Application of Supervaluation Semantics to Vaguely Defined Spatial Concepts

Brandon Bennett*

School of Computing, University of Leeds, Leeds LS2 9JT, UK
`brandon@comp.leeds.ac.uk`

Abstract. The paper examines ways in which the interpretation of spatial concepts is affected by vagueness and suggests mechanisms for taking account of this within spatial information systems. The theory of supervaluation semantics is explained and applied to the spatial domain and to particular problems of defining geographical concepts such as 'forest'.

Keywords: vagueness, supervaluation semantics, concept definitions, logic, spatial information systems

1 Introduction

Problems of indeterminacy, inaccuracy and imprecision of spatial information have been recognised as important aspects of *data quality* and much effort has been spent devising ways to handle such imperfections (Goodchild and Gopal 1989, Goodchild 1993, Heuvelink 1998, Burrough and Frank 1996). At the same time, high-resolution satellite images and sophisticated image interpretation software are yielding increasingly accurate geographical data. However, if this information is to be useful for high-level decision making about the environment, this detailed empirical information needs to be related to the vague natural concepts that we used to think and talk about the world.

Geographers, and more especially surveyors and cartographers, have long been aware of the difficulties of giving precise definitions of spatial features (see for example (Maling 1989, chapters 5 and 12)); but, although the phenomenon of linguistic vagueness has been studied by a number of philosophers and logicians, applications of theories of vagueness to practical problems are largely undeveloped. In the fields of AI and GIS, vagueness has often been seen as more or less the same as uncertainty and accounts such as fuzzy logic (Zadeh 1975, Wang and Brent Hall 1996) and rough sets (Orłowska 1997) are often supposed to encompass both phenomena. In the current paper I distinguish sharply between epistemic uncertainty of data and vagueness of the concepts in terms of which the data is expressed. Even when we have complete certainty about measurable information, we must still solve the problem of relating this data to the vague concepts of natural language.

---

* This work was supported by the EPSRC under grant GR/M56807.

The concerns of this paper overlap considerably with issues raised by those (e.g. (Frank 1997)) who have recognised the need for well defined *ontologies* for spatial information in order to avoid inconsistencies arising from the combination of incompatible representations. Whereas ontological formalisation replaces vague concepts with precise definitions, in treating vagueness itself, I am concerned with how to relate naturally vague concepts to possible precise interpretations. Problems of vagueness are also directly relevant to achieving *interoperability* (Vckovski 1997), where we want to combine data sources that may be based on different, perhaps incompatible, conceptualisations of spatial data.

I propose that, in order to achieve their full potential, the architecture of spatial information systems needs to include some kind of *Vagueness Reasoning Module* (VRM). This will be a layer of the query processing system which will provide a bridge between the natural concepts occurring in queries and the precise but unnatural quantitative database.

The structure of the paper is as follows. In the next section I examine the nature of vagueness and distinguish it from other phenomena that affect the interpretation of non-idealised information. Section 3 explains how vagueness can be modelled by *supervaluation semantics* and the sort of representations and implementations that can articulate the theory. In Section 4 I look at how vagueness relates to spatial extension and in Section 5 I apply the theory to definitions of geographical concepts. We then reach the conclusion.

2 Vagueness and Related Phenomena

I define vagueness as *a lack of clearly defined critera for the applicability of a concept*. Thus, it is a property of language not of the world itself. Typical examples of vague propositions are: 'All *mountains* are *very high*'; and 'That frog is *green*'. The words given in italics are the principal sources of vagueness. 'Mountains' is a vague classifier because there is no precise definition of the natural concept of a mountain. 'very high' is a vague adjectival phrase: it does tell us something about the actual measurable hight but does not fix any definite height range. The adjective 'green' is vague in (at least) two ways: firstly the exact range of colours that count as green is not precisely defined; and secondly, the concept of 'being green' is vague in that it is not clear exactly how much of the surface of an entity must be 'green' (the frog will almost certainly have some non-green parts).

I distinguish sharply between vagueness and *uncertainty*, which I regard as a distinct (though interacting) phenomenon. I take 'uncertainty' to mean *lack of exact knowledge about an object or situation*. So uncertainty is an *epistemic* state not a feature of language.[1] Although modelling of uncertainty is extremely important in the processing and interpretation of spatial information, it will not

---

[1] There is a view that vagueness is always epistemic, in that it arises from a lack of knowledge regarding the applicability of language (see e.g. (Williamson 1994)). But this is still compatible with a sharp distinction being made between uncertainty about language (i.e. vagueness) and uncertainty about the state of the world.

be considered in the current paper. I shall assume we are dealing with idealised data which is completely certain and accurate.

Vagueness can often lead to uncertainty in that where a concept such as 'forest', 'desert' or 'swamp' is vague we will in many cases be uncertain how to demarcate the spatial extension of the concept. On the other hand, if we are not completely certain of the exact details of some information we want to report, we may employ vagueness as a means of increasing the certainty of what we say, while at the same time conveying a sense of imprecision. For example, a statement such as 'The chair is *in the corner of* the room' is vague but can often be said with certainty, whereas an exact specification of the location of a chair (or even a range of possible locations) will typically be uncertain. This example also illustrates the fact that vagueness is not merely a defect of language; it also often facilitates communication without the cumbersome language required to achieve precision.

