UNCERTAINTY OF TOPOLOGICAL RELATIONS IN GIS

Stephan Winter *
Institut für Photogrammetrie, Universität Bonn, Germany
Stephan.Winter@ipb.uni-bonn.de

Intercommission Working Group III/IV

KEYWORDS: uncertainty, spatial relations, topology, data quality, geographical information systems.

ABSTRACT: We present a concept for representing uncertain topological relations and their derivation from uncertain sets, useful for spatial and temporal reasoning in GIS. The concept is based on the notion of a stochastic boundary of a geometric set and on tests performed to decide the validity of relations between the sets. It uses the power-function to derive the probabilities of the found relations. The concept is applicable to all questions where uncertain geometric queries or analysis have to be performed.

1 Introduction

1.1 Motivation

Geometric data are uncertain by nature. E.g. two points in a database representing the same real world object will nearly never be captured with identical coordinates for several reasons

- the abstraction of the real world into the nominal ground during the measurement process is uncertain (object identifiability),
- the data capturing from the nominal ground is burdened with random errors (imprecision),
- the representation of real numbers in a computer is discrete at least, due to data processing, rounding and due to data formats (resolution).

Spatial queries in GIS are based on the model of point sets and, if available, given topological relations. The quality of the result is hard to evaluate without knowledge of the quality of the given data, the quality of the used algorithm and of the sensitivity or the propagation of imprecision through the analysis process.

The demand for explicitely representing data quality in the data model of the GIS seems to be a consensus in the research community (cf. Chrisman 1988, Goodchild and Gopal 1989, Caspary 1992). Methods of error propagation from statistics are expensive and not adaptable to every type of (spatial) reasoning in GIS. However, statistical methods seem to be the only way to consistently reason about spatial or temporal objects and they do not appear to be fully exploited yet. Efforts are made to visualize quality of results (e.g. Beard *et al.* 1991, Fisher 1993, Kraus and Haussteiner 1993, van der Wel and Hootsmans 1993) and also to use simulations (Veregin 1994) instead of analytical error propagation.

If data uncertainty is explicitly given, for example by variances and covariances, or a covariance function, it also should be possible to represent the uncertainty of the topological relations. There are attempts to express uncertainty of data and relations in fuzzy constraints of space and time using the fuzzy set theory (Dutta 1991), however lacking an observation theory.

*Supported by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 350, "Wechselwirkungen kontinentaler Stoffsysteme und ihre Modellierung"

1.2 The Role of Topology

Topological relations have been found useful for speeding up spatial queries. At times they are sufficient for themselves and no geometric analysis of the the data need be performed. Often the analysis of topological relations may reduce the burden of geometric computations.

A formal analysis of relations between sets has been provided by Egenhofer in several publications (Egenhofer and Franzosa 1991, Egenhofer 1991, Egenhofer 1993, Egenhofer *et al.* 1994). The idea is to represent the mutual relations of two sets $\mathcal{A}$ and $\mathcal{B}$ by a 2 by 2 matrix, called the 4-intersection $\mathbf{F}$, which is given by

$$\mathbf{F} = \begin{pmatrix} \partial\mathcal{A} \cap \partial\mathcal{B} & \partial\mathcal{A} \cap \mathcal{B}^\circ \\ \mathcal{A}^\circ \cap \partial\mathcal{B} & \mathcal{A}^\circ \cap \mathcal{B}^\circ \end{pmatrix} \tag{1}$$

where $\partial\mathcal{A}, \partial\mathcal{B}$ denote the boundaries, $\mathcal{A}^\circ, \mathcal{B}^\circ$ the interiors of the regions $\mathcal{A}$ and $\mathcal{B}$, and the matrix elements can have the values *empty set* ($\emptyset$, 0) and *non-empty set* ($\neg\emptyset$, 1). From the $2^4 = 16$ possible matrices $\mathbf{F}$ only 8 can be realized by simply connected regions in 2D, which are DISJUNCT, MEETS, OVERLAPS, EQUALS, COVERS, COVEREDBY, CONTAINS, and CONTAINEDBY (cf. Table 1).

One goal is to predict possible values for a relation $R(\mathcal{A}, \mathcal{C})$ in case $R(\mathcal{A}, \mathcal{B})$ and $R(\mathcal{B}, \mathcal{C})$ are given (Egenhofer 1991). E.g. from CONTAINS($\mathcal{A}, \mathcal{B}$) and CONTAINS($\mathcal{B}, \mathcal{C}$) follows CONTAINS($\mathcal{A}, \mathcal{C}$) without geometric analysis.

The relations are assumed to be crisp. No order, link or relation between these relations themselves is established.

1.3 Scope of the Paper

The scope of this paper is to add uncertainty to the representations. This is done by two steps:

Relation R_i	$\mathbf{F}$	Example 1D
Disjunct	$\begin{pmatrix} \emptyset & \emptyset \\ \emptyset & \emptyset \end{pmatrix}$	$\mathcal{A}$ $\mathcal{B}$
Meets	$\begin{pmatrix} \neg\emptyset & \emptyset \\ \emptyset & \emptyset \end{pmatrix}$	
Overlaps	$\begin{pmatrix} \neg\emptyset & \neg\emptyset \\ \neg\emptyset & \neg\emptyset \end{pmatrix}$	
Covers	$\begin{pmatrix} \neg\emptyset & \emptyset \\ \neg\emptyset & \neg\emptyset \end{pmatrix}$	
CoveredBy	$\begin{pmatrix} \neg\emptyset & \neg\emptyset \\ \emptyset & \neg\emptyset \end{pmatrix}$	
Equals	$\begin{pmatrix} \neg\emptyset & \emptyset \\ \emptyset & \neg\emptyset \end{pmatrix}$	
Contains	$\begin{pmatrix} \emptyset & \emptyset \\ \neg\emptyset & \neg\emptyset \end{pmatrix}$	
ContainedBy	$\begin{pmatrix} \emptyset & \neg\emptyset \\ \emptyset & \neg\emptyset \end{pmatrix}$	

Table 1: Topologic relations in 1D and in 2D.

