Imprecision in Finite Resolution Spatial Data

MICHAEL WORBOYS

Department of Computer Science, Keele University, Staffs ST5 5BG UK

Abstract

An important component of spatial data quality is the imprecision resulting from the resolution at which data are represented. Current research on topics such as spatial data integration and generalisation needs to be well-founded on a theory of multi-resolution. This paper provides a formal framework for treating the notion of resolution and multi-resolution in geographic spaces. It goes further to develop an approach to reasoning with imprecision about spatial entities and relationships resulting from finite resolution representations. The approach is similar to aspects of rough and fuzzy set theories. The paper concludes by providing the beginnings of a geometry of vague spatial entities and relationships.

Keywords: uncertainty, vagueness, rough set, fuzzy set, resolution, spatial reasoning, data quality

1. Introduction

The notion of spatial resolution is fundamental to many aspects of the representation of spatial data, and a proper formulation of a multi-resolution data model is a key prerequisite for spatial data integration. Without a careful treatment of resolution and the consequent imprecise data representations, it is difficult to approach correctly scale and generalisation; both intimately connected with data quality issues.

This paper presents a formal treatment of resolution, including a model of spatial data entities that can exist in a multi-resolution setting. Observations and representations of spatial entities at any finite resolution will necessarily introduce bounds to the precision at which these entities can be handled. The paper discusses an approach to handling and reasoning with the uncertainty resulting from such representations using techniques which bear resemblance in places to rough set theory. The author hopes that such a treatment can be applied in areas such as the introduction of operators and reasoning arising from these kinds of spatial uncertainty in integrated, multi-resolution data models and database interaction languages.

The paper begins by providing motivation, background and a survey of some relevant literature around the topics of multi-resolution data models, uncertainty, vagueness and precision. It continues with a formal treatment of the foundational notion of resolution. A theory of representation of spatial entities with respect to multiple resolutions is developed. The paper concludes by showing some examples of the geometry of imprecise objects that emerges as an application of this theory, as well as indicating directions for further research. Throughout, we make the assumption that all sets are finite.

2. Background

Almost all the information that we possess about the real world is neither certain, complete, nor precise. Limitations upon geographic data quality are generated at all stages of the data life-cycle, from capture, through input, manipulation and analysis, to the presentation of results. It is now widely recognised (e.g. [12]) that data quality is an important component of the description of the data, that each data entity should carry information describing its quality, that each operation on the data should have associated error tracking procedures, and that systems should contain quality control mechanisms as standard. It is necessary to know the quality of the information in order to be able to use it effectively. Information on quality arises both in the modelling, representation and storage of data in databases, and also in their analysis, reasoning and visualization [17].

2.1 Components of spatial data quality

Data quality is multi-dimensional, with components ranging from subjective aspects such as fitness-for-purpose, to objective measurables like deviation from observed or otherwise known true values. Limitations on data quality associated with uncertainty can arise for a variety of reasons. They may be inherent in the real-world entities that are under observation (e.g. the vague notion of 'the South of England'). Some data (e.g. statistical data) are inherently imperfect, or are deliberately degraded for security reasons. Associated with data capture are inherent limitations of measuring instruments and erroneous readings. Quality degradation can arise at data input, or result from application of inappropriate data models and representations. It may be propagated by the computational operations performed on the data to achieve the required results, or can result from inappropriate presentation of the results of the computational processes.

Deficiencies in data quality, all leading to different kinds of uncertainty, may be the result of several factors:

- *Inaccuracy and error*: deviation from true values.
- *Vagueness*: imprecision in concepts used to describe the information (e.g. near and far are vague metric properties of a spatial object describing distance from the observer).
- *Incompleteness*: lack of relevant information.
- *Inconsistency*: conflicts arising from the information.
- *Imprecision*: limitation on the granularity or resolution at which the observation is made, or the information is represented.

Error as a component of data quality has been quite widely treated in the GIS literature. Hunter and Goodchild [14] survey means of communicating and reasoning with spatial data error, which include ignoring it, epsilon bands, misclassification matrices, map reliability diagrams, fuzzy logic, probability surfaces, and variability diagrams. They provide a case study of some possible treatments of error.

Couclelis [4] notes that vagueness can result from the inherent nature of the object ('the South of England' is inherently more vague than 'the county of Surrey', although both regions have some inherent vagueness); the method of observation; and the purpose and requirements of the user.

Inconsistency and incompleteness are present in many spatial data sets. A particular range of applications where these deficiencies in data quality cause particular difficulties is provided by highly dynamic situations, such as in transportation networks or battlefield scenarios, where observations are subject to noise, conflict and incompleteness [26].

The arguments contained in this paper focus on the contribution that imprecision makes to uncertainty. Given the nature of digital computation, all data, spatial or otherwise, can be represented at only a finite precision. The paper focuses on the concepts of resolution and multi-resolution. It should be emphasized that it is not just narrowly concerned with pixel resolution in a raster image, but takes a wider view, where any computational spatial data model is seen as based upon some resolution structure. The well-known triangulated irregular networks (TIN) representation, and the ‘realms’ representations of Güting and Schneider [13], both provide example of representations with respect to particular resolution structures. The hierarchical terrain models of De Floriani and her colleagues [11] give examples of representations with respect to multi-resolution structures. Other work that takes a formal approach to spatial and temporal resolution has been undertaken by Euzenat [9].

2.2 Formal treatments of spatial uncertainty and imprecision

A currently fashionable approach to locational uncertainty in the GIS literature is the application of fuzzy reasoning, originated in a more general setting by Zadeh [27], [28]. By far the most common application of fuzzy reasoning techniques is to reasoning about and representing the locations of boundaries. For example Leung [15] constructs a model in which a boundary is represented as a zone in which attribute values change continuously from values associated with one region to that of its adjacent region. Leung’s methods have something in common with the approach taken later in this paper. Wang and Brent Hall [25] describe Leung’s partitioning of a region into its *core* (‘the area whose characteristics are most compatible with the linguistic proposition characterising the region’) and its *boundary* (‘the area whose characteristics are more or less compatible with the linguistic proposition characterising the region’). The fuzzy approach to geographic regional boundaries (and thus to the regions themselves) has been researched by several authors (e.g. [7], [15], [16], [25]). The essence of the approach is that uncertainty of membership of a location in a region is indicated by a real number between 0 and 1, where a membership value of 0 indicates that the location is definitely not in the region, a value of 1 that the location definitely is in the region, and the magnitude of an intermediate value indicating a level of certainty that the location is in the region.

