Modellierung von Unschärfe in GIS

Gerhard Joos
AGIS – Arbeitsgemeinschaft GIS
Universität der Bundeswehr München
Neubiberg - Germany
E-mail: Gerhard.Joos@unibw-muenchen.de

Abstract

Geometric and semantic representation of spatial phenomena in GIS always differs from the actual situation in the real world. This is due to natural limitations of modelling and several other factors, like unclear boundaries of objects, necessary simplification of object boundaries (generalization) and limited precision of the digitizing process. Natural objects are not as well defined as we unconsciously assume they are or they have to be simplified because of the restrictions by the GIS software.

This paper gives an overview and a classification of uncertainties in GIS, and it shows the different approaches to model them. Adjustment of random uncertainties of measurement is based on a well known theory and developed to a very high level. The modelling of transition zones between different feature types is a subject of research.

The positional accuracy of spatial objects in vector GIS can be represented by ε-bands. These are based on statistical behaviour. The resolution of a geometrical representation, i.e. the smallest distance that can be discriminated, is related with the positional accuracy.

Fuzzy boundaries can be modelled using the fuzzy set theory. In this approach it is assumed that the boundary is represented by a transition zone, which is modelled by a membership function. The topological relationship between objects can than be enhanced by fuzzy operators. The number of states of relationships between features is increasing. A different approach of representing objects with fuzzy boundaries is the egg-yolk theory. The object is represented by a set of cascading polygons with different membership values.

Zusammenfassung

Die geometrische und die semantische Beschreibung von räumlichen Objekten in GIS wird sich grundsätzlich von der tatsächlichen Situation in der realen Welt unterscheiden. Das liegt an den natürlichen Einschränkungen bei der Modellierung und anderen Einflüssen wie unscharfen Abgrenzungen von Objekten, Vereinfachungen bei der Objekterfassung und eingeschränkter Erfassungsgenauigkeit. Natürliche Objekte sind nicht so eindeutig definiert, wie wir das unbewusst bei der Datenmodellierung von GIS voraussetzen. Außerdem müssen aufgrund der Einschränkungen der GIS-Software Vereinfachungen gemacht werden.

Dieser Artikel gibt einen Überblick über die Unsicherheiten von Geodaten und zeigt Ansätze, wie diese modelliert werden können. Die Behandlung von stochastischen Messunsicherheiten ist in der Ausgleichungsrechnung ausgereift und gut erforscht. Unscharf begrenzte Objekte werden erst seit Mitte der 80er in der Literatur behandelt.

Die Lagegenauigkeit von objektstrukturierten Geodaten kann mit Hilfe von ε-Bändern beschrieben werden. Diese entstehen durch die Fortpflanzung der Messunsicherheiten auf Punkte der Interpolationskurven. Die Auflösung und damit die Anzahl der Stützpunkte zur Erfassung eines Objektes ist mit der Genauigkeit korreliert.

Unscharfe Grenzen können mit Hilfe der Fuzzy Set Theorie beschrieben werden. Bei diesem Ansatz wird die Begrenzung nicht durch eine scharfe Linie wiedergegeben, sondern durch eine Übergangszone. Zur Beschreibung der topologischen Beziehung zwischen unscharf begrenzten Objekten muss die Anzahl der Operatoren zur Beschreibung der topologischen Beziehung erweitert werden. Ein anderer Ansatz zur Beschreibung von fuzzy Objekten ist durch die „Egg-Yolk“-Theorie (am besten vielleicht mit Ei-Dotter zu übersetzen) gegeben. Dabei wird ein flächenhaftes Objekt durch geschachtelte Polygone mit unterschiedlichen Zugehörigkeitswerten dargestellt.

