $r_{ij} \in [0,1]$, $i = 1,2,\ldots,m$; $j = 1,2,\ldots,n$.

If R is a fuzzy set, it depicts the fuzzy relation between A and B. The value of elements of R can be defined as

$$r_{ij} = \mu_R(a_i, b_j)$$

where $\mu_R(a_i, b_j)$ stands for the grade of membership in the universe of $A \times B$.

4.2.4 Defining the Membership Functions

The membership function of a fuzzy set, usually expressed as $f_A(x)$, defines how the grade of membership of x in A is determined. There are two possible ways of deriving these membership functions (Metternicht 1999). The first approach, called the similarity relation model, resembles cluster analysis and numerical taxonomy in that the value of the membership function is a function of the classifier used (Robinson 1988). A common version of this model is the fuzzy *k*-means or *c*-means method, which is used for soil grouping, remote sensing image classification of cloud cover, and vegetation analysis (McBratney and de Gruijter 1992; Wang 1990). The second approach, known as the semantic import model, uses an a priori membership function with which individuals can be assigned a membership grade. This model is useful in situations where users have a good qualitative idea of how to group data, i.e., the exact associations of the standard Boolean model (Burrough 1989).

4.3 Applications in Previous Studies

According to the analysis above, fuzziness in spatial data can be divided into three aspects: the distribution of geographical phenomena or the concept of geographical entity, spatial relationships, and spatial analysis and spatial reasoning operations. Since fuzzy set theory was introduced into GIS, many scholars have made efforts to clarify the problems in each of these three aspects.

4.3.1 Fuzzy Representation of Geographical Entities and Their Distribution

Generally speaking, there is an implicit assumption in geographical entity modeling that the scope or boundary of spatial phenomena or entities can be defined accurately.

In a vector structure, the geometric shape of a geographical entity is represented by a point, line, or polygon which can be described accurately. The values of the attributes of such geographical entities are constant within the whole space range, such as land parcels, houses, roads, etc. However, the traditional modeling method is not appropriate when dealing with geographical entities with fuzziness in their definition and geographical distribution (called fuzzy objects) (Du et al. 2007; Schneider 1999). Fuzzy modeling can properly express fuzzy geographical objects caused by the vagueness of their geographical distribution or fuzzy definition, including natural, social, and cultural phenomena with consecutive change attributes. Schneider (1999) proposed accurate definitions for a fuzzy point, fuzzy line, and fuzzy polygon based on vector structure.

A fuzzy object is closely related to a field-based model (Du et al. 2007; Zhao et al. 2005). There are two types of field-based models: the numeric type and the category type. The numeric type of field-based model is suitable for modeling geographical phenomena whose attribute values change consecutively with location, such as topography fluctuations, gradual permeations from grassland to desert, etc. A function can be constructed to denote such an attribute value at any position. This model is a kind of numerical value of expression, and is accurate. In digital representations, field-based data models can be represented as the following continuous two-order relationship on a 2-D plane N^2:

$$R = \int_{(x,y)} \frac{\mu_R(x,y)}{(x,y)} \quad x, y \in N^2 \tag{4.2}$$

where fuzzy membership value $\mu_R(x,y)$ represents the attribute density of a surface feature character at point (x,y). That is to say, it stands for the extent to which a point belongs to one class (object). If $\mu_R(x,y)$ equals any one of both numbers {0, 1}, all the objects in real-life have crisp boundaries. If $\mu_R(x,y)$ is a numerical value in the interval [0, 1], R becomes a fuzzy set and the model can represent fuzzy geographical phenomena. This relationship can be expressed with a 2D matrix in which the row and column numbers are the coordinates of the spatial surface feature. For example, an urban area can be represented as shown in Fig. 4.1.

The values of the cells stand for the extent to which the cell belongs to the urban area. 1.0 indicates that the cell belongs entirely to the classification of urban area; $0 <$ the value <1.0 means that the cell partly belongs to the urban area; 0, the cell cannot be characterized as urban area at all. The two-order relationship reflects the field view of geographic phenomena.

In the category type of field-based model, each position belongs to different types of attribute. As the attribute types are often qualitative and discrete, the key to this model is the classification system. That is to say, each position is given an attribute type. Each pixel in this model belongs to just one category, and the degree of membership is 1. Therefore, it cannot describe partial membership of fuzzy phenomena.

The numeric type of field-based model is relatively more suitable for describing a fuzzy object. However, because of the deficiency of functions in existing GIS to process fuzzy data, its application is not possible. Clementini and Di Felice (1997)

a

0	0	0	0	0	0.1	0.1	0.1	0	0	0	0
0	0	0	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0	0
0	0.1	0.2	0.2	0.3	0.4	0.4	0.4	0.3	0.3	0.2	0.1
0.1	0.2	0.4	0.5	0.5	0.7	0.7	0.6	0.5	0.5	0.3	0.2
0.1	0.3	0.4	0.6	0.7	0.9	0.9	0.8	0.7	0.6	0.4	0.3
0.3	0.4	0.6	0.7	0.9	1.0	1.0	1.0	0.8	0.7	0.5	0.3
0.3	0.5	0.7	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.7	0.5
0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.6
0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	0.6