The focus of this paper is an analysis of the effect that the resolution of representation has on uncertainty associated with the information. In particular, if the same or related information is represented at a collection of resolutions, how can we reason about the integration? A formalism on which some aspects of reasoning of this kind may be based is provided by the theory of rough sets. As rough sets have not been considered much in the GIS literature, we take some space to introduce the basic ideas.

The starting point of the theory of rough sets (e.g. [18], [19], [20]) is that entities can only be perceived by making observations about them, and that the observations provide information at differing degrees of precision and accuracy. Any particular observation is made at some granularity or resolution, in which collections of elements are indiscernible from each other. The higher the resolution (or lower the granularity), the better we discern differences between elements.

Formally, for any observation, assume an *indiscernibility relation* ρ on set S, where $t \rho s$ can be read as t is indiscernible from s.

An indiscernibility relation ρ on set S leads to a collection of subsets of S, defined by:

$$R(s) = \{t \in S \mid t \rho s\}$$

In most of the literature on rough sets, and for our considerations, it is assumed that ρ is an equivalence relation, so that sets of the form $R(s)$ for $s \in S$ form a partition of S. The set of equivalence classes of S with respect to ρ is denoted S / ρ.

We may use this definition of indiscernibility to define the following two rough set constructs:

$$L(T) = \{x \in S / \rho \mid x \subseteq T\}$$
$$U(T) = \{x \in S / \rho \mid x \cap T \neq \emptyset\}$$

called the *lower* and *upper approximations* to set T (with respect to indiscernibility relation ρ on set S). In this case the approximations are defined in terms of the equivalent classes. Sometimes, it is useful to view the approximations directly in terms of elements of the underlying set S. In this case, we define:

$$L^*(T) = \cup \{x \in S / \rho \mid x \subseteq T\}$$
$$U^*(T) = \cup \{x \in S / \rho \mid x \cap T \neq \emptyset\}$$

We note that $L^*(T) \subseteq T \subseteq U^*(T)$. It is the case that $L^*(T) = U^*(T)$ (or equivalently $L(T) = U(T)$), if and only if the set T can be defined precisely (crisply) with respect to ρ; otherwise T is defined approximately (roughly) with respect to ρ.

It can be seen that the rough set approach is about formalisation of reasoning under imprecision. In that sense, it is a more specialised tool than fuzzy logic, which is concerned with a more generic notion of uncertainty. It will be clear that a representation of spatial entities at any resolution results in an indiscernibility relation, and this fact is used in the next section.

3. Resolution and resolution objects

In this section we provide a rigorous treatment of precision in the context of spatial data representation, focusing on the notion of *resolution*. We have already noted that resolution is not to be narrowly viewed as solely concerned with pixel resolution in a raster image, but more generally, so that any computational spatial data model is seen as based on some resolution structure.

Even in the simplest case of a physical resolution dependent upon pixel size, problems arise from the uncertainty imposed by the lack of precision. Fisher [10] writes that 'the pixel, the elementary unit of analysis in remote sensing and the usual vehicle for integrating data between GIS and remote sensing, is a delusion which may become a snare for the unwary given the way it is treated in most modern software'. Fisher also notes that pixels will not in general be nicely positioned in geographic space so as to match its contents. He gives examples where a pixel might contain entities of sub-pixel size, a boundary (which might itself be uncertain), linear sub-pixel objects, or a continuous gradation from one value to another. Clearly the notion of imprecision caused by finite resolution needs to be tackled head on, and to this end the following formal definition is provided.

3.1 Resolution

Let S be a set of locations (which may be, but does not have to be, a connected region of the Euclidean plane). A *resolution* R of S is a finite partition of S. Alternatively, a resolution is defined by the equivalence relation ρ that gives rise to the finite partition R. As in [24] we call an element $x \in R$ a *resel*.

We note that a resolution is *any* partition of the underlying set into a finite number of subsets. The partition may arise from a pixellation of the space and be a regular square grid, or be more formed in more complex ways, for example from a triangulated irregular network (TIN).

3.2 Resolution objects

Let R be a resolution of S. A *resolution object* (*R-object)* is an object defined with respect to a particular resolution R. It is defined in a similar way to a rough set as the two-stage set $<L, U>$, where $L \subseteq U \subseteq R$. The intuition is that each resel in L is definitely part of the R-object; each resel in U may or may not be part of the R-object; and each resel not in U is definitely not part of the R-object.

Figure 1 shows an example of a resolution object with respect to an irregular triangular resolution. The darker shaded grey area represents the definite part of the object, while the lighter shaded grey areas contain those locations that may be part of the object.

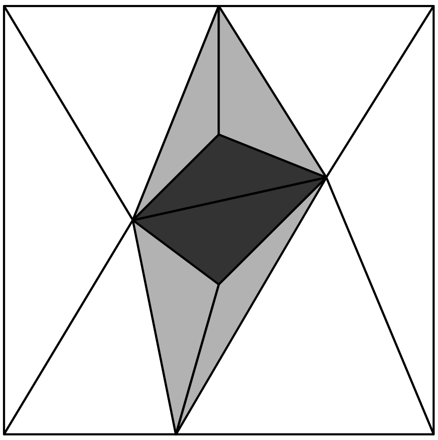

Figure 1: An example of a resolution object

It should be noted that we are not necessarily defining a resolution object with reference to any particular predefined subset of S. In this sense our treatment is different from rough set theory, where upper and lower approximations may be found if we are given a subset T of S, as in section 2.2. However, there is a sense in which the two-stage set $<L, U>$ (where there is no prior mention of an underlying subset) may be viewed as an approximation to some subset T of S with respect to the precision provided by resolution R. Rough set theory would provide the following interpretation of this approximation:

$$L(T) = \{x \in R \mid x \subseteq T\}$$

$$U(T) = \{x \in R \mid x \cap T \neq \varnothing\}$$

We note that $L(T) \subseteq U(T)$.

There is however a more general and maybe more useful interpretation of the two-stage set $<L, U>$ viewed as an approximation to a subset T of S with respect to the precision provided by resolution R.

$$L(T) \subseteq \{x \in R \mid x \subseteq T\}$$
$$U(T) \supseteq \{x \in R \mid x \cap T \neq \varnothing\}$$

Again $L(T) \subseteq T \subseteq U(T)$. The advantage of this interpretation is that it allows for the possibility of classes of $U(T)$ having *every* element in T, or having *no* element in T.

The intuition that both these interpretations are attempting to capture is that of an approximation to a spatial object T, provided by an observation of T at a particular resolution, where resels in $L(T)$ contain locations all of which are definitely in T, while resels in $U(T)$ contain locations some of which may be in T and some may not be in T, and resels outside $U(T)$ contain locations all of which are definitely not in T.

