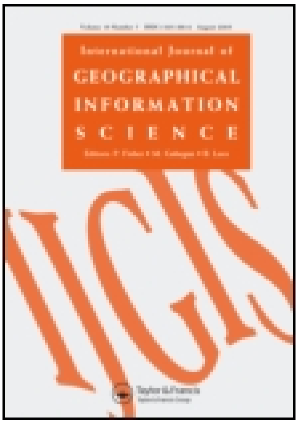


International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgis20

Positioning localities based on spatial assertions

Y. Liu [a b] , Q. H. Guo [b] , J. Wieczorek [c] & M. F. Goodchild [d]

[a] Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing 100871, China

[b] School of Engineering and Sierra Nevada Research Institute, University of California, Merced, CA 95344, USA

[c] Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA

[d] National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA

Available online: 23 Oct 2009

To cite this article: Y. Liu, Q. H. Guo, J. Wieczorek & M. F. Goodchild (2009): Positioning localities based on spatial assertions, International Journal of Geographical Information Science, 23:11, 1471-1501

To link to this article: http://dx.doi.org/10.1080/13658810802247114

International Journal of Geographical Information Science
Vol. 23, No. 11, November 2009, 1471–1501

Taylor & Francis
Taylor & Francis Group

Research Article

Positioning localities based on spatial assertions

Y. LIU*†‡, Q. H. GUO‡, J. WIECZOREK¶ and M. F. GOODCHILD§

†Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing 100871, China

‡School of Engineering and Sierra Nevada Research Institute, University of California, Merced, CA 95344, USA

§Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA

¶National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA

(*Received 25 September 2007; in final form 1 June 2008*)

In practice, descriptive localities are often communicated using named places and spatial relationships. Uncertainty associated with such descriptions of localities is inevitable, and knowledge of such uncertainty is normally not explicit. When translating descriptive localities into spatially explicit ones, it is critical to circumscribe locations and to estimate the associated uncertainty based on a set of appropriate spatial relationships. In conventional research on qualitative spatial reasoning (QSR), spatial relationships are modeled using formal logic. Unfortunately, QSR cannot deal with the uncertainty of a position. In this paper, based on the conceptual model of spatial assertions, we introduce the uncertainty field model to represent the probability distribution of a point locality. Using probability operations, we can combine a set of assertions to position a locality. Conflicts among assertions for a single locality can be detected based on the resulting field. Since spatial relationships play an important role in the uncertainty of target objects, we investigate conceptually the uncertainty fields associated with various types of spatial relationships (for example, topological, directional and metric). In a concrete application, these uncertainty fields can be customized and used without altering the proposed framework.

Keywords: Geographic information system; Spatial positioning; Probability; Uncertainty field; Spatial relationship

1. Introduction

Geographic information sciences and technologies have entered an era of massive spatial data sets with numerous algorithms available for a variety of applications. Three classes of geographic information can be identified according to their sources and forms: multiple (spatial, temporal, or spectral) resolution remotely sensed imagery, data sets derived from analog maps, and non-structured or semi-structured documents organized based on place names. The first two classes have been extensively studied. Many software packages, both proprietary and open source, have been developed to manage and analyze them. Due to the ubiquitous use of the

*Corresponding author. Email: liuyu@urban.pku.edu.cn

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810802247114

World Wide Web (WWW) (Goodchild 2007), it has become urgent to try to deal with text-based information, which represents human spatial knowledge in non-analytical and non-explicit forms. Tools and techniques such as digital gazetteers (Hill *et al.* 1999; Schlieder *et al.* 2001; Hill 2006), geographic information retrieval (GIR) (Silva *et al.* 2006; Purves *et al.* 2007; Jones and Purves 2008), and qualitative spatial reasoning (QSR) (Frank 1996; Cohn and Hazarika 2001; Renz 2002) are useful in dealing with such textual information. Geographic features (e.g. cities) and their associated place names (e.g. 'San Francisco') are essential for these techniques.

Place names roughly describe localities such that we can communicate geospatial knowledge based on our shared knowledge of them. In geographic information systems (GIS), a spatial reference system (SRS) provides a specific way to describe a position on the surface of the earth. Named places thus form a typical example of a qualitative SRS. Although rough and sometimes ambiguous (Longley *et al.* 2005), such an SRS provides an efficient approach to communicate spatial information in everyday life. In practice, spatial assertions that contain one or more place names and spatial relationships are also widely used (Yao and Thill 2006; Bennett and Agarwal 2007).

Compared with numeric geographic coordinates, textual locality descriptions can be seen as qualitative or semi-quantitative, representing a rough distribution range of the locality based on the understanding of the person who recorded the information and subject to the interpretation of the person reading it. Hence, uncertainty associated with these descriptions is inevitable. In many applications, it is necessary to determine localities and associated uncertainty based on a qualitative description. To address and make use of the vast amount of descriptive locality data in museum collections, Wieczorek *et al.* (2004) developed a point radius method to describe a locality using a point with its maximum error. Guo *et al.* (2008) refined the point-radius method by taking into consideration the actual shape of the reference object, and positioning the locality with the uncertainty probability surface. Spatial analysis without incorporating or considering uncertainty would be of limited utility (Fisher 1999). For example, Rowe (2005), who used locality uncertainty to study its impact on species distribution patterns, found that results without taking locality uncertainty into consideration would be misleading.

Spatial relationships play an important role in GIS as well as qualitative spatial reasoning (QSR), and are important common attributes of locality descriptions. As pointed out by Cohn and Hazarika (2001), QSR deals with spatial relationships to make predictions about spatial objects when precise quantitative information is not available. Based on a set of known constraints that include spatial relationship instances, QSR seeks to: (1) infer unknown spatial relationships; (2) check inconsistencies among the constraints; and (3) find consistent scenarios. Most current research on spatial reasoning is based on formal logic. In terms of spatial positioning, Dehak *et al.* (2005) proposed a probabilistic approach to deducing relative position based on incomplete information represented by distance and direction relationships. It encourages us that these problems can be solved in a probabilistic manner.

Two additional characteristics of spatial relationships should be considered in spatial positioning. First, some qualitative relationships, such as 'far', are inherently vague. Dutta (1990, 1991) developed a semi-quantitative approach to spatial reasoning using fuzzy logic. Du *et al.* (2004) argued that cardinal direction relationships also belong to the realm of fuzzy concepts, and established their corresponding membership functions. Second, a number of spatial predicates used in the interpretation of locality descriptions, such as those found in the descriptions 'between highway 99 and highway 5', and '10 km north of Merced along highway 99', are different from those studied in QSR, which focuses on binary relationships between two objects. Note that in this research, for demonstration purposes, we used several place names from the state of California in the United States (Figure 1); however, this choice will not influence our proposed methods.

Field and object models have gained acceptance as two alternative approaches to conceptualizing and modeling geographic phenomena (Goodchild 1992). Since the field model is more suitable for uncertain features (Goodchild 1989), Guo *et al.* (2008) proposed the concept of an uncertainty field to represent a locality with uncertainty in natural museum records (MaNIS 2001). Guo *et al.* (2008) studied six

Figure 1. Several place names in California that are mentioned in this research.

uncertainty factors that are related to museum collection data: spatial extent of reference objects, geodetic datum, imprecision of coordinates, map scale, imprecision in distance measurements and imprecision in direction measurements. In addition, they investigated the impact of these factors on estimating the uncertainty for descriptive localities, such as ‘offset from a feature (or a path) at a heading’ (e.g. 5 miles north of Berkeley). As mentioned earlier, spatial relationships based assertions, which are similar to locality descriptions in MaNIS, are widely used in representing geographic knowledge. In this research, we therefore extend the research by Guo *et al.* (2008), and aim to develop a general and more complete positioning method that is suitable for a variety of textual descriptions, specifically on the following two aspects:

(1) We propose a four-level conceptual model of the uncertainty in spatial assertions, and further investigate the characteristics and operations of uncertainty fields. Two operation methods, refinement and integration, were studied and compared. Refinement can be used to reduce uncertainty about the spatial extent of the locality based on a group of independent, certain assertions about it; while integration can be used to compose mutually exclusive uncertain predicates.
(2) Since a spatial relationship is a key component in the assertion, we discuss the context that an observer makes a predicate, such as ‘close to Berkeley’, for describing a locality. Two assumptions are proposed to deal with spatial relationships, which will result in different uncertainty fields. Following these two assumptions, uncertainty fields associated with topological, directional, and metric relationships are discussed. Additionally, we study two special cases of spatial assertions (i.e. the metric relationship in linear SRSs and a ternary relationship ‘between’) and present corresponding uncertainty fields.

To illustrate the concepts of uncertainty fields and their related operations, we developed a software toolbox to combine different types of spatial assertions and compute the probability distribution of a locality.

2. Modeling spatial assertions

2.1 *Reference objects, target objects and spatial relationships*

As argued by Guo *et al.* (2008), three components should be considered in spatial assertions to position a locality. They are reference objects (RO), target objects (TO) and spatial relationships.

Most locality descriptions are based on at least one named place, which acts as a reference object for positioning a locality. Bennett and Agarwal (2007) used the term ‘anchor’ for an RO. Theoretically, any natural or artificial feature with a name can be used as an RO. In practice, however, a feature should be significant and stable enough to be identified as an RO. From the perspective of topological dimension of an RO, it could be a point, a line, or an area object.

The locality for which a position is to be determined is called a TO, which is determined based on the position and the shape of an RO, as well as spatial relationships between the TO and the RO. In a two-dimensional space, TOs may be point, line, or area features. For linear or areal TOs, their shapes should ideally be considered explicitly. In practice, however, qualitative assertions seldom describe the actual shape of a TO. Hence, the TOs can be abstracted as point features, while the positioned point TO can be seen as the centroid of a line or area object (Guo *et al.* 2008). When the TOs are points, they can be further divided into two sub-classes: single-point TO and multi-point TO. A typical example of multi-point TO is 'two observation points close to Berkeley'. In practice, the multi-point TO problem can be decomposed to a group of single-point TO problems if these target objects are independent. In addition to the topological dimension, a TO may be typed or typeless according to the absence or presence of geographical type of the TO. An example of a typed TO is 'a hotel close to Berkeley'. Yao and Thill (2006) have studied such expressions. A typed TO may also be expressed based on a field, such as 'a point higher than 2000 feet in California'. Generally, positioning a typed TO should take geographical facts into account. In this research, we focus on typeless TOs, which are abstracted to simple geometries, especially points.

A spatial relationship provides a rough distribution range of the TO with respect to the RO. The uncertainty of a TO derives predominantly from the spatial relationships, especially when the RO is well defined. The common types of spatial relationships are topological, directional and metric (Renz 2002). Although there are some exceptions, the spatial relationships used in most locality descriptions are covered by these three categories as follows.

(1) In localities with topological relationships, such as 'inside California', the reference object describes the distribution range of the TO. This category is very common in textual locality descriptions. For example, 60.2% of the locality descriptions processed in the MaNIS project (MaNIS 2001) contained this locality type, which is simply called 'named place'. Note that in MaNIS, topological relationships are not expressed explicitly. For example, 'Springfield' implies 'inside Springfield'. Features without names, such as intersections or addresses, also fall into this category, where ROs can be viewed as point features.
(2) In localities with directional relationships, the directions might be external, such as 'west of Tucson', or internal (Liu *et al.* 2005), such as 'north part of Mono Lake', to the RO. External direction relationships can be viewed as a refinement of the topological relationship 'disjoint from' (or 'outside of'); while internal direction relationships refine the topological relationship 'contains' (or 'within').
(3) Metric relationships may be expressed by a quantitative ('5 km outside Calgary') or qualitative distance ('Big Bay vicinity'). In qualitative cases, a series of distinctions may be used, such as 'close', 'far', etc. In practice, the distance may be measured based on a straight connecting line (shortest distance) or on a real-world linear feature. At times it is unclear which of these two cases was meant in the original locality description. For the latter case, the shape of the linear RO should be considered in order to position the TO.