b

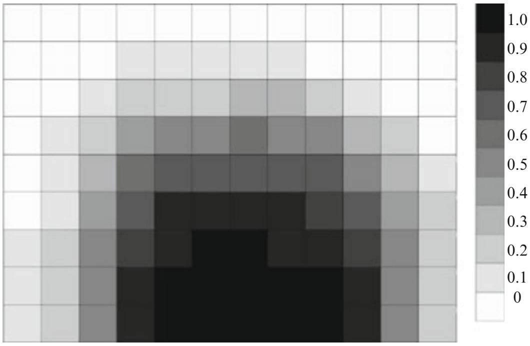


Fig. 4.1 Fuzzy representation of a spatial object—an urban area. (**a**) In numerical form. (**b**) In *gray* form. The hierarchy of *gray* values is shown on the *right* of (**b**)

put forward the concept of a broad boundary model. An object is formed by an interior, a broad boundary, and an exterior (Fig. 4.2a) (Clementini and Di Felice 1997). A broad boundary has a certain width and area, and is no longer a geometric line. The broad boundary model can usually be expressed as two areas: exterior and interior. The exterior area illustrates where the object may be located, and the interior area where the object must be located. The difference between the exterior area and the interior area is the broad boundary. This model uses a broad boundary to reflect the uncertainty of a fuzzy object. An object represented using the broad boundary model is called a broad boundary object. According to the complexity of the object, a broad boundary object can be defined as simple or complex. A simple broad boundary object is composed of a continuous interior, a continuous boundary, and a continuous exterior (see Fig. 4.2a), while a complex broad boundary object is a combination of several simple broad boundary objects. Cohn and Gotts (1996) advanced the "egg-yolk" approach to represent the uncertain area (Fig. 4.2b). The internal deep gray sub-region in the model is called the "egg-yolk", and the light gray sub-region outside is called the "egg-white" (Cohn and Gotts 1996). The "egg-white" stands for the uncertain part. The broad boundary model and the "egg-yolk" model can qualitatively describe the fuzzy extent, but cannot distinguish the membership of each pixel

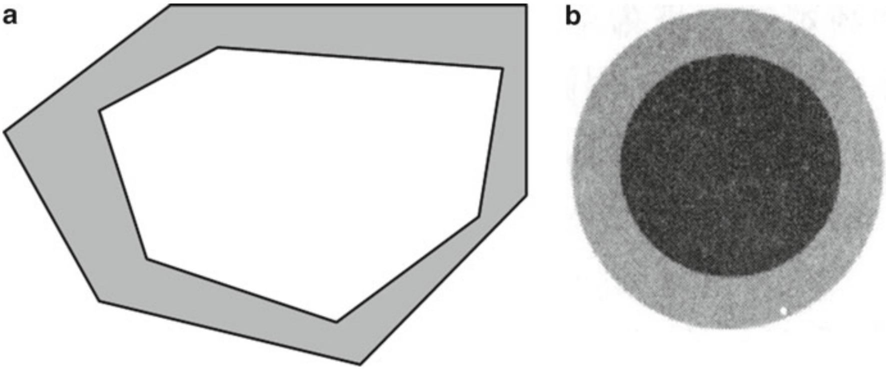


Fig. 4.2 Broad boundary region (**a**) and "egg-yolk" model (**b**) (Cohn and Gotts 1996)

belonging to the broad boundary or the "egg-white". A broad boundary region or "egg-white" can be obtained by a λ-level cut set for the fuzzy object (Fig. 4.2).

Du et al. (2007) summarized the source of a fuzzy object. There are three sources of fuzzy objects: the inherent model characteristics of the geographical phenomena (i.e., geographic distributions or geographic concepts are vague), the deficiency of spatial resolution (the geographic phenomena are accurate, but the spatial resolution is insufficient), and the derivation from existing fuzzy or non-fuzzy objects. The inherent fuzziness of geographical phenomena determines that a pixel does not completely belong to a certain category. There exists a certain transition or overlap area among categories. For the second source, as the spatial resolution in remote-sensing images is not high enough, fuzziness and hybrid pixels are generated. What the pixel represents on the ground is a synthesis of different adjacent objects. The third source comes from fuzzy operations of fuzzy or non-fuzzy objects (Stefanakis et al. 1999). The fuzzy operations include fuzzy overlay analysis, fuzzy buffer analysis, and fuzzy focus operations, etc. The result of these operations is also a kind of fuzzy object, just an outcome of the logical or arithmetic operations on the original objects. This derived object is similar to a fuzzy object in attributes and processing method except for the sources and meaning. Therefore, it can also be processed as a fuzzy object.

Cheng et al. (2001) classified fuzzy objects into three categories, fuzzy–fuzzy (FF), fuzzy–crisp (FC), and crisp–fuzzy (CF), and pixels can be classified into fuzzy objects according to different criteria (Fig. 4.3).