1. We investigate the effect of small changes of two sets on their relations. This will reveal a network of likely changes between the relations.
2. We present a probabilistic description of the uncertainty of the relations, which allows to derive the probability of a certain relation given the boundary of a set and its uncertainty.

2 A Partial Order of Topologic Relations

For dynamic systems, Galton discusses an order of topologic relations which a body (a region, a line, a point) must pass to change a position relatively to other bodies (Galton 1994). Although he argues within a framework for space and time *sequences* we can adopt his idea to a static situation in a GIS database because the uncertainty of a decision about the mutual relation between two sets can be seen to be a random perturbation of the sets within time.

Assume two regions (sets) $\mathcal{A}$ and $\mathcal{B}$ move towards each other (cf. Fig 1). When observing Meets($\mathcal{A}, \mathcal{B}$) it actually is uncertain whether still $\mathcal{A}$ is Disjunct from $\mathcal{B}$, whether $\mathcal{A}$ Meets $\mathcal{B}$ or whether already $\mathcal{A}$ Overlaps $\mathcal{B}$. On the other hand, if we state Disjunct($\mathcal{A}, \mathcal{B}$) by some observation then we implicitely state that Overlaps($\mathcal{A}, \mathcal{B}$) is very unlikely. This idea gives us the possibility to introduce a transitional order between Egenhofer's relations.

Galton introduces three notions which specify the mutual relation between two relations R_i and R_j.

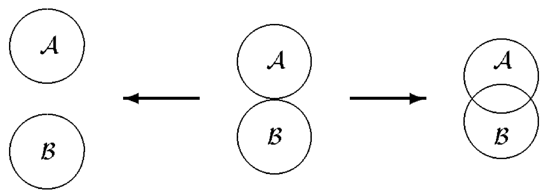

Figure 1: The time sequence Disjunct, Meets, Overlaps, with dominance of Meets indicated by the arrows.

1. A relation R_i is a *bounding relation* of R_j if R_i holds at an instant which bounds an interval throughout R_j holds.
2. A relation R_i is *dominant* if R_i is a bounding relation of R_j but R_j is no bounding relation to R_i.
3. *Perturbation*, if R_i and R_j are mutually dominant, a notion we do not use here.

Observe that Galton's definition of dominance does not refer to any uncertainty.

In the example of fig. 1 there will be an interval where Disjunct holds, one crisp point of meeting ($\{\partial\mathcal{A} \cap \partial\mathcal{B}\} = \neg\emptyset$) and next an interval Overlaps holds. We say the crisp relation Meets dominates the relations Disjunct and Overlaps.

If we now analyse the eight relations between two sets, we find conditions for a relation dominating another, which can be seen in the 4-intersections $\mathbf{F}_{ij}$ for the corresponding relations.

1. The element $\mathbf{F}_{11} = \partial\mathcal{A} \cap \partial\mathcal{B}$ changes from 1 to 0, thus the boundaries of $\mathcal{A}$ and $\mathcal{B}$ separate (e.g. Meets $\rightarrow$ Disjunct).
2. The elements $\mathbf{F}_{12} = \partial\mathcal{A} \cap \mathcal{B}^\circ$ or $\mathbf{F}_{21} = \mathcal{A}^\circ \cap \partial\mathcal{B}$ change from 0 to 1, thus the boundary of one set penetrates the interior of the other set (e.g. Covers $\rightarrow$ Overlaps).
3. The element $\mathbf{F}_{22} = \mathcal{A}^\circ \cap \mathcal{B}^\circ$ changes from 0 to 1, thus the interiors of $\mathcal{A}$ and $\mathcal{B}$ penetrate each other (e.g. Meets $\rightarrow$ Overlaps).
4. The combination of two of the previous transitions 1 or 2 reveal Equals also to dominate Contains and ContainedBy, and Equals to dominate Overlaps. However we will see that these dominations are weak ones. Table 2 shows the dominance relations as 8×8 matrix, Fig 2 as a directed graph.

3 Uncertain Geometry and Topology

Topologic relations are independent on the geometry. However, topologic relations usually are derived from geometric descriptions. This is also necessary when taking the uncertainty of the geometric entities into consideration which of course leads to *derived* quantities of uncertainty of the topologic relations.

Remark: Uncertainty of relations is already used in GIS during data acquisition. E.g. if two polygons as in fig. (3)

	DJ	ME	OV	CV	CVB	EQ	CT	CTB
DJ	$\odot$							
ME	$\otimes$	$\odot$	$\otimes$					
OV			$\odot$					
CV			$\otimes$	$\odot$			$\otimes$	
CVB			$\otimes$		$\odot$			$\otimes$
EQ			$\times$	$\otimes$	$\otimes$	$\odot$	$\times$	$\times$
CT							$\odot$	
CTB								$\odot$

Table 2: Transition matrix: the left columned relations fulfill the marked dominance conditions to the top line relations. $\odot$: $\mathbf{F}_{ij}$ identity, $\otimes$: relation dominates, $\times$: relation dominates by at least two conditions (weak dominance).