It is useful to distinguish between two kinds of vagueness, which I shall call *conceptual vagueness* and *sorites vagueness.* Conceptual vagueness occurs where there is no single completely adequate definition of conceptual term. Certain requirements may be clearly identifiable, whereas for other conditions it is arguable whether or not they are necessary. Certain combinations of these conditions may capture typical senses of the term but none is representative of all possible senses. Thus if we take the intersection of plausible definitions we get a concept that is much to strict (perhaps even unsatisfiable), whereas if we take their disjunction we get a concept that is overly general.

Conceptual vagueness is closely related to *ambiguity.* If a word is ambiguous, it has two or more distinct senses that are clearly distinguishable. However, a conceptually vague term corresponds to a complex cluster of many overlapping senses, such that we cannot say exactly what senses make up the cluster. Moreover, one can meaningfully use a conceptually vague term without being committed to any one of its possible precise interpretations. What we can say about these clusters will be made clear below when I introduce the theory of supervaluation semantics.

Sorites vagueness is the kind of indeterminacy that affects the thresholds at which we assert properties such as 'tall' or 'heavy'. Such predicates classify entities with respect to some relevant measurable quantity, without being committed to any specific boundary value.

In a pure case of sorites vagueness it is uncontroversial which factors are relevant to the ascription or how those factors should be measured, it is only the threshold that is at issue. However, many natural terms are affected by both types of vagueness. For example to precisely interpret the concept 'tall man' we have first to decide how we are to measure the height of a man: must he remove his shoes and hat? what about hair and prosthetic limbs? what about posture? Once we have resolved these conceptual issues we then still have to deal with sorites vagueness in setting the threshold for tallness.

Both types of vagueness also interact strongly with contextual phenomena of various kinds. Many concepts exhibit some form of *contextual variability.* For example ‘tall’ in the context of ‘tall man’ has a different interpretation from in the context of ‘tall child’. In cases such as this we see that a sorites concept may be affected by its context so that location of its albeit vague threshold is shifted. The range of possible interpretations for a conceptually vague concept can also be affected, not so much by their immediate syntactic context but by their more general context within an information source or exchange. Because it is largely independent of vagueness, issues of contextual vagueness will not be addressed in the current work. I shall assume that context can be eliminated or ignored.[2] Despite the fact that in many cases vagueness seems to be separable from context there may be cases where this distinction is blurred. A close connection between the phenomena is born out by the fact that formalisations of the logic of context (see e.g. (McCarthy 1993)) have much in common with the supervaluation approach to vagueness.

3 Supervaluation Semantics

On the *supervaluation* account of vagueness (Fine 1975) a vague language is one which can be made precise in many different and sometimes incompatible ways. A way of making a language precise is called a *precisification*. Each precisification p is identified with a precise interpretation, I_p, of the vocabulary of the language. In the simplest case this would be a classical propositional or 1st-order model. A supervaluation model then consists simply of a set of precisifications. Given a supervaluation model $\mathcal{V}$, a proposition which is true under every interpretation $I_p \in \mathcal{V}$ is called *super-true* or — in my own terminology — *unequivocally* true.

Supervaluation semantics by itself does not add anything interesting to logic at the object level. It is easy to see that those formulae that are unequivocally true in every model are just the classically valid formulae. However, the semantics does provide a framework within which we can define operators that articulate certain aspects of the logic of vagueness.

3.1 Modal Supervaluation Logic

One possibility is to take a modal approach and represent vagueness in terms of propositional operators (Bennett 1998). $\mathbf{U}\phi$ means that ϕ is *unequivocally* true — i.e. true for all precisifications; $\mathbf{S}\phi$ can be read ‘ϕ is *in some sense* true’ — i.e. true for some precisification. $\mathbf{S}$ is the dual of $\mathbf{U}$ and thus can be defined by $\mathbf{S}\phi \leftrightarrow \neg\,\mathbf{U}\,\neg\phi$. Logically $\mathbf{U}$ behaves as the $\Box$ operator of the modal logic $S5$ and $\mathbf{S}$ as its dual, $\Diamond$.

We can now qualify assertions according to whether they hold in some or all precisifications. For example $\mathbf{S}[\mathsf{Wood}(\textit{‘Woodsley Clough’})]$ or $\mathbf{U}[\mathsf{Wooded}(\textit{parcel1})]$.

---

[2] For instance we might suppose that some transformation can be carried out that replaces contextually variable concepts with non-contextual concepts and explicit constraints; or, we could just consider composite concepts such as ‘tall man’ and ‘tall child’ as if they were syntactically atomic.

We can also use these operators to specify dependencies between the meanings of vague concepts. Thus $\forall x[\mathsf{Copse}(x) \rightarrow \mathbf{S}\,\mathsf{Wood}(x)]$ means anything which is a copse (i.e. a small group of trees) is in some sense a wood. Similarly, $\forall x[\mathsf{Wood}(x) \rightarrow \mathbf{S}\,\mathsf{Forest}(x)]$ captures the intuition that any wood is arguably a (small) forest. If a copse is in some sense a wood and a wood is in some sense a forest, this does not mean that a copse is in some sense a forest; and indeed according to supervaluation semantics the formula $\mathbf{U}\,\neg\exists x[\mathsf{Forest}(x) \wedge \mathsf{Copse}(x)]$ is consistent with the previous two formulae. This illustrates the ability of the theory to model the blurring of concepts, while still maintaining certain strong constraints.