Direction and distance relationships are generally too rough to make a specific determination of a locality. Therefore, they are usually combined to form a more precise description in practice, as well as in QSR research (Clementini *et al.* 1997). A typical example is '10 km north of Kuala Lumpur'. In Sections 3.2–3.4, we will investigate these three types of relationships in detail.

Normally, only one RO is involved in a locality description, for example, 'Big Bay vicinity', which can be abstracted to a diagram as shown in Figure 2(a). However,

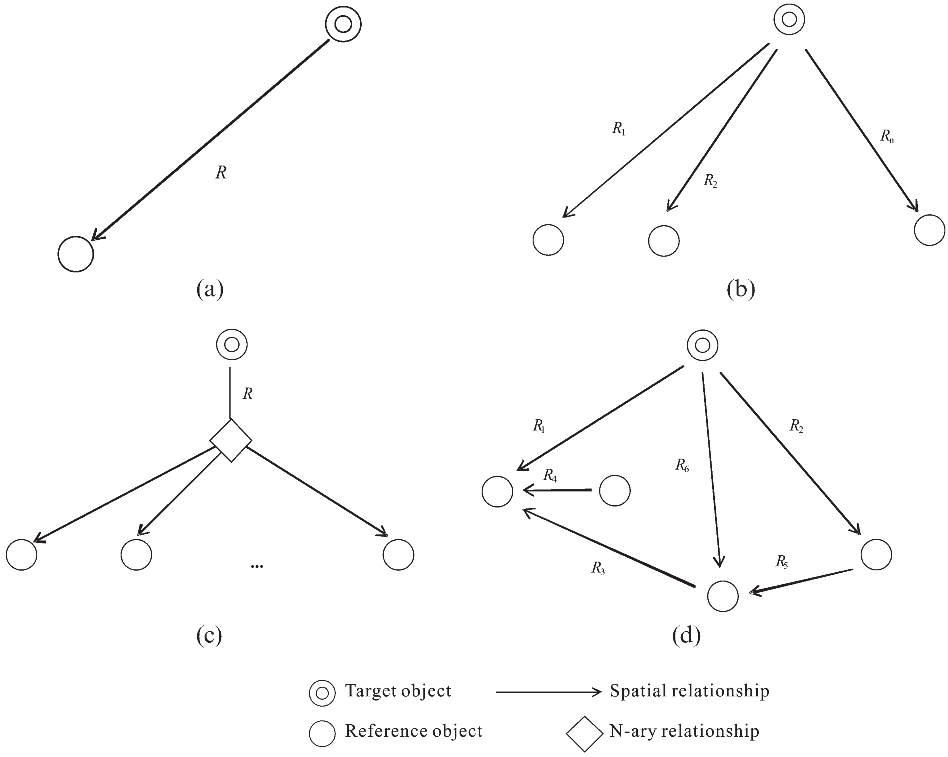


Figure 2. Conceptual diagrams of four types of spatial assertion.

the following four multi-RO cases are also commonly found in museum collection data (Guo *et al.* 2008) as well as other real-world locality descriptions.

(1) An RO is described by two or more other objects. For example, in the description 'area at junction of Black River and Oshetna River', the point reference object is specified by a topological operation (i.e. intersection) on two linear features. This case is actually equivalent to ordinary single RO assertions (Figure 2(a)).

(2) The assertion includes more than one independent predicate. Each predicate consists of one RO and at least one spatial relationship (Figure 2(b)). An example of this case is 'north of Merced and east of Atwater'. Two points need to be emphasized here. First, an RO may be repeated such that the assertion '10 miles north of Merced' belongs to this category. Second, these predicates may be conjunctively ('AND' operation-based) or disjunctively ('OR' operation-based) combined, although we seldom directly use the latter case.

(3) In the above two cases, assertions are expressed based on binary relationships. However, a description may include a ternary (or N-ary) spatial relationship and multiple ROs in natural language statements representing spatial knowledge (Figure 2(c)). For example, in the assertion 'between Point Reyes and Inverness', the ternary relationship 'between' is employed.

(4) Let us consider a predicate '10 km north of Merced along highway 99'. The metric relationship is based on an RO and bound to a linear feature. This assertion implies much richer semantics: 'highway 99' is a linear feature

whose major direction is north-south; 'highway 99' passes 'Merced'; the TO is close to 'highway 99' and north of 'Merced'; and the distance is measured along 'highway 99'. These constraints can be modeled using a propositional graph (like Figure 2(d)). The example predicate is widely employed in practice. However, we do not exclude other possible assertions that imply a propositional graph to locate a TO. They should be dealt with case-by-case. Note that the set of all consistent instances of a propositional graph may be viewed as an N-ary relationship. However, it is usually difficult to find a single appropriate word (or phrase), like 'between', to describe this relationship.

These four diagrams cover most spatial assertions that describe localities, from simple to complex. Moreover, they can be nested, that is, the RO in a diagram can be replaced by the TO in another diagram to represent more complex expressions. Such a replacement can be solved using the integration operation mentioned in Section 2.4. Generally, the first two cases can be transformed to ordinary predicates based on binary relationships. The third and fourth cases are special and their examples will be discussed in Section 3.5.

2.2 *Uncertainty due to spatial assertions*

The uncertainty of a TO based on the given assertions is inevitable. Wieczorek *et al.* (2004) identified six uncertainty sources. They can be classified into two aspects: the uncertainty due to ROs, and the uncertainty due to the spatial relationships involved. More generally, there are four categories of uncertainty that may influence the final distribution of the TO.

(1) Distribution ranges of target objects: generally, each spatial relationship used in a spatial assertion represents a possible distribution range for the target object. For instance, the predicate 'in the north part of California' will generate an areal range such that the TO is distributed inside it with some probability distribution, and we will investigate such distributions in Section 3.
(2) Imprecision and vagueness of spatial relationships: the spatial relationship involved in an assertion is often imperfect due to measurement errors in quantitative cases or due to conceptual vagueness in qualitative cases. Generally, a crisp spatial relationship will lead to a distribution range with a determinate boundary; while a vague spatial relationship may generate a region with a gradual boundary.
(3) Imperfection of reference objects: if the reference object has positional imperfection, then the derived uncertainty will be propagated when positioning the TO. The positional imperfection of an RO includes three aspects. First, if the RO is represented based on geographic coordinates, it is accompanied with measurement error. Second, an RO, such as 'Santa Barbara downtown' (Montello *et al.* 2003), does not have a determinate boundary due to its vagueness. Third, we usually express an RO using its name. The ambiguity of place names may cause positional imperfection, especially through time. For example, 'Springfield' may imply many cities inside (or even outside) the United States, and the size and shape of each place referred to by that name will have changed.
(4) Uncertainty of assertions: the last category of uncertainty is that based on the assertion itself. Examples of assertions with uncertainty are 'the

probability that the distance between A and B is less than 5 km is 0.8', or 'it is very likely that A is north of B'. For each assertion, we can assign a degree of belief to represent the uncertainty. In practice, the belief degree is often viewed as the probability that the corresponding assertion holds true.

From the perspective of TOs, the above four uncertainty factors can be viewed as a sequence including first-level uncertainty (distribution range), second-level uncertainty (imprecision and vagueness of spatial relationships), third-level uncertainty (imperfection of ROs), and fourth-level uncertainty (uncertainty of assertions) (Figure 3). Worboys and Clementini (2001) presented three interesting cases of making propositions that a particular location is in a given region. The first case is a proposition such as 'we are in France' without any uncertainty. The second case is an assertion based on scanty evidence, such as deciding whether the observer is in Switzerland. The third case is an assertion that the observer is in a place defined with vagueness, such as 'the south of England'. The uncertainties corresponding with these three propositions belong to the first, fourth and second levels respectively. Note that the uncertainty of ROs is not considered in the above three cases, since the three reference places (France, Switzerland and England) have crisp boundaries. Since uncertainty is propagated from higher level to lower level, this sequence should be followed to compute the uncertainty of the TO. In this paper, we primarily focus on the first two issues (distribution of target objects, and imprecision and vagueness of spatial relationships) with the temporary assumption that both the assertions and the reference objects are certain.

2.3 *Uncertainty fields of point target objects*

In single-point TO cases, it is essential to determine a position where the point is most likely to be located, and to measure the associated uncertainty. For instance, comparing the following two predicates: 'inside California State'; and 'inside Merced County, California', the former obviously has greater spatial uncertainty.

The uncertainty field concept is introduced as a probabilistic representation of the spatial distribution of a point TO along with its associated uncertainty. A two-dimensional uncertainty field (Guo *et al.* 2008) of the TO can be defined using its

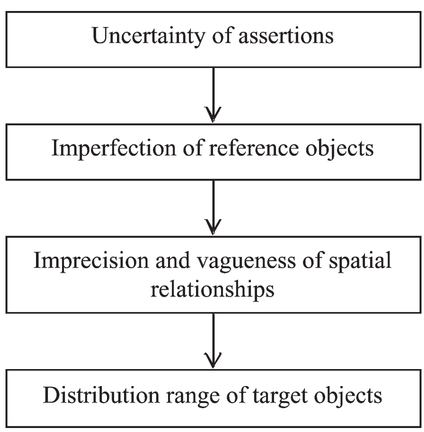


Figure 3. Diagram of the uncertainty due to spatial assertions showing the order in which the uncertainty should be computed in the determination of a target object.

probability density function (PDF), given by

$$z = p(x, y) \tag{1}$$

According to the definition of a PDF, the probability that a TO is inside a region R is $\iint_R p(x, y)\mathrm{d}x\mathrm{d}y$. In the single-point TO case, the following constraint should be satisfied.

$$\iint_D p(x, y)\mathrm{d}x\mathrm{d}y = 1 \tag{2}$$

where D is the domain of field $p(x,y)$. Equation (2) can be extended to deal with the case when the assertion is uncertain. If the probability that an assertion holds is P, then the constraint of the associated field becomes

$$\iint_D p(x, y)\mathrm{d}x\mathrm{d}y = P \tag{3}$$

Moreover, if there are two independent target points to be positioned based on one or more determinate assertions, since corresponding events are independent but not mutually exclusive, the probability that at least one point is inside a region R can be computed as

$$P_R = \iint_R p_1(x, y)\mathrm{d}x\mathrm{d}y + \iint_R p_2(x, y)\mathrm{d}x\mathrm{d}y - \iint_R p_1(x, y)\mathrm{d}x\mathrm{d}y \iint_R p_2(x, y)\mathrm{d}x\mathrm{d}y \tag{4}$$

where p_1 and p_2 are fields associated with those two target points. The corresponding uncertainty field can be obtained by employing a differential operation on equation (4). Actually, such an operation is not necessary in many applications, since uncertainty fields are usually managed based on a raster data format, and each pixel value stands for the probability that the TO is inside it such that equation (4) can be directly used. Cases that involve three or more TOs can be modeled in an approach similar to equation (4). However, if these points are dependent, such as in the description 'two target points are north of Santa Barbara, and the Euclidean distance between them is 10 km', establishing their PDFs may be complicated and is outside the scope of the current discussion.