The FF model represents objects with uncertain thematic attributes and spatial scope. It allows different objects to overlap with each other. The FC model describes objects with a certain thematic content but uncertain spatial scope. The CF model describes objects with a certain spatial scope but uncertain thematic content. The FC and CF models are suitable for describing fuzzy objects which are separated spatially. An FC object's boundary is fuzzy with a precise interior. FC objects can overlap with each other, but CF objects cannot. Therefore, the traditional accurate object (crisp–crisp) model just describes objects with a determinate spatial scope and attribute range. According to the fuzzy object classification, Cheng et al. (2001)

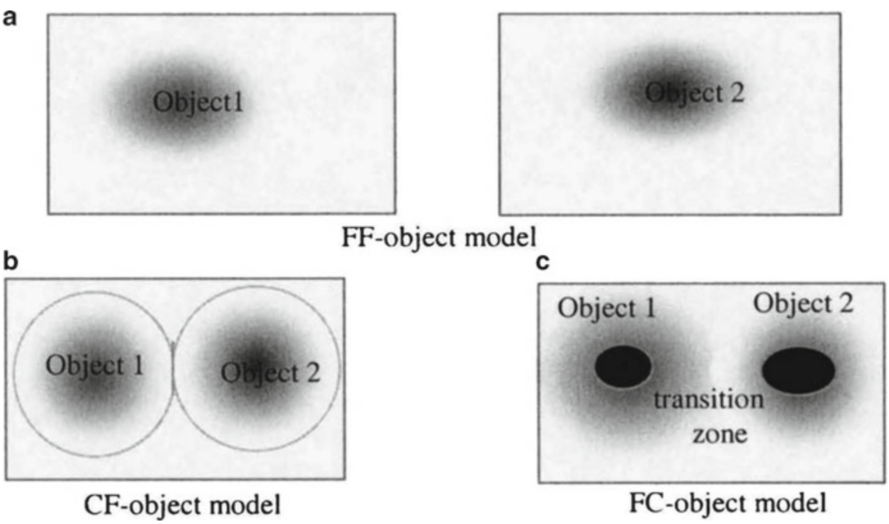


Fig. 4.3 Three fuzzy object models (**a**) FF-object model; (**b**) CF-object model; (**c**) FC-object model (Cheng et al. 2001)

defined different criteria using fuzzy and probability methods to extract fuzzy objects from the uncertainty classification results of remote sensing images.

4.3.2 Fuzzy Spatial Relationships

Spatial relationships may be caused by the geometric characteristics of spatial phenomena (the geographical position and shape of spatial phenomena) such as distance, direction, and connectivity, etc., or by the geometric and non-geometric characteristics of spatial phenomena together (including measurement attributes such as elevation value, slope values, etc., and the name attribute such as place names, etc.). For instance, the statistical correlation of spatially distributed phenomena, spatial autocorrelation, spatial interaction, spatial dependence, etc., belong to this kind of spatial relationship (Du et al. 2007). In qualitative spatial reasoning, it is common to consider the main spatial aspects of topology, direction, and distance, and to develop a system of qualitative relationships between spatial entities which cover this spatial aspect to some degree, and which appear to be useful from an application or cognitive perspective (Renz 2002). Therefore, this chapter chiefly focuses on the description of fuzzy spatial relationships such as topology, direction, and distance.

4.3.2.1 Fuzzy Description of Spatial Topology Relationships

A topological relationship refers to the property that remains the same in the process of topological transformation, such as translation, rotation, and scaling transformation, etc.

Topological relationships have always been the main content in spatial relationship research, and also an important component in spatial database queries and retrieval language. The 4-intersection model and the 9-intersection model are commonly used and accurate methods of describing topological relationships (Egenhofer and Franzosa 1991; Egenhofer and Herring 1991), and have received wide use and recognition in theoretical research and applications of GIS. The 9-intersection model can distinguish between 8 types of meaningful polygon–polygon topological relationships, 19 types of line–polygon topological relationships, and 33 types of line–line topological relationships. However, one shortcoming of the 4-intersection model and the 9-intersection model is that they can only describe topological relationships between determinate objects, while failing to describe the topological relationships of fuzzy objects.

Clementini and Di Felice (1997) replaced the mathematical boundary in the 9-intersection model with a broad boundary. The 9-intersection model derived from a combination of the interior, broad boundary, and exterior of two broad boundary objects is extended to describe fuzzy topological relationships. This is called the extended 9-intersection model. It can describe 44 topological relationships between simple broad boundary objects, and 56 topological relationships between complex broad boundary objects (Clementini and Di Felice 1997). Cohn and Gotts (1996) also proposed 46 types of topological relationships among fuzzy polygons base on the "egg-yolk" method.

In fact, a topological relationship can be formally described with a quintuple S_ Topologic (*U*, *V*, *F*, *H*, *C*) (Du et al. 2007). *U* is an object set, *V* is a conceptual set for the topological relationship, *F* is a function mapping set, *H* is a partition function set of object space, and *C* is the range of values of *F* and *H*. For any topological relationship concept $v_i \in V$, a corresponding mapping $f_i \in F$ always exists with it. The function $f_i : U \times U \rightarrow C$ represents the consistency between topological relationships and the conceptual meaning of a topological relationship for any object *A* and *B* in the set *U*. The description of a topological relationship is implemented by mapping the topological relationship of *A* and *B* through the function f_i to set *V*. Thus, topological relationships of any two objects in *U* can be described by the concept in set *V*. For instance, when a polygon–polygon topological relationship is described in the 9-intersection model, *V* = {*disjoint*, *meet*, *overlap*, *cover*, *covered-by*, *contain*, *inside*, *equal*}, each concept in *V* corresponds to a matrix. *F* is a binary logic function set used to determine which concept matrix in set *V* is the same as the topological relationships of objects *A* and *B*. The value range of C is {0, 1}. Set *H* of a partition function set is the definition of the interior point, exterior point, and boundary point in point set topology (Gaal 1964). This definition is determinate, and its range of values is {0, 1}.