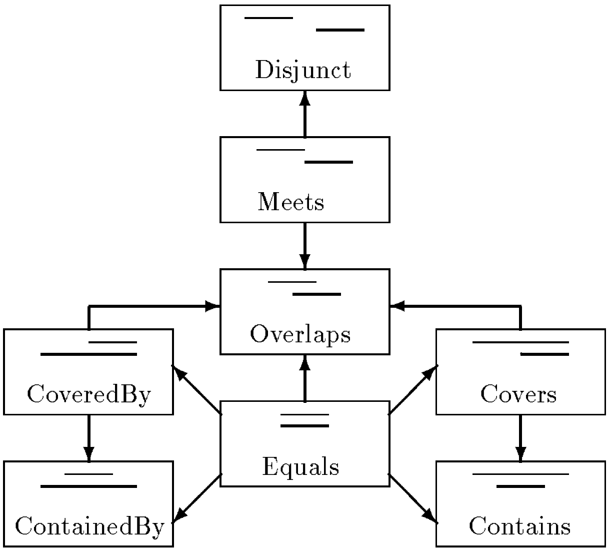

Figure 2: Directed graph of the partial order of topologic relations. The direction is derived from the dominance of relationships.

are given to a GIS it may report overlap between the two regions but also indicate that the error band of one polygon contains some part of the boundary of the other one, asking the user to decide on the true relation, MEETS or OVERLAPS. This indication is based on some tolerance, say δ, which the user can specify, which however is not really related to the accuracy of the data within the GIS. It obviously would be better to store the uncertainty of the polygon as meta-data and to use this information for a consistency check. This of course requires consistent representation and use of the uncertainty. □

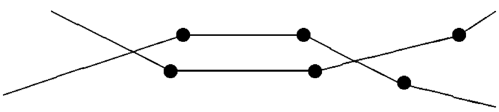

Figure 3: Are the two polygons overlapping, or meeting?

3.1 Uncertain Sets and Regions

A crisp set $\mathcal{A}$ may be defined by the indicator function

$$I(\mathbf{x}) = \begin{cases} 1 & : \ \mathbf{x} \in \mathcal{A} \\ 0 & : \ \mathbf{x} \notin \mathcal{A} \end{cases} \qquad (2)$$

Similarly, we are able to define an *uncertain set* $\underline{\mathcal{A}}$:

Definition 1: An *uncertain set* $\underline{\mathcal{A}}$ is defined by the probability function

$$I(\underline{\mathbf{x}}) = P(\underline{\mathbf{x}} \in \underline{\mathcal{A}}) \qquad (3)$$

with $P \in [0,1]$.

In case of $P \in \{0,1\}$ this reduces to a crisp set $\mathcal{A}$. Definition 1 implies that $P(\underline{\mathbf{x}} \notin \underline{\mathcal{A}}) = 1 - P(\underline{\mathbf{x}} \in \underline{\mathcal{A}})$.

This allows to define uncertain regions. We assume a region be a realization of a stochastic process describing the set of all regions of that geometry but with uncertain boundary. Therefore we define

Definition 2: An *uncertain region* is a random vector or process describing the *boundary* of the region in some suitable representation.

Example 1: A connected uncertain region $\underline{\mathcal{A}} = \mathcal{A}(\underline{x}_1, \underline{x}_2)$ in 1D is described by the random vector $\underline{\mathbf{x}} = (\underline{x}_1, \underline{x}_2)^T$ together with the joint probability function $P_x(x_1, x_2)$.

Example 2: A connected uncertain polygon $\underline{\mathcal{A}}$ in 2D is described by the random vector $(\underline{x}_1, \underline{y}_1, \ldots, \underline{x}_n, \underline{y}_n)^T$ of its boundary points together with their covariance matrix, assuming the coordinates to be normally distributed.

Example 3: A connected uncertain region $\underline{\mathcal{A}}$ in 2D with a smooth boundary is described by the possibly infinite number of coefficients of its Fourier descriptors, together with their variances, assuming normally distributed coefficients.

We follow the notion of a random variable in statistics: The actual data $\mathcal{A}$ are certain without any ambiguity, as they are treated as an outcome of a certain experiment (in statistical terms). The uncertainty relates to the variable $\underline{\mathcal{A}}$ defined by the experiment, i.e. to the set of all possible outcomes of the experiment. In the above mentioned example a measured region might be $\mathcal{A} = [3.1, 4.9]$ whereas the uncertain region is $\underline{\mathcal{A}} = [\underline{x}_1, \underline{x}_2]$ with $\underline{\mathbf{x}} \sim N(\mu_x, \Sigma_{xx})$ and $\mu_x = (3.0, 5.0)$, $\Sigma_{xx} = \mathrm{Diag}(0.25, 0.25)$. An element belongs to $\mathcal{A}$ or to $\neg\mathcal{A}$, $x \in \mathcal{A}$ xor $x \in \neg\mathcal{A}$, however these *statements* are uncertain. This is in contrast to Fuzzy-Logic where the *degree* of x *belonging* to $\mathcal{A}$ is measured by $\mu(x \in \mathcal{A})$, there for $\mu(x \in \mathcal{A}) = 0.6$ does not necessarily yield to $\mu(x \in \neg\mathcal{A}) = 1 - 0.6 = 0.4$.

3.2 Uncertain Relations

We are now prepared to define the notion of an uncertain relation:

Definition 3: An *uncertain relation* is a relation which exists with a certain probability.