There are two special cases of uncertainty fields. If the spatial extent of non-zero values is a one-dimensional curve, then the constraint becomes

$$\int_L p(x, y)\mathrm{d}\sigma = 1 \tag{5}$$

where L is the integral curve. Moreover, if the TO is point-distributed, then the PDF is in discrete form, and the constraint is

$$\sum_{i=1}^{n} p(x_i, y_i) = 1 \tag{6}$$

where n is the number of discrete locations where the point might be found. The probability of the point being found at location (x_i,y_i) is $p(x_i,y_i)$. Clearly, a position

Table 1. Three possible distributions of single point target objects.

Distribution	Example
Area	'The TO is inside California', 'The TO is north of Los Angeles'…
Line	'The TO is on Highway 1, between Santa Barbara and Ventura',…
Point	'The TO is possibly located in one of the following three cities: Merced, Fresno, and Modesto,'…

with a probability of 1 is equivalent to a certain target point in the single-point TO case. Table 1 presents some examples of the above three cases. Note that features in the second and third examples are abstracted to line and point objects, that is, their widths or sizes are neglected.

Let the support of $f(x,y)$ be F, which is a subset of the domain of $f(x,y)$ and satisfies $f(x,y)>0$ for any point (x,y) inside F. A function $S(f(x,y))$ can be introduced to represent equations (2), (5) and (6) in an integrated manner. $S(f(x,y))$ is defined as

$$S(f(x,y)) = \begin{cases} \iint\limits_F f(x,y)\mathrm{d}s, & \text{if } F \text{ is two} - \text{dimensional} \\ \int\limits_F f(x,y)\mathrm{d}\sigma, & \text{if } F \text{ is one} - \text{dimensional} \\ \sum\limits_{i=1}^{n} p(x_i, y_i), & \text{if } F \text{ consists of } n \text{ points} \end{cases} \tag{7}$$

We thus have $S(f(x,y))=1$ for a certain assertion. Using numerical methods to describe the uncertainty, the field must be discretized to a raster map (Guo *et al.* 2008) and the sum operation is used instead of integration. In other words, there is no difference at the implementation level. We will use the area integral form in the following equations for the sake of generality.

2.4 *Operations on uncertainty fields*

Two operations, refinement and integration, can be used to handle cases where more than one predicate or reference object is involved. Suppose two uncertainty fields $p_1(x,y)$ and $p_2(x,y)$ are derived from two independent certain predicates on the same TO. Intuitively, more predicates could reduce the uncertainty of the TO since more clues are provided to constrain the TO. Thus, we introduce the refinement operation to deal with multi-predicate cases. Refinement generates a new field using the following equation

$$p(x,y) = p_1(x,y)p_2(x,y) \tag{8}$$

Based on the field $p(x,y)$, we can define the probability that the two predicates are not conflicting as

$$P_{NC} = S(p(x,y)) = S(p_1(x,y)p_2(x,y)) \tag{9}$$

where the function S is defined by equation (7). If $P_{NC}=0$, then there must be a conflict between the input predicates. For example, 'south of Los Angeles, California' and 'north of San Francisco, California' are two conflicting statements, since Los Angeles is south of San Francisco and as such we cannot find a position that satisfies both statements. Under the condition that p_1 and p_2 are not conflicting, the resulting uncertainty field is

$$p'(x, y) = \frac{p_1(x, y)p_2(x, y)}{S(p_1(x, y)p_2(x, y))} \tag{10}$$

According to equation (10), we have

$$S(p'(x, y)) = 1 \tag{11}$$

which is consistent with equations (2), (5) and (6). Equation (10) can thus be viewed as a normalization operation allowing equation (11) to hold. The following example provides a demonstration of the use of equation (10) for spatial positioning.

Figure 4 shows two independent certain predicates: 'Euclidean distance R_1 from RO point A' and 'Euclidean distance R_2 from RO point B'. For each assertion, the probability of the TO is distributed on a circle (Figure 4(a)). According to equation (5), the probability densities of points on these two circles are $1/2\pi R_1$ and $1/2\pi R_2$ respectively. These two fields can be combined using a multiplication operation (equation (10)). This results in two points of non-zero probability, O_1 and O_2, which are the two points of intersection between the two circles. The probability density after multiplication is $1/4\pi^2 R_1 R_2$ at both points. Finally, since we have the constraint expressed by equation 6, their probabilities are both normalized to be 0.5 using equation (10) (Figure 4(b)). This example is somewhat like the principles of the global positioning system. One predicate refines the existing knowledge derived from the other, and decreases the uncertainty. Note, in the case shown in Figure 4, that the TO cannot be positioned if $R_1 + R_2 < D$, where D is the distance between A and B.

Integration (Guo *et al.* 2008) can be used to solve the following problem: suppose an RO has n possible positions with probabilities q_i ($1 \leqslant i \leqslant n$), and the conditional probability density function associated with the ith possible position of the RO is $p_i(x,y)$. The corresponding joint probability density function is $q_i p_i(x,y)$. Consequently, the final uncertainty field of the TO can be obtained by summing these n fields using equation (12)

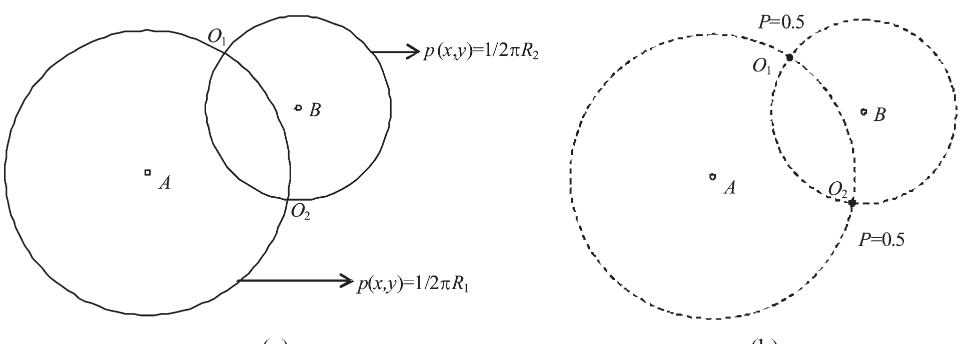


Figure 4. Refinement of two uncertainty fields in which the continuous possibilities of the independent assertions are refined to two discrete points of intersection.

$$p(x, y) = \sum_{i=1}^{n} q_i p_i(x, y) \tag{12}$$

We can extend the integration operation from the following two aspects. First, assuming there are infinite possible positions for the RO, and $R(RO)$ is a representative point, such as the centroid, of RO, equation (12) becomes

$$p(x, y) = \iint_D q(u, v) p(x, y, u, v) \mathrm{d}u \mathrm{d}v \tag{13}$$

where D is the distribution range of $R(RO)$, $q(u,v)$ is the probability distribution of $R(RO)$, and $p(x,y,u,v)$ describes the probability distribution of the TO when the $R(RO)$ is at point (u,v).

Integration can also be used to simulate the composition [or weak composition, following Renz and Ligozat (2005)] of relationships to answer questions such as 'given that an object A and the relationships between C and B, and between B and A, are known, where is C?' A similar approach was proposed by Moratz and Wallgrün (2003) based on purely geometric computation. According to object A and the relationship between B and A, the uncertainty field of B is obtained. Then, using equation (13), the uncertainty field of C can be computed based on the relationship between C and B.

Second, in equation (12), $q_i p_i(x,y)$, $i=1, 2,\ldots, n$ are actually generated based on a group of mutually exclusive uncertain assertions. Equation (12) can thus deal with uncertainty predicates that should be disjunctively combined.

In sum, the differences between the above two operations (refinement and integration) can be demonstrated using the following example. Suppose we have two assertions: 'the TO is north of San Francisco, California' and 'the TO is south of Los Angeles, California'. If they are independent and certain, we cannot locate the TO since they are conflicting (the resulting uncertainty field has a zero probability density everywhere). However, if they are both uncertain and mutually exclusive (for example, their confidence degrees are both 0.5), the possible distribution of the TO is two separate regions according to equation (12). In this research, we will focus on developing a positioning framework where assertions are assumed, for the sake of simplicity, to be certain.

2.5 *Sampling method based on uncertainty fields*

According to Guo *et al.* (2008), two-dimensional uncertainty fields can be managed in a raster model. Suppose a cell covers a square region C, the uncertainty value of this cell can be calculated by

$$v = \iint_C p(x, y) \mathrm{d}x \mathrm{d}y \tag{14}$$

where $p(x,y)$ is the uncertainty field. The uncertainty values, which vary in the interval [0, 1], are the probabilities that the TO is located in the corresponding cell (or pixel) (Figure 5(a)).

Monte Carlo methods are a class of algorithms that rely on iterative random sampling and probability statistics to investigate problems. They are widely used for

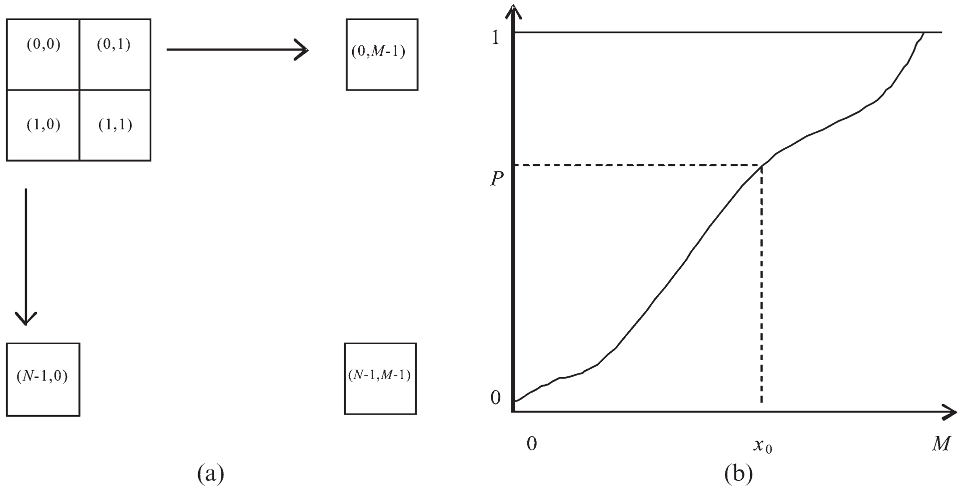


Figure 5. Diagrams representing (a) a raster uncertainty field and (b) the cumulative probability curve.

simulating behavior in various fields (Fishman 1996). In GIS applications, Monte Carlo methods are commonly used to deal with spatial uncertainty (Dutton 1992, Jacquez and Jacquez 1999, Ratcliffe 2004). In this study, we propose a sampling method to simulate the distribution of TOs with raster uncertainty fields as inputs. In order to generate random numbers following a specific distribution, inverse transform sampling, which is based on the cumulative distribution function of a PDF, is often employed. This method can be easily extended for two-dimensional discrete cases, since the x and y coordinates of a position are independent.

Let the size of a raster field be N rows × M columns. We can compute the sum of each column using

$$S_j = \sum_{i=0}^{N-1} G(i,j), j = 0, 1, \ldots, M-1 \tag{15}$$

where $G(i,j)$ is the pixel value at the ith row and jth column. The cumulative function can thus be defined as

$$f(k) = \sum_{i=0}^{k} S_i, k = 0, 1, \ldots, M-1 \tag{16}$$

Figure 5(b) illustrates such a function. In order to obtain the x-coordinate of a sampling point, we can generate a random number P, which follows a uniform distribution in [0, 1], and the value x_0 that makes $f(x_0) = P$ is the x-coordinate of the target point (Figure 5(b)). The y-coordinate can be similarly computed based on the cumulative function of the $[x_0]$th column, where $[x_0]$ stands for the integer part of x_0. Shi (2007) proposed a restricted Monte Carlo method based on Post Office Box addresses in environmental health studies. This can also be handled by the above method, if we establish an uncertainty field for each Post Office Box address.

