Integration of Imperfect Spatial Information

Michael F Worboys
Department of Computer Science
Keele University, Staffordshire
UK ST5 5BG

Eliseo Clementini
Department of Electrical Engineering
University of L'Aquila I-67040 Poggio di Roio
L'Aquila Italy

Abstract

The theme of this paper is integration of information arising from observations of spatial entities and relationships. The assumption is that observations are imperfect; in particular, that they are imprecise and inaccurate. Each observation is made in a context that among other things provides a level of resolution. So, a treatment of integration of observations of this type must take account of multiresolution spatial data models. After an introduction, the paper discusses an ontology of imperfection, focusing on imprecision and inaccuracy. The paper goes on to consider logics that are appropriate for integration of information arising from imperfect observations. Two case studies, showing some of the facets of this treatment are developed in greater detail. The first case study considers integration of imperfect (inaccurate and imprecise) observations of a single spatial region. The second case study develops the theory of regions with broad boundary to address the issue of integrating imprecise observations of spatial relationships.

1 Introduction

This paper is motivated by the need to properly understand the issues involved when integrating data from multiple data sources. Each data source results from an observation of a real world phenomenon. Different data sources are associated with different observations of the same or different phenomena. Before data can be meaningfully retrieved and integrated from multiple data sources in response to a query, the contexts of the separate data sets need to be articulated and understood. There are many separate issues here, including handling various kinds of heterogeneity within the data sources. An impediment to effective integration arises when the data sources are semantically heterogeneous, lacking uniformity in 'meaning, interpretation or intended use' [SL90]. An important aspect of semantic heterogeneity is the difference in quality indicators for the data sets. Data quality has many dimensions and has been studied by many researchers (see, for example, [Goo93]). Aspects of quality that have received attention include accuracy, timeliness, relevance, completeness, consistency, and precision. It is probably not overstating the case to offer as a proposition that all observations of natural spatial phenomena have some associated level of imperfection in their quality. The challenge is to provide a framework for explicit representation of these classes of imperfection,

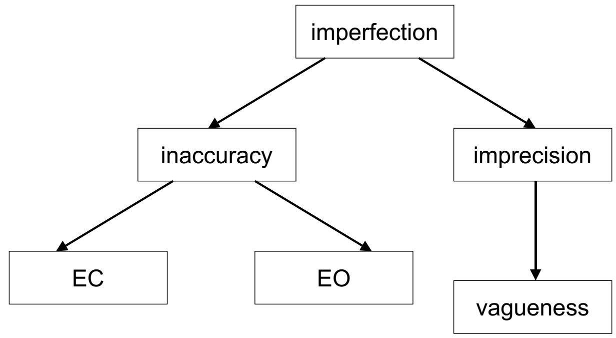

Figure 1: Ontology of imperfection

and to go further to provide a means of computing meaningfully with imperfect data, and to return results with known levels or intervals of imperfection characteristics. The use to which information is put is of course key, and the context of observation and use together with imperfection characteristics will determine the level of fitness for use in specific applications.

Classification of imperfection in databases has received some attention by researchers [Par96]. In this paper we study that form of imperfection that is classed as imprecision, and falls into the even more specialized class of vagueness. Integration of spatial information at a multiplicity of resolutions is a subject of research under the title of multiresolution spatial data models. This paper contributes and extends that effort by also allowing the possibility of error in the data. After considering the ontology of information imperfection in a general and more specific spatial sense, and briefly describing a framework for spatial multiresolution, we go on to look at logics for integrating imprecise and inaccurate spatial information. We conclude with two case studies on integrating information about spatial objects and about spatial relationships, to demonstrate the application of our theories in principle. Industry-scale applications to real world situations based on this framework are part of the European project REVIGIS, and will be reported in later publications.

2 Ontology of imperfection

Imperfection is an essential aspect of almost all observations of natural phenomena. This is partly because many observational and measurement devices have some inherent level of imprecision and inaccuracy, but also because many representations of the natural world are inherently vague. In the next few paragraphs, we expand upon these ideas and refine our notions of imperfection through the construction of a simple ontology of imperfection (see figure 1).

Researchers use the words ‘inaccuracy’ and ‘imprecision’ in different ways. In our ontology, there is a clear distinction between inaccuracy and imprecision, and in fact these concepts are orthogonal. An *inaccurate* proposition is one that lacks correlation with the actual state of affairs, whereas an *imprecise* proposition lacks specificity. Thus, we might propose that the time of writing this sentence is January 1st in the year 1900, at 1530 in the afternoon – fairly precise, but quite inaccurate. On the other hand, we might record the same time by merely observing that the sentence was written in the year 2000, actually accurate but imprecise.

Vagueness is that particular kind of imprecision where there are borderline cases for which it is difficult to decide whether they are covered by the concept or not. Many concepts

(Bertrand Russell [Rus23] has argued that all of language) contain vagueness. So, for example, 'tallness' is vague, because there will be people for whom is it difficult to decide whether they are tall or not. However, not all imprecise propositions are vague; For example, we could say that we are located in Europe, which is not a vague statement (assuming for the moment a crisp boundary to Europe), but is quite imprecise.

There are two types of errors that are distinguished in our ontology. The distinction here is similar to the distinction between Type I and Type II error in statistical significance testing. In Type I error, a true null hypothesis is incorrectly rejected, while in Type II error, a false null hypothesis fails to be rejected. A Type II error is only an error in the sense that an opportunity to reject the null hypothesis correctly is lost. It is not an error in the sense that an incorrect inference is made, as no conclusion is drawn when the null hypothesis is not rejected.

By analogy with Type I and II errors in statistical significance testing, we distinguish between an *error of commission* (EC) and an *error of omission* (EO). An error of commission (EC) is made when we conclude that a proposition is definitely the case when in fact it is either undetermined or definitely not the case. An error of omission (EO) is made when we do not declare a proposition to be the case when it is the case. For an example of EC, we may be using a remotely sensed image to determine land use at a location, and despite cloud cover, we wrongly assume the location to be a forest because some neighbouring pixels might suggest forest. For an example of EO, we might fail to infer that we were in Switzerland because we judged some scanty evidence based on mountain scenery to be insufficient.

In summary, an ontology has been constructed that contains the following components of imperfection in observations and the data resulting from them. Inaccuracy and imprecision are proper and disjoint subconcepts of imperfection, and vagueness is a proper subconcept of imprecision. Also distinguished are the types of inaccuracy referred to as errors of omission and errors of commission. An error of commission occurs when something is declared to be the case when in fact it is not the case. An error of omission occurs when something is not declared to be the case when in fact it is the case.