For fuzzy topological relationships, set *V* is identical to the 9-intersection model. The difference between fuzzy topological relationships and classic ones exists in *F*, *H*, and *C*. Three functions $h_1(x, y)$, $h_2(x, y)$, $h_3(x, y)$ in *H* define the membership of point (*x*, *y*) belonging to the interior, exterior, and boundary respectively, and each range of values is extended to [0, 1]. The value range of *C* in the function *F* $f_i : U \times U \rightarrow C$ is [0, 1], and the function $f_i : U \times U \rightarrow C$ is used to determine the

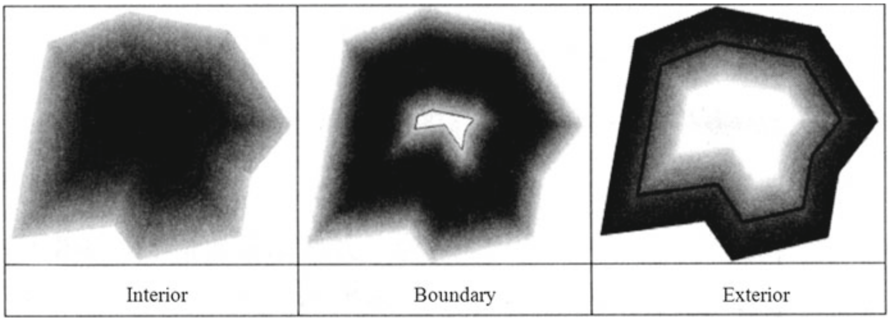


Fig. 4.4 Fuzzy partition of a polygonal object in a topological universe (Du et al. 2007)

topological relationships of objects *A* and *B* and the membership of each concept in set *V*.

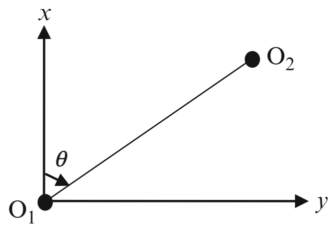
In the case of polygonal objects, the definitions of interior point, exterior point, and boundary point in point set topology can be used to divide fuzzy space into three fuzzy sets as a boundary region, an interior region, and an exterior region (Du et al. 2007) (Fig. 4.4). As influenced by the uncertainty of spatial data, the boundary of a polygonal object is not simply referred to as its boundary with coordinates arrayed in a vector structure or a sequence of grids after the rasterization of the boundary in the raster structure. The boundary is extended to become a region spreading inward and outward to the object. The farther away from the boundary of an object a pixel is, the smaller the degree of membership of belonging to the object boundary becomes. The maximum membership degree is 1.0 at the polygon boundary.

4.3.2.2 Fuzzy Representation of Spatial Direction

Directional—also called orientational—relationships of spatial entities with respect to other spatial entities is usually given in terms of a qualitative category such as "*to the north of*" rather than using a numerical expression such as "*12 degrees*" (which is certainly more common in technical communications such as aviation). These are important and common-sense linguistic and qualitative properties used in everyday situations and qualitative spatial reasoning (Frank 1996). The direction of spatial entities is a ternary relationship depending on the located object, the reference object, and the frame of reference, which can be specified either by a third object or by a given direction. In the literature, one distinguishes between three different kinds of frames of reference, extrinsic ("external factors impose a direction on the reference object"), intrinsic ("the direction is given by some inherent property of the reference object"), and deictic ("the direction is imposed by the point of view from which the reference object is seen") (Hernàndez 1994). Given the frame of reference, directions can be expressed in terms of binary relationships with respect to the frame.

Most approaches to dealing with direction qualitatively are based on points as the basic spatial entities and consider only two-dimensional space. Frank (1991)

Fig. 4.5 Definition of azimuth θ from object O_1 to O_2



suggested different methods for describing the cardinal direction of a point with respect to a reference point in a geographic space, i.e., directions are in the form of [north, east, south, west] depending on the granularity (Frank 1991). Zhao et al. (2005) proposed a spatial direction model based on trigonometric functions. In this model, all objects are considered as a point—even those with irregular shape and size—and follow Frank's suggestion about directions (a centroid-based method, where the direction between two objects is determined by the angle between their centroids) (Fig. 4.5). The azimuth θ from object O_1 to object O_2 is computed. This angle, denoted by $\theta(O_1, O_2)$, takes values in $[0, 2\pi]$, which constitutes the universe on which primitive directional relations are defined. $Sin^2(\theta)$ and $cos^2(\theta)$ are chosen as fuzzy membership functions to describe the direction [north, east, south, west] with reference to the relative position relation functions proposed by Miyajima and Ralescu (1994) (Fig. 4.6). Miyajima and Ralescu (1994) used the square trigonometric function to illustrate the relative position relations [above, right, below, left] of segmented images. Square trigonometric functions are also suitable for directions in the form of [north, east, south, west] (Miyajima and Ralescu 1994). For instance, in Figure 4.5, if $\theta = 50°$, then the direction relationship is [0.4132, 0.5868, 0, 0] in the form of [north, east, south, west] according to Eqs. (4.3)–(4.6). This means that object O_2 is located to the north of object O_1 with 0.4132 of membership degree, and to the east with 0.5868 of membership degree. That is, $\mu_{\text{north}}(O_1, O_2) = 0.4132$, $\mu_{\text{east}}(O_1, O_2) = 0.5868$, $\mu_{\text{south}}(O_1, O_2) = 0$, $\mu_{\text{west}}(O_1, O_2) = 0$, and $\mu_{\text{north}}(O_1, O_2) + \mu_{\text{east}}(O_1, O_2) + \mu_{\text{south}}(O_1, O_2) + \mu_{\text{west}}(O_1, O_2) = 1$. Therefore, fuzzy membership functions not only show the characteristics of transition of the directional relationship, but also ensure the integrity of the definition of the direction for any target object.