In this section, we introduced the concept of reference objects, target objects, uncertainty fields, and refinement and integration operations on them. Generally, four levels of factors influence the uncertainty field of a TO: distribution ranges of target objects, imprecision and vagueness of spatial relationships, imperfections of reference objects and uncertainty of assertions. Of these, uncertainties associated with spatial relationships require detailed consideration. In terms of imperfections of reference objects, Guo *et al.* (2008) adopted a probability approach to representing imperfections of reference objects in the context of spatial positioning for museum collection data. In the future we plan to investigate such uncertain assertions.

3. Uncertainty fields associated with spatial relationships

3.1 *Spatial relationship issues*

A spatial assertion generally provides a rough distribution range of a TO. Several issues should be considered in deciding such a range according to a given relationship. We will take an expression 'the TO is north of a point RO *B*' as an example to investigate these issues.

First, suppose we have an experiment where subjects are asked to plot points satisfying 'north of *B*'. Can the distribution of these points be viewed as the potential distribution range of the TO? The answer is negative. In this experiment, subjects will tend to dot at prototypical points, according to prototype theory within cognitive psychology (Rosch and Mervis 1975). Thus, the range obtained in the experiment is too narrow to determine the TO, since some positions that can make the expresser to declare 'north of *B*' are excluded.

Second, with the description 'the TO is north of a point RO *B*', we should assess under what conditions the expresser would declare such an assertion. Two situations are possible: the expresser believes that the predicate is correct, and the expresser believes that 'north' is the most appropriate cardinal direction relationship to describe the TO among a set of candidate relationships, which include north, northeast, northwest, east and west, etc. If 'north' is a crisp concept, then there is no difference between these two assumptions. However, the vagueness of spatial relationships makes them different. For example, if the TO is actually northeast of *B*, the observer may give a positive answer to the question 'is the TO north of *B*?' (a similar investigation on the concept 'above' is presented by Crawford *et al.* 2000), but might choose northeast for another question, such as 'which is the most appropriate relationship between the TO and *B*?' The first assumption will generally lead to a wider range than the second, and the uncertainty fields of two neighbouring relationships are overlapped. These two fields will not overlap when following the second assumption.

In this research, the two assumptions are named 'assumption one' and 'assumption two': assumption one is that the position makes the assertion true; while assumption two is that the assertion is the more suitable to describe the TO than others. We generally believe that assumption two is more reasonable, since in practice a number of spatial relationships and ROs will be available such that the expresser can choose a most appropriate relationship according to his (or her) perception, instead of a relationship that merely makes the assertion hold. However, assumption one is still acceptable when there is a need to use a wider range for the potential distribution of the TO. This point will be discussed in detail in Sections 3.3 and 3.4. Following assumption two, the uncertainty fields of spatial relationships are represented based on existing models, which allow us to establish a set of jointly exhaustive and pairwise disjoint relationships (Renz 2002; Renz and Ligozat 2005).

In addition to the rough range described by spatial relationships, the vagueness and imprecision that accompanies many spatial relationships make TOs uncertain at a higher level. For a qualitative relationship, the uncertainty comes from the vagueness during its conceptualization, such as 'north' or 'far'. The boundary between two neighboring relationships, such as 'far' and 'near', is indeterminate. In terms of topological relationships, although they are theoretically determinate (Egenhofer and Herring 1991; Randell *et al.* 1992), a topological relationship might be uncertain in the context of natural language description (Mark and Egenhofer 1994). For example, Egenhofer and Shariff (1998) propose quantitative indices to further discriminate the topological relationships belonging to the same category, such as 'overlap'. Quantitative spatial relationships often encounter precision issues. Distances are commonly recorded with few or no significant digits, or even with fractions (Wieczorek *et al.* 2004). For instance, in the description 'nine kilometers north of Bakersfield', the true distance might be an arbitrary value in [8.5, 9.5] due to the rounding operation.

In the following sections, we will discuss the uncertainty field associated with different types of spatial relationships. It should be noted that we focus on typeless TOs and spatial expressions themselves without further considering much contextual information. For instance, given a statement 'a traffic accident occurred in California', the accident point is assumed to be uniformly distributed inside California. The uniform distribution assumption can thus be viewed as a reasonable initial assumption about the TO's location. With the knowledge that a traffic accident can only happen on transportation lines, the PDF is evenly linear and satisfies equation (5). Moreover, considering that there is a correlation between probabilities of accident and traffic conditions, the distribution can be adjusted, and further refined by taking into account the prior probability distribution of traffic accidents in California. In this research, several probabilistic distribution assumptions are made in order to describe the theoretical framework. Those probabilistic distribution assumptions could be refined if contextual information becomes available, yet the proposed framework will still be valid for different distribution assumptions.

3.2 *Topological relationships*

The topological relationships vary depending on the types of RO geometry. When the RO is a point, the possible topological relationships are relatively straightforward. They are 'equal' and 'disjoint'. The former implies a certain position, while the latter does not contribute to determining the TO. In this study, we emphasize the cases in which the RO is linear or areal. Much research has been done on the topological relationships between two spatial objects (Egenhofer and Herring 1991; Randell *et al.* 1992). Since the TO is abstracted to a point object, the topological relationships on spatial positioning are relatively simple.

If the RO is linear, then the possible topological relationships could be inside, coincident with the two end points, and disjoint. In the first case, the TO can be assumed to be uniformly distributed along the line (Figure 6(a)). According to equation (5), the probability density is $1/length(RO)$, where $length(RO)$ is the length of the linear RO. In the second case, the possible location contains one (when the linear RO is simple) or more (when the linear RO is complex) points. Their associated probabilities for each point are $1/n$, where n is the number of end points. In the last case, the TO is uniformly distributed on the geometric complement of the

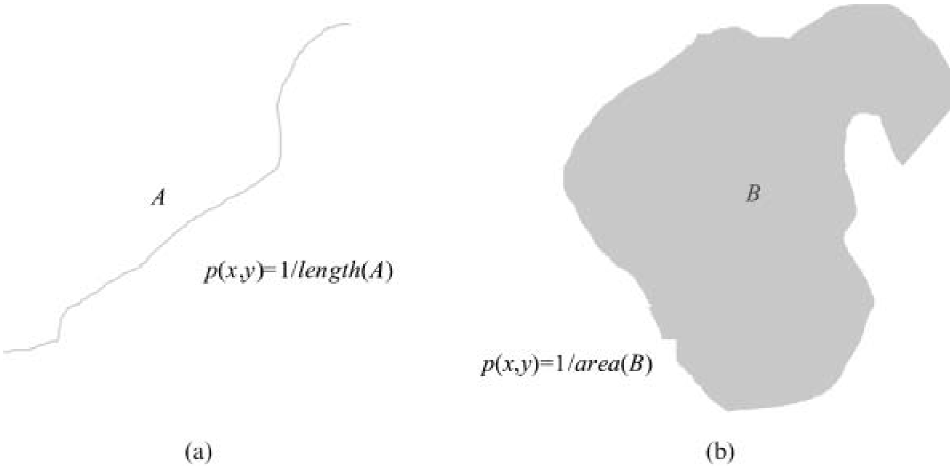


Figure 6. Topological relationship (i.e. 'inside') based locality description. (a) linear reference object; (b) areal reference object.

RO. In theory, the range is infinite. In practice, we usually specify a finite searching space to position the RO, which will be discussed in detail in the next section when dealing with directional relationships.

If the RO is a region, the possible topological relationships could be inside, on the boundary, or disjoint. In the first and the third cases, the uniform probability density is $1/area(RO)$ or $1/area(RO^{-})$, where *area* is a function to compute the area of the RO (Figure 6(b)) or its complement (i.e. RO^{-}). Meanwhile, the second case is the same as the 'inside a line' case for a line discussed above.

In locality descriptions, the disjoint relationship is seldom used directly. As argued by Mark (1999), disjoint relationships are generally refined in practice by using direction relationships and distance relationships. In other words, the disjoint relationships are not usually stated explicitly.

3.3 *Directional relationships*

The directional relationship can be represented either quantitatively or qualitatively. In practice, the locality descriptions are often qualitative. There have been numerous models developed for cardinal direction relationships (CDR), including the cone, project (Frank 1991), double-cross (Freksa 1992), and minimal bounding rectangle (MBR) (Goyal and Egenhofer 2001; Skiadopoulos *et al.* 2004) models. Generally, the cone- and project-based models are suitable for point reference objects (Figure 7(a)), while the MBR-based model is more applicable for linear or areal objects (Figure 7(b)).

Let the reference object be A. Each model partitions A^{-} (the complement of A) into a number of subsets. The TO can be assumed to be uniformly distributed inside them. Additionally, four relevant issues should be considered when uncertainty fields are introduced to model CDRs.

3.3.1 Internal cardinal relationships. Liu *et al.* (2005) proposed the internal cardinal direction (ICD) model to represent the refined spatial relationship when the TO is inside an areal RO, for instance, 'northern part of Merced City'. The ICD-9 model, which includes nine atomic base relationships, is often employed in practice

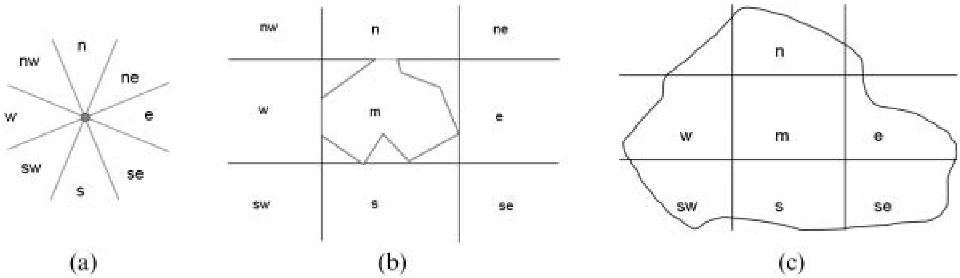


Figure 7. (a) Cone-based CDR model, (b) MBR-based CDR model and (c) MBR-based ICD models.

(Figure 7(c)). It is similar to the conventional MBR-based model in regarding the middle part as a reference object.