3 Imperfect spatial information

3.1 Semantics of spatial uncertainty

The general argument of the previous section is briefly illustrated using the assertion that a particular location is in a given region. This section discusses some of the interpretations that can be given to such an assertion, and leads on to a discussion on how two or more such assertions can be integrated. In the course of the discussion we use the concept *uncertainty* in a rather specialized way to indicate a cognitive state induced by imperfect observations and states of knowledge.

Case 1: Suppose that, following an observation of our location, we assert the proposition that we are in France. Then, assuming that we are not near enough the French boundary for there to be any intrinsic uncertainty about our position, it is clear enough that our assertion is true or false depending on whether or not our observation was accurate (and that we choose to tell the truth). In this case, at the scale in which the assertion takes its context, the region named France is crisp, and there is no uncertainty due to an observation of a vague phenomenon.

Case 2: Suppose that we are on a train travelling from Paris to Zurich, fall asleep and then wake up to see a mountain landscape from the window. We might use the evidence provided by the scenery to infer that we are in Switzerland. Uncertainty is greater here than the previous proposition, not due to any intrinsic vagueness in the location of the region named 'Switzerland' (at least not at this scale) but because of our imperfect observation of location. In this case, a truth value of 'maybe' arises because of the imperfection of the observation due to lack of evidence.

Case 3: Suppose that we are sitting in the Bodleian Library in Oxford. A tourist asks us whether we are in the south of England, and we answer that we are unsure. The reason for the uncertainty is not due to any lack of knowledge about our physical location in England (we might have GPS receiver that will tell us our position in terms of latitude and longitude with great accuracy and precision), but because of inherent vagueness in the concept 'the south of England'.

In summary, Case 1 shows the classical paradigm where complete observations are made of phenomena represented by crisp concepts. This case can be well formalized by a two-valued classical logic. For Case 2, the phenomena are still represented by crisp concepts, but this time the observations are not sufficient for us to be fully specific. Reasoning about imprecise observations of crisp phenomena will require at least a three-valued logic. In Case 3, the phenomena are represented by inherently vague concepts, and even though the observation is as complete as possible, we may be in a borderline case where it is ultimately not possible to be certain. This case will also require at least a three-valued logic, but not necessarily the same logic as for Case 2.

3.2 Regions with broad boundary

In the case when the spatial phenomenon under observation is something as simple as a region, then the 3-valued indeterminacy of location in the region has been represented by several authors as a region with broad boundary, also known as the egg-yolk diagram [CF96, CG96a, CG96b, Sch96, ES97, CF97]. Broad boundaries can be seen as a geometric model that approximate many different situations related to uncertainty. A representation of a spatial phenomenon as a region with broad boundary can have several semantic interpretations, paralleling the cases in the previous section. In what follows we will try to be careful in our use of language, reserving words such as feature, phenomenon, or real world object for the real world occurrence or manifestation, and representation and object as some conceptualization or description of the real world object or event. Some examples of how regions with broad boundary can arise now follow.

Example 1: incomplete representation of a feature. An example of this case is the situation sometimes arising in vector databases, where in a vector representation of a feature there might be missing sides of a polygon or entirely missing polygons, due to omissions in digitization or imperfect data conversion. The shaded area (broad boundary) in figure 2 can be introduced to represent the possible location of the missing line. This can be an example of Case 2 in the previous section, as it is an incomplete observation of an assumed crisp region.

Example 2: conflicting representations of a feature. This might arise in the case of cadastral data, or representations of political boundaries: in general, the case of existing data that provide approximations to some real objects. There might be both errors of commission and omission depending on which processes took place to obtain the representations. Different logics can be applied during the amalgamation of two resolutions. In the case of ignorance

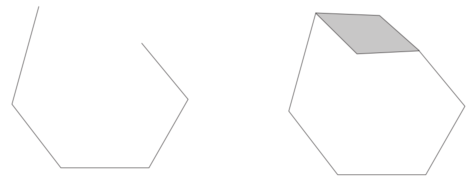

Figure 2: Correction of an incomplete boundary with a broad boundary

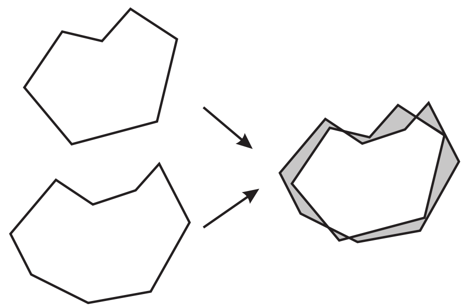

Figure 3: Inconsistent objects and the corresponding merged object.

of the origin of the data, the most pessimistic view would need to be adopted (both types of errors in the two resolutions). Figure 3 shows an example of this, where the shaded broad boundary results from the merging of the two objects on the left, and may be interpreted as consisting of locations that may or may not belong to the region.

Example 3: changing representations of a dynamic spatial phenomenon. Variability over time applies when we have different observations of the same geographic phenomenon taken at different times. Figure 4 represents a region that varies its shape and size over times t_1 and t_2. The core region in the integrated region on the right may be interpreted as being the collection of locations that are in the region at all (in this case, both) times. The broad boundary may be interpreted as the collection of locations that are in the region at some, but not all, times. The exterior of the outer integrated region represents the region where at no instant of time are the locations part of the time-varying region. Different logics will result, depending on our assumptions about errors in the observations or representations.

Example 4: imprecise observation of a spatial phenomenon. This is the case

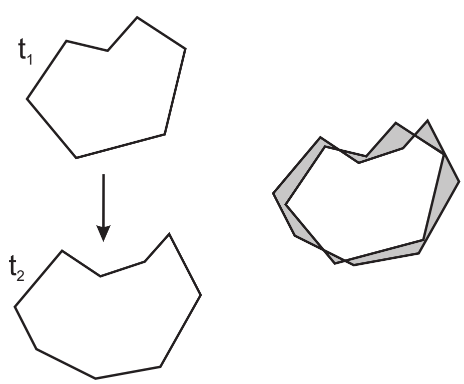

Figure 4: A region changing through time, and its integration

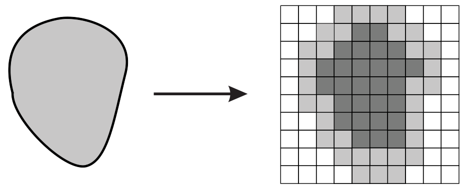

Figure 5: A rough region resulting from the observation of a crisp region under conditions of granularity

treated in [Wor98b] in the purely spatial case and in [Wor98a] in the case where there is both spatial and semantic imprecision. The region with broad boundary represents an imprecise observation of a region feature at a certain resolution. A resolution is intended here as a partition of the space in which locations indiscernible from the observation are grouped into the same elementary unit of the space. These elementary units impose a granularity of the underlying space. Figure 5 shows a region observed with respect to a granularity shown as a partition of the plane into squares. As is discussed in [Wor98b, Wor98a], such observations result in rough sets (see also [Paw82, Paw91, PS99] for detailed discussion about the concepts and applications associated with rough set theory. Representation of a multiplicity of such observations requires a multi-resolution data model.