4. An integrated model of multiple resolutions

In Earth observation, sensors exist at a wide range of different resolutions, from low resolutions of more than one kilometre (e.g. NOAA AVHRR) to fine granularities of about 10m (e.g. SPOT HVR) [10]. Integration of images represented at these different resolutions with each other, and with other forms of data in a GIS, is dependent on an appropriate model of multi-resolution spatial data. Multi-resolution spatial data models are clearly important for all forms of spatial data integration.

The representation of a single data set at a range of resolutions and with consequent varying levels of generalization is necessary if resolution is to be dependent on the context in which the data set is required. If the context is dependent on user role, then resolution may well vary with user role. For example, a cyclist would require information presented at a different resolution from a car driver. When the context is provided by user location, then a user would often require more detailed data available about the current location, with less detail at the periphery of interest. In this case, resolution for different areas may vary as the user moves around the geographic space.

4.1 The resolution space

A formal model that allows an integrated general treatment of multiple resolutions is necessary to provide a foundation for a rigorous multi-resolution spatial data model. The development of such an integrated model of multiple resolutions is the subject of this section.

Let **R** be the set of all resolutions of a set S. A partial order $\leq$ may be imposed on **R** as follows. For R_1 and R_2 belonging to **R**, $R_1 \leq R_2$ if and only if $\forall x \in R_1, \exists y \in R_2$ such that $x \subseteq y$. In terms of equivalence relations ρ_1 and ρ_2, $\rho_1 \leq \rho_2$ if and only if $\forall u, v \in S$, $u \, \rho_1 \, v$ implies $u \, \rho_2 \, v$.

The idea behind these definitions is that the smaller resolutions in the partial ordering provide the finer granularities. We note that the ordering is not total, for in figure 2, neither $R_1 \leq R_2$ nor $R_2 \leq R_1$. We show that **R** with the partial order just defined is a lattice.

For any two resolutions R_1 and R_2 of S, it is always possible to form their greatest lower bound $R_1 \wedge R_2$ and least upper bound $R_1 \vee R_2$ (see figure 2 for an example).

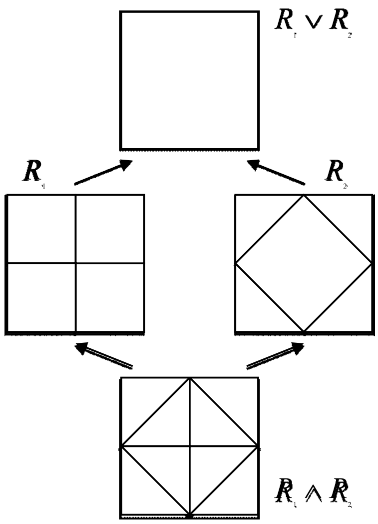

Figure 2: The greatest lower bound and least upper bound of two resolutions.

The greatest lower bound $R_1 \wedge R_2$ of resolutions R_1, R_2 is given by:

$$R_1 \wedge R_2 = \{x \cap y \mid x \in R_1, y \in R_2, \text{ and } x \cap y \neq \emptyset\}$$

or, in terms of equivalence relations, the greatest lower bound $\rho_1 \cap \rho_2$ of resolutions ρ_1, ρ_2 is given by:

For all $u, v \in S$, $u\ (\rho_1 \wedge \rho_2)\ v$ if and only if $u\ \rho_1\ v$ and $u\ \rho_2\ v$.

It is easy to show that these definitions are equivalent and that indeed $R_1 \cap R_2$ is the greatest lower bound of R_1 and R_2.

The definition may be generalised to the greatest lower bound of any finite set $R = \{R_1, \ldots, R_n\}$ of resolutions, as follows:

$$\wedge R = \{x_1 \cap \ldots \cap x_n \mid x_1 \in R_1, \ldots, x_n \in R_n \text{ and } x_1 \cap \ldots \cap x_n \neq \emptyset\}$$

It would be natural do define the least upper bound ρ of resolutions ρ_1, ρ_2 as $u\ \rho\ v$ if and only if $u\ \rho_1\ v$ or $u\ \rho_2\ v$. However, this will not work as the resultant is not necessarily transitive and so may not be an equivalence relation. We need to ensure transitivity by the following definition. The least upper bound $\rho_1 \vee \rho_2$ of resolutions ρ_1, ρ_2 is given by:

For all $u, v \in S$, $u\ (\rho_1 \vee \rho_2)\ v$ if and only $\exists\ w_1, \ldots, w_m \in S$ such that
$u\ \alpha_0\ w_1, w_1\ \alpha_1\ w_2, \ldots, w_m\ \alpha_m\ v$
where $\alpha_i \in \{\rho_1, \rho_2\}$ for $0 \leq i \leq m$.

The least upper bound $R_1 \vee R_2$ of resolutions R_1 and R_2 is given by the equivalence classes resulting from the definition immediately above.

The definition may be generalised to the least upper bound $\rho_1 \vee \ldots \vee \rho_n$ of any finite set $\{\rho_1, \ldots, \rho_n\}$ of resolutions, as follows:

For all $u, v \in S$, $u\ (\rho_1 \vee \ldots \vee \rho_n)\ v$ if and only $\exists\ w_1, \ldots, w_m \in S$ such that
$u\ \alpha_0\ w_1, w_1\ \alpha_1\ w_2, \ldots, w_m\ \alpha_m\ v$
where $\alpha_i \in \{\rho_1, \ldots, \rho_n\}$ for $0 \leq i \leq m$.

The least upper bound $\vee R$ of the set $R = \{R_1, \ldots, R_n\}$ of resolutions is given by the equivalence classes resulting from the definition immediately above.

Thus, **R** is a lattice. Furthermore, **R** has top and bottom elements. The top element $\top$ is the resolution consisting of the single resel, that is S. The bottom element $\perp$ is the resolution whose resels are the singleton sets $\{s\}$ for each $s \in S$.

Any sublattice of **R** (not necessarily with the same top and bottom elements) will be termed a *resolution space*. Of course, any resolution space will be closed with respect to the lattice operations defined above.

We can note that in general a resolution lattice is not modular (and therefore not distributive [6]). A modular lattice L has the property that for all $x, y, z \in S$, if $z \leq x$ then $x \wedge (y \vee z) = (x \wedge y) \vee z$

To see that a resolution lattice is in general not modular, let $S = \{a, b, c, d\}$, and

$$R_1 = \{\{a, b\}, \{c, d\}\}$$
$$R_2 = \{\{a, c\}, \{b, d\}\}$$
$$R_3 = \{\{a\}, \{b\}, \{c, d\}\}$$

Then, $R_3 \leq R_1$ and $R_1 \wedge (R_2 \vee R_3) = \{\{a, b\}, \{c, d\}\} = R_1$, but
$(R_1 \wedge R_2) \vee R_3 = \{\{a\}, \{b\}, \{c, d\}\} = R_3$. So, in this case,
$R_1 \wedge (R_2 \vee R_3) \neq (R_1 \wedge R_2) \vee R_3$

5. Multi-resolution objects

Having defined a setting in which multiple resolutions may be handled in an integrated manner, the next step is to consider the type of entities that may inhabit such a resolution space. We have already discussed how an entity may be represented as a two-stage set with respect to a single resolution. In a multi-resolution context, we need a way of determining how an object given at one resolution is represented at another, and also we must be able to amalgamate the representations at several resolution of an object into a single more precise representation.