This definition does not specify how to determine the existence nor its probability.

The goal now is to show how to derive the uncertainty of topological relations from the uncertainty of the underlying sets.

Deriving the validity of a topologic relation from uncertain regions is a *classification problem*, as any procedure leads to an assignment to one of the eight relations which itself is uncertain, i.e. may be wrong. The 8×8 confusion matrix $\mathbf{C} = P(\hat{R}_i \mid R_j)$ contains the probabilities of obtaining a certain relation $\hat{R}_i$ while R_j holds true.

In case the boundary of the sets concerned is uncertain without any limit, e.g. in case of assuming a Gaussian for the coordinates of a polygon, *all* relations will always be valid with a probability > 0. This may be relevant in case the uncertainty of the boundary or the position of a set is large compared with its size. Typically, however, the uncertainty of the boundary is small compared with the size of the region, a condition we intuitively used in the argumentation above. We therefore use the following

Assumption: The uncertainty of the boundary is small compared with the size of the region, i.e. $\underline{\mathcal{A}}\ominus\mathcal{D}_\varrho$ and $\underline{\mathcal{A}}\oplus\mathcal{D}_\varrho$ are topologically equivalent to $\underline{\mathcal{A}}$, where $\ominus$ and $\oplus$ denote to erosion and dilation, and $\mathcal{D}_\varrho$ a disk with a sufficiently small radius ϱ.

This is to ensure local analysis of spatial relations.

Based on this assumption, the confusion matrix $\mathbf{C}$ contains entries with very small values $< \eta$, e.g. $\eta = 10^{-6}$, as it could happen e.g. for $P(\hat{R}_i =$ EQUALS $\mid R_j =$ DISJUNCT). Replacing these low probabilities by 0 leads to a sparse confusion matrix $\mathbf{C}$ with entries at the same places as the transition matrix in fig. 2, which may be represented by a undirected graph (i.e. the graph of fig. 2 without directions). The edges of this graph connect neighbouring relations, i.e. relations which may be confused during classification.

Moreover, the relations are of different type with respect to their classification: MEETS, COVERS and COVEREDBY require at least one point of the two boundaries $\partial\mathcal{A}$ and $\partial\mathcal{B}$ to be equal, EQUALS requires all boundary points to be equal. On the other hand, the relations DISJUNCT, OVERLAPS, CONTAINS and CONTAINEDBY put no geometric restriction on the boundary. This suggests to treat the relations MEETS, COVERS, COVEREDBY and EQUALS differently than the other relations, and to develop the classification scheme sequentially, starting from the relation which put the strongest condition onto the geometry of the two sets in concern, using *hypothesis testing.*

4 Classifying Uncertain Topological Relations

We now want to investigate the derivation of probabilities of relations for connected sets in 1D. For this purpose we define crisp and uncertain sets in a unified manner. The classification of a situation into the relation MEETS$(\mathcal{A},\mathcal{B})$ and its neighbouring relations DISJUNCT$(\mathcal{A},\mathcal{B})$ and OVERLAPS$(\mathcal{A},\mathcal{B})$ is analysed in detail using an hypothesis test on the distance between the borders of $\mathcal{A}$ and $\mathcal{B}$ and the corresponding power functions. The tests on COVERS, COVEREDBY and EQUALS then can easily be derived and analysed similarily. The analysis will suggest the use of the power of the test on a point x sitting at the border $\partial\mathcal{A}$ as probability $P(x \in \partial\mathcal{A})$, and consequently lead to explicit expression for the probability the eight relations hold.

4.1 Crisp and Uncertain Sets in 1D

We define two crisp sets $\mathcal{R}_a(x)$ and $\mathcal{L}_a(x)$ in the following manner:

$$\mathcal{R}(a) = \{x \mid x \geq a\} \tag{4}$$
$$\mathcal{L}(a) = \{x \mid x \leq a\} \tag{5}$$

with the corresponding indicator functions

$$I_{\mathcal{R}(a)}(x) = H(x-a) \tag{6}$$
$$I_{\mathcal{L}(a)}(x) = 1 - H(x-a) \tag{7}$$

where $H(x)$ is the Heavyside-function

$$H(x) = \begin{cases} 1 & : \quad x > 0 \\ 1/2 & : \quad x = 0 \\ 0 & : \quad x < 0 \end{cases} \tag{8}$$

being the integral of the δ-function

$$H(x) = \int_{-\infty}^{x} \delta(t)dt. \tag{9}$$

$\mathcal{R}$ and $\mathcal{L}$ are sets with elements right and left of $x = a$. In order to be symmetric we use $I(a) = 1/2$ instead of $I(a) = 1$ (cf. eq. (2)), which is irrelevant when applying this concept.

A crisp set $\mathcal{A}(b,e)$

$$\mathcal{A}(b,e) = \{x \mid b \leq x \leq e\} \tag{10}$$

can be written as

$$\mathcal{A}(b,e) = \mathcal{R}(b) \cap \mathcal{L}(e) \tag{11}$$

with indicator function

$$I_{\mathcal{A}(b,e)}(x) = I_{\mathcal{R}(b)}(x) \cdot I_{\mathcal{L}(e)}(x) \tag{12}$$
$$= H(x-b)\cdot(1-H(x-e)) \tag{13}$$

We now interprete $I_{\mathcal{R}(a)}(x) = H(x-a)$ (eq. (6)) as probability function:

$$I_{\underline{\mathcal{R}}(a)}(x) = P(\underline{a} < x) = P(x \in \underline{\mathcal{R}}(a)) \tag{14}$$

measuring the probability that for a given x the stochastic position $\underline{a}$ of the boundary of the set $\underline{\mathcal{R}}$ is left of x, which is equivalent to saying x can be found to be an element of $\underline{\mathcal{R}}(a)$. Because of the steepness of the Heavyside function the transition is from the left side to the right side of $x = a$, thus changing the probability from 0 to 1 is instantaneous.