Example 5: inherently vague representations of real world spatial objects. This case includes features (such as a mountain, a lake, etc.) whose boundaries are not sharp. Most geographical objects which are not man-made artifacts or conventions fall in this category. The problem is that, as the representation, linguistic or otherwise, of the object is a result of human definition and creation, it is likely to be inadequate. Such objects may be well represented by fuzzy sets, if a reasonable method of quantifying the membership function for the fuzzy set is available. A representation of such objects with regions with broad boundary is interpreted as an approximation in which the broad boundary of the region represents the part of the object where the membership function gradually decreases from 1 to 0. We can distinguish two special cases of inherent vagueness in real world objects.

Sub-case 5.1: inherently vague representations of real world spatial objects resulting from scale change. A special category of inherently vague representations of real world spatial objects are those that only exist at coarser resolutions (small scale), but are constructed from more finely grained real world objects (large scale). Such objects may be obtained by semantic generalization, aggregating other objects of a different nature that exist at a larger scale: this is the case of urban settlements that are seen as an aggregation of other objects (such as houses, streets, subways, bridges, and so on), or woods that are made up of trees. The broad boundary of such objects represents a peripheral zone where the density of the elements composing the aggregate has an intermediate value. This presents an issue for multi-resolution spatial databases, where different object classes may only exist at certain resolution levels.

Sub-case 5.2: inherently vague representations of real world spatial objects resulting from variation of context. These include objects such as ‘the south of England’ or ‘the outskirts of Rome’. The linguistic proposition defining such real world objects is usually made up of a geographic location plus a qualitative modifier. The representation results in an inherently vague spatial object because a sharp boundary does not exist but is replaced by a transition from the object to its surroundings. The characteristic of this sub-case is that the vagueness of the concept is mainly due to the qualitative modifier of the linguistic proposition. As a consequence, the boundary of a region as 'the south of England' is dependent on the context in which the linguistic proposition is placed and its intended use. It might be a different region if the proposition is placed in a historic, architectural, or biological context; finally, it might even depend on different opinions of individuals. Also, the qualitative modifier, like all qualitative spatial terms [Coh95], has a different meaning depending on the granularity with which it is defined: for example, in the two domains for qualitative modifiers {south, centre, north} and {south, north}, the proposition the 'south of England' would subsume a rather different underlying region. However, even if all context issues are resolved, so that the context is absolutely clear, it might still be the case that the boundary is not crisp.

4 Spatial imprecision, vagueness, and multi-resolution

4.1 Spatial vagueness

As has become clear from the preceding sections, there is a notion of vagueness, which may not be clarified by any amount of observation. This notion is to be distinguished from imprecision arising from incomplete or imperfect observation or representation of a crisp concept.

The notion of vagueness has been the subject of considerable discussion amongst philosophers [Wil94, KS96, Smi98]. Vague predicates admit borderline cases for which it is not clear whether the predicate is true or false. Spatial and temporal phenomena present us with many examples of vagueness, as the examples in the previous section illustrate.

When reasoning with vague predicates, we come up against the sorites paradox, where classical induction breaks down. To illustrate this point, consider Mount Everest. We can all agree that Everest is a mountain. We can also all probably agree that if we have a real world object that is by comment consent a mountain, then removal of a small stone from the summit will not change its status as a mountain. If we now start with Everest and begin removing stones one at a time, then by our assumptions and the law of classical induction we always have a mountain. But eventually we will have removed so much matter that we have razed Everest to the ground. So Everest is both a mountain (by induction) and not a mountain (by observation). This example provides a rationale to explore non-classical logics for handling spatial vagueness, as strict application of traditional logics lead to the kinds of problems just described.

A fundamental question regarding vagueness is where it lies. Is it a characteristic of the language in which the world is described, or could the world itself be vague? Could there be vague real world objects? Does the vague spatial object 'the South of England' actually exist as a spatial object in the world? The majority of the philosophical literature has dealt with linguistic vagueness. For example, Russell has stated "Vagueness and precision alike are characteristics which can only belong to a representation, of which language is an example" [Rus23].

Another approach to this issue, would be to concentrate on the *use* of a concept, rather than its semantics. Returning to case 3 in section 3.1, a common-sense procedure would be to start a Wittgensteinian language game [BH80]. What was the purpose of the tourist's question about the south of England? Is the tourist writing a postcard home, or maybe he is bored with his sight-seeing and seeks some diversion. For several possible uses it may be unnecessary to deliberate too much on the answer, since the question was only meant to keep a conversation running.

Before we leave the general treatment of vagueness we should note a fundamental difficulty in representing spatial vagueness by regions with broad boundaries. Is the boundary between the definite region and the broad boundary crisp? Is the region between the broad boundary and the exterior crisp? We see that regions with broad boundaries provide only a first-order approximation to what is a higher-order construction. Most of the work that follows will be concerned with first-order vagueness only.

4.2 Representing and reasoning with spatial vagueness

Treatments of spatial vagueness may be divided into quantitative and qualitative methods. An example of a quantitative approach is the theory of fuzzy sets and fuzzy regions. The theory of fuzzy sets was initiated by Zadeh [Zad65], and has been extensively developed and applied to spatial uncertainty in many works, with early examples including [Leu87, MC89]. The degree of membership of an element in a fuzzy set is represented by a real number in the closed interval $[0, 1]$. This generalizes the membership function for a classical set, which takes values from the set $\{0, 1\}$, indicating whether the element is or is not in the set. The same ideas carry over to fuzzy regions, where we are concerned about the membership of a location in a region. In many cases, fuzzy concepts are useful, but difficulties arise when it is unclear how to assign the degree of membership in a way that leads to useful results as we apply fuzzy analysis. For example, what degree of membership should we apply to the location 'Oxford' in the South of England?

Qualitative approaches to vagueness, while having a substantial philosophical literature, are less commonly applied. The *epistemic view* of vagueness is that it is a type of ignorance (e.g. [Wil94]). This seems to the authors, at least on face value, to be implausible in the case of concepts such as the south of England. The epistemic view would be that the boundary of the south of England is actually crisp: it is just that we do not know where it is. This kind of vagueness is considered here for example in Case 2 above.