5.1 The embedded object space

As with rough approximations in section 2.2, a resolution object is to be distinguished from its embedding in the set S. Any subset X of a resolution R of set S has an embedding defined by:

$$e(X) ::= \cup X$$

Then the embedding of a resolution object $O = \langle L, U \rangle$, where $L \subseteq U \subseteq R$, is given by:

$$e(O) ::= \langle e(L), e(U) \rangle$$

and we note that $e(L) \subseteq e(U)$.

Define a relation ~ on the set of all resolution objects in a resolution space as follows. Let $O_1 = <L_1, U_1>$ be an R_1-object, and $O_2 = <L_2, U_2>$ be an R_2-object. Then $O_1 \sim O_2$ if and only if $e(O_1) = e(O_2)$. It is immediate that ~ is an equivalence relation. The set of equivalence classes is called *the embedded object space*.

Now, we define an ordering in the embedded object space. Let T_1 and T_2 be elements of the embedded object space. Let $T_1 = [\ O_1\ ]$ and $T_2 = [\ O_2\ ]$, where R_1-object $O_1 = <L_1, U_1>$ and R_2-object $O_2 = <L_2, U_2>$ are representatives of equivalence classes of resolution objects in the embedded object space. Define $T_1 \leq T_2$ if and only if $e(L_1) \supseteq e(L_2)$ and $e(U_1) \subseteq e(U_2)$. It can be shown that ≤ is well-defined in the embedded object space, and in that space is a partial ordering. The intuition behind this ordering is that $T_1 \leq T_2$ expresses the fact that T_1 is a less imprecise approximation to some subset T of S than T_2. That is, T_1 provides at least as much information as T_2 about sets of elements that are certainly contained in T, and sets of elements that are certainly not contained in T.

5.2 Changing resolutions

The question to be addressed in this subsection is, if we know the representation of an entity at one resolution, what is its representation at a second resolution?

Let $O = <L, U>$ be an R-object representing an entity in geographic space, and let a second resolution R' be given. Then, define the R'-object $O' = <L', U'>$ to give the best possible representation of the same entity, in the following way. For $x \in R'$:

$x \in L'$ if and only if $e(x) \subseteq e(L)$

$x \in U'$ if and only if $e(x) \cap e(U) \neq \emptyset$

We note that $O \leq O'$, and this accords with intuition, as we would not expect to gain precision by transferring the representation of a geographic entity from one resolution to another, without additional information being provided. Figure 3 shows object O_1 represented as object O_2 at a different resolution.

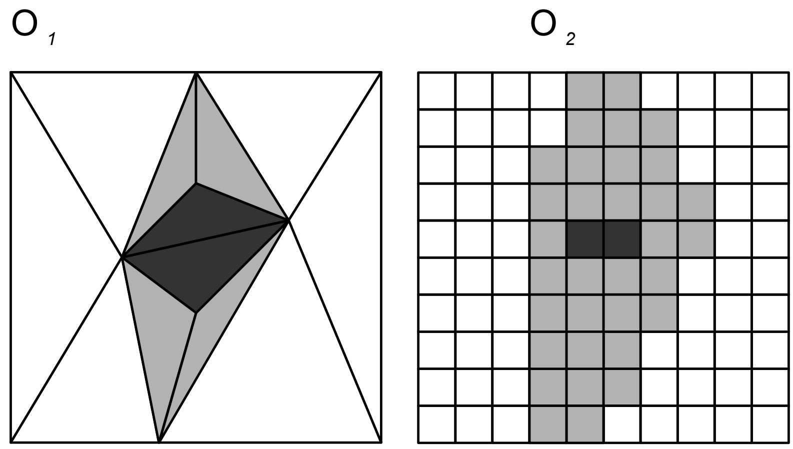

Figure 3: Object O_1 represented as object O_2 at a different resolution

5.3 Compatible resolution objects

In order that a set of observations of an object at different resolutions may be amalgamated, they must have a level of compatibility provided by the following definition.

Suppose we have a set $R = \{R_1, \ldots, R_n\}$ of resolutions and a set $\mathbf{O} = \{O_1, \ldots, O_n\}$ of resolution objects, where for $1 \leq i \leq n$, $O_i = \langle L_i, U_i \rangle$ is an R_i-object. Then **O** is defined to be a *compatible* collection of resolution objects if and only if for $1 \leq i, j \leq n$: $e(L_i) \subseteq e(U_j)$.

The idea is that if the set of resolution objects together define an object, then it cannot be the case that at one resolution a particular location is definitely part of the object while at another resolution the same location is definitely not part of the object. Figure 4 shows an example of two compatible resolutions objects, O_1 and O_2. The reader is invited to verify that the compatibility conditions are satisfied in this case. We note that the compatibility relation really acts in the embedded object space.

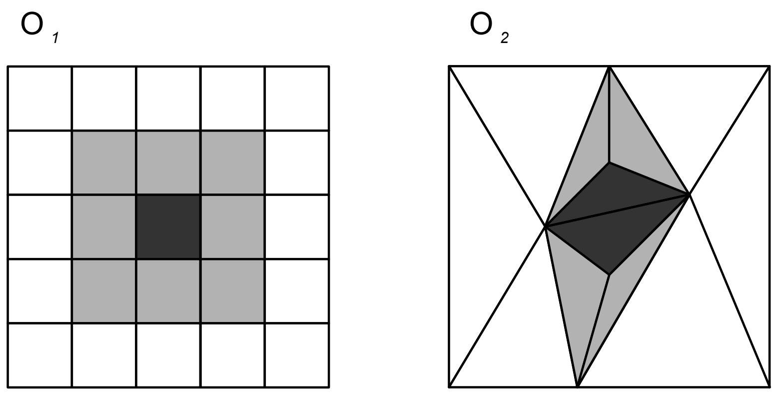

Figure 4: Two compatible resolution objects O_1 and O_2.