In the context of a computer database we will often be dealing with concepts that are precise sharpened version of natural terms. The supervaluation operators enable us to relate these artificial concepts to their vague natural language counterparts. For instance, the following formula asserts that $\mathsf{Forest1}$ is a more precise version of the concept Forest:

$$\forall x[\mathsf{Forest1}(x) \rightarrow (\mathbf{S}\,\mathsf{Forest}(x))] \wedge \forall x[(\mathbf{U}\ \mathsf{Forest}(x)) \rightarrow \mathsf{Forest1}(x)]$$

By specifying such axioms, ‘soft’ constraints are placed on the meanings of natural concepts. Classifications in terms of artificial concepts can be combined with information containing natural concepts.

Supervaluation semantics allows one to specify a number of different entailment relations of varying strength. Bennett (1998) defines the following (together with three weaker forms of ‘reliable’ entailment):

$$\begin{array}{llll}
\phi_1, \ldots, \phi_n \models_{\text{arguable}} \psi & \text{iff} & \models_{S5\mathbf{U}} & \mathbf{S}((\phi_1 \wedge \ldots \wedge \phi_n) \rightarrow \psi) \\
\phi_1, \ldots, \phi_n \models_{\text{global}} \psi & \text{iff} & \models_{S5\mathbf{U}} & (\mathbf{U}\phi_1 \wedge \ldots \wedge \mathbf{U}\phi_n) \rightarrow \mathbf{U}\psi \\
\phi_1, \ldots, \phi_n \models_{\text{local}} \psi & \text{iff} & \models_{S5\mathbf{U}} & \mathbf{U}((\phi_1 \wedge \ldots \wedge \phi_n) \rightarrow \psi) \\
\phi_1, \ldots, \phi_n \models_{\text{reliable}} \psi & \text{iff} & \models_{S5\mathbf{U}} & (\mathbf{S}\phi_1 \wedge \ldots \wedge \mathbf{S}\phi_n) \rightarrow \mathbf{U}\psi
\end{array}$$

The weakest entailment is ‘arguable’ which holds if there is any sense of the concepts in the formulae under which the implication corresponding to the entailment holds. This gives us entailments that hold under very flexible (perhaps even inconsistent) interpretations of the concepts involved. The strongest is ‘reliable’ entailment, which holds if: whatever senses the premisses are interpreted under, the conclusion holds in every sense. This can be used to derive entailments which must hold despite the presence of vagueness. For instance $\mathbf{S}\,\mathsf{Desert}(x) \rightarrow \mathbf{U}\,\neg\mathsf{Marsh}(x)$ might hold even where Desert and Marsh are very vague predicates. This ability to derive secure consequences involving vague concepts is perhaps the main advantage of supervaluation semantics over fuzzy logic, where fuzzy concepts cannot support completely reliable inferences.

3.2 Reified Precisifications

Although modal operators allow many logical properties of vague concepts to be expressed, they do not provide any way of referring directly to individual precisifications. However, in the environment of a GIS we will often want to record information about which interpretations can be given to information in particular datasets. To do this we need a language in which names of precisifications can be related explicitly to spatial formulae.

In a general *reified* supervaluation semantics we could associate arbitrary propositions with precisification variables and constants. Thus, $\mathsf{InPrec}(p, \phi)$ would assert that ϕ is true according to precisification p. At the expense of some elegance we can achieve the same expressive power by simply supplementing each predicate and function of an ordinary 1st-order language by an additional argument place. For instance a predicate $\mathsf{Swamp}(x)$ would be replaced by a relation $\mathsf{Swamp}(p, x)$ saying that, in precisification p, x is a swamp. If we use this approach we need not worry about axiomatising the logical predicate InPrec. Logical relationships between vague concepts that hold whatever reasonable way they are interpreted can now be represented by quantifying over the (possibly infinite) space of precisifications.

Since a precisification fixes the meanings of all the vague vocabulary of a language, a classification which makes precise only part of the vocabulary may be common to a class of precisifications. In a formalism with reified precisifications, we can model this by introducing predicates of precisifications. For example, $\mathsf{UNESCOF}(p) \leftrightarrow \Phi(p)$ might mean that the predicate $\mathsf{UNESCOF}$ applies to those precisifications satisfying some precise formal specification Φ of the UNESCO forestation classification given below in Table 1.

The use of a language with reified precisifications is also motivated by the analysis of vague nominal expressions given in the next section, which seems to be difficult to express within a modal framework.

3.3 Vague Nominal Expressions

The established theory of supervaluation semantics models a situation where we wish to reason with vague predicates, which are applied to a perfectly definite domain of objects. The question of whether the objects referred to by a language can themselves be vague has also received attention from philosophers (Evans 1978, Hughes 1986, McGee 1997). There is no consensus on this issue but the more popular view seems to be that vagueness is a feature of language and not of the objects it describes. But even if this is so, one could still argue that the nominal expressions of a language may be vague in that they may not unequivocally refer to a unique entity, but rather may be understood in different senses as applying to different physical entities.