$$\mu_{north}(\theta) = \begin{cases} \cos^2(\theta), & \frac{3\pi}{2} \le \theta \le 2\pi \; or \; 0 \le \theta \le \frac{\pi}{2} \\ 0, & \frac{\pi}{2} \le \theta \le \frac{3\pi}{2} \end{cases} \tag{4.3}$$

$$\mu_{east}(\theta) = \begin{cases} \sin^2(\theta), & 0 \le \theta \le \pi \\ 0, & \pi \le \theta \le 2\pi \end{cases} \tag{4.4}$$

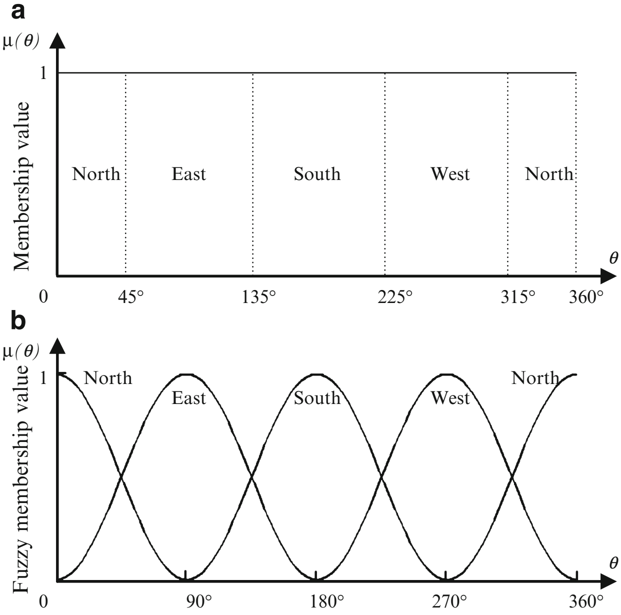


Fig. 4.6 An example of the classic (**a**) and fuzzy (**b**) classification of direction. θ stands for the azimuth from object O_1 to O_2

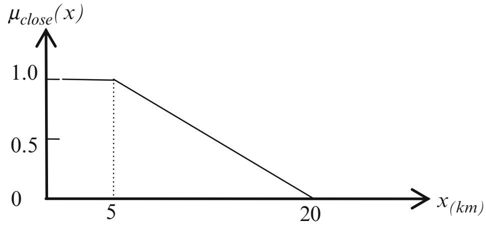
$$\mu_{south}(\theta) = \begin{cases} \cos^2(\theta), & \frac{\pi}{2} \le \theta \le \frac{3\pi}{2} \\ 0, & \frac{3\pi}{2} \le \theta \le 2\pi \; or \; 0 \le \theta \le \frac{\pi}{2} \end{cases} \tag{4.5}$$

$$\mu_{west}(\theta) = \begin{cases} \sin^2(\theta), & \pi \le \theta \le 2\pi \\ 0, & 0 \le \theta \le \pi \end{cases} \tag{4.6}$$

4.3.2.3 Fuzzy Description of Spatial Distance

In a spatial decision-making process, the distance relation between spatial entities always plays a key role. Dealing with distance is an important cognitive ability in our everyday life (Renz 2002). When representing distance, we usually use qualitative categories such as "*A is close to B*" (binary constraint) or qualitative distance comparatives such as "*A is closer to B than to C*" (ternary constraint), but sometimes also numerical values such as "*A is about 20 m away from B*". One can distinguish between absolute distance relations (the distance between two spatial entities) and relative distance relations (the distance between two spatial entities as compared with the distance to a third entity) (Guesgen and Albrecht 2000; Renz 2002).

Fig. 4.7 Membership function of "close to a city"



The choice of which relation should be used depends on the application universe and the requirements posed by decision-makes. For two individual locations *A* and *B*, which in general are abstracted as points, the Euclidean distance is given by the formula

$$d(A,B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \tag{4.7}$$

where (x_A, y_A) and (x_B, y_B) denote the coordinates of two locations *A* and *B,* respectively.

Qualitative absolute distance relations are obtained, for example, by dividing the real line of distance into several sectors such as "*very close*," "*close*," "*commensurate*," "*far*," and "*very far*" depending on the chosen level of granularity (Hernàndez et al. 1995). In practice, we usually use one of the sectors. For instance, Figure 4.7 represents the "*close*" degree from a point on a map to a city.

$$\mu_{close}(x) = \begin{cases} 1, & x \le 5 \\ (20 - x)/15 & 5 < x \le 20 \\ 0, & x > 20 \end{cases} \tag{4.8}$$

where *x* denotes the distance (in kilometers) from the location to the city. The division values such as 5 km and 20 km are designed arbitrarily by decision-makers according to the understanding of their definition of a "*close*" degree.