5.4 Amalgamation objects

Suppose we have a set $R = \{R_1, \ldots, R_n\}$ of resolutions, and a compatible set $\mathbf{O} = \{O_1, \ldots, O_n\}$ of resolution objects, where for $1 \le i \le n$, $O_i = \langle L_i, U_i \rangle$ is an R_i-object. This section discusses the representation of the resolution object that results. We are able to define an $\wedge R$-*amalgamation object* $\wedge\mathbf{O}$ as follows:

$\wedge\mathbf{O} = \langle L, U \rangle$, where, for $x \in \wedge R$,

$x \in L$ if and only if $e(x) \subseteq e(L_1) \cup \ldots \cup e(L_n)$, and

$x \in U$ if and only if $e(x) \subseteq e(U_1) \cap \ldots \cap e(U_n)$

Figure 5 shows the amalgamation object $O_1 \wedge O_2$, where O_1 and O_2 are given in figure 4.

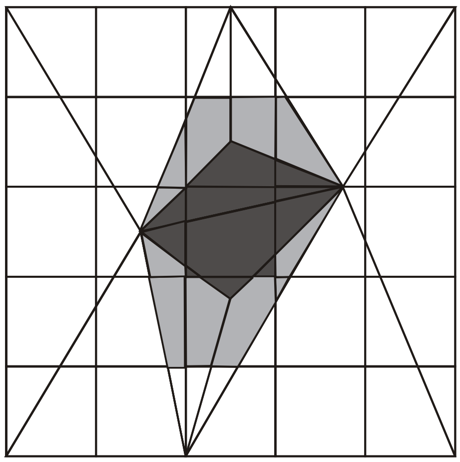

Figure 5: The amalgamation object $O_1 \wedge O_2$

In order to show that in the general case, the amalgamation object $\wedge\mathbf{O}$ is a well-defined resolution object, we just have to show that $L \subseteq U$. So, assume that $x \in L$. Then, by definition $e(x) \subseteq e(L_1) \cup \ldots \cup e(L_n)$. By compatibility, for $1 \le i \le n$,

$e(L_i) \subseteq e\ (U_1)$,

therefore

$e(L_1) \cup \ldots \cup e(L_n) \subseteq e(U_1)$.

In a similar way

$$e(L_1) \cup \ldots \cup e(L_n) \subseteq e(U_2),$$

$$\ldots,$$

$$e(L_1) \cup \ldots \cup e(L_n) \subseteq e(U_n),$$

and so $e(L_1) \cup \ldots \cup e(L_n) \subseteq e(U_1) \cap \ldots \cap e(U_n)$. Thus, $x \in U$.

■

It can also be seen that the amalgamation object $\wedge\mathbf{O}$ is more precise (contains more information) than any of its constituent objects, and this is reflected in the embedded object space, where $[\wedge\mathbf{O}] \leq [O_i]$ for all $1 \leq i \leq n$. To see this, suppose $s \in e(L_i)$. Then there exists $x \in \wedge R$ and $y \in L_i$ such that $x \subseteq y$. Thus, $e(x) \subseteq e(y) \subseteq e(L_i) \subseteq e(L_1) \cup \ldots \cup e(L_n)$, and by the definition of $\wedge\mathbf{O}$, $x \in L$ and $s \in e(L)$. A similar argument works for $e(U) \subseteq e(U_n)$.

The amalgamation construction formalises the notion of providing better and more precise refinements by amalgamating a collection of compatible observations at multiple resolutions.

5.5 Generic objects

The previous subsection provided a process for combining imprecise observations of an spatial entity by amalgamation to provide a more precise single representation - the amalgamation object. This section considers the more general situation where the set of observations may not be globally compatible (that is, the observations may not all be of the same spatial entity), but there will be local compatible subsets of observations. In this case, it is a matter of extracting from the set of observations those subsets that may define single spatial entities, namely the compatible subsets, but also finding those which provide as much precision as possible. To formalise this idea, let $\mathbf{R}$ be a resolution space, and Ω be a finite set of resolution objects observed with respect to resolutions in $\mathbf{R}$. A subset of resolution objects, $\mathbf{O} \subseteq \Omega$, is said *maximal compatible* if:

1. $\mathbf{O}$ is a compatible set of resolution objects.

2. For all $\mathbf{O'} \subseteq \Omega$, if $\mathbf{O} \subseteq \mathbf{O'}$ and $\mathbf{O'}$ is compatible, then $\mathbf{O} = \mathbf{O'}$.

The maximal compatible sets of resolution objects are the largest compatible sets of resolution objects in Ω, and in a sense provide the most precise collections of observations of spatial entities within the observational data that is available. If $\mathbf{O} = \{O_1, \ldots, O_n\}$ is a maximal compatible set of resolution objects, with respect to resolutions $R_1, \ldots, R_n$ in $\mathbf{R}$, then the amalgamation object $\wedge\mathbf{O}$ with respect to resolution $R = R_1 \wedge \ldots \wedge R_n$ (which may not itself belong to Ω) is termed an **R**-*generic object*. The interesting question now is what is the structure of the geometry of **R**-generic objects?

6. Geometry of embedded R-generic objects

Earlier parts of the paper have shown how different representations of the same object may be handled in a space of multiple resolutions. This section discusses some of the properties of the spatial relationships between different objects. There is little in the literature on this topic, although a recent paper [8] contains some similar underlying ideas to those set out below, but in a different formal framework.

As usual, let S be a set, and **R** be a resolution space on *S*. We assume throughout this section that all objects under discussion are **R**-generic, and therefore represented at best possible precision given the constraints of the finite resolution space **R** (see section 5.4). We will not explicitly mention the resolutions to which particular objects are represented. Instead, we work only with the embeddings of the **R**-generic objects, so each embedded object will be of the form $O = \langle A, B \rangle$ where $A \subseteq B \subseteq S$, and A and B are the embeddings respectively of the lower and upper constituents of an **R**-generic object.

This section sets out to show how some basic spatial relationships can be treated in the context of the kinds of imprecise spatial objects under discussion in this paper. The relationships will themselves inherit a degree of imprecision from the imprecision of the constituent spatial objects.

Formally, let ρ be some binary spatial relationship between crisp entities in a geographic space. For example, for crisp spatial entities O_1 and O_2, $\rho(O_1, O_2)$ might indicate that O_1 is connected to O_2, or that O_1 is a part of O_2. The issue is how ρ may be extended to vague geographic objects.

Let $O_1 = \langle A_1, B_1 \rangle$ and $O_2 = \langle A_2, B_2 \rangle$ be two embedded **R**-generic objects. Define

$$\rho^*(O_1, O_2) = \langle \rho(A_1, A_2), \rho(A_1, B_2), \rho(B_1, A_2), \rho(B_1, B_2) \rangle$$

Then ρ^* will provide an extension of the crisp spatial relation ρ. It may be noted that ρ^* takes values in the space $\mathbf{B}^4$, where **B** is the Boolean space of truth values **true** and **false**. This extension in the value space provides the required mechanism for expressing imprecision.