Another approach is supervaluationism [Fin75]. The idea is that a vague predicate may admit many *precisifications*, each having a crisp boundary. Thus a precisification of the south of England might be all locations in England south of the line 52° latitude. In general, there will be many precisifications of a vague concept. Precisifications must meet certain constraints, such as not altering truth values of propositions. Thus, the proposition that Paris is not in the south of England is true also of the 52° latitude precisification above. Classical logic can be extended with modal operators of the form 'in all precisifications' and 'there exists a precisification'.

For vague spatial predicates, various theories of regions with broad boundaries have been developed [CF96, CG96a, CG96b, Sch96, ES97, CF97]. These were discussed earlier, and our case study below focuses on the approach of Clementini and Di Felice.

4.3 Multiresolution as a framework for integrating imprecise data

A key component of a useful theory of integration of imperfect spatial data is a multiresolution spatial data model that has been extended to take account of errors. Multiresolution spatial data models have been the subject of much research [PD95, RS95, SW98, Wor98a, Wor98b],

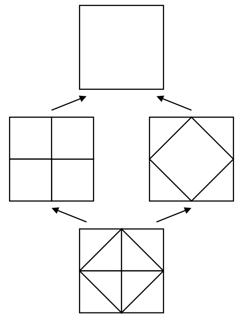

Figure 6: A simple granularity lattice showing the join and meet of two granularities

but no attempt has yet been made to integrate a treatment of error with that of imprecision/scale.

A multiresolution model that excludes the presence of errors is quite restrictive because it does not allow for changes quite common in practice such as a small object in a finer resolution being dropped in a coarser resolution, and all other changes that go under the name of semantic generalization. A generalized resolution is obtained by deliberately introducing imperfections in order to obtain a more abstract view of a given resolution. Typical processes include the creation of new features (seen as an aggregation of smaller units) and the elimination of other features that were not considered important.

Scale is connected to generalization, but they are two different concepts. Two different resolutions (a coarser one and a finer one) may refer to objects represented at the same scale. Conversely, objects at different scale may use the same resolution. Generally speaking, a coarser resolution is associated with a smaller scale. As was seen in sub-case 5.1 in section 3.2, objects existing at a small scale may sometimes be obtained by the aggregation of objects of a different nature that exist at a larger scale.

The work on multiresolution spatial data models in [SW98, Wor98a, Wor98b] is based upon the notions of resolution, granularity and indiscernibility. The main construction that we will use in the second case study below is that of a *granularity lattice*, and so we briefly revise this concept here. Further details of these concepts can be found in the above cited references. A *granularity* is a partition of space, with the idea that objects or locations belonging to the same block of the partition are indiscernible from each other at the given granularity. Suppose we fix the underlying space as S, then the set of granularities on S, $\mathcal{G}(\mathcal{S})$ may be structured as a lattice, where the underlying partial order is the usual partial order on equivalence relations on a set (see [DP90]). Figure 6 gives a very simple example of a granularity lattice. The partition at the bottom is the finest of all the granularities, and as we move up arrows of the lattice, partitions above are coarser than their descendants. The notion of a granularity lattice may be used as a framework to which the multiplicity of spatial datasets is referenced [SW98]. Thus, in figure 7, with each granularity g is associated a collection of spatial representations, the *map space* for g, where each representation is made at granularity g. In [SW98, Wor98a, Wor98b] is developed a theory of transformations of maps from one granularity to another, either at the object level, or at the level of the whole map. Some of these themes are discussed further in the case studies that follow.

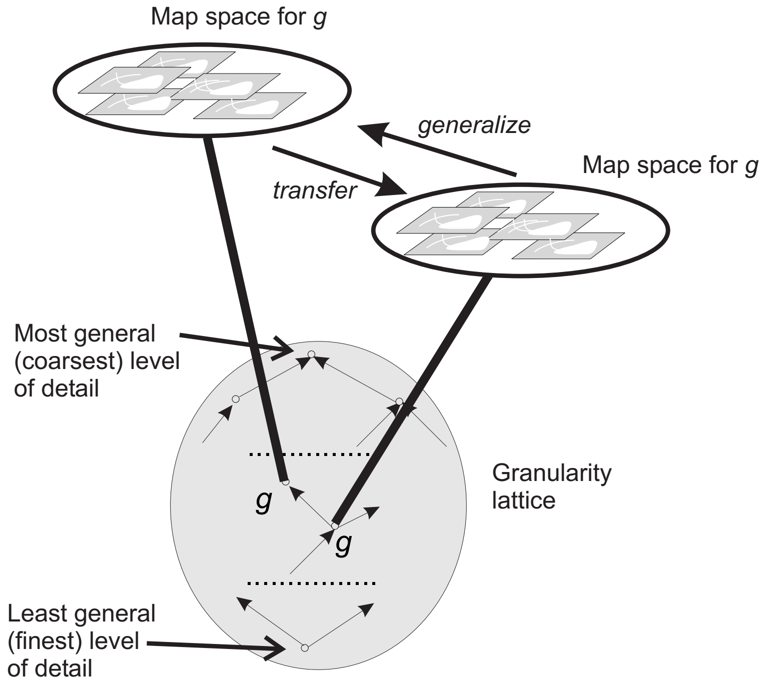

Figure 7: Granularity lattice as framework for map spaces

5 Case study 1: Integration of observations of regions

5.1 Accurate observations of regions

As has been seen, the context of an observation impacts on the approach to reasoning about it, and integrating inferences based upon it with other inferences based on observations made in different contexts. There are various possible logics that can formalise the integration of more than one observation about location in a region, as can be seen from the following example. Assume that we are in Oxford and ask two people nearby whether we are in the south of England. It is perfectly possible for one to answer yes and the other no, without either person being in error. This is because the intrinsic nature of the region allows room for more than one opinion (otherwise it would not be vague). On the other hand, if we are on that train between France and Switzerland and ask two fellow passengers whether we are in Switzerland, then assuming that they are not in error, they cannot answer yes and no at the same time (and place). So, assuming no errors, at least two different truth tables result from these cases. Table 1 gives a possible logic for the integration of observations of location within an intrinsically vague region, where * gives the combined truth table, and in this case we have made the fairly arbitrary decision to give greater weight to an uncertain response. Table 2 gives a possible logic for the integration of two imprecise (but not inaccurate) observations of location within a crisp region, where the X indicates an impossible case. In this case, '?' has a different semantics, indicating incomplete information to make a decision. Here it might seem more reasonable to give a greater weight to a definite response over an indefinite response. How to weight positive, negative and indefinite response depends on the semantics of the context as well as what is known about the accuracy of the observations.