3.3.2 Vagueness of CDRs. Cardinal direction relationships are abstracted from concrete angle values between two objects. Due to the vagueness of a CDR, the assumption that the TO is uniformly distributed in the corresponding range does not hold. As shown in Figure 8(a), point *B* is not as directly 'north' as point *A*. Although several research results (Dutta 1991; Montello and Frank 1996; Du *et al.* 2004) have been reported on the vagueness of CDRs, it remains very challenging to establish a widely accepted model. Additionally, the situation is much more complicated when a CDR is used to describe a locality. First, a couple of CDR models are available when an individual makes a predicate such as 'north of Santa Barbara'. Even if he (or she) chose the cone-based model, the number of cones could be 4, 8 or 16. Without additional contextual information, we cannot tell what angle range the relationship 'north' stands for. Second, there may be more than one candidate RO to describe a locality in practice. As shown in Figure 8(b), three candidate ROs: *A*, *B* and *C*, can be used to describe the position of the TO. If the observer chooses a four-sectored cone-based model, the TO can be expressed as 'south of *A*', 'north of *B*', or 'west of *C*'. However, since the TO is more closely matched by a cardinal direction 'south of *A*', it is reasonable to assume that the observer would choose the first expression. Note, in this case, that the distance (or more general, accessibility) to each RO, and the weight (e.g. population of a city) of each RO, will obviously affect the decision of the observer. For simplicity, however, we do not consider these

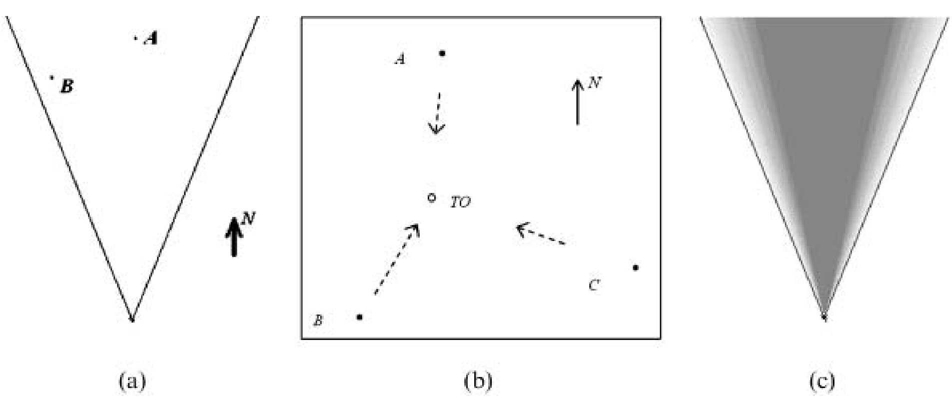


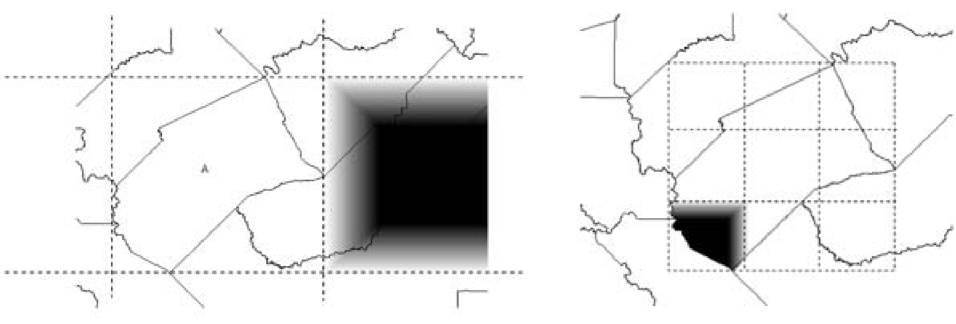
Figure 8. Vagueness associated with cone-based cardinal direction relationships.

factors in this research. Therefore, angles become the major factor in deciding a qualitative direction relationship.

In summary, the uncertainty associated with CDRs in spatial positioning is complicated. Given an assertion '*A* is north of point *B*', one assumption could be that the *A* is more likely to occur around the centerline of the corresponding cone than away from the centerline (Guo *et al.* 2008), that is, the uncertainty field is not homogeneous. Consequently, Figure 8(c) represents the pattern of the uncertainty field due to direction relationships in point-RO cases. In the toolbox mentioned in Section 4, we use this pattern as the default setting. However, users can customize it as well as the uncertainty fields associated with other spatial relationships. Similarly, the conceptual probability distributions associated with MBR-based CDRs and ICD relationships are represented, as shown in Figure 9.

3.3.3 Searching space. In QSR, spatial relationships are often defined in a Euclidean space. Therefore, the possible distribution range is effectively infinite for a given expression 'north of an RO', and it is thus hard to define an uncertainty field that satisfies equation (2). Fortunately, the searching space provides a constraint during spatial positioning. The searching space is often known in most GIS applications. For instance, in research on an endemic California species, it is unreasonable to locate the TO inside Washington State based on the expression 'north of San Francisco', even if the assertion holds true. Generally, the largest searching space for most geographic applications is the Earth's surface even without any contextual information.

Another factor that affects the searching spaces associated with direction relationships has been mentioned earlier, that is, more than one RO can be used to describe the TO's position, and the observer will select the most appropriate one. For instance, without the constraint of 'inside California', a TO in Washington State may be described as 'north of Seattle', or other reasonable expressions, instead of 'north of San Francisco', if we follow assumption two (please refer to Section 3.1). It implies that each RO has an influence range in expressions with directional relationships when there is more than one available RO. On this topic, Kettani and Moulin (1999) introduced the notion of an influence area to model spatial conceptual maps and to generate natural language descriptions of routines, and Wang *et al.* (2005) proposed a Voronoi diagram based model for spatial referencing. It should be noted that selecting the most appropriate expression (i.e. assumption



(a) (b)

Figure 9. Uncertainty fields associated with (a) MBR-based CDR – east, and (b) ICD relationship – southwest, where the partition schemata are depicted using dash lines.

two) is based on the assumption that the observer has enough spatial knowledge of the study area, and is able to freely choose the RO. However, there are some exceptions. For example, in the case of georeferencing textual descriptions of museum collection data (Guo *et al.* 2008), the description 'five miles north of Merced' means the observer traveled five miles north from Merced, and the TO is constrained by the location where the observer started to collect the specimens.

3.3.4 Starting point. As mentioned earlier, MBR-based CDR models can be employed when the reference object is a line or a region. It is a simplified model such that an inferrable algebraic system can be established. However, it is not consistent with commonsense knowledge, especially when the RO is a linear feature. For instance, Figure 10 (a) shows a reasonable relationship '*B* is north of *A*.'; while we cannot obtain such a result according to the MBR-based model. Consequently, the potential TO (i.e. *B*) cannot be positioned correctly, especially when following assumption one. In this paper, since the uncertainty field is introduced to represent a spatial relationship, we can assume that the CDR is measured based on a point that is uniformly distributed in the RO. According to the cone-based model, a cone can be obtained for each starting point (Figure 10(b)). Then the cones can be integrated to get the final probability distribution based on equation (13), or equation (12) in discretized cases. Schmidtke (2001) investigated direction relationships between extended spatial objects, and two models, region-based sector and side-based sector, are formally defined. The probabilistic approach proposed in this paper is similar to the region-based sector model. We prefer the first model for the following two reasons: most ROs do not have prominent sides in practice, and an overlap is 'safer' than a gap in many spatial positioning applications. The region-based sector model may yield a gap between two neighbouring directions.

3.4 *Metric relationships*

The metric relationships in locality descriptions may be qualitative or quantitative. The relationships themselves and their associated uncertainty lead to the uncertainty of the TO. In qualitative cases, 'close to' or similar terms, such as 'near', are often used in practice to express qualitative metric relationships. Metric relationships such as 'far from' are seldom employed, since they provide a coarser clue about the TO than 'near'. There are two issues related to qualitative distance relationships. First, they are dependent on the sizes of the involved objects as well as the scale of the

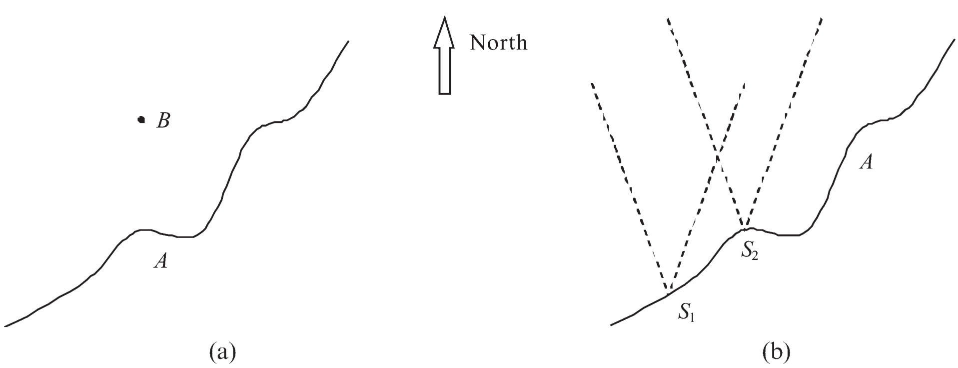


Figure 10. Cardinal direction relationship for linear reference object.

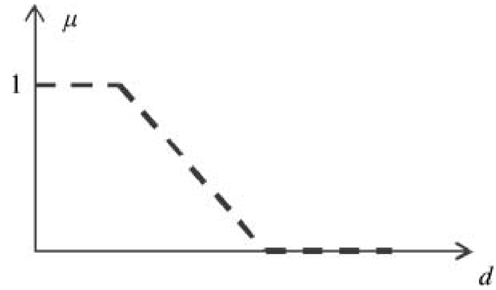


Figure 11. Conceptual membership function of qualitative metric relationship 'near'. The greater the distance d from the reference object, the less likely the location would be described as near.

context. For instance, the concept 'near' at the urban scale may be shorter than the same concept at the country scale. Second, they are concepts with vagueness. Figure 11 depicts the conceptual function between the membership degree of 'near' and the distance between two objects. Worboys (2001) provided a comprehensive investigation on the vagueness of nearness relationships. There is much controversy about the relationship between fuzziness and probability (Zadeh 1995; Ross *et al.* 2002). As argued by Dubois and Prade (1993), given a population of individuals and a fuzzy concept F, each individual is asked whether a given element $u \in U$ can be called an F or not. The membership can be represented by the proportion of individuals who answered 'yes' to the question. In other words, the membership degree can be defined in a probabilistic approach. Following the assumption that the probability and membership degree are positively related, if the membership degree at a specific position p is high, then the probability that the target point is located in a small area surrounding p is also high. Consequently, given an RO and a qualitative metric relationship, such as 'near', the conceptual probability distribution of the TO can be obtained (Figure 12(a)).

Note, in the above discussion, that we actually followed assumption one, that is, a position that makes the assertion 'the TO is close to the RO' true is acceptable. However, if other relationships are available and assumption two is adopted, the corresponding uncertainty field will change. Suppose 'far' is the conceptually neighbouring relationship of 'close'. Obviously, these two relationships are both vague. Without considering other factors, their membership functions, $\mu_c(t)$ and $\mu_f(t)$, which are defined on the quantitative distance t between two objects, are depicted in Figure 13(a). The intersection point of these two functions is $t=t_0$. Note that the intersection point may not exist if there are more distinctions of qualitative metric relationships. For instance, Clementini *et al.* (1997) considered five distinctions, and the relationship 'commensurate' was defined between 'far' and 'close'. For the sake of simplicity, we assume the intersection point exists. For two objects, say A and B, with distance $t_1>t_0$, if we ask 'is A close to B', the answer may be 'yes' since $\mu_c(t_1)>0$. This has been validated in the experiment presented by Worboys (2001). Meanwhile, if we ask 'is A close to or far from B', the answer is likely to be the latter, since $\mu_f(t_1)>\mu_c(t_1)$. Consequently, following assumption two, the boundary of a vague relationship seems to be crisp, at the intersection point of two membership functions. Due to the subjectivity of vagueness, however, different observers have similar but slightly different membership functions upon the same concept. As shown in Figure 13(b), the intersection point varies in an interval $[t_1,t_2]$.