6.1 Vague extensions to the connection and part-whole relations

Two fundamental spatial relationships are *part of* (a mereological relationship) and *connection* (a topological relationship). Mereology has been discussed by several authors (e.g. [22], [23]). Clarke's calculus of individuals [1], [2] has been set in a many-sorted first-order logic [5], the focus of the theory being a reflexive, symmetric, connection relation between spatial regions. Cohn and Gotts [3] provide some ideas for the vague treatment of connection. A unified treatment of mereology and topology is undertaken in [23].

To illustrate our approach, we discuss the vague extension to the connection and part-whole relationships.

Example 1: Vague spatial connection

Suppose that $C(O_1, O_2)$ is the usual connection relation between crisp spatial objects O_1 and O_2. Then the extended connection operator C^* is defined by:

$$C^*(O_1, O_2) = \langle C(A_1, A_2), C(A_1, B_2), C(B_1, A_2), C(B_1, B_2) \rangle$$

There are some dependencies between the constituents of this ordered quadruple.

$C(A_1, A_2)$ implies $C(A_1, B_2)$
$C(A_1, A_2)$ implies $C(B_1, A_2)$
$C(A_1, B_2)$ implies $C(B_1, B_2)$
$C(B_1, A_2)$ implies $C(B_1, B_2)$

These dependencies arise from the relationships

$A_1 \subseteq B_1$

$A_2 \subseteq B_2$

and the intuition that any well-defined connection relationship between crisp regions must have the property that if two regions are connected, then any super-regions of which the original regions are parts must also be connected. These dependencies are shown in figure 6.

$$\begin{tikzcd}
 & C(B_1, B_2) & \\
C(A_1, B_2) \arrow[ur] & & C(B_1, A_2) \arrow[ul] \\
 & C(A_1, A_2) \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Figure 6: Dependencies between constituents of the vague connection operator

These dependencies may now be lifted to give the sublattice of $\mathbf{B}^4$ induced by the connection operator C^*, shown in figure 7, where the quadruples of truth values are the permissible values of the constituents of C^*, constrained by the dependencies in figure 6.

$$\begin{tikzcd}
 & \mathrm{TTTT} & \\
 & \mathrm{FTTT} \arrow[u] & \\
\mathrm{FTFT} \arrow[ur] & & \mathrm{FFTT} \arrow[ul] \\
 & \mathrm{FFFT} \arrow[ul] \arrow[ur] & \\
 & \mathrm{FFFF} \arrow[u] &
\end{tikzcd}$$

Figure 7: Sublattice of $\mathbf{B}^4$ *induced by the connection operator* C^*

In order to reason with C^*, it is necessary to decide how the sublattice in figure 7 is to be interpreted. It may be that the full sublattice, with all the fine distinctions that are made in it, is required. Another more plausibly useful interpretation is to encapsulate all the intermediate values in the lattice, to indicate a state of 'maybe connected'. This allows an expression of the imprecision of the connection relationship between imprecise objects in a 3-valued logic, illustrated in figure 8.

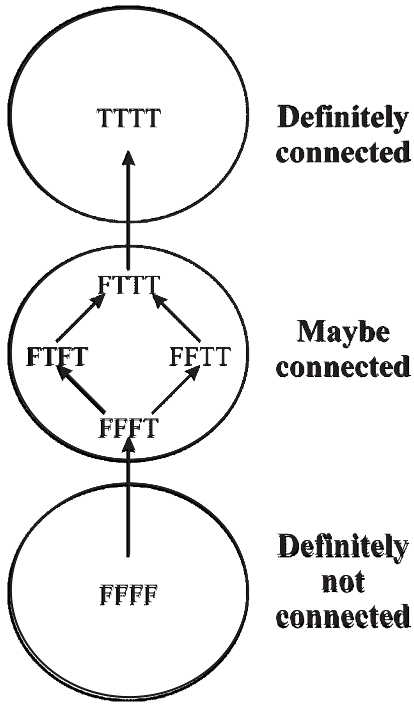

Figure 8: A 3-valued interpretation of the connection operator C^*

Example 2: Vague part-whole relationship

Suppose that $P(O_1, O_2)$ is to be interpreted as that crisp object O_1 is a part of crisp object O_2. Then the extended part-whole operator P^* is defined by:

$$P^*(O_1, O_2) = <P(A_1, A_2), P(A_1, B_2), P(B_1, A_2), P(B_1, B_2)>$$

The following dependencies exist between the constituents of this ordered quadruple

$P(A_1, A_2)$ implies $P(A_1, B_2)$

$P(B_1, A_2)$ implies $P(B_1, B_2)$

$P(B_1, A_2)$ implies $P(A_1, A_2)$

$P(B_1, B_2)$ implies $P(A_1, B_2)$

as any well-defined part-whole relationship between crisp regions must have the property that if a region x is a part of a region y, then x is a part of any super-region of region y, and any sub-region of x is a part of y. These dependencies are shown in figure 9.

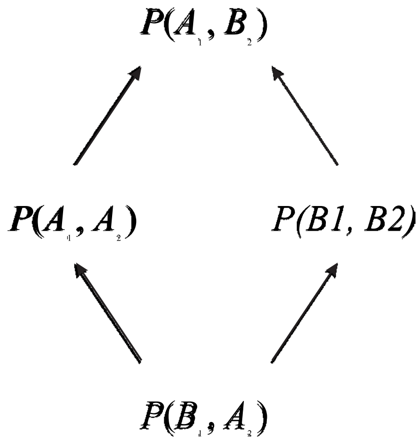

Figure 9: Dependencies between constituents of the vague part-whole operator

These dependencies may now be lifted to give the sublattice of $\mathbf{B}^4$ induced by P^*, shown in figure 10, where the quadruples of truth values are the permissible values of the constituents of P^*, constrained by the dependencies in figure 9.