A further consideration is what credence should be given to the resulting amalgamated

*	1	?	0
1	1	?	?
?	?	?	?
0	?	?	0

Table 1: Inherently vague region, no error in either source, weighted towards indefinite response

*	1	?	0
1	1	1	X
?	1	?	0
0	X	0	0

Table 2: Crisp region, incomplete observation, no error in either source, weighted to definite response

observation. In the cases in tables 1 and 2, as no errors were introduced, and assuming that no incompatibility was found, we can assume that the results contain no errors.

5.2 Inaccurate observations of regions

In this section are presented some three-valued logics that will handle integration of observations of crisp regions that are assumed neither to be precise nor accurate. As earlier, we make a distinction between errors of commission and errors of omission. Table 3 provides a catalogue of the logics under consideration. Let us suppose that an observation ω is made in order to determine whether a location is part of a region. The logic is three-valued, and so the possible outcomes are:

1: the location is definitely in the region

?: the location may or may not be in the region

0: the location is definitely not in the region

Let Ω be a set of observations of a location in a region. When we come to integrate two observations ω_1 and ω_1 in Ω, then various possibilities emerge depending on the contexts of the observations. The first two columns of table 3 give the truth values, 1, ? or 0, of the two observations. The remaining columns give the truth tables for integrations acting under different contexts. The contexts discussed here are:

***n*:** observation has no errors

***o*:** observation may contain errors of omission

***c*:** observation may contain errors of commission

***b*:** observation may contain both types of errors

ω_1	ω_2	1	2	3	4	5	6	7	8	9	10
		$n{*}n$	$n{*}o$	$n{*}c$	$n{*}b$	$o{*}o$	$o{*}c$	$o{*}b$	$c{*}c$	$c{*}b$	$b{*}b$
1	1	1	1	1	1	1	1	1	1	1	1
1	?	1	1	1	1	1	1	1	?	?	?
1	0	X	1	X	1	1	X	1	0	?	?
?	1	1	1	?	?	1	?	?	?	?	?
?	?	?	?	?	?	?	?	?	?	?	?
?	0	0	?	0	?	?	0	?	0	?	?
0	1	X	X	0	0	1	?	?	0	0	?
0	?	0	0	0	0	?	?	?	0	0	?
0	0	0	0	0	0	0	0	0	0	0	0

Table 3: Logics of integration in the presence of inaccuracy and imprecision

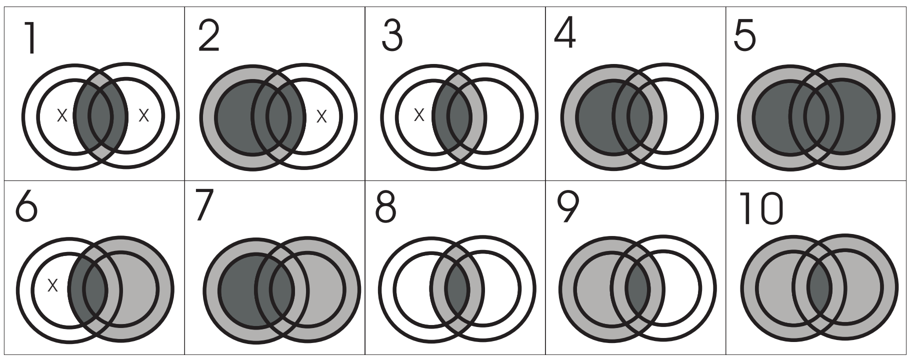

Figure 8: Visualizing the integrated regions for different contexts

Let $\mathcal{C} = \{n, o, c, b\}$. If $x, y \in \mathcal{C}$, then ${}_x{*}_y$ denotes the binary integration operator, where the first and second operands have contexts x and y, respectively. For completeness, the earlier logic (table 2) is given in Table 3 as logic 1. These logics of integration may also be visualized by reference to figure 8. For each context, the two imperfect observations of the regions are shown as the left and right regions with broad boundary. An area of the figure is shaded dark, medium, or light depending on whether the area is definitely, maybe, or definitely not part of the region, respectively. Symbol 'X' indicates that the labeled part of the area cannot exist in the configuration.

In the model of this section, each observation ω has a context $\gamma(\omega) \in \mathcal{C}$. We can unify the collection ${}_x *_y$ $(x, y \in \mathcal{C})$ of integration operators into a single context-dependent operator $*$ by the following formal construction.

$$\omega_1 * \omega_2 = \omega_1 \ {}_{\gamma(\omega_1)} *_{\gamma(\omega_2)} \ \omega_2$$

The integrated observation $\omega_1 * \omega_2$ will itself have a context $\gamma(\omega_1 * \omega_2)$. The contexts of integrated observations may be deduced from the three valued logics above, and are enumerated in table 4. The context of an integrated observation may sometimes allow a reduction in uncertainty. For example, the integration of observations with errors of commission and errors of omission can result in a context with no errors. However, the down-side to this is that more uncertainty in the form of truth-value '?' may be introduced into the body of the

$\gamma(\omega_1)$	$\gamma(\omega_2)$	$\gamma(\omega_1 * \omega_2)$
n	n	n
n	o	n
n	c	n
n	b	n
o	o	o
o	c	n
o	b	o
c	c	c
c	b	c
b	b	b

Table 4: Contexts of integrated observations

truth-table.

6 Case study 2: Integration of observations of topological relationships between spatial regions

6.1 Spatial relationships

The previous section discussed logics for integrating more than one imperfect observation of a single region. This section discusses integration of observations of topological relationships between regions. As with the previous case study, this would be an important issue when integrating observations in a multiresolution context, where the observations are being made at different granularities.

Work in [CF96, CF97] catalogues 44 realizable topological relationships between two simple regions with a broad boundary (see figure 9). Each relationship between regions A and B has an associated matrix, shown in equation 1. Here, $A^\circ, \Delta A, \cap A^-$ denote respectively the interior, broad boundary, and exterior of A. This is an extension of the 9-intersection model for topological relationships between crisp regions [EH94].

$$M = \begin{pmatrix} A^\circ \cap B^\circ & A^\circ \cap \Delta B & A^\circ \cap B^- \\ \Delta A \cap B^\circ & \Delta A \cap \Delta B & \Delta A \cap B^- \\ A^- \cap B^\circ & A^- \cap \Delta B & A^- \cap B^- \end{pmatrix} \quad (1)$$