The current paper concentrates primarily on vague concepts rather than nominals. However, I suggest that the vagueness nominals can be explained as derived from the concepts that they exemplify. Thus, an expression such as 'Sherwood Forest' refers to a particular instance of the vague sortal predicate 'forest' and the range of possible extensions that might be assigned to Sherwood Forest are determined by the range of possible precise senses that can be given to the concept 'forest'.

3.4 Implementing Supervaluation Semantics

The reified precisification approach to vagueness lends itself well to implementations in established logic programming languages such as Prolog. Alternatively one could use some *description logic* system (Calvanese, Lenzerini and Nardi 1998). In either case one could explicitly add precisification variables to the definitions of vague concepts.

For most applications one would probably want to hide the apparatus of precisifications from the user and employ a more intuitive way of showing the results of vague queries. For instance, this might be achieved by expanding a query such as '`?- show(forest).`' to a form something like

```
?- setof(R, (forest(P,F), Ext(P,F,R)), PossExts),
   illustrate(PossExts).
```

`P` is a precisification variable, which parameterises both the sense in which 'forest' is interpreted and the sense in which its extension is determined by the `Ext` relation. '`illustrate(PossExts)`' produces informative graphical output about the distribution of possible extensions under different interpretations.

This example glosses over certain difficulties. In particular it assumes that a finite number of senses of forest are defined, whereas the space of possible senses might be better described in terms of certain continuous parameters (such as average tree height). To handle this a much more sophisticated procedure would be needed which might display the extensions of a range of possible interpretations taken at sample points within the space of precisifications.

4 Spatial Vagueness

Supervaluation semantics is a very general approach to vagueness but it can only be useful for reasoning about a specific domain if the peculiar logic of that domain is adequately modelled. With the exception of (Kulik 2000) I am not aware of any other work applying supervaluation semantics to spatial concepts.

4.1 Spatial Concepts and Extensions

The spatial properties that are easiest to understand semantically are those that can be defined in terms of properties of points — i.e. their extension consists of all points satisfying some given condition. Examples of such concepts are 'the region of the Earth that is more than 1000m above sea-level. However, in general, a 'region property' will be associated with a property of the whole set of points, which cannot be explicitly reduced to properties of individual points. For example a 'lake' is not simply made up of the set of points which are covered by water, it is rather a particular *maximal connected* set of water covered points.

Although maximal connectedness is one of the most important factors in the individuation of geographical features, only very basic types of feature can be regarded simply as maximal connected sets of points exhibiting a given property.

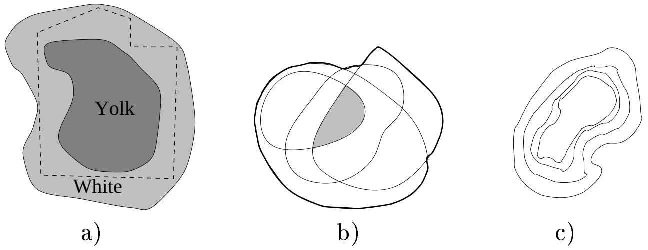

Fig. 1. Models of Vagueness and Extension

Typically, whether a set of spatial points can be taken as the extension of a feature of a given type, is dependent on much more complex constraints on the structure of this set (consider e.g. how we differentiate lakes from other hydrological features or how we might characterise a 'building'). Some of these constraints may not even relate to physical properties (e.g. a 'listed building').

I say that a region property is: *integral* iff when it is true of a region it is not true of any proper parts of that region; and, *divisible* iff whenever it holds of some region it holds of all parts of that region. According to this analysis, concepts describing geographical features, such as 'lake', 'river', 'forest' are integral; whereas land-type concepts such as 'wooded region', 'marshy region' are divisible. Many properties are neither integral nor divisible — e.g. 'r encompasses a lake'. Whether properties are divisible is important in determining whether a conjunction of several spatial concepts applies to the intersection of their extensions of different spatial concepts (or of the same concept under different precisifications)

A further issue that complicates the identification of regions with sets of points is the status of boundary points. For many spatial concepts it is not clear whether boundary points should be counted as included in the region to which they refer. This may be regarded as an example of conceptual vagueness. However, it is also a general ontological issue which applies to spatial concepts that are not in other ways vague.

4.2 Extensions of Vague Concepts

From the perspective of spatial information, the most important property of any object is its spatial extension. Hence, it is vital to model the way in which vagueness affects the attribution of extensions.

The 'Egg-Yolk' theory (Lehmann and Cohn 1994, Cohn and Gotts 1996a, Cohn and Gotts 1996b) directly models the notion of a vague region in terms of its maximal and minimal possible extensions. The maximal extension is called the 'egg' and the minimal is the 'yolk', which is required to be a part of the egg (see Figure 1a). (The case where the yolk is equal to the egg is allowed, such cases corresponding to 'crisp' regions.) This analysis is simple and supports an account of some significant inferences involving relationship between vague regions. However, it cannot handle complex constraints on a region's possible extensions between its maxima and minima. For instance, although a vague region such as an area of marshland might have maxima and minima as illustrated in Figure 1a, the area within the dotted line might not correspond to any reasonable precise interpretation of 'marshland'.