4.3.3 Fuzzy Operations on Spatial Reasoning

Spatial reasoning is an approach for reasoning out unknown spatial relationships based on the determinate spatial relationships of objects. The reasoning is implemented through a symbolic operation based on implicit knowledge and rules. Spatial reasoning is a hot topic in fields such as GIS, artificial intelligence, and computer vision, and many spatial reasoning methods have been put forward. Du et al. (2007) systematically summarized the classification systems of spatial reasoning methods. Existing spatial reasoning methods are mainly concentrated in a single spatial relationship (Du et al. 2007) such as topological relationship reasoning by topological relationships, direction relationship reasoning by direction relationships, and so on. Fewer combinatorial spatial reasoning methods were presented. In addition, most of these spatial reasoning methods were developed for determinate geographical objects or spatial relationships, and sometimes it is difficult to adopt these methods for spatial relationships with uncertainty. Therefore, these methods need an extension and supplement for fuzzy geographic objects or fuzzy spatial relationships.

Hong et al. (1995) have researched combinatorial spatial reasoning methods for direction and distance relationships. The distance and direction between *A* and *C* can be reasoned according to the distance and direction between *A* and *B* and those between *B* and *C* (Hong et al. 1995). Sharma (1996) used a projection model to describe direction relationships, a 9-intersection model to describe topological relationships, and an interval model to describe qualitative distance relationships. Then single, combinatorial, and integrated spatial reasoning methods were proposed (Sharma 1996). As he used a projection model to describe a direction relationship between polygonal objects, this method cannot exactly express the actual directional relationship of such polygonal objects. In particular, because he focused on spatial reasoning for two single direction relationships rather than for single to multinomial direction relationships, multinomial to single direction relationships, and multinomial to multinomial direction relationships, the efficiency and accuracy of spatial reasoning results are limited.

A broad boundary model has been used to describe fuzzy objects (Clementini and Di Felice 1997; Worboys and Clementini 2001). Clementini and Di Felice (1997) proposed 44 types of topological relationships using the extended 9-intersection model based on the broad boundary model, and Cohn and Gotts (1996) put forward 46 types of topological relationships using the "egg-yolk" theory. As the number of topological relationships based on the broad boundary model or the "egg-yolk" theory goes beyond the range of cognition of a human being, such methods are inconvenient for topological relationship reasoning. Some scholars have also put forward advanced models to extend the fuzzy spatial reasoning method. Du et al. (2007) proposed a quadruple model to describe the topological relationship of objects with broad boundaries, and topological relationship reasoning was implemented based on the quadruple model.

Zhao et al. (2005) proposed a field-based integrated spatial reasoning model for the case of the constraint satisfaction problem (CSP). They argued that knowledge about spatial entities or about the relationships between spatial entities are often given in the form of constraints. Ordinarily, binary constraints such as "*the primary school should be laid out in the north of the residential area*," ternary constraints such as "*the primary school should be laid out between residential area A and residential area B*," or in general, *n*-ary constraints restrict the universe of 2, 3, or *n* variables. Problems like these are formalized as a constraint satisfaction problem: given a set of variables *R* over a universe *D* and a set *A* of constraints on the variables *R* (Renz 2002). CSP is a powerful general framework in which a variety of combinatorial problems can be expressed (Creignou et al. 2001; Marriott and Stuckey 1998). The aim of CSP is to assign values to the variables subject to specified constraints. In fact, it is the most popular reasoning

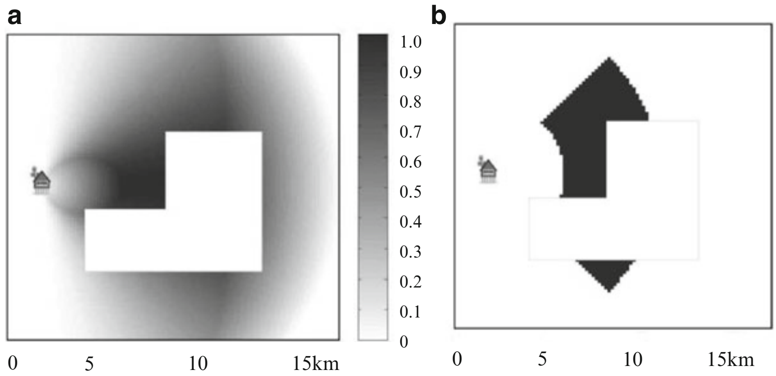


Fig. 4.8 The result of combinatorial spatial reasoning (**a**) and of traditional spatial reasoning (**b**)

method used in qualitative spatial reasoning (Renz 2002), and a common problem in spatial decision-making processes such as the examples described above.

Ladkin and Maddux (1994) formulated binary CSPs as relation algebras as developed by Tarski (1941). This allows binary CSPs to be treated in a uniform way. In a fuzzy domain, the relation algebras constitute fuzzy logic reasoning. Fuzzy logic reasoning, one of the application domains of fuzzy relationship generalization, is the fundamental basis of fuzzy spatial reasoning. It implements tasks through logical operations based on normal relation algebra theory. The operations can be extended to *n* sets of fuzzy relationships, i.e., the operation is applicable to multiple fuzzy sets. Assuming that there are *n* sets of fuzzy relationships, operations can be expressed uniformly as

$$\mu_R(z) = \bigotimes_{i=1}^{n} \mu_{A_i}(z) \quad z \in C \tag{4.9}$$

where ⊗ denotes operators *union*, *intersection*, and c*omplement,* respectively, and A_i stands for multiple fuzzy relationships.