In [CF97], spatial relations are organized as a graph having a node for each relation and an arc for each pair of matrices at minimum topological distance (figure 10). Such a distance is measured in terms of the number of different values in the corresponding matrices. This kind of graph has been called closest topological relation graph in [EAT92]. Another way of interpreting the graph is to consider the arc between two relations as a smooth transition that can transform one relation to the other and vice versa. This kind of graph under the latter interpretation has been called the conceptual neighborhood graph in [Fre92] and is useful for the analysis of deformations that can affect topological relations during motion or changes over time. (The same graph is called a continuity network in [CCR93]). A sequence of elementary deformations corresponds to a path in the graph. The graph is also useful for

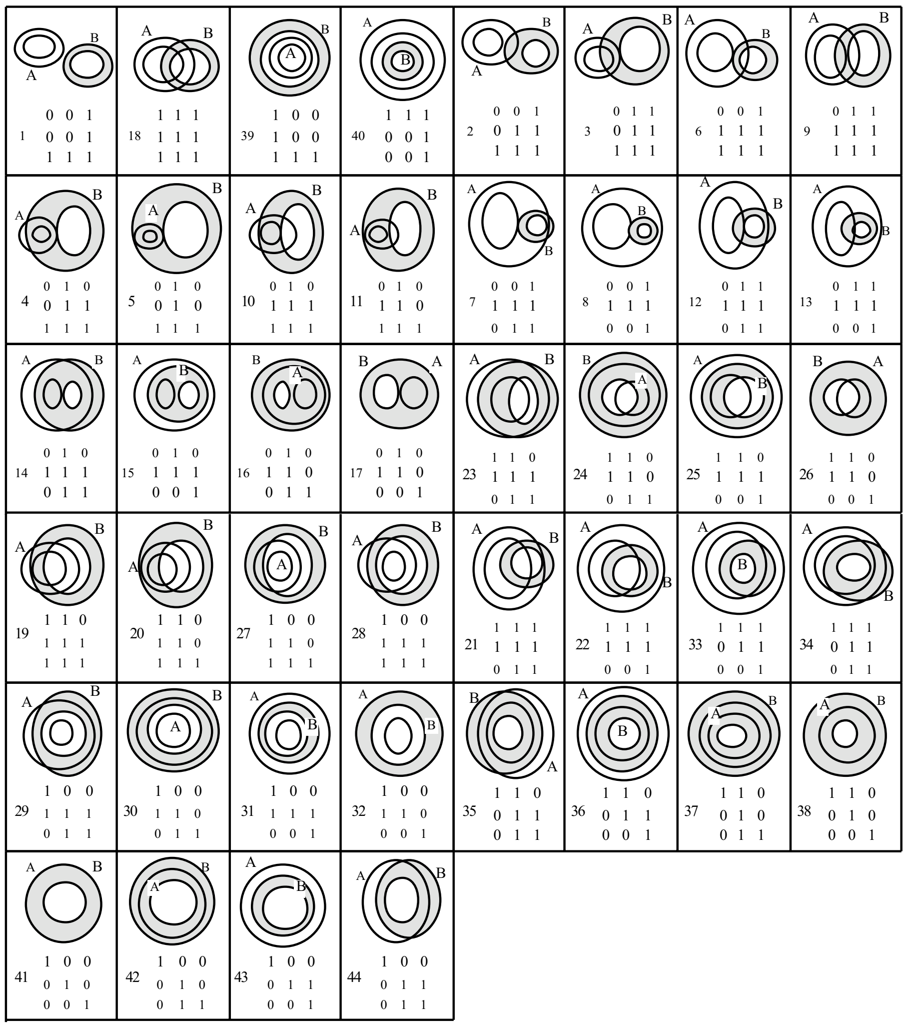

Figure 9: Topological relationships between regions with broad boundaries

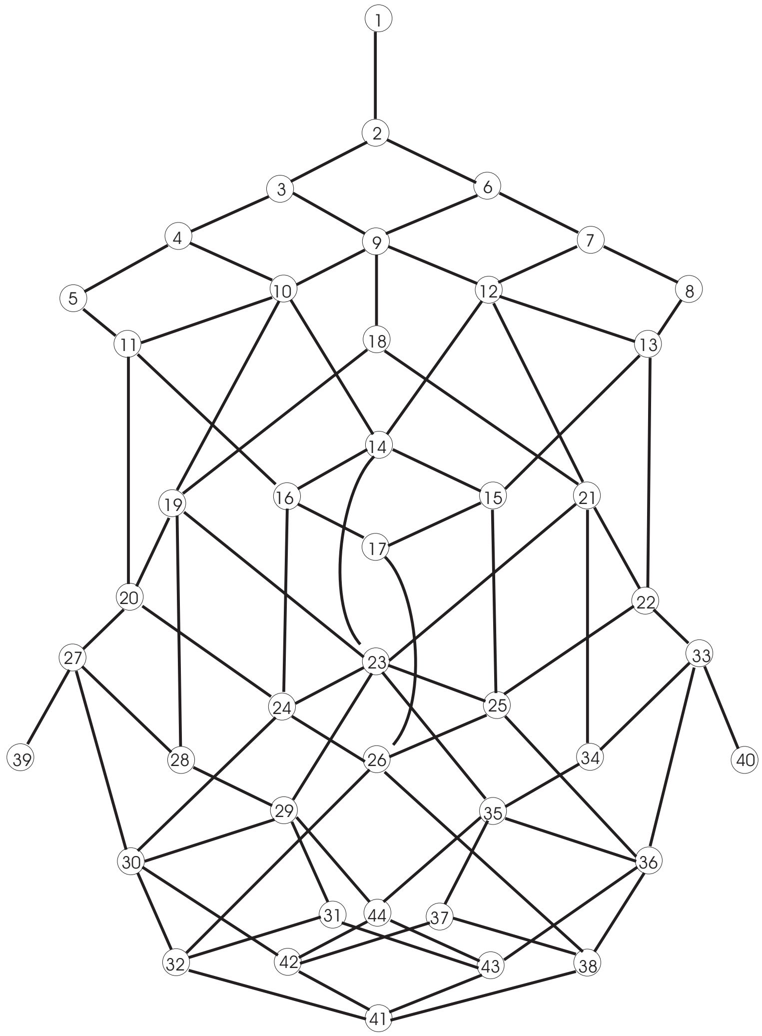

Figure 10: Minimum topological distance graph for regions with broad boundaries

spatial reasoning and for the process of grouping relations in clusters. The clustering process can be done with respect to a formal basis or with human subjects testing: it has been suggested that both approaches are in substantial agreement over the significant groupings [ME94].