Supervaluation semantics is much more general in that it has the potential to model arbitrary constraints on the distribution of possible extensions, as illustrated in Figure 1b. However, the possible extensions of natural vague concepts will not be completely chaotic since, according to supervaluation theory, they correspond to a cluster of precise concepts with similar meanings. In the case of a purely sorites vague concept, where the vaguenss is in the choice of a suitable threshold for some observable, the possible extensions will typically (though not necessarily) be contoured as shown in Figure 1c. Each contour corresponds to a more or less strict sense of a spatial concept. For instance different definitions of 'marshland' may require more or less water to be present. Where we have mixed vagueness we will have several sets of contours each corresponding to varying the threshold for some conceptually unambiguous but still sorites vague concept.

5 Geographical Concepts and Features

Vagueness is pervasive in spatial and geographical concepts and tends to persist even where steps are taken to give them precise definitions. For example, in (Ordnance Survey 2000), the guide book to the Ordnance Survey's *Land-Line* data set a road is defined as: "A metalled way for vehicles." This does tell us something about what is meant by 'road' but the definition is still vague in many respects. We may be unsure about what surfaces count as 'metalled'. Neither condition of the surface nor any restrictions on its spatial extent is specified. The term 'way' could be understood in many more or less general ways. 'Vehicle' is also too general a term for what is intended. The OS definition of road would seem to apply to bicycle paths, which may not be intended. It also seems to rule out cobbled or flagged streets which one might expect to be classified as roads.

5.1 When is Vagueness an Issue?

To illustrate how our theoretical approach might be applied to practical problems we focus in this paper on the example of the concept of a 'forest', looking at different ways in which the term can be interpreted and how these effect the determination of the spatial extension of a forest.

Inventories of the properties and extent of woodland are vital to efficient forest management. However, problems of definition are not a major concern in the literature on forestry (the standard text book (Husch, Miller and Beers 1963) does not mention any definitional problems). This is not surprising when we consider the nature of forestry and the kinds of information it requires. For most purposes a forester can assume that a forest consists of a collection of *stands* whose boundaries are well-defined. The properties of each stand can then be determined by random sampling techniques; and from these measurements, economically important quantities such as 'forest volume' can then be derived by simple computations or by the use of empirically verified tables. The problem of determining boundaries is not of great importance because the statistical approach to measurements works with any reasonable bounding of the forest area, and, in all but exceptional cases, mitigates the effect of any uncertainty in this boundary.

Although meta-questions about the nature of a forest may be largely irrelevant to the narrowly defined concerns of industrial forestry, they are certainly relevant to more general problems of determining and allocating land types. For instance, if we want to answer a question such as 'How rapidly is the forested area of the earth shrinking?' the problem of demarcating forest areas is central.

Similar problems apply to the identification and classification of 'deserts' (and the problem of measuring and monitoring the progress of desertification). For instance, the Global Change Data Base produced by NOAA includes multiple data sets on topics such as soil, precipitation, vegetation, temperature, land cover, etc. Several of these data sets have 'desert' as a specific class, each with its own method of compilation and concept of what actually constitutes a desert — absence of vegetation, annual rainfall below a particular (and varying) threshold, number of months per year exceeding a precipitation threshold, type of soil, ecosystem characteristics, etc. The result is a set of maps that produces a very different distribution of deserts according to which classification you choose to use at any one time. Moreover, the concept of desert can itself be variously classified into sub-types (e.g. 'desert, mostly bare', 'sand desert, partly blowing', 'other desert and semi-desert', 'polar desert', 'tropical desert').

A further example of the importance in environmental modelling of clarifying vague terms is provided by Alker, Joy, Roberts and Smith (2000) who consider issues in defining the concept of a 'Brown-field' which is often used in formulating development policies.

5.2 Defining 'Forest'

I now examine the range of possible definitions which may be used to specify a precise concept corresponding to some reasonable sense of the natural language concept of 'forest'. I start by considering a number of questions, each of which addresses one of the main aspects of vagueness associated with the term, and hence has no clear-cut answer:

1. Is a forest a natural feature or one determined by convention and legality?
2. Does 'forest' refer to an integral feature or can it be applied to an arbitrary region of land?
3. What type of vegetation can constitute a forest? (i.e. what species and how big must they be?)
4. How dense must the vegetation be?
5. How large an area must a forest occupy?
6. Are there any constraints on its shape?

7. Must a forest be self connected, or can it consist of several disjoint parts?
8. Must it be maximal or could it share a border with another region of forest?
9. Is a clearing a part of or a hole in a forest?
10. Are roads and paths going through a forest parts of the forest?
11. How should seasonal and other temporal variations be taken into account?
12. If part of a forest is felled and subsequently re-grown, does it remain part of the forest throughout?[3]

In the following subsections I shall suggest how the issues underlying these questions can be clarified by differentiating between many possible precise senses of 'forest'. Preliminary analysis of some of these sense variations in terms of supervaluation theory will also be given.