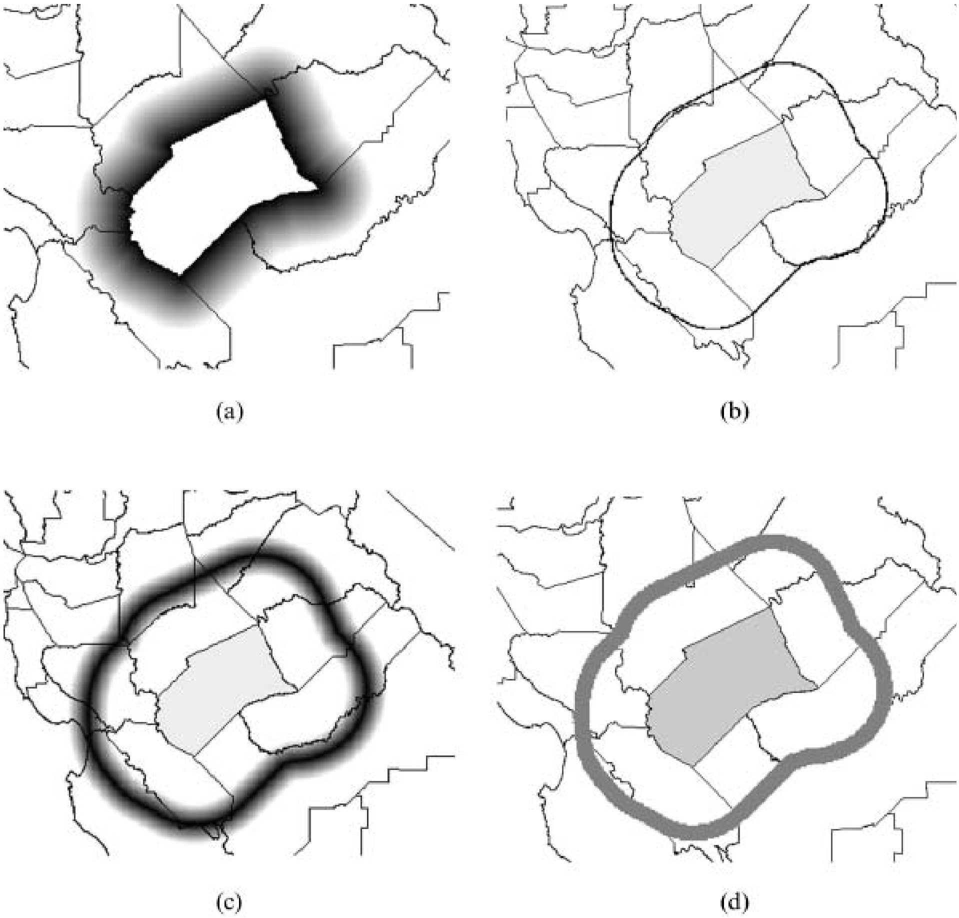


Figure 12. Uncertainty fields associated with different metric relationships (a) close to the RO; (b) the Euclidean distance between the TO and the RO is *d*, which is accurate and measured from the boundary of the RO; (c) similar to case b, but the distance is imprecise and the error follows a normal distribution; (d) similar to case c, but the error follows a uniform distribution.

Hence, the boundary of the relationship 'close to' is still vague and Figure 12(a) also depicts the associated uncertainty field even if we use assumption two.

In contrast to qualitative distances, quantitative distances are specified in practice using numeric distance values, which may be precise or uncertain. The TO will be distributed in a linear ring around the RO when the distance is precise (Figure 12(b)). More commonly, the distance may be uncertain due to measurement error or record imprecision. In such cases, the TO will be distributed in a band around the RO. Moreover, the probability distributions are different in these two cases. If the uncertainty is caused by measurement error, the true distance value could be normally distributed (Figure 12(c)). If the uncertainty is caused by record imprecision, however, then the actual distance could be assumed to be uniformly distributed in an interval (Wieczorek *et al.* 2004). For instance, if the offset distance is ten kilometers with a precision of 0.1 km, then we can compute a band such that the distance between a point inside it and the RO varies in [9.95~10.05]. If this

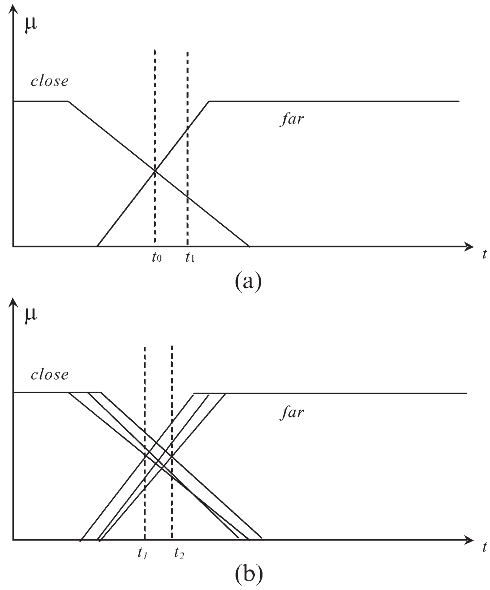


Figure 13. (a) Comparison of assumption one and assumption two; (b) assumption two leads to uneven distribution when there are two neighbouring vague relationships involved in spatial positioning.

interval is narrow compared with the distance, the TO can be approximately represented by a uniform distribution inside this band (Figure 12(d)).

In the above discussions, the starting point is assumed to be on the boundary of an areal RO. As argued by Wieczorek *et al.* (2004), however, the starting point may be any point inside the RO. Consequently, the extent of the RO will influence the uncertainty of the TO. This problem can be solved using the integration approach mentioned in Section 2.4. Another situation arises when the RO is vague and thus has no crisp boundary. Since a vague object can be modeled using a fuzzy set, the metric relationship can be computed using the method proposed by Dilo *et al.* (2007).

3.5 *Two special spatial relationships*

3.5.1 Metric relationships in linear spatial reference systems. In Section 3.4, the distance is measured based on a straight connecting line between the RO and TO. In practice, the TO might be described based on the distance along a linear feature, such as a road or river, which forms a linear spatial reference system (SRS). There are three issues related to the uncertainty field based on this type of assertion:

(1) If the distance is exact, then a predicate such as '10 km from *A* along *L*' can lead to two possible locations, one on either side of *A*, each with probability of 0.5. Moreover, if we can add a direction constraint, yielding a locality such as '10 km north of *A* along *L*', the TO will be a certain point without considering the uncertainty of the distance and the linear feature. Note that the direction relationships used in such expressions are different from those defined in a two-dimensional space. In linear SRSs, we decide a direction based on the major direction of the linear RO instead of the actual angle.

(2) If the distance is qualitative (such as 'near *A* on *L*'), or quantitative but imprecise (such as 'roughly ten kilometers from *A* on *L*'), then the distribution range of the TO consists of one or more sections of the linear feature (Figure 14(a)). Due to the vagueness of qualitative distances or the error distribution of quantitative distances, the boundaries of those sections are gradual.
(3) If the linear reference feature has positional uncertainty, it can be modeled using an error band (Chrisman 1982, Caspary and Scheuring 1992) or an epsilon band (Shi 1998). In this case the distribution range of the TO consists of one or more arc zones if the distance value is exact, and one or more band zones surrounding the linear feature if the distance is not precise (Figure 14(b)).

3.5.2 'Between' relationships. A special relationship, 'between', is sometimes used to describe localities. The relationship 'between' is ternary, that is, three objects are involved in an assertion, for instance, 'the target point is between highway 99 and highway 5'. Generally speaking, if a point TO *O* is between *A* and *B*, the following properties hold:

(1) the topological relationship between *A* and *B* is disjoint;
(2) the topological relationships between *O* and *A*, and *O* and *B* are both disjoint;
(3) the directional relationships between *O* and *A*, and *O* and *B* are inverse.

In order to position the target object based on the 'between' relationship, we should find a region that satisfies the three conditions above. If at least one RO has spatial extent, then the region satisfies the constraint that, for any point inside it, say *O*, there exists a straight line passing through *O* that intersects the two ROs, and the

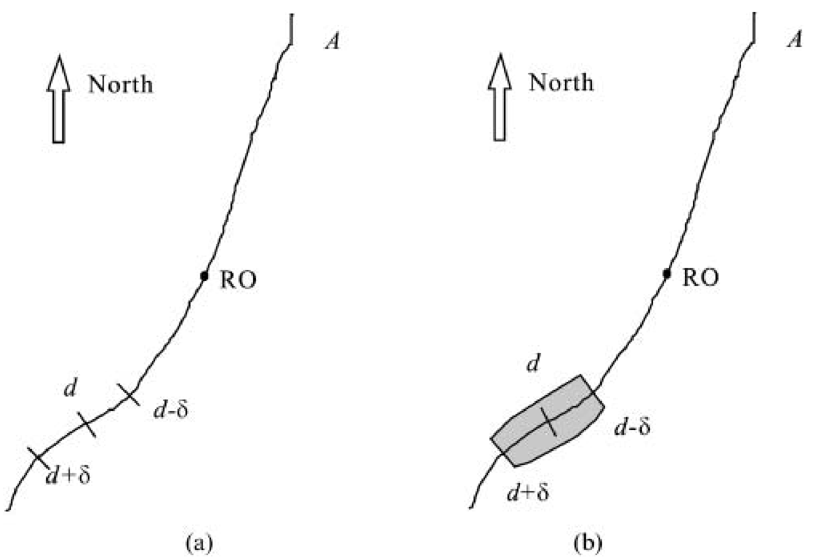


Figure 14. Metric relationships with uncertainty in linear SRS: the TO is *d* km (or other metric units) south of the RO along line *A*, without (a) and with (b) positional uncertainty.

intersection points are located on opposite sides of O (Figure 15(a) and (b)). If the two ROs are both points, it is a little complicated. A straightforward approach is to follow the constraint. Hence, the region will degenerate to a straight connecting line segment. A point close to this line, however, is also viewed to be between those two point ROs according to common perception. A diamond region is thus adopted instead of a line segment (Figure 15(c)). Between relationships are also accompanied with vagueness. Bloch *et al.* (2006) investigated the vagueness of 'between' relationships between spatially extended objects. Hence, the associated uncertainty fields can be adjusted accordingly. In the last case, that is, 'between' two point ROs, the probability can be defined according to the deviation angle from the connecting line when taking vagueness into account.

4. Case study

We designed a software toolbox, Spatial Positioning Probabilistically (SPPro), to calculate the uncertainty fields associated with various spatial relationships. In this toolbox, the uncertainty fields discussed in Section 3 are configured to be default settings for supported spatial relationships. Users can customize the configuration of an uncertainty field associated with a spatial relationship. One can load reference maps, which contain geographic features (such as cities, rivers, and administrative units) that may be potential reference objects in spatial assertions, and then designate predicates to position a TO. For each predicate, SPPro provides two modes to compute the associated uncertainty field. They are 'calculate new' to generate a new uncertainty field and 'refine existing' to refine an existing field. Figure 16 demonstrates a case study based on the administrative map of California State (Figure 16(a)). The three predicates are: southeast of Merced County; middle part of Fresno County; 50 miles (record precision is ten miles, that is, the distance is between 45 and 55 miles) from San Benito County.