Motivation

Die Behandlung von Messunsicherheiten beim Einmessen von scharf definierten Punktobjekten hat in der geodätischen Wissenschaft eine lange Tradition. Mit Hilfe von stochastischen Ansätzen werden in der Ausgleichungsrechnung die Widersprüche zwischen den Beobachtungen und dem funktionalen Modell verteilt. Man geht davon aus, dass es den richtigen Wert gibt und dieser wird innerhalb des funktionalen Modells anhand der Beobachtungen und deren Messunsicherheit optimal geschätzt. Wenn dieser Wert aber nicht existiert, weil die Größe, die ermittelt werden soll, nicht genau bestimmt ist, so hilft der Ansatz der Ausgleichungsrechnung nicht weiter. Diese Modellierungseinschränkungen liegen aber bei vielen Naturphänomenen vor. So lässt sich z.B. die Uferlinie eines Meeres oder die Fußlinie eines Berges nicht genau lokalisieren. Jedenfalls erheblich ungenauer als die Messtechnik es zulassen würde. Immer wenn zwei Objektklassen ineinander übergehen, besteht das Problem, dass die Grenzen nicht scharf definiert sind. Üblicherweise wird das Problem umgangen, indem in der Erfassungsanweisung zur Digitalisierung von Geoobjekten ein Grenzwert eingeführt wird, ab wann ein Gebiet zu einer Objektklasse gehört. Bei dieser Art der Modellierung begeht man allerdings den Fehler, dass Teile des Objektes unterschlagen werden. Würde man so erfassen, dass alles hinzugenommen wird, was noch irgendwie zum dem Objekt dieser Klasse gehört, so ist die Gesamtflächenbilanz verletzt und es kommt zu Überlagerungen von benachbarten Objekten. Egal welchen Ansatz man wählt, man wird der natürlichen Situation nicht gerecht. Was benötigt wird, ist eine Erweiterung der Modellierung zur Darstellung von Unschärfe bzw. Übergangszonen.

Die Unterscheidung zwischen Wahrscheinlichkeitsrechnung und Fuzzy Set Theorie wird bei Fisher (1996) ausführlich diskutiert.

Messunsicherheiten

Fortpflanzung der Varianzen von Koordinaten der Stützpunkte auf interpolierte Punkte

In GIS werden die Stützpunkte der Geometrie von Geoobjekten erfasst. Die Erfassung ist ein Messvorgang und dadurch mit Messunsicherheiten behaftet. Jede einzelne Messung ein Zufallsexperiment. Die Streuung der Messung ist statistischer Natur. Durch Wiederholungsmessungen und durch Redundanzen in unabhängigen Messanordnungen werden die Parameter der statistischen Verteilung geschätzt. In einem GIS soll allerdings das ganze Objekt repräsentiert werden, d.h. auch alle Begrenzungspunkte die zwischen den Stützpunkten liegen. Mit der Geometrie wird auch eine Interpolationsvorschrift zwischen den Stützpunkten angegeben. Durch die Varianzenfortpflanzung können die Varianzen und Kovarianzen der Zwischenpunkte ermittelt werden. In vielen Fällen ist die Interpolation linear. Dieser Fall ist ausführlich untersucht worden (Caspary und Scheuring, 1992). Die Einhüllende aller Fehlerellipsen kann durch sogenannte ε-Bänder dargestellt werden.

Eine Visualisierung dieser Fortpflanzung von Messunsicherheiten in den Stützpunkten auf ein geschlossenes Polygon zeigt Abbildung 1.

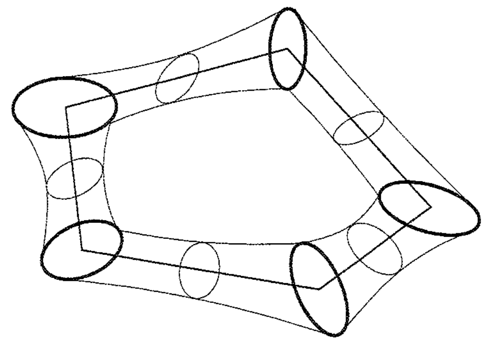

Abbildung 1: ε-Bänder um ein geschlossenes Polygon.

Geometrievereinfachungen

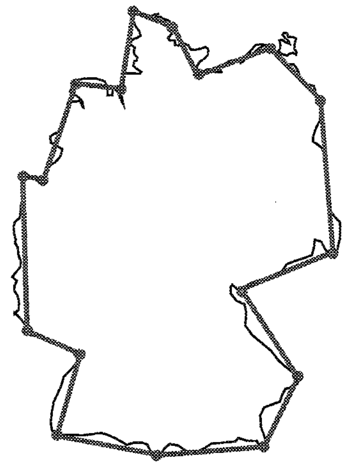

Abbildung 2: Unterschiedliche Auflösungen.

Durch das Setzen von Stützpunkten erfolgt eine Vereinfachung der Geometrie. Die Auswahl der zu setzenden Stützpunkte ist von der Auflösung, also der Angabe, welche Details in den Daten noch aufgelöst werden können, abhängig. Um ein Objekt repräsentativ zu erfassen, kann eine manuelle Erfassung u. U. mit wenig Stützpunkten auskommen. Zuviel erfasste Stützpunkte können mit Linienvereinfachungsalgorithmen (z.B. Douglas-Poiker-Algorithmus, Douglas und Peuker, 1973) ausgedünnt werden. Die Modellvereinfachung ist mit der Lagegenauigkeit korreliert, da die interpolierten Punkte der groben Geometrie den Genauigkeitsanforderungen entsprechen müssen. Aus diesem Grund müssen Auflösung und Genauigkeitsziele immer in Zusammenhang stehen.