5.3 Natural *vs.* 'Fiat' Forests

One of the most important aspects of the conceptual vagueness of the term 'forest' is the ambiguity between forests conceived of as a natural feature and forests as parcels of land upon which is legally or conventionally conferred the status of being a forest. Although it may be argued that forests are always originally identified with some natural feature, once they are named (and thus probably also owned) additional conventional and legal mechanisms may be employed to individuate forests. Smith (1995) has investigated the ontology of conventional regions of this kind, which he calls *fiat* regions.

In axiomatising the vague term 'forest' it is clear that the natural and fiat interpretations will obey rather different axioms. Hence, any ontology of geographical concepts should split the concept into two specialisations. The following axioms ensure that in any precisification Fiat_Forest and Natural_Forest are sub-concepts of Forest and that all forests are of one of these two types (they do not rule out the possibility that something may be both):

- $\forall px[\mathsf{Fiat\_Forest}(p,x) \rightarrow \mathsf{Forest}(p,x)]$
- $\forall px[\mathsf{Natural\_Forest}(p,x) \rightarrow \mathsf{Forest}(p,x)]$
- $\forall px[\mathsf{Forest}(p,x) \rightarrow (\mathsf{Fiat\_Forest}(p,x) \vee \mathsf{Natural\_Forest}(p,x))]$

Though free from a certain ambiguity, the predicates Fiat_Forest and Natural_Forest are still extremely vague, each will correspond to a wide range of possible senses and further subdivisions and axioms will be required to explicate these. In the rest of the analysis I shall dealing only with 'natural' forests, since these seem to be vague in a greater variety of ways; however, the semantics of fiat forests is no doubt also very complex. Henceforth the predicate Forest shall be used to mean Natural_Forest.

---

[3] Accompanying its 'Land Usage of the World' data the web site `www.ecoworld.com` gives the following definition of forest: "Forest: Land under natural forests or planted stands of trees. Also includes logged areas to be replanted in the near future, after logging."

5.4 Forest as Feature or Land Type

In clarifying the concept of '(natural) forest' we immediately encounter a second fundamental ambiguity that affects this and many similar geographical concepts. When used with an article ('a forest' or 'the forest') the term typically refers to a particular integral feature whose boundary (albeit vague) is determined by the meaning of the concept. However, it can also be used in an adjectival sense to describe an arbitrary region as 'forest'. These two uses are not really due to vagueness but rest on a logical distinction that ought to be explicit in the ontology of any GIS that supports high-level queries.

Though ontologically distinct, features and corresponding land-type concepts have strong logical interdependence which must be formally specified (see (Eschenbach 2000)). Let us use the predicate Forest as a vague feature type and Forested as the corresponding vague land-type classifier and see what axioms one would expect to link the two concepts. We might be inclined to say that a region is 'forested' iff it is part of some forest. However, this definition suffers from a problem of granularity, since forests may contain pockets which are not at all forested. We can avoid this problem by taking forested as the more basic property. Using P for parthood and CON for connected we can define a forest as a maximal connected wooded region:

$$\begin{aligned}\mathsf{Forest}(p,x) \ \equiv_{def} \ & (\mathsf{Forested}(p,x) \wedge \mathsf{CON}(x) \wedge \\ & \neg\exists y[\mathsf{CON}(y) \wedge \mathsf{P}(x,y) \wedge \mathsf{Forested}(p,y)]) \ .\end{aligned}$$

The scope of the precisification variable p ensures that under any given precisification the meaning of Forest is logically determined by the meaning of Forested under that same precisification. This would support various patterns of reliable inference that hold whatever reasonable sense we give the concepts.

5.5 Classifying Vegetation

Having defined 'forest' in terms of 'forested' we need to consider how observable measurements of the physical world determine which regions should be deemed 'forested'; or rather, given our supervaluation methodology, we need to elucidate how these observables relate to different precise interpretations of 'forested'.

Table 1 shows a *physiognomic* classification, of levels of forestation that was proposed in (UNESCO 1973) and later adopted in (USGS 1994b). The range of different terms employed in the table illustrates how a precisification (or class of precisifications) is not merely associated with a collection of senses of individual terms but with a complex system of logical constraints concerning the meanings of multiple interrelated concepts.

This classification carries with it a lot of implicit conceptual baggage which may not be compatible with other ways of defining forests. For instance, any precisification satisfying it must enforce the constraint that woodland and shrubland are necessarily disjoint. There is also some lack of specificity in the classification. It is not clear whether a population of fairly widely spaced tall trees growing among a dense cover of small shrubs should be counted as 'sparse woodland' or

	Percent Canopy Cover of Vascular Vegetation			
Plant-form/Height	100%–60% (interlocking)	60%–25% (touching)	25%–10% (spaced)	10%–1%
Trees >5m	Forest	Woodland	Sparse Woodland	Sparsely Vegetated
Shrubs/Trees 0.5–5m	Shrubland		Sparse Shrubland	
Shrubs <0.5m	Dwarf Shrubland		Sparse Dwarf Shrubland	
Herbs	Herbaceous			

Table 1. A physiognomic classification of vegetation types (UNESCO 1973)

'shrubland'. So some precisifications satisfying UNESCOF might require height to take precedence over density while others could require the converse.