It is now easy to change to uncertain sets by replacing the δ-function defining the Heavyside-function in (9) by a density function $f(x)$ with the corresponding distribution function $F(x)$.
Thus we obtain the uncertain set

$$\underline{\mathcal{R}}(f_a) = \{\{x \mid \underline{a} \leq x\}, \underline{a} \sim f_a(x)\} \tag{15}$$

where we assume $E(\underline{a}) = \mu_a$ is the mean value for the left boundary of $\underline{\mathcal{R}}(f_a)$. Eq. 15 states the probability that for a given x we will find $x \in \mathcal{R}(a)$ where a is a sample of $\underline{a}$, thus $\mathcal{R}(a)$ is a sample of $\underline{\mathcal{R}}(f_a)$.

The indicator function for $\underline{\mathcal{R}}(f_a)$ is now given by

$$I_{\underline{\mathcal{R}}(f_a)} = F_a(x) = \int_{-\infty}^{x} f_a(t)\,dt \qquad (16)$$

cf. Fig. 4a with f_a being a box function.
In the following we will write

$$I_{\underline{\mathcal{R}}(a)} \doteq I_{\underline{\mathcal{R}}(f_a)} \qquad (17)$$

in short. Furthermore we will restrict the discussion to f_a being a Gaussian. With

$$\varphi_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp(-\frac{(x/\sigma)^2}{2}) \qquad (18)$$

and

$$\Phi_\sigma(x) = \int_{-\infty}^{x} \varphi_\sigma(t)\,dt \qquad (19)$$

the indicator function reads as

$$I_{\underline{\mathcal{R}}(a)} = \Phi_\sigma(x-a) \qquad (20)$$

Analogeously we obtain the indicator function of the uncertain set for a left-sided region

$$I_{\underline{\mathcal{L}}(a)} = 1 - \Phi_\sigma(x-a) \qquad (21)$$

and of the *uncertain set* $\underline{\mathcal{A}}(b,e)$ (cf. Fig. 4)

$$\begin{aligned} I_{\underline{\mathcal{A}}(b,e)} &= \Phi_\sigma(x-b) \cdot (1 - \Phi_\sigma(x-e)) \\ &= P(\underline{b} < x) \cdot (1 - P(\underline{e} < x)) \\ &= P(\underline{b} < x) \cdot P(x < \underline{e}) \\ &= P(x \in \underline{\mathcal{A}}(b,\,e)) \end{aligned} \qquad (22)$$

showing the indicator function to represent the probability that for a given x we will find x to be right of the beginning point b and left of the end point e thus element of $\underline{\mathcal{A}}(b,\,e)$.

We assumed equal standard deviations for b and e and independence for simplicity.

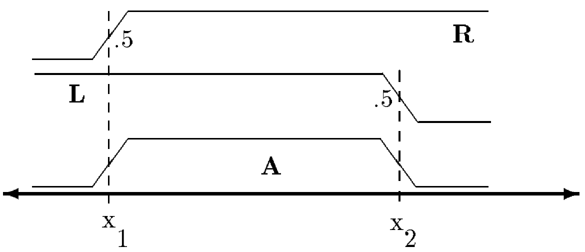

Figure 4: An uncertain 1D region derived from one-sided regions (here using a box function ($x_1 = b, x_2 = e$) for f_a).

We now are able to test dominant relations with respect to dominated relations in order to arrive at probabilities for relations.

4.2 Test for Meets

We first want to test the relation Meets against the relations Disjunct and Overlaps.

Let set $\mathcal{A}(b_\mathcal{A}, e_\mathcal{A})$ be left of set $\mathcal{B}(b_\mathcal{B}, e_\mathcal{B})$ then Meets$(\mathcal{A},\mathcal{B})$ can be tested using the distance

$$s = b_\mathcal{B} - e_\mathcal{A} \qquad (23)$$

Thus the test scheme is as follows:

$$H_0 = \text{Meets}(\mathcal{A},\mathcal{B}) \;:\; s = 0 \qquad (24)$$

is tested against the two alternatives

$$H_{a_1} = \text{Overlaps}(\mathcal{A},\mathcal{B}) \;:\; s < 0 \qquad (25)$$

and

$$H_{a_2} = \text{Disjunct}(\mathcal{A},\mathcal{B}) \;:\; s > 0 \qquad (26)$$

The test uses the test statistic

$$T = \frac{s}{\sigma_s} \sim N(0,1) \qquad (27)$$

assuming σ_s to be derivable from the given standard deviations $\sigma_{b_\mathcal{B}}$ and $\sigma_{e_\mathcal{A}}$.