Die Auflösung, also Stützpunktabstand und Mindestgrößen für die Erfassung, und Kenngrößen wie Umfang eines Objektes oder die Anzahl der erfassten Objekte eines bestimmten Gebietes können einer fraktalen Analyse unterzogen werden. Interessanterweise besteht ein empirisch sehr gut nachweisbarer fraktaler Zusammenhang zwischen diesen Größen (Gatrell, 1991). Sarjakoski (1996) hat für Seen und Inseln in Finnland Untersuchungen durchgeführt.

Der Zusammenhang kann über die Formel

$$N(A > a) = Fa^{-D/2}$$

ausgedrückt werden. N gibt dabei die Anzahl der Seen in Finnland mit einer Fläche A an, die größer ist als ein Grenzwert a. F ist eine Konstante, die aus den Beobachtungen geschätzt wird, und D ist die fraktale Dimension. Für Seen und Inseln nimmt D in diesem Fall den Wert 1.7 an. Eine einfache Konsequenz aus diesen Untersuchungen ist, dass alle quantitativen Aussagen, die aus einem GIS abgeleitet werden, mit der Auflösung der Daten in Zusammenhang stehen. Dies muss bei der Interpretation von Ergebnissen berücksichtigt werden.

Unscharfe Begrenzung

Fuzzy Set Theory

Die Abgrenzung zwischen zwei Objekten ist in vielen Fällen nicht genau definiert. Zum Beispiel kann durch Buschwerk und vereinzelte Bäume nicht zwischen Wald und Wiese unterschieden werden. Ab einer bestimmten Entfernung vom Wald, kann man klar sagen, dass eine Wiese vorliegt, auch wenn einzelstehende Bäume vorhanden sind. Genauso kann man im Wald sagen, dass es sich um einen Wald handelt, auch wenn am Waldboden Gras wächst. Nur in der Zwischenzone kann niemand genau festlegen, wo der Wald anfängt und wo die Wiese aufhört. Tatsächlich würde man sagen, dass in diesem Bereich sowohl Wald als auch Wiese vorhanden ist, und die Zugehörigkeit zu einer bestimmten Objektklasse nimmt zum Inneren des Objektes zu. Es handelt sich bei dieser Zugehörigkeit nicht um eine Wahrscheinlichkeit.

Diese anschauliche Festlegung kann mit der Fuzzy Set Theorie formalisiert werden, indem man eine Zugehörigkeitsfunktion einführt, die Werte zwischen 0 und 1 annimmt. Der Zugehörigkeitswert 0 bedeutet, das dieser Ort nicht zur Objektklasse gehört und der Wert 1 besagt, dass dieser Ort vollständig und ausschließlich der Objektklasse angehört. Bei Werten zwischen 0 und 1 ist die Zugehörigkeit sowohl zu der einen als auch einer anderen Objektklasse gegeben. Die Summe aller Zugehörigkeitswerte für einen Ort muss wieder genau 1 ergeben, sonst ist die Flächendeckungsbedingung nicht gegeben.

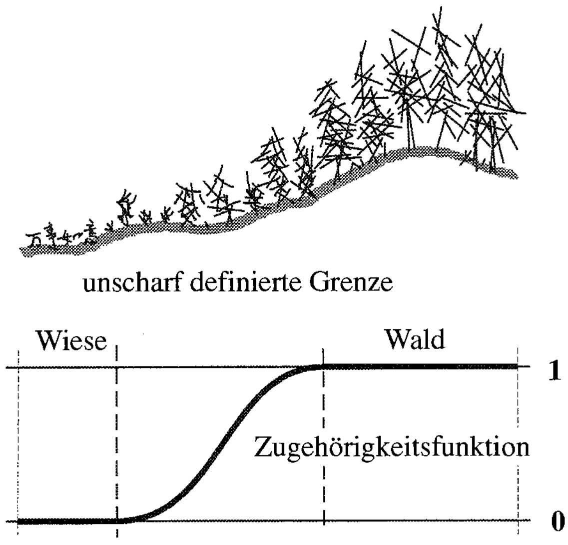

Abbildung 3: Zugehörigkeitsfunktion und Übergangszone zwischen verschiedenen Objektklassen.