A crisping of a region with broad boundary is defined as follows. Let A, B be regions with broad boundaries. B is a *crisping* of A iff $\Delta B \subseteq \Delta A$. The relationships between regions with broad boundary shown in figure 10 may now be structured in an alternative way, as in figure 11. An arrow is drawn from node n to node n' if and only if under some crisping of the regions in the relationship labeled by n we can arrive at the relationship labeled by n'. In figure 11, an edge with no arrow stands for an arrow going both ways. In this case, clusters of nodes emerge, and are shown in shaded blocks, and labeled by e, i, o, c, d to indicate the crisp spatial relationships equality, inclusion, overlapping, containment, and disjointness, respectively. Figure 12 shows a simplified version of figure 11, where the nodes that are in two-way association with each other by the crisping order are identified.

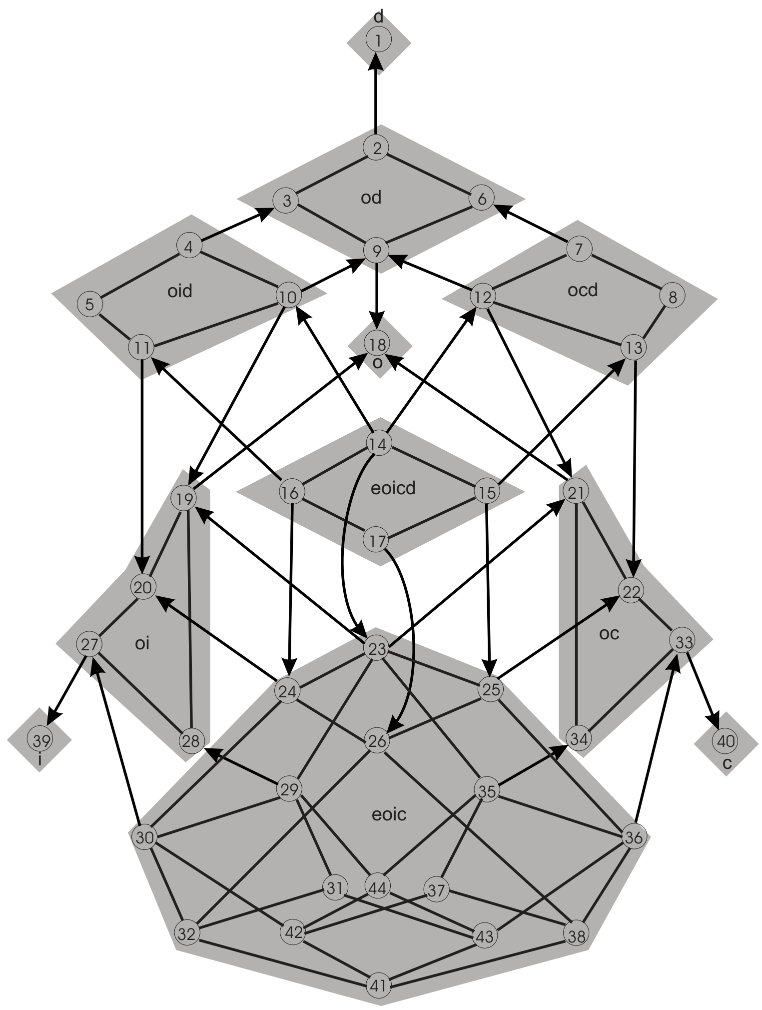

Figure 11: Crisping lattice associated with the topological relationships between regions with broad boundaries

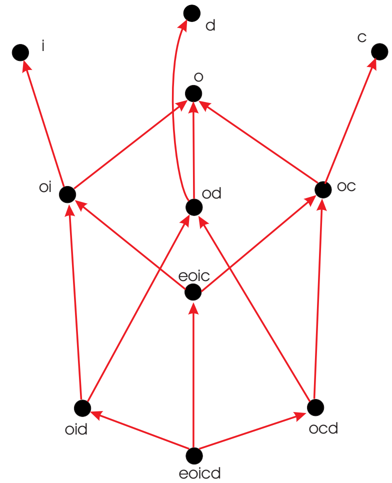

Figure 12: Blocked crisping lattice associated with the topological relationships between regions with broad boundaries

6.2 Integration of observations of spatial relationships

We consider the case where the regions with broad boundary do not model inherent vagueness, but an underlying crisp region exists. The uncertainty of broad boundary is due to imprecision of representation. Therefore, a topological relation between two regions with broad boundary (taken from the set of the 44 relations) could be made more precise, if a more accurate observation would be carried out, obtaining one of the relations between crisp regions.

The main objective of this case study is to show that in some cases a reduction of uncertainty in the topological relation can be obtained with the integration of two observations of a spatial relationship between two broad boundary regions. The method takes advantage of the graph structures recalled in previous subsection to find out which are the crisp topological relations that are compatible in the two observations after a crisping action on the regions.

Let us illustrate the method by an example. Suppose that we make two accurate but imprecise observations of the relationship between two crisp regions A and B. Suppose that our first observation tells us that the regions are related as regions with broad boundaries by relation 5, using the classification in figure 9, while the second tells us that the regions are related by relation 12. Reference to figure 11 shows that the first observation provides the information that the topological relationship between regions A and B is in the set $\{o, i, d\}$. A similar reference tells us that the topological relationship between regions A and B is in the set $\{o, c, d\}$. Therefore, we can infer from the integration of the observations that the topological relationship between regions A and B is in the set $\{o, i, d\} \cap \{o, c, d\} = \{o, d\}$, as shown in figure 12.

7 Conclusion

This paper has discussed some techniques for handling the integration of imperfect spatial observations. The work extended previous work in two principal ways.

1. The treatment of imperfection takes account of both inaccurate and imprecise observations.

2. The discussion of integration encompassed not only observations of spatial regions, but also relationships between regions.

The method and examples provided in the paper were simplified in several ways: the logic was never more than three valued, inaccuracy was classified in a rather primitive manner, and the relationships considered between regions were only binary and topologically not complex. This provides us with some possibilities for extending the work. In particular, we would like to develop techniques for integration of observations of spatial *scenes*, defined here to be configurations of locations and relationships between several spatial objects. The techniques developed in this paper provide the basis for some less simplified situations.

The paper provides a theoretical framework for integration of imperfect spatial information, and by means of the examples and case studies hints at applications. This framework is forming one of the bases for the European project REVIGIS, that is tackling the problem of data integration by database revision. The particular areas of application are land use classification (e.g. in the coastal region of the Netherlands, and cadastral determination in France and Austria.

8 Acknowledgements

Mike Worboys acknowledges the support of the UK Engineering and Physical Sciences Research Council Grant for the project 'Vagueness, uncertainty and granularity in spatial information'. He also acknowledges the support of the Italian MURST project 'Metodologie e tecnologie per la gestione di dati e processi su reti Internet e Intranet' for a visit to the University of L'Aquila during the early stages of this work. Both authors thank Paolino di Felice for useful discussions on this topic. The authors are grateful for the constructive suggestions of the referees and in particular for one referee pointing out a connection between our treatment of vagueness and Wittgenstein's philosophy of language.