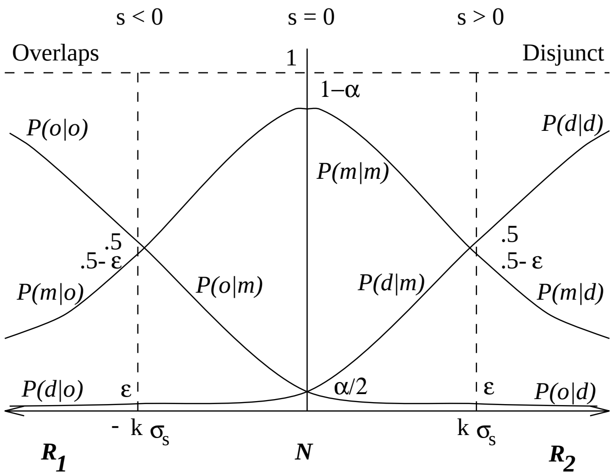

Figure 5: The power of testing Meets vs. Disjunct and Overlaps: graphic of probabilities for all decisions.

Specifying a significance number α leads to an acceptance region $\mathcal{N} = [-k(\alpha), k(\alpha)]$ and two rejection regions $\mathcal{R}_1 = (-\infty, -k(\alpha))$ and $\mathcal{R}_2 = (k(\alpha), +\infty)$. $k(\alpha)$ is the critical value depending on α. For the discussion we use $\alpha = 0.05$ thus $k = 1.96 \approx 2$. The decision of the test yields

$$T \in \mathcal{N} \;\rightarrow\; \hat{R} = m \qquad (28)$$

$$T \in \mathcal{R}_1 \;\rightarrow\; \hat{R} = o \qquad (29)$$

$$T \in \mathcal{R}_2 \;\rightarrow\; \hat{R} = d \qquad (30)$$

where m, o, d stand for Meets, Overlaps, and Disjunct. The probability of correct and incorrect decision is collected in the 3×3 confusing matrix shown in table (3) being a submatrix of **C**. Figure (5) shows the probabilities $P(\underline{\hat{o}} \mid s)$, $P(\underline{\hat{m}} \mid s)$ and $P(\underline{\hat{d}} \mid s)$ depending on the distance s. Obviously the decision is (nearly) a maximum likelihood decision as the most likely relation is selected by the hypothesis test:

- The correct decisions are selected with probabilities between 0.5 and 1. The relation Meets is only selected with a probability up to $1 - \alpha\,(= 0.95)$.
- In case the two regions meet ($R = m, s = 0$) the likelihood of erroneously deciding them to overlap or as being Disjunct is $\alpha/2\,(= 0.025)$.

$\hat{R} \setminus R$	$R = o\,(s < -k\sigma_s)$	$R = m\,(\mid s \mid \leq k\sigma_s)$	$R = d\,(s > k\sigma_s)$
$\hat{R} = o$	$P(\underline{\hat{o}} \mid o(s)) > 0.5$ correct	$P(\underline{\hat{o}} \mid m(s)) \geq \frac{\alpha}{2}$ incorrect	$P(\underline{\hat{o}} \mid d(s)) < \varepsilon$ incorrect
$\hat{R} = m$	$P(\underline{\hat{m}} \mid o(s)) < 0.5 - \varepsilon$ incorrect	$P(\underline{\hat{m}} \mid m(s)) \leq 1 - \alpha$ correct	$P(\underline{\hat{m}} \mid d(s)) < 0.5 - \varepsilon^*$ incorrect
$\hat{R} = d$	$P(\underline{\hat{d}} \mid o(s)) < \varepsilon$ incorrect	$P(\underline{\hat{d}} \mid m(s)) \geq \frac{\alpha}{2}$ incorrect	$P(\underline{\hat{d}} \mid d(s)) > 0.5$ correct

$^*\varepsilon = \Phi_\sigma(-2k(\alpha))$, for $\alpha = 0.05$ we obtain $\varepsilon < 0.0001$.

Table 3: Confusion matrix $\mathbf{C}$ for test on MEETS, OVERLAPS and DISJUNCT.

- The relation DISJUNCT can be detected with a probability $> 50\%$ only if $s > k\sigma_s$. Therefore if $\mid s \mid < k\sigma_s$ it is more likely that we will decide the regions to meet. The probability of erroneously deciding MEETS through $0 < s < k\sigma_s$ is between $\alpha/2$ and 0.5.
- If the relation DISJUNCT holds, it is extremely unlikely that the test yields $\hat{R} =$ OVERLAPS. This probability is less than
$$\varepsilon = \Phi_\sigma(-2k(\alpha)) \tag{31}$$
which for $\alpha = 0.05$ is $\varepsilon < 0.0001$. This goes with intuition and is the reason for exploiting the sparseness of the confusion matrix.
- The reasoning for $R =$ OVERLAPS follows a similar line of thought.

We now can use the power function in the areas where the decisions are correct as probabilities that the relation actually holds. Strictly we must take into account the probabilities of the two other boundary points of $\mathcal{A}$ and $\mathcal{B}$ respectively not being near to the tested boundary points, a situation we explicitely excluded. These probabilities are assumed to be 1. Thus we say

- The probability that OVERLAPS$(\mathcal{A}, \mathcal{B})$ holds is $P(\underline{\hat{o}} \mid o(s)) = P(\underline{\hat{o}} \mid s)$ with $s < -k\sigma_s$. .
- The probability that MEETS$(\mathcal{A}, \mathcal{B})$ holds is $P(\underline{\hat{m}} \mid m(s)) = P(\underline{\hat{m}} \mid s)$ with $\mid s \mid < k\sigma_s$.
- The probability that DISJUNCT$(\mathcal{A}, \mathcal{B})$ holds is $P(\underline{\hat{d}} \mid d(s)) = P(\underline{\hat{d}} \mid s)$ with $s > k\sigma_s$.