Für die Zugehörigkeitsfunktion können verschiedene funktionale Zusammenhänge gewählt werden. Am einfachsten ist auch hier ein linearer Anstieg. Bei Kurven höherer Ordnung wie in Abbildung 3 besteht das Problem, dass zusätzliche Parameter über die Flankensteilheit ermittelt werden müssen. Anhand von Messwerten sind diese zusätzlichen Parameter schwer zu bestimmen, es müssten dann Erfahrungswerte gewählt werden, die den natürlichen Verlauf hinreichend gut repräsentieren. Für alle möglichen Objektklassenpaare müssten dann solche Erfahrungswerte im Datenschema verankert werden. Die Breite des Übergangsbereiches ist der wichtigste Parameter bei der Modellierung mit Zugehörigkeitsfunktionen. Diese können entweder bei der Erfassung bestimmt werden, oder sind wieder über empirische Werte festzulegen.

„Egg-Yolk" (Ei-Dotter) – Ansatz

Seinen Namen verdankt dieser Ansatz dem Aussehen, indem nämlich die zugehörigen Polygone zu der 0- und der 1-Grenze aus dem Fuzzy Set Ansatz erfasst werden. Damit entstehen 2 ineinander geschachtelte Flächen, die an ein aufgeschlagenes Ei erinnern. Die innere Fläche darf zwar die äußere berühren, nämlich genau dort, wo eine scharfe Abgrenzung vorliegt, aber ein Schneiden dieser Flächen ist nicht erlaubt.

Cohn und Gotts (1996) zeigen, wie mit diesem Ansatz topologische Beziehungen beschrieben werden können. Da die beiden Polygonpaare jeweils entweder disjunkt, sich schneiden oder kongruent sein können, und bestimmte Konstellationen auszuschließen sind, weil sich die inneren Polygone nicht schneiden können, wenn die äußeren disjunkt sind, ergeben sich 46 mögliche topologische Beziehungen.

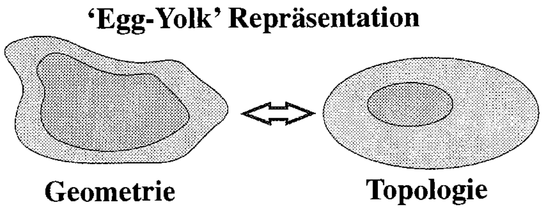

Abbildung 4: Der 'egg-yolk' Ansatz zur Darstellung von unscharfen Objekten

Neben dem allgemeinen Fall der topologischen Beziehungen zwischen zwei unscharf begrenzten Objekten kann das Modell auch zwischen einem scharf und einem unscharf begrenzen Objekt angewandt werden. Die Beziehungen im 9-Intersection Modell (Egenhofer et al., 1994) ergeben sich dabei als Spezialfall für zwei klar abgegrenzte Objekte.

Die Stufen der 7 möglichen Fälle von topologischen Beziehungen, die ein scharf und ein unscharf begrenztes Objekt miteinander eingehen können, werden bei Clementini und Di Felice (1996) definiert und mit intuitiv verständlichen Namen (disjoint, nearly meet, meet, close, nearly coverdBy, coveredBy und inside) bezeichnet. Abbildung 5 zeigt wie diese Operatoren graphisch interpretiert werden können.

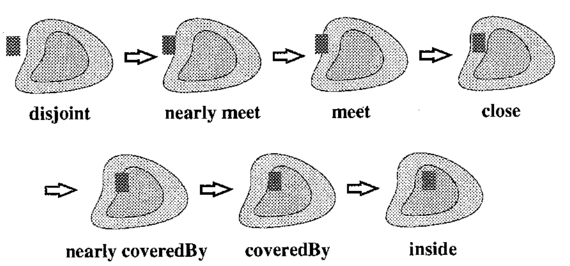

Abbildung 5: Topologische Beziehungen zwischen einem unscharf begrenzten Objekt wie z.B. Wald und einem gut bestimmten Haus (Clementini and Di Felice, 1996)

Semantische Unschärfe

Neben der geometrischen Unschärfe von Objekten ist die Zugehörigkeit zu einer Objektklasse nicht immer eindeutig. Nachdem ein Objekt nur einer Klasse angehören kann und diese Zuordnung entweder richtig oder falsch ist, führt dies zu Verzerrungen bei der semantischen Interpretation. Die Unschärfe von Abgrenzungen der Objektklassen im Anwendungsschema kann zu solchen falschen Zuordnungen führen. Beispielsweise führt die Koexistenz der Objektklassen „Feuchtwiese", „Nasswiese", „Sumpf" und „Wiese" zu Problemen. Die Abgrenzung ist nicht klar genug, dass jeder Erfasser und jeder GIS-Anwender die selben Vorstellungen hat. Oder haben Wanderer mit verschiedenen alpinistischen Ambitionen unterschiedliche Erwartungen an Objekte der Klassen „Bergpfad" und „Klettersteig". Was für den einen Wanderer noch ein leicht begehbarer Pfad mit leichten ausgesetzten Stellen ist, wird für manch anderen zur gefährlichen Kletterpartie. Auch zeitliche Schwankungen der Zugehörigkeit machen eine eindeutige Zuordnung schwierig.