According to equation (10), the product of three fields corresponding to the above predicates can be obtained as the resulting field (Figure 16(d)). We can input new predicates to further refine the TO's distribution, until any inconsistencies occur among the predicates. For instance, if we input 'north of San Francisco County' as the fourth predicate, all pixel values in the uncertainty field will be zero. In other

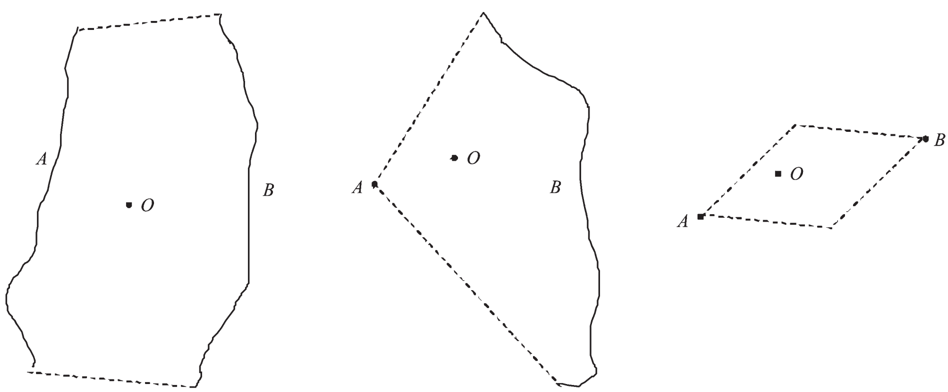


Figure 15. Ternary spatial relationship – 'between': (a) between two extended objects; (b) between a point object and an extended object; (c) between two point objects.

words, the TO that simultaneously satisfies those four conditions cannot be found. It should be noted that the result is computed based on the default uncertainty field configuration of direction relationships. As shown in Figure 16(d), a point *p*, which is outside the possible distribution area of the TO, intuitively satisfies the above three predicates. From Figure 16(b), we can see that the MBR-based CRD model excludes *p* as a potential TO. This implies that MBR-based models are not very suitable for spatial positioning, as discussed in Section 3.3. However, the framework for probabilistic spatial positioning can remain unchanged if a new CDR model is employed.

It should be noted that the toolbox can deal with more spatial predicates than the above three types. Some other examples were provided in Guo *et al.* (2008). Based

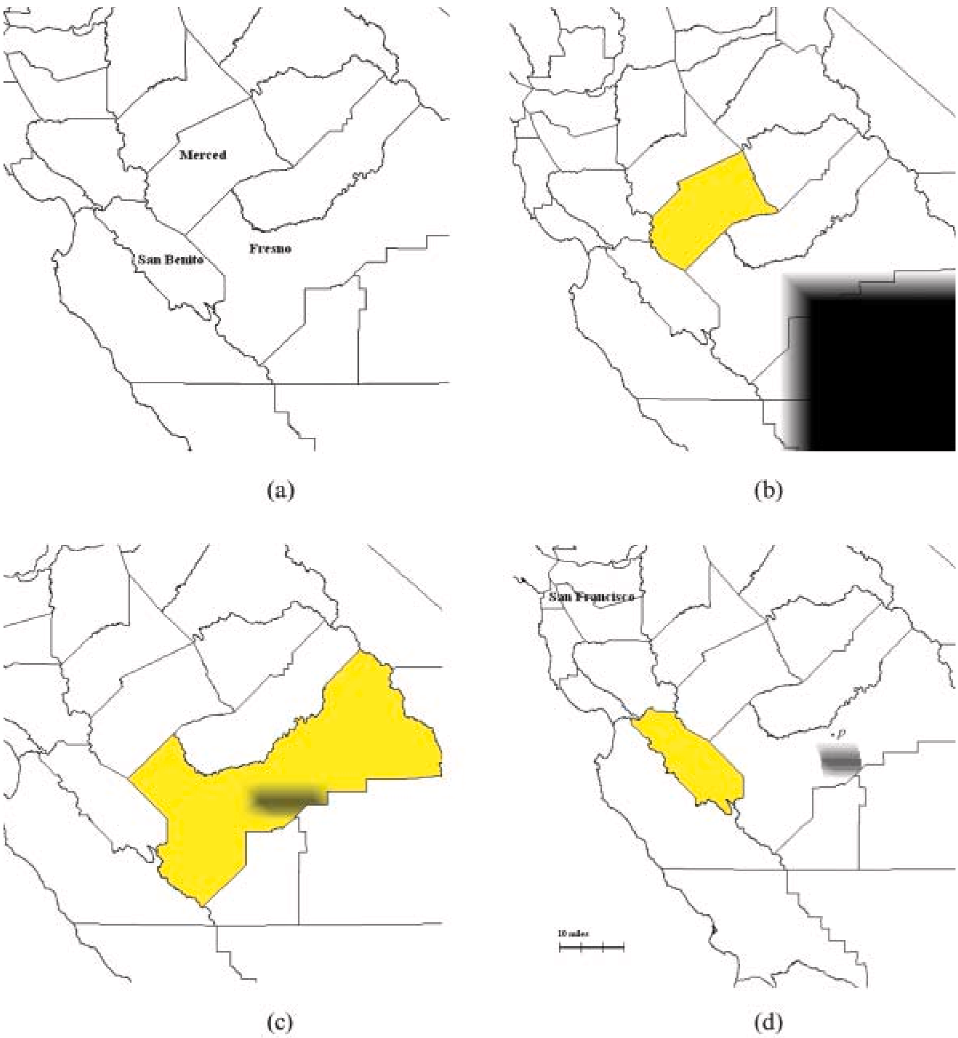


Figure 16. Positioning a TO based on three predicates. (a) reference map: part of the administrative map of California State; (b) uncertainty field generated by 'southeast of Merced County'; (c) uncertainty field refined by 'middle part of Fresno County'; (d) uncertainty field further refined by '50 miles from San Benito County, associated record precision is 10 miles'.

on the uncertainty fields, several uncertainty measures such as maximum error and confidence envelope (Guo *et al.* 2008) can be computed and stored as alternative approaches to represent spatial uncertainty (Wieczorek *et al.* 2004).

5. Conclusions

To position a locality and estimate the associated uncertainty based on qualitative or semi-quantitative descriptions is a critical issue in a range of geographic applications where spatial knowledge is represented and communicated verbally or textually. In conventional QSR research, the spatial relationships are processed qualitatively such that new relationships can be generated and the inconsistencies can be checked. However, it is difficult to determine the position of a point locality based on formal relationship operations. Furthermore, the vagueness associated with spatial relationships makes the positioning computation not a purely qualitative process.

Theoretically, there are four levels of uncertainty associated with the spatial assertions. They are rough distribution ranges of target objects, imprecision and vagueness of spatial relationships, imperfection of reference objects, and uncertainty of assertions. With the assumption of point target objects, we develop the concept of an uncertainty field, which is a two-dimensional probability density function, to represent an uncertain TO described by an assertion. According to probability theory, there are two possible operations on the uncertainty field: refinement and integration. Refinement can be used to summarize a group of predicates to obtain a refined distribution of the TO, as well as to check inconsistencies among different predicates. The objective of integration is to compute the distribution of the TO based on an uncertain RO, and to simulate relationship compositions.

A spatial relationship together with reference objects describes a rough range of the target object, which is the first-level uncertainty of the TO. For a relationship-based assertion, two assumptions are available for deciding the possible position of a TO. Assumption one is that the position makes the assertion true; while assumption two is that the assertion is the more suitable to describe the TO than others. Two aspects, the real distribution of reference objects and the vagueness of involved relationships, make the two assumptions different. Generally, assumption one will lead to a wider distribution range and thus greater uncertainty. In this study, we choose to adopt assumption two in spatial positioning problems, but assumption one can be used to interpret declarative geographical propositions.

Three categories of binary spatial relationships are often found in locality descriptions. They are topological relationships, directional relationships and metric relationships. The associated conceptual uncertainty fields are established according to the nature of each type of relationship. Due to the vagueness of a qualitative relationship or the measurement error of a metric relationship, the probability distribution of the target object is not always uniform. Since a vague concept is dependent on the corresponding context, it is necessary to specify the parameters of the uncertainty field in a concrete application. In addition to these three categories of binary relationships, the metric relationship in linear SRS and a special ternary relationship, 'between', often found in locality descriptions, have been investigated. A software toolbox, SPPro, has been implemented to demonstrate the uncertainty fields of the above relationships and their operations, and is available on request. In SPPro, the user can input a group of spatial assertions and the distribution of a TO can be obtained if they are consistent. The toolbox provides a software framework for spatial positioning, to which one can add new spatial relationships or customize existing relationships if needed.

The proposed approach focuses on descriptions about point objects. According to Bennett and Agarwal (2007), such a description can be viewed as an expression for a place. Let us reconsider the example of 'inside the south of England'. We actually have two approaches to modeling this predicate. First, the RO is 'England' and the spatial relationship is an internal cardinal direction 'south'. Second, the RO is 'the south of England' and the spatial relationship is 'inside'. In the second case, the RO is equivalent to the vague place described by the predicate. Without considering the uncertainty of spatial assertions, each spatial positioning problem can be simplified to a predicate that the TO is inside a place, which is presented by the corresponding assertion. Suppose the place is modeled by a two-dimensional membership function $m(x,y)$, and the uncertainty field associated with the TO is $p(x,y)$. Three heuristic rules can be identified as follows: (1) the membership degree of a point is positively relevant to the corresponding probability density; (2) for a point (x_0,y_0), $p(x_0,y_0)=0$ means $m(x_0,y_0)=0$, that is, $m(x,y)$ and $p(x,y)$ share the same support; (3) a uniform distribution implies an areal place with crisp boundary. In GIR systems, a query can be a formalized by triplet of <theme, spatial relationship, location> (Jones and Purves 2008), where a geographical range (or a place) is specified by given spatial relationships and reference locations to search web pages about a theme. Such a triplet is similar to the typed TO cases presented in this research. The above three rules indicate that uncertainty fields can be applied to compute geographical relevance scores in GIR.

In terms of spatial positioning, three issues remain for future research: projections, uncertainty of reference objects and assertions as well as spatial extents of TOs:

(1) In the current version of SPPro, the computation space is assumed to be an ideal two-dimensional Euclidean space $\boldsymbol{R}^2$. Clearly, in order to maintain consistency with common GIS, it should be possible to determine the uncertainty fields associated with the CDRs and metric relationships depending on the projection and the datum, since these two relationships measured in $\boldsymbol{R}^2$ are different from those measured in the real world.

(2) In this research, we focused on spatial relationships, and briefly discussed the imperfection (such as vagueness, ambiguity, and imprecision) of reference objects and uncertainty of assertions, while Guo *et al.* (2008) have investigated the imprecision of ROs. As mentioned in Section 2.2, the two aspects should be considered for integrating the four levels of uncertainty.

(3) The point TO assumption could be a limitation in some applications. Finding a method to position an object with spatial extent is therefore a problem to be solved. With a given size, the object can be abstracted to a circle (or a square). The problem is thus transformed to positioning its center. However, the shape characteristics still cannot be neglected, which makes the problem challenging. Note that we did not discuss some special relationships that are available for spatially extended TOs. For example, the term 'along' (Takemura *et al.* 2005) can be used to describe a linear TO. We plan to study these relationships in the future.

Acknowledgements

We thank the reviewers for their comments that helped strengthen the paper. This research is supported by the BioGeomancer Project funded by the Gordon and Betty

Moore Foundation, the National High Technology Development 863 Program of China (Grant No. 2007AA12Z216), and NSFC (Grant No. 40701134 and 40629001).

References

Bennett, B. and Agarwal, P., 2007, Semantic categories underlying the meaning of 'place'. *Lecture Notes in Computer Science*, **4736**, pp. 78–95.

Bloch, I., Colliot, O. and Cesar, R.M., 2006, On the ternary spatial relation "between"". *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, **36**(2), pp. 312–327.