In all cases $s = b_\mathcal{B} - e_\mathcal{A}$ is taken from the actual values of the sets $\mathcal{A}(b_\mathcal{A}, e_\mathcal{A})$ and $\mathcal{B}(b_\mathcal{B}, e_\mathcal{B})$.

4.3 Tests for COVERING

The tests for COVERING work similarly. E.g. the test for COVERS$(\mathcal{A}(b_\mathcal{A}, e_\mathcal{A}), \mathcal{B}(b_\mathcal{B}, e_\mathcal{B}))$ uses the distance

$$s = b_\mathcal{B} - b_\mathcal{A} \tag{32}$$

if $e_\mathcal{B} \ll e_\mathcal{A}$ (cf. table (1)), or

$$s = e_\mathcal{B} - e_\mathcal{A} \tag{33}$$

if $b_\mathcal{A} \ll b_\mathcal{B}$.

4.4 Test for EQUALS

The test for $\mathcal{A} = \mathcal{B}$ can be accomplished by testing the identity of the complete boundary thus $\partial\mathcal{A} \equiv \partial\mathcal{B}$. In our case of 1D this reduces to the test

$$H_0 = \text{EQUALS}(\mathcal{A}, \mathcal{B}) : \begin{pmatrix} b \\ e \end{pmatrix}_\mathcal{A} = \begin{pmatrix} b \\ e \end{pmatrix}_\mathcal{B} \tag{34}$$

against the alternative

$$H_a = \neg\text{EQUALS}(\mathcal{A}, \mathcal{B}) : \begin{pmatrix} b \\ e \end{pmatrix}_\mathcal{A} \neq \begin{pmatrix} b \\ e \end{pmatrix}_\mathcal{B} \tag{35}$$

An optimal test is based on the difference vector $\underline{\mathbf{s}} = (b_\mathcal{B} - b_\mathcal{A},\, e_\mathcal{B} - e_\mathcal{A})^T$ and its covariance matrix, leading to the test statistic

$$T = \frac{\mathbf{s}^T\,\mathbf{C}_{ss}^{-1}\,\mathbf{s}}{2} \sim F_{2,\infty} \tag{36}$$

In case of $T < F_{2,\infty,\alpha}$ we have no reason to reject the hypothesis EQUALS$(\mathcal{A}, \mathcal{B})$. If $T > F_{2,\infty,\alpha}$ no indication on the alternatives is available from T alone. This needs to be taken from individual tests on the boundary.

Observe that the dimension of this test is 2 in 1D and equal to the number of boundary points for regions in 2D. Therefore an adaption of critical values or significance levels would be necessary (cf. Baarda 1967, Baarda 1968) in order to obtain the same power of the test on MEETS etc. and EQUALS.

It seems however to be better to perform two independent tests on the two boundary points. With the distances

$$\begin{aligned} s_b &= b_\mathcal{B} - b_\mathcal{A} \\ s_e &= e_\mathcal{B} - e_\mathcal{A} \end{aligned} \tag{37}$$

we have the two tests

$$\begin{aligned} H_{01} &: s_b = 0 \\ H_{a1} &: s_b \neq 0 \end{aligned} \tag{38}$$

$$\begin{aligned} H_{02} &: s_e = 0 \\ H_{a2} &: s_e \neq 0 \end{aligned} \tag{39}$$

which can use the reasoning from section 4.2.

In case both hypothesis are accepted the probability of EQUALS$(\mathcal{A}, \mathcal{B})$ is given by

$$P(\text{EQUALS}(\mathcal{A}, \mathcal{B})) = P(H_{01} \mid s_b) \cdot P(H_{02} \mid s_e) \tag{40}$$

This idea can be generalized to derive the probability for all relations.

4.5 Probability for Relations in 1D

We now can give explicit expressions for the probability of all relations between two connected 1D-sets. We refer to table (1) where the sets $\mathcal{A}(b_\mathcal{A}, e_\mathcal{A})$ and $\mathcal{B}(b_\mathcal{B}, e_\mathcal{B})$ are shown with a thin and a thick line respectivly. We obtain

$$\begin{aligned} P(\text{DISJUNCT}(\mathcal{A},\mathcal{B})) &= P(e_\mathcal{A} < b_\mathcal{B}) \\ P(\text{MEETS}(\mathcal{A},\mathcal{B})) &= P(e_\mathcal{A} = b_\mathcal{B}) \\ P(\text{OVERLAPS}(\mathcal{A},\mathcal{B})) &= P(b_\mathcal{A} < b_\mathcal{B}) \cdot \\ & \quad P(b_\mathcal{B} < e_\mathcal{A}) \cdot \\ & \quad P(e_\mathcal{A} < e_\mathcal{B}) \\ P(\text{COVERS}(\mathcal{A},\mathcal{B})) &= P(b_\mathcal{A} = b_\mathcal{B}) \cdot \\ & \quad P(e_\mathcal{B} < e_\mathcal{A}) \\ P(\text{EQUALS}(\mathcal{A},\mathcal{B})) &= P(b_\mathcal{A} = b_\mathcal{B}) \cdot \\ & \quad P(e_\mathcal{A} = e_\mathcal{B}) \\ P(\text{CONTAINS}(\mathcal{A},\mathcal{B})) &= P(b_\mathcal{A} < b_\mathcal{B}) \cdot \\ & \quad P(e_\mathcal{B} < e_\mathcal{A}) \end{aligned} \quad (41)$$

COVEREDBY and CONTAINEDBY are treated similarily. Observe that the number of conditions which have to hold is different for the different relations. This results from the assumption on the relative uncertainty of the boundaries and the condition $e > a$. The probabilities on the right side are again taken from the power of tests. The power only depends on the significance number of the tests, which has to be specified by the user.
The weak dominance of e. g. EQUALS over OVERLAPS stems from the fact that

$$P(b_\mathcal{A} < b_\mathcal{B} \mid b_\mathcal{A} = b_\mathcal{B}) \cdot P(e_\mathcal{A} < e_\mathcal{B} \mid e_\mathcal{A} = e_\mathcal{B}) \le (\alpha/2)^2 \quad (42)$$

which with e. g. 0.000625 is an order smaller than $\alpha(= 0.05)$.