Zhao et al. (2005) used this model to implement a combinatorial fuzzy reasoning including direction and distance relationships. A task in combinatorial fuzzy reasoning is to find a suitable location for a special factory given certain constraining factors.

(a) The factory must be located to the east of the environmental monitoring station.
(b) The factory must not be far from the environmental monitoring station.
(c) The factory must not be situated on land suitable for agriculture.

They compared the results of the combinatorial fuzzy reasoning model (Fig. 4.8a) with those of traditional spatial reasoning (Fig. 4.8b). It is evident that the information in the combinatorial fuzzy reasoning result is more abundant and more detailed than that in the result of the traditional approach. The combinatorial fuzzy reasoning model gives decision-makers more chances to choose a suitable result as it provides the degree of suitability to the proposition proposed by the users.

4.4 Conclusion and Future Prospects

Fuzziness is an inherent characteristic in nature and in the language of human beings. As GIS can model and digitize the real world, it needs to be able to deal with the fuzziness existing in the real world. In addition, the model of the real world in GIS is based on the cognition of humans, and the concept and methods of modeling and analysis embedded in GIS software are always influenced by human cognition. Spatial decision-making support provided by GIS should relate to human cognition. Accordingly, the representation and analysis of the real world in GIS should be geared to human language, so that the representation, analysis process, and results can completely satisfy the natural mode of expression of humans.

Scholars began to pay attention to the quality of the spatial data in GIS soon after the emergence of GIS in the 1960s, and they adopted fuzzy set theory to deal with spatial geographical phenomena and spatial analysis methods. In the early years, the application of fuzzy set theory in GIS emphasized the analysis of the sources of fuzzy phenomena in spatial data and spatial analysis methods. These sources include the fuzziness in three aspects of geographical distributions, or the concept of spatial entity, spatial relationships, and spatial reasoning. The exploration of sources indicated the research directions for the application of fuzzy set theory in GIS. After that, methods of representing fuzzy geographical objects in GIS were explored based on vector and raster structures. However, most of these representational methods can be deemed to be concept models, and do not deviate from traditional spatial data structure. In fact, field-based raster models or vector models of fuzzy geographical objects are always more complex than traditional data structure. Therefore, such models cannot be widely used for storing fuzzy spatial data. As spatial relationships and spatial reasoning are based on determinate geographical objects, research into them was also influenced by representational methods of spatial relationships and spatial reasoning. Researchers have not abandoned the struggle with fuzzy spatial data. The broad boundary model and the "egg-yolk" theory were proposed in order to extend traditional topological relationships, and the qualitative directional and distance relationships of fuzzy objects were also researched using fuzzy set theory and membership functions. In particular, some fuzzy reasoning methods have been put forward to solve fuzzy geographical problems.

These, however, are only the beginning of longer term collaborative research efforts. The application of fuzzy set theory in GIS is still developing. Both fuzzy mathematics itself and the concept of fuzzy geographical objects and analysis need innovation.

1. Further work will have to emphasize research into the measurement of fuzzy uncertainty. Based on this idea, a membership function model should be constructed for the fuzzy uncertainty of spatial data derived from different sources.
2. Based on geographical ontology, formalized methods of expression should be explored for fuzzy objects. In addition to traditional vector and raster structures, we may find a new integrative data structure in which fuzzy information can be stored and visualized conveniently in computer systems.

3. Further research should also focus on the formal definitions of combinatorial fuzzy spatial reasoning operations and predicates, with the integration of fuzzy spatial data types into query languages, and with aspects of implementation leading to sophisticated data structures for efficient algorithms for the operations.
4. The promulgation of fuzzy information in the process of spatial reasoning is a key point for users who pay attention to the results of spatial analysis. We will have to explore the law of promulgation and to assess its effect on the results of spatial analysis. In that way we could provide users with a confidence range of fuzzy spatial reasoning results.

References

Burrough PA (1989) Fuzzy mathematical methods for soil survey and land evaluation. J Soil Sci 40:477–492

Burrough PA, Frank AU (1996) Geographic objects with indeterminate boundaries. Taylor & Francis, London

Burrough PA, Heuvelink GBM (1992) The sensitivity of Boolean and continuous (Fuzzy) logical modeling to uncertain data In: Proceedings of 3rd European conference on geographical information systems, Munich, Germany, pp 1032–1041

Burrough PA, van Gaans PFM, Hootsmans R (1997) Continuous classification in soil survey: spatial correlation, confusion and boundaries. Geoderma 77:115–135

Cheng T, Molenaar M, Lin H (2001) Formalizing fuzzy objects from uncertain classification results. Int J Geogr Inf Sci 15:27–42

Clementini E, Di Felice P (1997) Approximate topological relations. Int J Approx Reason 16:173–204

Cohn A, Gotts N (1996) The "Egg-Yolk" representation of regions with indeterminate boundaries. In: Burrough P, Frank AU (eds) Geographic objects with indeterminate boundaries. Taylor and Francis, London, pp 171–187

Creignou N, Khanna S, Sudan M (2001) Complexity classifications of Boolean constraint satisfaction problems. SIAM Monographs on Discrete Mathematics and Applications, Philadelphia