Die semantische Unschärfe lässt sich nur durch klare Abgrenzungen bei den Definitionen der Objektklassen vermeiden oder gering halten. Trotzdem werden gerade bei einer feinen Granularität der Modellierung Unschärfen bleiben und Verwechslungen bzw. Zuordnungen zu mehreren oder zu keiner eindeutigen Klasse vorkommen.

Räumliche Operationen für fuzzy Geometrien

Für viele geometrischen Operationen eines GIS, wie z.B. Verschneidungen, müssen Erweiterungen für unscharf begrenzte Objekte eingeführt werden. Die Lösungen sehen für die Fuzzy Set Theorie anders aus als für den ‚egg-yolk' Ansatz. Die Ergebnisse geometrischer Operationen können sehr komplex werden. Bei dem ‚egg-yolk' Ansatz ergeben sich im ungünstigsten Fall (konvexe Objekte und keine Löcher vorausgesetzt, sonst können sich noch kompliziertere Gebilde ergeben) 8 Teilflächen, die alle zu unterscheiden sind, während im Fall von scharfen Begrenzungen nur 3 Teilflächen entstehen (Abbildung 6).

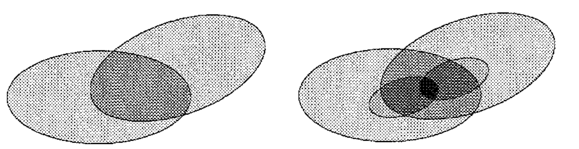

Abbildung 6: Verschneidungsoperation mit fuzzy Objekten als 'egg-yolk' modelliert

Bei der Modellierung der unscharfen Grenze mit Zugehörigkeitsfunktionen ergeben sich schon bei einem linearen Übergang komplizierte Funktionen. Werden Kurven höherer Ordnung verwendet, ist ein erheblicher numerischer Rechenaufwand erforderlich.

Literatur

Burrough, P. A. and A. U. Frank (Eds), 1996: Geographic objects with indeterminate boundaries. Taylor & Francis

Caspary, W. und R. Scheuring, 1992: Error-bands as measures of geometrical accuracy. Proceedings of EGIS '92, Utrecht, pp. 226-233

Clementini, E. and P. Di Felice, 1994: A comparison of methods for representing topological relationships. Information Sciences 80, pp. 1-34

Clementini, E. and P. Di Felice, 1996: An algebraic model for spatail objects with indeterminate boundaries. In: Burrough and Frank (1996), pp. 155-169

Cohn, A. G. and N. M. Gotts, 1996: The 'Egg-Yolk' representation of regions with indeterminate boundaries. In: Burrough and Frank (1996), pp. 171- 187

Douglas, D. H. and T.K Peuker, 1973: Algorithms for the reduction of the number of points required to represent a digitised line or its caricature. The Canadian Cartographer, Vol. 10. pp. 112-122#

Egenhofer, M., D. M. Mark und J. Herring, 1994: The 9-intersection: Formalism and ist use for natural-language spatial predicates. NCGIA Technical Repot 94-1, Santa Barbara, CA, Buffalo, NY, Orono, ME, USA

Fisher, P., 1996:Boolean and Fuzzy Regions. In: Burrough and Frank (1996), pp. 87-94

Gatrell, A. C., 1991: Concepts of space and geographical data. In: D. J. Maguire, M. F. Goodchild and D. W. Rhind, Geographical information systems. Longman Scientific and Technical, pp. 119-134

Molenaar, M., 1998: An introduction to the theory of spatial object modelling for GIS. Taylor & Francis

Molenaar, M., 1999: Spatial objects without boundaries. Proceedings of The International Symposium on Spatial Data Quality '99, The Hong Kong Polytechnic University, pp. 479-486

Sarjakoski, T.,1996: How many lakes, islands and rivers are there in Finland? In: Burrough and Frank (1996), pp. 299-312