Another way of describing vegetation is the so-called *floristic* approach, where vegetation types are specified in terms of the plant species found within a region (see e.g. (USGS 1994b)). Given that an area may contain limitless possible combinations and distributions of different plants there is the potential for a huge number of different classifications. Although I have not investigated these in any detail I am confident that the supervaluation approach could be used to formalise relationships among floristically defined concepts.

5.6 Determining Boundaries for Dense Forestation

The diversity of views taken in the literature on vegetation classification indicates that no generally applicable solution has yet been found to the problem of demarcating boundaries for vegetation types and similar land classifications. (USGS 1994a) surveys a number of approaches by which the demarcation of stand boundaries is elicited from other relevant and more directly measurable factors, such as climate, and topography. In relation to soil-type boundaries Mark and Csillag (1989) note that boundaries are often introduced on the basis of surface features that are correlated with but not essential to soil classification. Similar influences of inessential features on forest stand demarcation from aerial photographs are hinted at by the results of Johnston and Lowell (2000).

Although indirect methods may be effective for many purposes they do not elucidate how to partition vegetation-types in terms of properties of the vegetation itself and hence, from an ontological point of view they are inadequate. If we consider the problem from the point of essential properties of forested regions we have at least the following options:

- *Atomic area based:* divide an area up into cells. Apply some measurement procedure to each cell to determine whether is is a 'forest' area.
- *Tree population based:* first determine the distribution of trees (e.g. as a set of point coordinates) and then apply some algorithm to determine the forest boundary.

Most data in GIS and other land surveys is implicitly based on atomic areas (e.g. raster cells) as a minimal unit for which a land-type is determined. Using this ontology the difficulty of assigning a boundary to an intricate natural object is largely avoided. Instead, measurements are applied to whole cells (or random samples from cells) and a land-type inherits its boundary from the already given boundaries of a group of cells. This is fine as long as we always take a coarse view of the world, where we can treat the cells as atomic units. However, if we are in the business of accounting for the different senses of a term like 'forest' we will also want to account for perspectives that go right down to the level of individual trees. For example we might have data that tells us that a garden is within a dense forest but we cannot infer that the garden contains trees unless the atomic units that have been classified as forest are smaller than the garden.

The alternative of defining 'forested' in terms of individual trees is difficult but may lead to a fruitful analysis. There are certainly a large number of incompatible ways in which this could be done and the lack of any reason to chose a particular approach is perhaps the main reason why this kind of classification has not been widely studied. However, using the framework of supervaluation theory one can explore many possible definitions without committing to any one.

A simple example is the following definition, in which a forested region is defined as one such that all of its points are within a certain distance of a tree:

$$\mathsf{Forested}(p,x) \;\equiv_{def}\; \forall(\pi \in x)[\exists t[\mathsf{Tree}(p,t) \wedge \mathsf{dist}(\pi,t) \leq \delta]]$$

This actually gives a family of possible precise concepts determined by the rather arbitrary parameter δ. There is a strong logical constraint on this family in that the extension of this concept with a small δ is always a subset of the extension determined by any larger δ. Hence we might get a contoured distribution of extensions similar to that shown in Figure 2. An obvious problem with this definition is that each isolated tree constitutes a small forest. Perhaps a better approach is to consider the sum of *convex hulls* of sets of 'sufficiently close' trees. A more detailed examination of possible definitions of 'forested' can be found in (Bennett 2001).

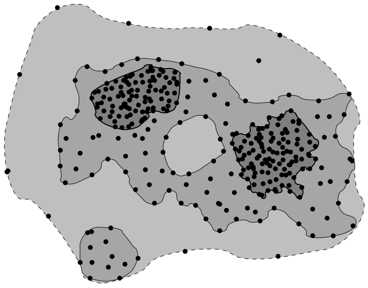

Fig. 2. Possible forest demarcations for a given tree distribution

6 Conclusion

The paper has highlighted the effect of vaguely defined concepts on the interpretation of spatial data. The idea of a space of possible precisifications has been suggested as a powerful tool for describing and manipulating many aspects of vagueness, and its use illustrated by a variety of examples relevant to GIS. However, certain phenomena of spatial vagueness may be beyond the expressive capabilities of the simple languages that have been presented here. More complex logics with additional operators are described in (Bennett 1998).

Although the supervaluation approach is extremely general, applications of the theory are at a very early stage of development. To use the formalism within a GIS one would need to provide detailed axioms describing the logical relationships between the meanings of a significant family of vague geographical concepts. Further work in this direction can be found in (Bennett 2001). Additionally one would need to implement practical functionality for manipulating vague concepts within a query system or data analysis package. Each of these tasks is a major project in itself but one that must be undertaken if the detailed quantitative information stored in GIS is to be connected with the vague intuitive concepts of natural language.

References

Alker, S., Joy, V., Roberts, P. and Smith, N.: 2000, The definition of brownfield, *Journal of Environmental Planning and Management* **43**(1), 49–69.

Bennett, B.: 1998, Modal semantics for knowledge bases dealing with vague concepts, *in* A. G. Cohn, L. Schubert and S. Shapiro (eds), *Principles of Knowledge Representation and Reasoning: Proceedings of the 6th International Conference (KR-98)*, Morgan Kaufman, pp. 234–244.

Bennett, B.: 2001, What is a forest? on the vagueness of certain geographic concepts, *Topoi* **20**(2).