5 Conclusions and Future Work

We have presented a technique for representing the uncertainty of uncertain sets and for deriving the uncertainty of the topological relations between points of sets. The basic idea was to treat the boundary as stochastic process, in 1D as stochastic variable, and to use the power of the test on identity of boundary points as probabilities of the decisions. Combining the outcomes of different tests lead to explicit expressions for the probability of a relation.

The concept is a first step towards a joint treatment of continous and discrete variables for representing complex patterns in space and time.
Various extensions are obvious and have to be worked out:

- The extension to non-connected sets in 1D is simple, as all boundary points can be treated as has been shown in section (4). This especially may be meaningful in modelling uncertain time-intervals.
- The extension to 2D-regions is more involving. In case of polygons the tests can easily be adapted. The derivation of probabilities, however, has to take multiple connections of the boundaries of the sets into account.
- The symbolic reasoning on more than two uncertain sets, e.g. to predict the relation $R(\mathcal{A},\mathcal{C})$ and its uncertainty from $R(\mathcal{A},\mathcal{B})$ and $R(\mathcal{B},\mathcal{C})$ (Egenhofer 1991), will lead to similar structures which may speed up the rigorous evaluation of uncertainties.
- A raster implementation may simplify the evaluation of the probabilities which may be useful in case of given raster data.
- It has to be investigated in how far the simplification about independence, or the elimination of relations with very low probability lead to serious disturbations or biases in the results.
- The close relation to Fuzzy-Sets revealed by Fig. 4 has to be investigated.

References

Baarda, W. (1967): *Statistical Concepts in Geodesy*, Band 2 der Reihe New Series. Netherlands Geometric Commission, Delft, 1967.

Baarda, W. (1968): *A Testing Procedure for Use in Geodetic Networks*, Band 2 der Reihe New Series. Netherlands Geodetic Commission, Delft, 1968.

Beard, M. Kate; Buttenfield, Barbara P.; Clapham, Sarah B. (1991): NCGIA Research Initiative 7 — Visualization of Spatial Data Quality. *NCGIA Technical Paper*, 1991.

Caspary, Wilhelm (1992): Qualitätsmerkmale von Geo-Daten. *ZfV*, 117(7):360–367, 1992.

Chrisman, Nicholas R. (1988): Digital Cartographic Data Quality. *The American Cartographer*, 15(1), 1988.

Dutta, Soumitra (1991): Topological Constraints: A Representational Framework for Approximate Spatial and Temporal Reasoning. In: Günther, O.; Schek, H.-J. (Eds.), *Advances in Spatial Databases*, pages 161–180, Berlin, 1991. 2nd Symposium SSD '91, Springer Verlag.

Egenhofer, Max J.; Franzosa, Robert D. (1991): Point-set topological spatial relations. *Int. Journal of Geographical Information Systems*, 5(2):161–174, 1991.

Egenhofer, Max J.; Clementini, Eliseo; di Felice, Paolino (1994): Topological relations between regions with holes. *International Journal of Geographical Information Systems*, 8(2):129–142, 1994.

Egenhofer, Max J. (1991): Reasoning about Binary Topological Relations. In: Günther, O.; Schek, H.-J. (Eds.), *Advances in Spatial Databases*, pages 143–160. 2nd Symposium SSD '91, Springer, 1991.

Egenhofer, Max J. (1993): A Model for Detailed Binary Topological Relationships. *Geomatica*, 47(3 & 4):261–273, 1993.

Fisher, P. F. (1993): Algorithm and implementation uncertainty in viewshed analysis. *Int. Journal of Geographical Information Systems*, 7(4):331–347, 1993.

Galton, Anthony (1994): Perturbation and Dominance in the Qualitative Representation of Continous State-Spaces. Technical Report 270, Department of Computer Science, University of Exeter, Exeter, 1994.

Goodchild, Michael; Gopal, Sucharita (Eds.) (1989): *The Accuracy of Spatial Databases.* Taylor & Francis, London, 1989.

Kraus, Karl; Haussteiner, Karl (1993): Visualisierung der Genauigkeit geometrischer Daten. *GIS*, 6(3):7–12, 1993.

van der Wel, Frans; Hootsmans, Rob (1993): Visualization of Quality Information as an Indispensable Part of Optimal Information Extraction from a GIS. In: *Proc. of 16th Int. Cartographic Conference*, pages 881–897, Köln, 1993.

Veregin, Howard (1994): Integration of Simulation Modeling and Error Propagation for the Buffer Operation in GIS. *Photogrammetric Engineering & Remote Sensing*, 60(4):427–435, 1994.

Reference for this paper

This paper will appear in: *Proceedings of ISPRS Commission III Symposium on Spatial Information from Digital Photogrammetry and Computer Vision.* Munich, September 1994. Edited by SPIE.