Doucette M, Paresi C (2000) Quality management in GDI. In: Groot R, MacLaughlin J (eds) Geospatial data infrastructure: concepts, cases, and good practice. Oxford University Press, Oxford, pp 85–96

Du S, Wang Q, Qin Q (2007) Representation and combined reasoning of spatial relationships. Science Press (in Chinese), Beijing

Egenhofer MJ, Franzosa R (1991) Point-set topological spatial relations. Int J Geogr Inf Syst 5:161–174

Egenhofer MJ, Herring J (1991) Categorizing binary topological relations between regions, lines and points in geographic databases. Technical Report, Department of Surveying Engineering, University of Maine

Fisher P (2000) Sorites paradox and vague geographies. Fuzzy Set Syst 113:7–18

Fisher PF (1999) Models of uncertainty in spatial data. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW (eds) Geographical information systems: principles and technical issues, vol 1. John Wiley & Sons, New York, pp 191–205

Frank AU (1991). Qualitative spatial reasoning about cardinal directions. In: Proceedings of Auto-Carto 10, Baltimore, Maryland, pp 148–167

Frank AU (1996) Qualitative spatial reasoning: cardinal directions as an example. Int J Geogr Inf Syst 10:269–290

Gaal SA (1964) Point set topology. Academic, Duluth

Goodchild MF (1992) Geographical information science. Int J Geogr Inf Syst 6:31–46
Goodchild MF (1999) Future directions in geographic information science. Geogr Inform Sci 5:1–8
Goodchild MF, Dubuc O (1987) A model of error for choropleth maps with applications to geographic information systems. In: Proceedings of AUROCARTO8, Falls Church, pp 165–174
Goodchild MF, Gopal S (1989) Accuracy of spatial database. Taylor & Francis, London
Goodchild MF, Jeansoulin R (1998) Data quality in geographic information – from error to uncertainty. Hermes, Paris
Guesgen HW, Albrecht J (2000) Imprecise reasoning in geographic information systems. Fuzzy Sets Syst 113:121–131
Hernàndez D (1994) Qualitative representation of spatial knowledge. Sparger-Verlag, Berlin
Hernàndez D, Clementini E, Di Felice P (1995) Qualitative distances. Lect Notes Comput Sci 998:45–57
Heuvelink G, Burrough P, Stein A (1989) Propagation of errors in spatial modelling with GIS. Int J Geogr Inf Syst 3:303–322
Hong J, Egenhofer M, Frank AU (1995) On the robustness of qualitative distance and directions reasoning. In: The 12th international symposium on computer assisted cartography, Charlotte, North Carolina, pp 301–310
International Organization for Standardization (ISO) (1995) Guide to the expression of uncertainty in measurement. ISO, Geneva
Ladkin PB, Maddux R (1994) On binary constraint problems. J ACM 41:435–469
Marriott K, Stuckey P (1998) Programming with constraints: an introduction. MIT Press, Cambridge
McBratney A, de Gruijter J (1992) A continuum approach to soil classification by modified fuzzy k-means with extra grades. J Soil Sci 43:159–175
Metternicht G (1999) Change detection assessment using fuzzy sets and remotely sensed data: an application of topographic map revision. ISPRS J Photogramm Remote Sens 54:221–233
Miyajima K, Ralescu A (1994) Spatial organization in 2D segmented images: representation and recognition of primitive spatial relations. Fuzzy Set Syst 65:225–236
Morrison JL (1995) Spatial data quality. In: Guptill SC, Morrison JL (eds) Elements of spatial data quality. Elsevier Science, Tokyo, pp 1–12
Mowrer TH, Congalton RG (2000) Quantifying spatial uncertainty in natural resources: theory and applications for GIS and remote sensing. Taylor & Francis, London
Östman A (1997) The specification and evaluation of spatial data quality. In: Proceedings of the 18th ICA international cartographic conference. Swedish Cartographic Society, Sweden, pp 836–847
Renz J (2002) Qualitative spatial reasoning with topological information. Springer, Berlin
Robinson VB (1988) Some implications of fuzzy set theory applied to geographic databases. Comput Environ Urban Syst 12:89–97
Schneider M (1999) Uncertainty management for spatial data in database: fuzzy spatial data types. Lect Notes Comput Sci 1651:330–351
Sharma J (1996) Integrated spatial reasoning in geographic information systems: combining topology and direction. Ph.D. Dissertation, Department of Surveying Engineering, University of Maine
Shi W, Fisher P, Goodchild MF (2002) Spatial data quality. CRC Press, London
Stefanakis E, Vazirgiannis M, Sellis T (1999) Incorporating fuzzy set methodologies in a DBMS repository for the application domain of GIS. Int J Geogr Inf Sci 13:657–675
Tarski A (1941) On the calculus of relations. J Symb Log 6:73–89
Wang F (1990) Improving remote sensing image analysis through fuzzy information representation. Photogramm Eng Remote Sens 56:1163–1169
Worboys M, Clementini E (2001) Integration of imperfect spatial information. J Vis Lang Comput 12:61–80
Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
Zhao Y, Murayama Y, Zhang Y (2005) Field-based fuzzy spatial reasoning model for geographical information systems: case of constraint satisfaction problem. Theory Appl GIS 13:21–31