Burrough, P. and Frank, A. (eds): 1996, *Geographical Objects with Undetermined Boundaries*, GISDATA, Taylor and Francis.

Calvanese, D., Lenzerini, M. and Nardi, D.: 1998, Description logics for conceptual data modelling, *in* J. Chomicki and G. Saake (eds), *Logics for Databases and Information Systems*, Kluwer, chapter 1.

Cohn, A. G. and Gotts, N. M.: 1996a, The 'egg-yolk' representation of regions with indeterminate boundaries, *in* P. Burrough and A. M. Frank (eds), *Proceedings, GISDATA Specialist Meeting on Geographical Objects with Undetermined Boundaries*, Francis Taylor, pp. 171–187.

Cohn, A. G. and Gotts, N. M.: 1996b, Representing spatial vagueness: a mereological approach, *in* J. D. L C Aiello and S. Shapiro (eds), *Proceedings of the 5th conference on principles of knowledge representation and reasoning (KR-96)*, Morgan Kaufmann, pp. 230–241.

Eschenbach, C.: 2000, Ontology, predicates and identity, *in* S. Winter (ed.), *Geographical Domain and Geographical Information Systems*, Vol. 19 of *Geoinfo*, Institute for Geoinformation, Vienna University of Technology, pp. 33–36.

Evans, M.: 1978, Can there be vague objects?, *Analysis* **38**, 208. reprinted in his *Collected Papers*, Oxford, Clarendon Press 1985.

Fine, K.: 1975, Vagueness, truth and logic, *Synthèse* **30**, 263–300.

Frank, A.: 1997, Spatial ontology: a geographical information point of view, *in* O. Stock (ed.), *Spatial and Temporal Reasoning*, Kluwer, Dordrecht, chapter 5, pp. 321–352.

Goodchild, M. F.: 1993, 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.

Goodchild, M. F. and Gopal, S. (eds): 1989, *Accuracy of Spatial Databases*, Taylor and Francis, London.

Heuvelink, G. B.: 1998, *Error Propagation in Environmental Modelling with GIS*, Taylor and Francis.

Hughes, C.: 1986, Is a thing just the sum of its parts?, *Proceedings of the Aristotelian Society* **86**, 213–233.

Husch, B., Miller, C. and Beers, T.: 1963, *Forest Mensuration*, Wiley. 3rd edition 1982.

Johnston, D. and Lowell, K.: 2000, Forest volume relative to cartographic broundaries and sample spacing, unit size and type, *Geographical and Environmental Modelling* **4**(1), 105–120.

Kulik, L.: 2000, Vague spatial resoning based on supervaluation, *in* S. Winter (ed.), *Geographical Domain and Geographical Information Systems*, Vol. 19 of *Geoinfo*, Institute for Geoinformation, Vienna University of Technology, pp. 73–80.

Lehmann, F. and Cohn, A. G.: 1994, The EGG/YOLK reliability hierarchy: Semantic data integration using sorts with prototypes, *Proc. Conf. on Information Knowledge Management*, ACM Press, pp. 272–279.

Maling, D.: 1989, *Measurements from Maps: principles and methods of cartometry*, Pergamon Press.

Mark, D. and Csillag, F.: 1989, The nature of boundaries on 'area-class' maps, *Cartographica* **26**, 65–78.

McCarthy, J.: 1993, Notes on formalizing context, *Proceedings of the thirteenth international joint conference on artificial intelligence (IJCAI-93).*

McGee, V.: 1997, Kilimanjaro, *Canadian Journal of Philosophy* pp. 141–195. supplementary volume 23.

Ordnance Survey: 2000, *Land-Line User Guide v3.0.* http://www.o-s.co.uk/downloads/vector/landline/Lline_w.pdf.

Orłowska, E. (ed.): 1997, *Incomplete Information – rough set analysis*, Physica-Verlag, Heidelberg.

Smith, B.: 1995, On drawing lines on a map, *in* A. Frank and W. Kuhn (eds), *Spatial Information Theory - Proceedings of COSIT'95*, number 988 in *Lecture Notes in Computer Science*, Springer, Berlin, pp. 475–484.

UNESCO: 1973, International classification and mapping of vegetation, A report of the United Nations Educational, Scientific and Cultural Organization (UNESCO), Series 6, Ecology and Conservation.

USGS: 1994a, *Field Methods for Vegetation Mapping*, United States Geological Survey, NPS Vegetation Mapping Program. http://biology.usgs.gov/npsveg/fieldmethods/index.html.

USGS: 1994b, *Standardized National Vegetation Classification System*, United States Geological Survey, NPS Vegetation Mapping Program. http://biology.usgs.gov/npsveg/classification/index.html.

Vckovski, A.: 1997, Interoperability and spatial information theory, *Proceedings of Interop97, International Conference on Interoperating Information Systems.*

Wang, F. and Brent Hall, G.: 1996, Fuzzy representation of geographical boundaries in GIS, *International Journal of GIS* **10**(5), 573–590.

Williamson, T.: 1994, *Vagueness*, The problems of philosophy, Routledge, London.

Zadeh, L. A.: 1975, Fuzzy logic and approximate reasoning, *Synthese* **30**, 407–428.