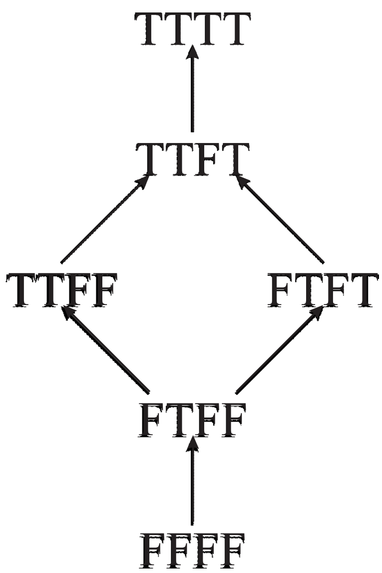

Figure 10: Sublattice of $\mathbf{B}^4$ *induced by the part-whole operator* P^*

As before, it may suit our purpose to reason with the full sublattice. On the other hand, a plausible 3-valued interpretation can be made, as follows. $P^*(O_1, O_2)$ takes truth value **definitely true** if and only if $P(B_1, A_2)$. $P^*(O_1, O_2)$ takes truth and **definitely false** if and only if **not**($P(A_1, B_2)$). The amalgamation of lattice elements into a 3-valued structure is shown figure 11. It is interesting to note the isomorphism between the connection and part-whole sub-lattices.

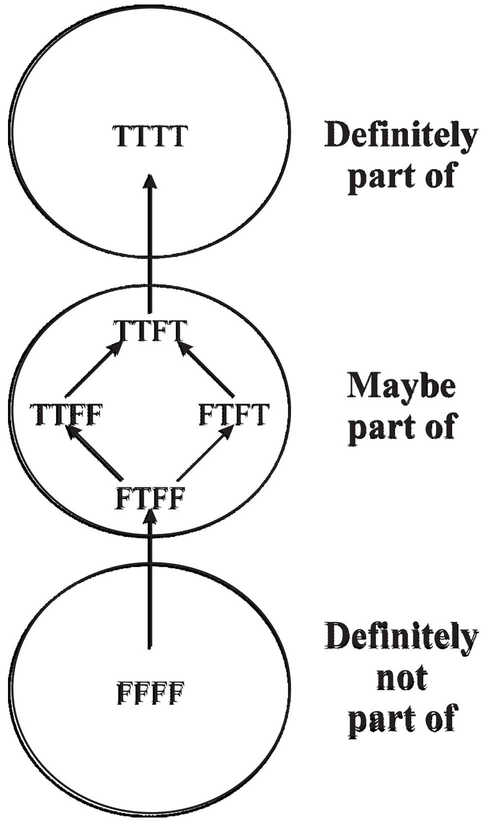

Figure 11: A 3-valued interpretation of the part-whole operator P^*

6.2 Some properties of vague spatial relationships

We conclude this section by giving some examples of the type of geometrical reasoning that we can undertake on imprecise objects using the formalism developed above. The extended vague operators, of which C^* and P^* are examples, are no longer Boolean, but take values in the lattice $\mathbf{B}^4$. In order to reason with these operators we need to decide on the appropriate logic for $\mathbf{B}^4$, or work with an appropriate derived 3-valued logic. A full working-out of this theme will be the subject of another paper, but we can give examples of the type of reasoning involved.

A simple example is the transitivity of the part-whole relation to vague regions. The classical situation is that for all crisp regions O_1, O_2 and O_3, if $P(O_1, O_2)$ and

$P(O_2, O_3)$, then $P(O_1, O_3)$. To extend to P^*, we need to interpret the logical conjunction and implication operators in $\mathbf{B}^4$. In this case, for x, $y \in \mathbf{B}^4$, the interpretation of conjunction is that $x \wedge y \in \mathbf{B}^4$ is the meet of x and x in the sublattice of $\mathbf{B}^4$ shown in figure 10. The interpretation of implication is that for $x = <x_1, x_2, x_3, x_4>$, and $y = <y_1, y_2, y_3, y_4> \in \mathbf{B}^4$

$x \rightarrow y = <x_1 \rightarrow y_1, x_2 \rightarrow y_2, x_3 \rightarrow y_3, x_4 \rightarrow y_4> \in \mathbf{B}^4$,

where inside the quadruple, $\rightarrow$ is interpreted as the usual classical implication operator.

Under these extensions to Boolean logic, the transitivity result carries through.

Theorem 1

For all embedded **R**-generic objects O_1, O_2 and O_3,

$$(P^*(O_1, O_2) \wedge P^*(O_2, O_3)) \rightarrow P^*(O_1, O_3)$$

Proof

Follows immediately from the definition of P^*.

■

A fundamental result concerning the connection and part-of relations for crisp sets (indeed used as a defining property in some formalisms [1], [2], [5]) is the following:

$$P(O_1, O_2) \text{ if and only if, for all } O,\ C(O, O_1) \text{ implies } C(O, O_2) \qquad (*)$$

An analogue of this result for the extended operators C^* and P^* is the following.

Theorem 2

Let O_1 and O_2 be two embedded **R**-generic objects. Then $P^*(O_1, O_2) = <\mathrm{T}, \_, \_, \mathrm{T}>$ if and only if for all O, $C^*(O, O_1) \rightarrow C^*(O, O_2)$

Proof

Suppose $P^*(O_1, O_2) = <\mathrm{T}, \_, \_, \mathrm{T}>$. Then $P(A_1, A_2)$ and $P(B_1, B_2)$, and so by (*):

$$\text{For all } X,\ C(X, A_1) \text{ implies } C(X, A_2) \qquad (1)$$

$$\text{For all } X,\ C(X, B_1) \text{ implies } C(X, B_2) \qquad (2)$$

Take an arbitrary $O = \langle A, B \rangle$, with $A \subseteq B$. Then two applications each of (1) and (2) give:

$C(A, A_1)$ implies $C(A, A_2)$
$C(A, B_1)$ implies $C(A, B_2)$
$C(B, A_1)$ implies $C(B, A_2)$
$C(B, B_1)$ implies $C(B, B_2)$

So,
$\langle C(A, A_1), C(A, B_1), C(B, A_1), C(B, B_1) \rangle \rightarrow \langle C(A, A_2), C(A, B_2), C(B, A_2), C(B, B_2) \rangle$.
That is, $C^*(O, O_1) \rightarrow C^*(O, O_2)$.

Conversely, suppose for all O, $C^*(O, O_1) \rightarrow C^*(O, O_2)$. In particular, for any A, when $O = \langle A, A \rangle$, we have
$\langle C(A, A_1), C(A, B_1), C(A, A_1), C(A, B_1) \rangle \rightarrow \langle C(A, A_2), C(A, B_2), C(A, A_2), C(A, B_2) \rangle$.
So, $C(A, A_1) \rightarrow C(A, A_2)$, and $C(A, B_1) \rightarrow C(A, B_2)$. Therefore, by (*) $P(A_1, A_2)$ and $P(B_1, B_2)$. Therefore, $P^*(O_1, O_2) = \langle \mathrm{T}, \_, \_, \mathrm{T} \rangle$.