Caspary, W. and Scheuring, R., 1992, Error-bands as measures of geometrical accuracy. In *Proceedings of EGIS'92*, 23–26 March 1992, Munich, Germany, pp. 226–233.

Chrisman, N.R., 1982, A theory of cartographic error and its measurement in digital databases. *AutoCarto*, **5**, pp. 159–158.

Clementini, E., Felice, P. and Hernández, D., 1997, Qualitative representation of positional information. *Artificial Intelligence*, **95**, pp. 317–356.

Cohn, A.G. and Hazarika, S.M., 2001, Qualitative spatial representation and reasoning: an overview. *Fundamenta Informaticae*, **46**, pp. 1–29.

Crawford, L.E., Regier, T. and Huttenlocher, J., 2000, Linguistic and non-linguistic spatial categorization. *Cognition*, **75**, pp. 209–235.

Dehak, S.M.R., Bloch, I. and Maître, H., 2005, Spatial reasoning with incomplete information on relative positioning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(9), pp. 1473–1484.

Dilo, A., de By R.A. and Stein, A., 2007, A system of types and operators for handling vague spatial objects. *International Journal of Geographical Information Science*, **21**(4), pp. 397–426.

Du, S., Wang, Q. and Yang, Y., 2004, Fuzzy description of fuzzy direction relations and their similarities. In *Proceedings of 12th International Conference on Geoinformatics – Geospatial Information Research: Bridging the Pacific and Atlantic*, 7–9 June 2004, Gävle, Sweden, pp. 496–502.

Dubois, T. and Prade, H., 1993, Fuzzy sets and probability: misunderstandings, bridges and gaps. In *Proceedings of 2nd IEEE Conference on Fuzzy Systems*, 4–5 March 1993, San Francisco, CA, USA, 1993, pp. 1059–1068.

Dutta, S., 1990, Qualitative spatial reasoning: a semi-quantitative approach using fuzzy logic. *Lecture Notes in Computer Science*, **409**, pp. 345–364.

Dutta, S., 1991, Approximate spatial reasoning: integrating qualitative and quantitative constraints. *International Journal of Approximate Reasoning*, **5**(3), pp. 307–330.

Dutton, G., 1992, Handling positional uncertainty in spatial databases. In *Proceedings of 5th International Symposium on Spatial Data Handling*, 3–7 August 1992, Charleston, South Carolina, pp. 460–469.

Egenhofer, M.J. and Herring, J., 1991, *Categorizing Binary Topological Relations between Regions, Lines and Points in Geographic Databases*, Technical Report, p. 28 (Orono, ME: Department of Surveying Engineering, University of Maine).

Egenhofer, M.J. and Shariff, A.R., 1998, Metric details for natural-language spatial relations. *ACM Transactions on Information Systems*, **16**(4), pp. 295–321.

Fisher, P.F., 1999, Models of Uncertainty in Spatial Data. In P. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind (Eds), *Geographical Information Systems: Principles and Applications*, 2nd edn, Vol. 1, pp. 191–205 (New York: John Wiley & Sons).

Fishman, G.S., 1996, *Monte Carlo: Concepts, Algorithms, and Applications*, p. 698 (New York: Springer).

Frank, A.U., 1991, Qualitative spatial reasoning about cardinal directions. In D. Mark and D. White (Eds), *Proceedings of the 7th Austrian Conference on Artificial Intelligence*, pp. 157–167 (Baltimore, MA: Morgan Kaufmann).

FRANK, A.U., 1996, Qualitative spatial reasoning: cardinal directions as an example. *International Journal of Geographical Information Systems*, **10**(3), pp. 269–290.

FREKSA, C., 1992, Using orientation information for qualitative spatial reasoning. *Lecture Notes in Computer Science*, **639**, pp. 162–178.

GOODCHILD, M.F., 1989, Modeling error in objects and fields. In M.F. Goodchild and S. Gopal (Eds), *Accuracy of Spatial Databases*, pp. 107–114 (London: Taylor & Francis).

GOODCHILD, M.F., 1992, Geographical data modeling. *Computers and Geosciences*, **18**(4), pp. 401–408.

GOODCHILD, M.F., 2007, Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research,* 2007, **2**, pp. 24–32.

GOYAL, R. and EGENHOFER, M.J., 2001, Similarity of cardinal directions. *Lecture Notes in Computer Science*, **2121**, pp. 36–55.

GUO, Q., LIU, Y. and WIECZOREK, J., 2008, Georeferencing locality descriptions and computing associated uncertainty in a probabilistic approach. *International Journal of Geographical Information Science*, **22** (10), pp. 1067–1090.

HILL, L.L., 2006, *Georeferencing – The Geographic Associations of Information* (Cambridge: MIT Press).

HILL, L.L., FREW, J. and ZHENG, Q., 1999, Geographic names: the implementation of a gazetteer in a georeferenced digital library. D-Lib Magazine, **5**(1), available online at: http://www.dlib.org/dlib/january99/hill/01hill.html.

JACQUEZ, G.M. and JACQUEZ, J.A., 1999, Disease clustering for uncertain locations. In A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel and R. Bertollini (Eds), *Disease Mapping and Risk Assessment for Public Health Decision Making*, pp. 151–168 (London: John Wiley & Sons).

JONES, C.B. and PURVES, R.S., 2008, Geographical information retrieval. *International Journal of Geographical Information Science*, **22**(3), pp. 219–228.

KETTANI, D. and MOULIN, B., 1999, A spatial model based on the notions of spatial conceptual map and of object's influence areas. *Lecture Notes in Computer Science*, **1661**, pp. 401–416.

LIU, Y., WANG, X., JIN, X. and WU, L., 2005, On internal cardinal direction relations. *Lecture Notes in Computer Science*, **3693**, pp. 283–299.

LONGLEY, P.A., GOODCHILD, M.F., MAGUIRE, D.J. and RHIND, D.W., 2005, *Geographic Information Systems and Science*, 2nd edn, p. 535 (New York: John Wiley & Sons).

MaNIS (MAMMAL NETWORKED INFORMATION SYSTEM) 2001, available online at: http://manisnet.org (accessed 1 November 2007).

MARK, D.M., 1999, Spatial representation: a cognitive view. In P. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind (Eds), *Geographical Information Systems: Principles and Applications*, 2nd edn, **1**, pp. 81–89 (New York: John Wiley & Sons).

MARK, D.M. and EGENHOFER, M.J., 1994, Modeling spatial relations between lines and regions: combining formal mathematical models and human subjects testing. *Cartography and Geographical Information Systems*, **21**(3), pp. 195–212.

MONTELLO, D.R. and FRANK, A.U., 1996, Modeling directional knowledge and reasoning in environmental space: testing qualitative metrics. In J. Portugali, (Ed), *The Construction of Cognitive Maps*, pp. 321–344 (Dordrecht: Kluwer Academic Publishers).

MONTELLO, D.R., GOODCHILD, M.F., GOTTSEGEN, J. and FOHL, P., 2003, Where's downtown? Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation*, **3**, pp. 185–204.

MORATZ, R. and WALLGRÜN, J.O., 2003, Spatial reasoning about relative orientation and distance for robot exploration. *Lecture Notes in Computer Science*, **2825**, pp. 61–74.

PURVES, R.S., CLOUGH, P., JONES, C.B., ARAMPATZIS, A., BUCHER, B., FINCH, D., FU, G., JOHO, H., SYED, A.K., VAID, S. and YANG, B., 2007, The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet. *International Journal of Geographical Information Science*, **21**(7), pp. 717–745.

RANDELL, D.A., CUI, Z. and COHN, A.G., 1992, A spatial logic based on regions and connection. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference (KR'92)*, pp. 165–176 (San Francisco, CA: Morgan Kaufmann).

RATCLIFFE, J.H., 2004, Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science*, **18**(1), pp. 61–72.

RENZ, J., 2002, Qualitative spatial reasoning with topological information. *Lecture Notes in Artificial Intelligence*, **2293**, p. 223.

RENZ, J. and LIGOZAT, G., 2005, Weak composition for qualitative spatial and temporal reasoning. *Lecture Notes in Computer Science*, **3709**, pp. 534–548.

ROSCH, E. and MERVIS, C.B., 1975, Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, **7**, pp. 573–605.

ROSS, T., BOOKER, J.M. and PARKINSON, W., 2002, *Fuzzy Logic and Probability Applications, Bridging the Gap*, p. 409, (Philadelphia, PA: SIAM Press).

ROWE, R.J., 2005, Elevational gradient analyses and the use of historical museum specimens: a cautionary tale. *Journal of Biogeography*, **32**(11), pp. 1883–1897.

SCHLIEDER, C., VÖGELE, T. and VISSER, U., 2001, Qualitative spatial representation for information retrieval by gazetteers. *Lecture Notes in Computer Science*, **2205**, pp. 336–351.

SCHMIDTKE, H.R., 2001, The house is north of the river: relative localization of extended objects. *Lecture Notes in Computer Science*, **2205**, pp. 415–430.

SCHNEIDER, M., 1999, Uncertainty management for spatial data in databases: fuzzy spatial data types. *Lecture Notes in Computer Science*, **1651**, pp. 330–351.

SHI, W., 1998, A generic statistical approach for modelling error of geometric features in GIS. *International Journal of Geographical Information Science*, **12**(2), pp. 131–143.

SHI, X., 2007, Evaluating the uncertainty caused by Post Office Box addresses in environmental health studies: a restricted Monte Carlo approach. *International Journal of Geographical Information Science*, **21**(3), pp. 325–340.

SILVA, M.J., MARTINS, B., CHAVES, M., AFONSO, A.P. and CARDOSO, N., 2006, Adding geographic scopes to web resources. *Computers, Environment and Urban Systems*, **30**(4), pp. 378–399.

SKIADOPOULOS, S., GIANNOUKOS, C., VASSILIADIS, P., SELLIS, T. and KOUBARAKIS, M., 2004, Computing and handling cardinal direction information. *Lecture Notes in Computer Science*, **2992**, pp. 329–347.

TAKEMURA, C.M., CESAR JR., R.M. and BLOCH, I., 2005, Fuzzy modeling and evaluation of the spatial relation 'along'. *Lecture Notes in Computer Science*, **3773**, pp. 837–848.

TANG, X., 2004, *Spatial Object Modeling in Fuzzy Topological Spaces with Applications to Land Cover Change*, PhD thesis, International Institute for Geo-Information Science and Earth Observation.

VAN NIEL, T.G. and MCVICAR, T.R., 2002, Experimental evaluation of positional accuracy estimates from a linear network using point- and line-based testing methods. *International Journal of Geographical Information Science*, **16**(5), pp. 455–473.

WANG, X., LIU, Y., GAO, Z. and GAO, Y., 2005, Landmark-based qualitative reference system. In *Proceedings of IGARSS 2005*, Vol. 2, pp. 932–935 (Piscataway, NJ: IEEE Press).

WIECZOREK, J., GUO, Q. and HIJMANS, R.J., 2004, The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, **18**(8), pp. 745–767.

WORBOYS, M.F. and CLEMENTINI, E., 2001, Integration of imperfect spatial information. *Journal of Visual Languages and Computing*, **12**, pp. 61–80.

WORBOYS, M.F., 2001, Nearness relations in environmental space. *International Journal of Geographical Information Science*, **15**(7), pp. 633–651.

YAO, X. and THILL, J.-C., 2006, Spatial queries with qualitative locations in spatial information systems. *Computers, Environment and Urban Systems*, **30**(4), pp. 485–502.
ZADEH, L.A., 1995, Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics*, **37**(3), pp. 271–276.