References

[BH80] G.P. Baker and P.M.S. Hacker. *Analytical Commentary on the Philosophical Investigations. Volume 1: Wittgenstein: Understanding and Meaning.* Oxford: Oxford University Press, 1980.

[CCR93] Z. Cui, A.G. Cohn, and D.A. Randell. Qualitative and topological relationships in spatial databases. In D.J. Abel and B.C. Ooi, editors, *Advances in Spatial Databases - Third International Symposium, SSD '93, Singapore*, volume 692 of *Lecture Notes in Computer Science*, pages 296–315. Berlin: Springer-Verlag, 1993.

[CF96] E. Clementini and P. Di Felice. An algebraic model for spatial objects with indeterminate boundaries. In P.A. Burrough and A.U. Frank, editors, *Geographic Objects with Indeterminate Boundaries*, GISDATA Series, pages 155–169. London: Taylor & Francis, 1996.

[CF97] E. Clementini and P. Di Felice. Approximate topological relations. *International Journal of Approximate Reasoning*, 16:173–204, 1997.

[CG96a] A.G. Cohn and N.M. Gotts. The 'egg-yolk' representation of regions with indeterminate boundaries. In P. Burrough and A. Frank, editors, *Geographic Objects with Indeterminate Boundaries*, pages 171–187. London: Taylor & Francis, 1996.

[CG96b] A.G. Cohn and N.M. Gotts. Representing spatial vagueness: A mereological approach. In *Principles of Knowledge Representation and Reasoning: Proc 5th International Conference (KR96)*. San Francisco: Morgan Kaufmann, 1996.

[Coh95] A G Cohn. The challenge of qualitative spatial reasoning. *Computing Surveys*, 27(3):323–327, 1995.

[DP90] B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order*. Cambridge: Cambridge University Press, 1990.

[EAT92] M.J. Egenhofer and K. Al-Taha. Reasoning about gradual changes of topological relationships. In A.U. Frank, I. Campari, and U. Formentini, editors, *Theories and Models of Spatio-Temporal Reasoning in Geographic Space*, volume 639 of *Lecture Notes in Computer Science*, pages 196–219. Berlin: Springer-Verlag, 1992.

[EH94] M.J. Egenhofer and J. Herring. Categorising topological spatial relations between point, line, and area objects. In D. Mark and J. Herring, editors, *The 9-Intersection: Formalism and its use in natural-language spatial predicates*. Technical Report 94-1. Santa Barbara, NCGIA, 1994.

[ES97] M. Erwig and M. Schneider. Vague regions. In M. Scholl and A. Voisard, editors, *Advances in Spatial Databases - Fifth International Symposium, SSD'97*, volume 1262 of *Lecture Notes in Computer Science*, pages 298–320. Berlin: Springer Verlag, 1997.

[Fin75] Kit Fine. Vagueness, truth and logic. *Synthese*, 30:265–300, 1975.

[Fre92] C. Freksa. Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54:199–227, 1992.

[Goo93] M.F. Goodchild. Data models and data quality: problems and prospects. In B.O. Parks M.F. Goodchild and L.T. Steyaert, editors, *Visualization in Geographical Information Systems*, pages 141–149. New York: John Wiley, 1993.

[KS96] Rosanna Keefe and Peter Smith, editors. *Vagueness: A Reader*. Cambridge MA: MIT Press, 1996.

[Leu87] Y. Leung. On the imprecision of boundaries. *Geographical Analysis*, 19:125–151, 1987.

[MC89] D.M. Mark and F. Csillag. The nature of boundaries on 'area-class' maps. *Cartographica*, 26:65–77, 1989.

[ME94] D.M. Mark and M.J. Egenhofer. Modeling spatial relations between lines and regions: combining formal mathematical models and human subjects testing. *Artificial Intelligence*, 21:195–212, 1994.

[Par96] S. Parsons. Current approaches to handling imperfect information in data and knowledge bases. *IEEE Transactions on Knowledge and Data Engineering*, 8:353–372, 1996.

[Paw82] Z. Pawlak. Imprecision in finite resolution spatial data. *Int. Journal of Inf. and Comp. Sci*, 11:341–356, 1982.

[Paw91] Z. Pawlak. *Rough Sets - Theoretical Aspects of Resoning about Data.* Netherlands: Kluwer, 1991.

[PD95] E. Puppo and G. Dettori. Towards a formal model for multiresolution spatial maps. In *Proceedings Fourth International Symposium, SSD'95*, volume 951 of *Lecture Notes in Computer Science*, pages 152–169. Berlin: Springer Verlag, 1995.

[PS99] S.K. Pal and A. Skowron, editors. *Rough Fuzzy Hybridization: A New Trend in Decision-Making.* Berlin: Springer Verlag, 1999.

[RS95] P. Rigaux and M. Scholl. Multi-scale partitions: Applications to spatial and statistical databases. In *Proceedings Fourth International Symposium, SSD'95*, volume 951 of *Lecture Notes in Computer Science*, pages 170–183. Berlin: Springer Verlag, 1995.

[Rus23] Bertrand Russell. Vagueness. *Australasian Journal of Philosophy and Psychology*, 1:84–92, 1923.

[Sch96] M. Schneider. Modelling spatial objects with undetermined boundaries using the realm/rose approach. In P. Burrough and A. Frank, editors, *Geographic Objects with Indeterminate Boundaries*, pages 141–152. London: Taylor & Francis, 1996.

[SL90] A.P. Sheth and J. Larsen. Federated database systems for managing distributed, heterogeneous and autonomous databases. *Association for Computing Machinery Computing Surveys*, 22:183–236, 1990.

[Smi98] Barry Smith, editor. *The Monist*, volume 81. The Hegeler Institute, April 1998.

[SW98] J.G. Stell and M.F. Worboys. Stratified map spaces. In T. Poiker and N. Chrisman, editors, *Proceedings of the 8th International Symposium on Spatial Data Handling, British Columbia*, pages 180–189. London: Taylor & Francis, 1998.

[Wil94] T. Williamson. *Vagueness.* London: Routledge, 1994.

[Wor98a] M.F. Worboys. Computation with imprecise geospatial data. *Journal of Computers, Environment and Urban Systems*, 22:85–106, 1998.

[Wor98b] M.F. Worboys. Imprecision in finite resolution spatial data. *Geoinformatica*, 2:257–280, 1998.

[Zad65] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.