■

7. Conclusion

This paper has discussed aspects of the foundations for a theory of spatial imprecision arising from observations of spatial entities and relationships at multiple finite resolutions. We have noted that such imprecision is inherent in much spatial data, and is an important component of spatial data quality. Such imprecision, and the vagueness and uncertainty that follows from it, should not be ‘hidden under the carpet’, as with most current systems, but acknowledged and made fully explicit. This is essential if an appropriate treatment of concepts such as generalisation and multi-resolution are to be well-founded.

Our approach has provided part of the formal framework for multi-resolution geographic spaces. Reasoning with the resulting vagueness has been treated using ideas related to fuzzy logic, and in particular rough set theory. In this paper we have only hinted at the kind of geometry that results. The two examples provided on connection and part-whole will be followed in later work by a more complete collection of spatial operators and relationships. For example, we have not here discussed topological properties of spatial regions, such as connectedness, nor numeric properties such as area and perimeter. Another concept that needs further development in the context of spatial data quality indicators is 'degree of imprecision' of spatial entities and relationships resulting from a collection of finite resolutions. Extensions are under investigation that apply the Dempster-Shafer theory of evidence [21] to spatial reasoning with variable precision and belief levels.

Applications investigated at this stage have been limited to reasoning, under conditions of imprecise observation, about the presence, absence or change of objects (such as buildings or vegetation type). The theory developed in this paper is quite general in that partitions in the resolution spaces can be of spatial or semantic origin (or a hybrid). Further work in the process of publication specializes the approach to semantic domains, while ongoing work is considering types of geometric precision, which we hope will lead to a theory in which the special characteristics of the semantic and spatial domains may be brought together in a unified theory.

Acknowledgements

The author thanks John Stell, Chris Johnson, Jenny Lingham, and Peter Fletcher, of the Keele University GIS Research Group, and Ivo Duentsch of the University of Ulster, for helpful discussions relating to aspects of the work described in this paper.

References

1. B.L. Clarke. A calculus of individuals based on 'connection'. *Notre Dame Journal of Formal Logic* **22**: 204-218, 1981.

2. B.L. Clarke. Individuals and points. *Notre Dame Journal of Formal Logic* **26**: 61-75, 1985.

3. A.G. Cohn and N.M. Gotts. The 'egg-yolk' representation of regions with indeterminate boundaries. In P. Burrough, and A. Frank, (eds.) *Geographic Objects with Indeterminate Boundaries*, Taylor and Francis, London, pp. 171-187, 1996.

4. H. Couclelis. Towards an operation typology of geographic entities with ill-defined boundaries. In P. Burrough and A. Frank, (eds.) *Geographic Objects with Indeterminate Boundaries*, Taylor and Francis, London, pp. 45-55, 1996.

5. Z. Cui, A.G. Cohn, and D.A. Randell. Qualitative and topological relationships in spatial databases. In D. Abel, B.C. Ooi, (eds.), *Advances in Spatial Databases, Proceedings of SSD'93, Singapore*, Lecture Notes in Computer Science 692, Springer, Berlin, pp. 296-315, 1993.

6. B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order.* Cambridge University Press, 1990.

7. T.J. Davis, and C.P. Keller. Modelling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: slope stability prediction. *Int. Jour. of GIS*, 11(5): 409-434, 1997.

8. M. Erwig, and M. Schneider. Vague regions. In Proceedings of the 5th Int. Symp. in Spatial Databases (SSD’97), Lecture Notes in Computer Science 1262, Springer, Berlin, pp. 298-320, 1997.

9. J. Euzenat. An algebraic approach to granularity in qualitative time and space representation. *Proc. Int. Joint Conference on AI,* ACM Publications, Montreal (CA), pp. 894-900, 1995.

10. P. Fisher. The pixel: a snare and a delusion. *Int. Jour. Remote Sensing* **18**(3): 679-685, 1997.

11. L. de Floriani, P. Marzano, and E. Puppo. Hierarchical terrain models: survey and formalization. In *Proceedings SAC'94, Phoenix, AR, USA,* pp. 323-327, 1994.

12. M.F. Goodchild. Data models and data quality: problems and prospects. In M.F. Goodchild, B.O. Parks, and L.T. Steyaert, (eds.), *Visualization in Geographical Information Systems.* John Wiley, New York, pp. 141-149, 1993.

13. R.H. Güting, and M. Schneider. Realms: A foundation for spatial data types in database systems. In D. Abel, B.C. Ooi, (eds), *Advances in Spatial Databases, Proceedings of SSD'93, Singapore, Lecture Notes in Computer Science 692.* Springer-Verlag, Berlin, Germany, pp. 14-35, 1993.

14. G.J. Hunter, and M.F. Goodchild. Dealing with error in spatial databases: A simple case study. *Photogrammetric Engineering and Remote Sensing*, **61**(5): 529-537, 1995.

15. Y. Leung. On the imprecision of boundaries. *Geographical Analysis*, **19**: 125-151, 1987.

16. D.M. Mark, and F. Csillag. The nature of boundaries on 'area-class' maps. *Cartographica* **26**: 65-77, 1989.

17. S. Parsons. Current approaches to handling imperfect information in data and knowledge bases, *IEEE Transactions on Knowledge and Data Engineering,* **8**(3): 353-372, 1996.

18. Z. Pawlak. Rough sets. *Int. Journal of Inf. and Comp. Sci*., **11**(5): 341-356, 1982.

19. Z. Pawlak. *Rough Sets - Theoretical Aspects of Resoning about Data*, Kluwer, 1991.

20. Z. Pawlak. Hard and soft sets. In Alagar, V.S., Bergler, S. and Dong, F.Q. (eds.), *Proceedings of RSSC'94, The Third International Conference on Rough Sets and Soft Computing*, San Jose State University, San Jose, CA, USA, 1993.

21. G. Shafer. A Mathematical Theory of Evidence . Princeton University Press, New Jersey, 1976.

22. P. Simons, Parts. A Study in Ontology. Clarendon Press, Oxford, 1987.

23. B. Smith. Mereotopology - A theory of parts and boundaries. *Data and Knowledge Engineering*, **20**: 287-303, 1996.

24. W.R. Tobler. Application of image processing techniques to map processing. In *Proc. 1st Int. Symp. on Spatial Data Handling,* Universität Zurich-Irchel, Zurich **1**: 140-144, 1984.

25. F. Wang, and G. Brent Hall. Fuzzy representation of geographical boundaries in GIS. *Int. Jour. of GIS* 10(5): 573-590, 1996.

26. G.J. Williams. Templates for spatial reasoning in responsive geographical information systems. *Int. Jour. of GIS*, **9**(2): 117-131, 1995.

27. L.A. Zadeh. Fuzzy sets. *Information and Control* **8**: 338-353, 1965.

28. L.A. Zadeh. Fuzzy logic. *IEEE Computer* **21**: 83-93, 1988.